

Künstliche Intelligenz wird zweifellos die Hochschullehre und Forschung nachhaltig verändern. Während sich die aktuelle Debatte hauptsächlich auf die Lehre fokussiert, insbesondere mit Blick auf Prüfungsmodalitäten, kommen Herausforderungen und Chancen in Bezug auf die Forschung erst langsam zur Sprache. In diesem Beitrag soll aufgezeigt werden, wie der Einsatz von textgenerierender künstlicher Intelligenz in der Forschung in Einklang mit den Regeln der guten wissenschaftlichen Praxis (GWP) gebracht werden kann. Neben bereits bestehenden Stellungnahmen, Gutachten und Empfehlungen stützt sich dieser Artikel auf einen Expertenworkshop, organisiert vom Dialogforen-Projekt des *Ombudsman für die Wissenschaft*, in dem deziert GWP-relevante Themen wie Transparenz, Autorschaftskriterien und GWP-Verstöße adressiert wurden.

Artificial intelligence will undoubtedly lead to permanent changes in teaching and research at universities. The recent debate has focused mainly on teaching, especially concerning examination modalities, however the challenges and opportunities associated with research are now slowly coming to the fore. The aim of this paper is to show how the use of artificial intelligence to generate research texts can be aligned with the rules of good research practice (GRP). The article draws on existing statements, expert opinions, recommendations, and on an expert workshop organised as part of a Discussion Hubs project of the *German Research Ombudsman* devoted to GRP-related topics such as transparency, authorship criteria and GRP violations.

KATRIN FRISCH, FELIX HAGENSTRÖM, NELE REEG

Textgenerierende KI und gute wissenschaftliche Praxis

Mit der Veröffentlichung von ChatGPT Ende 2022 nahm die Debatte um den Einsatz von textgenerierender künstlicher Intelligenz (TKI) in der Wissenschaft spürbar an Fahrt auf. Die Unsicherheit im Umgang mit (T)KI¹ ist wie in vielen anderen Bereichen auch in der Wissenschaft groß. Die Vielzahl an Veranstaltungen zu diesem Thema spiegelt den großen Diskussionsbedarf wider. Fokus solcher Veranstaltungen ist insbesondere die Aufklärung zur Funktionsweise, sich daraus ergebende Limitationen, Nutzungsrisiken und der Erfahrungsaustausch. Für den Hochschulbetrieb steht dabei vor allem der Umgang in der Lehre im Mittelpunkt, insbesondere Konsequenzen für Prüfungsleistungen.

In anderen Bereichen der Gesellschaft, über diverse Wirtschaftszweige bis hin zum Gesundheitswesen, wird KI bereits seit einigen Jahren angewendet. Herausgestellt wurde in den begleitenden Diskussionen immer zweierlei: das Potenzial von KI und die Gefahren ihrer Anwendung. Schnell wurde klar, dass es Regularien zum Einsatz von KI geben muss. Innerhalb der EU wurde 2017 eine Regulation von KI angestoßen.² Diese Initiative mündete schließlich in den vom Europaparlament im Juni 2023 verabschiedeten *Artificial Intelligence Act*; die finale Gesetzesform soll Ende 2023 vorliegen. Das Gesetz bezieht sich auf den Umgang mit KI im Allgemeinen und bildet somit angesichts der aktuellen Entwicklungen im Bereich von TKI einen bedeutenden Rahmen. Als Antwort auf den technologischen Fortschritt im Bereich TKI erstellte auch das Büro für Technikfolgen-Abschätzung beim Deutschen Bundestag

(TAB) für den Ausschuss für Bildung, Forschung und Technikfolgen-Abschätzung des Deutschen Bundestages einen Hintergrundbericht, der im April 2023 veröffentlicht wurde.³ Kurz darauf folgte ein öffentliches Fachgespräch. Der Bericht umfasst den technologischen Entwicklungsstand von TKI, enthält einen Überblick über Möglichkeiten und Grenzen und bespricht die Anwendung unter anderem in Unternehmen, im Gesundheits- und im Rechtswesen. Darüber hinaus widmet er sich auch den Auswirkungen von TKI auf Bildung und Forschung. Der Deutsche Ethikrat adressiert die neuesten Entwicklungen in seiner Stellungnahme *Mensch und Maschine – Herausforderungen durch Künstliche Intelligenz*, wobei darin der Fokus weniger auf der Forschung liegt als vielmehr auf anderen Bereichen wie etwa (schulische) Bildung, Entscheidungsprozesse der öffentlichen Verwaltung oder die Auswirkungen auf Kommunikation und Meinungsbildung. Das Projekt KI:edu.nrw der Ruhr-Universität Bochum erarbeitete in Zusammenarbeit mit dem Juristen Thomas Hoeren (Münster) das in der Wissenschaft vielbeachtete Gutachten *Didaktische und rechtliche Perspektiven auf KI-gestütztes Schreiben in der Hochschulbildung*.⁴ Es erörtert wichtige Aspekte der Anwendung von TKI, konzentriert sich allerdings eher auf die Lehre als auf die Forschung. Auch im Verlagswesen wirft die Nutzung von TKI zahlreiche Fragen auf, was Verlage zu verschiedenen Stellungnahmen veranlasste.⁵

Gegenstand der erwähnten Stellungnahmen, Positionspapiere, Gutachten und Richtlinien sind insbesondere Fragen zu Autorschaft, Transparenz und der

guten wissenschaftlichen Praxis (GWP). Auch wenn sich zu diesen drei Themen Punkte herauskristallisieren, zu denen Konsens herrscht, wird auch deutlich, dass wichtige Fragen noch offen sind. Dabei wurden viele Fragen bislang nicht detailliert aus der Perspektive der GWP betrachtet. Auf dem vom *Ombudsman für die Wissenschaft* (OfdW) veranstalteten Ombudssymposium im Februar 2023 wurde deutlich, dass es unter Ombudspersonen viel Unsicherheit gibt.⁶ Der *Ombudsman für die Wissenschaft* und das dort angesiedelte Dialogforen-Projekt nahmen dies zum Anlass, im Juni 2023 einen Workshop zum Verhältnis von TKI zu guter wissenschaftlicher Praxis zu veranstalten. Dafür wurden Expert*innen aus Forschung, universitärer Infrastruktur und dem Ombudswesen eingeladen, um entlang von Leitfragen zu den eingangs erwähnten Themen Autorschaft, Transparenz und Abweichungen zur GWP aktuelles Wissen zusammenzutragen, offene Fragen zu klären und Desiderate zu benennen. Im Fokus stand die Frage, wie der GWP-konforme Umgang mit TKI in der Wissenschaft aussehen kann und ob zusätzliche GWP-Empfehlungen notwendig sind. Abschließend gab es Raum, um über zukünftige Entwicklungen nachzudenken.⁷ Der vorliegende Beitrag trägt den sich anbahnenden Konsens in diesen Punkten zusammen und diskutiert einige offene Fragen, ohne dabei Anspruch auf Vollständigkeit zu erheben. Er soll vielmehr sensibilisieren und einen Austausch fördern.

TKI und wissenschaftliche Autorschaft

Angesichts der Fähigkeit von TKI, auf Basis von *prompts* Text zu generieren, stellt sich grundlegend die Frage: Kann TKI die Kriterien für wissenschaftliche Autorschaft erfüllen? Erste Veröffentlichungen mit Nennung von ChatGPT als Ko-Autor machten zumindest schnell die Runde.⁸ Eher unbemerkt von weiten Teilen der Wissenschaftsgemeinschaft wurde bereits 2019 das Buch *Lithium-Ion Batteries: A Machine-Generated Summary of Current Research* bei Springer veröffentlicht – gar in Alleinautorschaft im Namen von Beta Writer, einer Software.⁹ Benedikt Fecher zeigte ferner mit Kolleg*innen am Alexander von Humboldt Institut für Internet und Gesellschaft in einer Umfrage unter Forschenden aus dem Bereich der Science and Technology Studies kürzlich auf, dass 15,4 % der Befragten der Aussage »ChatGPT and other large language models should be able to be listed as co-authors in the texts they help produce« zustimmten; weitere 9,6 % äußerten gar eine hohe Zustimmung, 17,3 % äußerten weder Zustimmung noch Ablehnung.¹⁰ Doch wie steht die GWP zu dieser Haltung? Zur Beantwortung dieser Frage lohnt sich ein Blick in den nationalen Verhaltenskodex zu guter wissenschaftlicher Praxis, denn er enthält die von der Wissenschaftsgemeinschaft akzeptierte Definition von Autorschaft auf wissenschaftlichen Veröffentlichungen. In Deutschland sind das die *Leitlinien zur Sicherung*

guter wissenschaftlicher Praxis der Deutschen Forschungsgemeinschaft (DFG). Leitlinie 14 zu Autorschaft besagt: »Autorin oder Autor ist, wer einen genuinen, nachvollziehbaren Beitrag zu dem Inhalt einer wissenschaftlichen Text-, Daten- oder Softwarepublikation geleistet hat. Alle Autorinnen und Autoren stimmen der finalen Fassung des Werks, das publiziert werden soll, zu. Sie tragen für die Publikation die gemeinsame Verantwortung, es sei denn, es wird explizit anders ausgewiesen.«

Zur Erfüllung der Autorschaftskriterien sind also drei Dinge erforderlich: Die Erbringung eines genuinen, nachvollziehbaren Beitrags, die Äußerung von Zustimmung sowie die Übernahme von Verantwortung. In Anbetracht dessen muss die eingangs gestellte Frage, ob TKI gemäß den Regeln der GWP die Kriterien für wissenschaftliche Autorschaft erfüllt, klar verneint werden. Die nächsten Abschnitte verdeutlichen die Gründe hierfür.

Zustimmung und Verantwortung

So weist etwa der Deutsche Ethikrat bei der Frage, ob KI Verantwortung übernehmen kann, klar aus, dass »eine Verantwortungsübernahme (als Verantwortungssubjekt) nur Personen als verantwortlichen Wesen möglich ist, beispielsweise den Individuen, die Technik entwickeln und herstellen, die ihren Einsatz etwa in der Politik oder Unternehmen ermöglichen und fördern, oder denjenigen, die Technologien einsetzen«.¹¹ Dieser Grundsatz lässt sich auch auf den Bereich der Forschung erweitern und im Speziellen auf die Anwendung von TKI beziehen. TKI verarbeitet Sprache, gibt Texte aus, verfügt aber – auch wenn das Wort »Intelligenz« in der Bezeichnung dies suggerieren mag – über kein Verständnis für den erstellten Text.¹² Mohammad Hosseini weist in der Debatte darauf hin, dass für die Übernahme von Verantwortung ein freier Wille notwendig sei – eine Eigenschaft, die TKI nicht zugesprochen werden kann.¹³ Auch viele Verlage weisen in ihren mittlerweile erschienenen Statements auf die Unmöglichkeit der Übernahme von Verantwortung hin.

Was manchmal in diesen Statements unterzugehen scheint, ist das zweite eingangs erwähnte Kriterium für wissenschaftliche Autorschaft, nämlich das der Zustimmung: TKI kann auch keine Zustimmung zur Veröffentlichung eines Textes äußern, denn sie ist nicht als legale Entität anzusehen. So betont denn auch das Committee on Publication Ethics (COPE) in seinem Statement, dass TKI weder die An- oder Abwesenheit von Interessenkonflikten bestätigen noch Copyrights oder Lizenzvereinbarungen zustimmen bzw. beachten könne.¹⁴

Auch unabhängig von der Frage, ob TKI einen genuinen, nachvollziehbaren (oder wie oft auch beschrieben: wesentlichen) Beitrag zu einer Veröffentlichung leisten kann, zeigen diese beiden Punkte eindeutig auf, dass die Nennung von TKI als Autorin nicht in Ein-

klang mit den Leitlinien zu GWP steht. Auch lohnt sich ein Blick auf die rechtliche Perspektive zu der hier behandelten Frage hinsichtlich Autorschaft. Der Jurist Thomas Hoeren kommt in seinem Rechtsgutachten mit Blick auf § 7 UrhG und § 2 Abs. 2 UrhG zu dem Schluss: »Programmen wie ChatGPT kann [...] keine Urheberschaft und damit auch keine Autorschaft an einem von ihnen generierten Text zuerkannt werden, unabhängig davon, ob die Schöpfungshöhe erreicht wurde oder nicht«.¹⁵ Die Sicht der GWP und die rechtliche Einschätzung decken sich also in der Frage zu Autorschaft.

Obwohl sich die Regeln der GWP und juristische Einschätzungen darin einig sind, dass einer TKI keine Autorschaft zusteht, können sich Forschende, wie eingangs aufgezeigt, sehr wohl vorstellen, eine TKI in die Autorenzeile aufzunehmen. Zielführender als eine Anpassung oder Erweiterung der Leitlinie 14 des DFG-Kodex scheint zu sein, die bestehende Leitlinie im Kontext von TKI auszulegen. Ideal wäre eine allgemein gültige Auslegung, zum Beispiel in Form einer Erläuterung auf der dritten Ebene des DFG-Kodex. Damit wäre sie zum einen für alle Forschenden zugänglich. Zum anderen könnte sie bei eventuellen Konflikten als Orientierung und Lösungsbasis dienen. Die Diskrepanz zeigt zudem: Nicht nur Studierende sind unsicher; auch bei vielen Forschenden (aller Statusgruppen) besteht Aufklärungsbedarf für den Umgang mit TKI. Eine GWP-konforme Auslegung der bestehenden Leitlinie 14 ist also wichtig für das eigene wissenschaftliche Arbeiten, aber auch für das Arbeiten anderer, denn gerade Seniorforschende geben als Multiplikator*innen ihr Wissen an den Nachwuchs weiter.

Wenn also TKI keine Verantwortung für den Inhalt einer Veröffentlichung übernehmen kann, dann fällt die Verantwortung den Autor*innen zu, die TKI genutzt haben: sie tragen die Verantwortung für den Einsatz und dessen Ergebnis. Welche Aspekte dabei zu beachten sind, wird später noch einmal aufgegriffen. Hier soll zunächst auf das laut DFG-Kodex erste Kriterium für wissenschaftliche Autorschaft eingegangen werden, nämlich das Erbringen eines »genuine[n], nachvollziehbaren Beitrags«.

Genuiner, nachvollziehbarer Beitrag

Betrachtet man dieses Kriterium für wissenschaftliche Autorschaft aus Leitlinie 14, so schließen sich zwei Fragen an. Erstens: Kann TKI einen genuine[n], nachvollziehbaren Beitrag zu einer wissenschaftlichen Publikation leisten? Und zweitens: Welchen genuine[n], nachvollziehbaren Beitrag leisten Forschende zu einer wissenschaftlichen Publikation, wenn sie TKI nutzen?

Um diesen beiden Fragen nachzugehen, sei zunächst gezeigt, bei welchen Tätigkeiten KI-Tools im Bereich des Schreibens eingesetzt werden können. Peter Salden, Nadine Lordick und Maike Wiethoff demonstrieren in

ihrer Einführung *KI-basierte Schreibwerkzeuge in der Hochschule* des Bochumer Gutachtens die Vielfalt von bereits vorhandenen Tools: Neben der Generierung von Text zählen dazu die Korrektur von Text, das Paraphrasieren und Umschreiben, das Übersetzen, das Anregen von Kreativität, die Literaturrecherche und letztlich die Textbewertung.¹⁶ Es ist wahrscheinlich (und zum Teil schon eingetreten), dass Tools mit weiteren Funktionen und erhöhter Leistungsfähigkeit in diesen Bereichen hinzukommen.

Angesichts dieser Vielfalt von Schreibtools ist eine neue Rollenverteilung zu erwarten: Tätigkeiten werden an KIs ausgelagert. Dabei erübrigen sich die Kompetenzen zum Ausführen dieser Tätigkeiten jedoch keinesfalls; denn Forschende müssen nach wie vor in der Lage sein, das Ergebnis bewerten zu können. Grundsätzlich ist diese Konstellation nicht neu, wie Teilnehmende unseres Workshops hervorhoben. Eine ähnliche Situation bestehe bereits bei bisherigem kollaborativem Schreiben in Forschungsvorhaben. Hier entstehen Ideen im Austausch von Forschenden, Aufgaben werden gemäß der jeweiligen Expertise verteilt. Der Erkenntnisprozess und die Zuweisung von Autorschaft ist entsprechend unscharf und Eigenleistungen sind schwer voneinander abzugrenzen.¹⁷ Bislang handelte es sich jedoch um Mensch-Mensch-Kollaborationen. Mit dem Einsatz von TKI bekommen wir es vermehrt mit Mensch-Maschine-Kollaborationen zu tun. TKI sei dabei als Schreibpartnerin anzusehen, aber – und das ist ein wichtiger Unterschied zur Kollaboration unter Menschen – gleichzeitig als Werkzeug einzustufen. Doris Weßels betonte, die Rolle der Nutzer*innen sei so zu verstehen, dass sie das Werkzeug TKI steuern. Zu dieser Aussage herrschte Konsens unter den Teilnehmenden. Weßels verwies auf ihren Vorschlag, die Eigenständigkeitserklärungen bei Abgabe von Studienarbeiten wie folgt anzupassen: »Bei der Erstellung dieser Studienarbeit habe ich durchgehend eigenständig und beim Einsatz IT-/KI-gestützter Schreibwerkzeuge steuernd (Hervorhebung durch Verf.) gearbeitet«.¹⁸ Dieser Ansatz lässt sich auch auf die Forschung und wissenschaftliche Veröffentlichungen übertragen. Folglich bestünde die GWP-konforme Nutzung von TKI u.a. darin, die Werkzeuge mit Sachverstand und Expertise zu steuern. Aus dieser steuernden Funktion heraus ergibt sich auch: Forschende tragen die Verantwortung für deren Einsatz. Die Konsequenzen werden im nächsten Abschnitt behandelt.

Die Frage nach dem geleisteten eigenen Beitrag hängt somit von der Leistungsfähigkeit der TKI sowie vom Maß an Steuerung durch die Forschenden ab. Zudem beeinflusst auch das Autorschaftsverständnis der Fachdisziplin und der dazugehörige Stellenwert von Text die Antwort und Einschätzung: Disziplinen, in denen Textarbeit einen hohen Stellenwert hat, werden zu einer anderen Einschätzung gelangen als Disziplinen, in denen Textarbeit eine eher untergeordnete Rolle spielt. So ist

denn auch in den Geisteswissenschaften vordergründig der Text Ort des Ausdrucks von Originalität und Kreativität, während in anderen Fächern wie den Naturwissenschaften dem Text oftmals eher eine deskriptive oder dokumentierende Funktion zukommt und er in der Tendenz einer Standardisierung unterliegt. Kreativität im engeren Sinne kann eine TKI nicht leisten, daher erscheint ein Beitrag von TKI in geisteswissenschaftlichen Texten wenig wertvoll bzw. sinnvoll, während in den Naturwissenschaften oder Fächern mit ähnlichem Textverständnis bestimmte Schritte des Schreibprozesses eher ausgelagert werden können. So verdeutlichte Nadine Lordick, dass gerade für die Geisteswissenschaften, welche vor allem Sorge vor dem Einsatz von KI äußern, die neusten Entwicklungen zu TKI nicht so bedrohlich seien. In diesem Zusammenhang betonte zudem Katharina Beier, dass etwa für die Verschriftlichung von deskriptiven Teilen zu Ergebnissen die Eingabe von (erhobenen) Daten erforderlich sei. Den Fokus vorzugeben, eine Auswahl und Gewichtung zu treffen, all dies könnte TKI nicht übernehmen, sondern obliege den Forschenden. Je nach Stellenwert und Funktion des Textes ergeben sich unterschiedliche, fachspezifische Konsequenzen für das Verständnis einer GWP-konformen Nutzung von TKI und was als Abweichung hiervon eingestuft wird. Auch die Auswirkungen von TKI auf die Qualität von Forschung sollte dabei nicht aus dem Blick geraten.

Wie ist der Einsatz von TKI im Allgemeinen einzuschätzen? Stellt die Nutzung und Steuerung von TKI gar in einigen Disziplinen einen neuen wissenschaftlichen Skill dar, der dem wissenschaftlichen Nachwuchs im Studium ebenso wie Forschenden im Allgemeinen vermittelt werden sollte, während er in anderen Disziplinen als Verstoß gegen die GWP gewertet werden sollte? Zu dieser Frage wären fachspezifische Debatten und Stellungnahmen begrüßenswert, denn aufgrund der Unterschiede in Stellenwert und Funktion von Textarbeit werden zur Beantwortung dieser Frage fachspezifische, keine fachübergreifenden Antworten notwendig sein. Daran schließt sich auch die Frage an, wie der Einsatz von TKI von Drittmitgeber*innen gewertet wird. Auch hierzu sind fachspezifische Debatten notwendig. TKI könnte dabei in einigen Disziplinen als Chance angesehen werden, sich auf wissenschaftliche Kernkompetenzen abseits des (Nieder-)Schreibens zu fokussieren, wie Guido Juckeland verdeutlichte.

Verantwortung von Autor*innen bei Nutzung von TKI

Forschende tragen als Autor*innen für den Inhalt einer wissenschaftlichen Veröffentlichung die Verantwortung – unabhängig davon, ob TKI zum Einsatz kam oder nicht. Bei der Nutzung von TKI ist zusätzlich zu beachten, dass es in der Verantwortung von Autor*innen liegt, die von einer TKI erstellten Inhalte stets zu überprüfen und ggf. anzupassen. TKI wie ChatGPT sind da-

für bekannt, zu »halluzinieren«, d.h., Fakten, Zitate und Quellen zu erfinden. Blindes Vertrauen in die von einer TKI erstellten Texte widerspricht der GWP. Die verantwortungsvolle Nutzung muss einige Punkte beachten. Autor*innen müssen etwa sicherstellen, dass Zitate korrekt wiedergegeben oder paraphrasiert werden. Die dazugehörigen Angaben der Quelle müssen überprüft werden. Darüber hinaus müssen Autor*innen sicherstellen, dass der Stand der Wissenschaft korrekt dargestellt wird. Schreibunterstützende KI kann als Tool genutzt werden, um geeignete Literatur herauszusuchen. Aber es ist Aufgabe der Forschenden, die Literatur selbst zu lesen, einzuordnen und zu analysieren, wie auch Debora Weber-Wulff hervorhob. Entstehen durch den Einsatz von TKI Verstöße gegen die gute wissenschaftliche Praxis, so tragen die Autor*innen des Textes dafür die Verantwortung.

Was kennzeichnet eine GWP-konforme Nutzung von TKI?

Es stellen sich zwei grundlegende Fragen, die je nach GWP-Kontext unterschiedlich beantwortet werden können. Erstens: Ist generell eine Nutzung von TKI zulässig? Falls ja, schließt sich die zweite Frage an: Wie sieht eine GWP-konforme Nutzung aus?¹⁹

Zur ersten Frage: Nur wenige öffentliche Stellungnahmen schließen die Nutzung von TKI für die Forschung und insbesondere für die Erstellung wissenschaftlicher Publikationen generell aus. Stattdessen fordern die meisten, auch wenn sie den Einsatz von TKI nicht befürworten, den Umgang zu regeln. Auch aus den Leitlinien der GWP lässt sich kein grundsätzlicher Ausschluss der Nutzung von TKI ableiten.²⁰ Es bedarf also klarer Regeln, um einen wissenschaftlich integren Umgang mit der neuen Technologie sicherzustellen. Das TAB verdeutlicht hierzu: Eine Nutzung von KI-Systemen im Forschungsprozess wie auch beim Verfassen einer Publikation ist dagegen zulässig, sofern diese – entsprechend der Regeln guter wissenschaftlicher Praxis – transparent gemacht wird.²¹

Doch was heißt »entsprechend den Regeln der GWP« genau? Über diese Frage diskutierten die Expert*innen im Workshop äußerst rege. Konsens scheint darin zu bestehen, dass der Einsatz von TKI grundsätzlich offenzulegen bzw. ein von TKI stammender Text zu kennzeichnen ist. Diese Schlussfolgerung lässt sich auch aus dem DFG-Kodex ableiten: Leitlinie 1 fordert, »strikte Ehrlichkeit im Hinblick auf die eigenen und die Beiträge Dritter zu wahren«.²² Um dieser Forderung nachzukommen, ist insbesondere Transparenz hinsichtlich der Frage, wie ein Text entstanden ist, notwendig. Transparenz muss dabei auf mehreren Ebenen erfolgen. Die Leitlinien 7, 11, 12 und 13 stellen hierbei weitere Bezugspunkte dar. So erfordert Leitlinie 7 zwecks Qualitätssicherung die Zitation von Originalquellen oder eine Replizierbarkeit; Leitlinie 13 zur Herstellung von

öffentlichen Zugang zu Forschungsergebnissen, dass Ergebnisse »vollständig und nachvollziehbar«²³ zu beschreiben seien. Sie führt hierbei aus: »Dazu gehört es auch, soweit dies möglich und zumutbar ist, die den Ergebnissen zugrunde liegenden Forschungsdaten, Materialien und Informationen, die angewandten Methoden sowie die eingesetzte Software verfügbar zu machen und Arbeitsabläufe umfänglich darzulegen.«²⁴

Leitlinie 11 zu Methoden und Standards fordert hinsichtlich neuer Methoden »besonderen Wert auf die Qualitätssicherung und Etablierung von Standards« zu legen.²⁵ Wie sollten diese Leitlinien bei Nutzung von TKI genau umgesetzt werden? Auf diese Frage gibt es derzeit keine abschließende Antwort, vielmehr viele ungeklärte Fragen und eine Reihe von Vorschlägen. Die Forderung der Leitlinie 1 nach Ehrlichkeit erfordert dabei ein Maximum an Nachvollziehbarkeit des Forschungsprozesses für Rezipient*innen der Ergebnisse. Demgegenüber steht eine unübersichtliche Verschmelzung von TKI-Tools und intellektueller Eigenleistung der Wissenschaftler*innen (wie im Abschnitt zu Autorschaft ausgeführt).

Aus der Leitlinie 12 lässt sich ferner die Dokumentations- und Kennzeichnungspflicht ableiten: »Wissenschaftlerinnen und Wissenschaftler dokumentieren alle für das Zustandekommen eines Forschungsergebnisses relevanten Informationen so nachvollziehbar, wie dies im betroffenen Fachgebiet erforderlich und angemessen ist, um das Ergebnis überprüfen und bewerten zu können.«²⁶

Was dabei auch bedacht werden sollte, ist, dass die Art der angemessenen Offenlegung davon abhängt, wie TKI genutzt wurde. Es gilt also: Die Dokumentationspflicht muss der tatsächlichen Arbeitsweise mit TKI bzw. der tatsächlichen Nutzung von TKI gerecht werden. Die notwendigen Dokumentations- und Offenlegungsanforderungen sollten so ausfallen, dass der Aufwand so gering wie möglich gehalten wird. Unnötig aufwendige Vorgaben gilt es zu vermeiden. Sie hätten eine lähmende Ressourcenverschwendug zur Folge, sei es beispielsweise in Form von überbordendem Zeitaufwand für Forschende oder vermeidbarer Belastung von Infrastrukturen durch lange Supplements. Dabei darf aber auch nicht das notwendige Maß an Nachvollziehbarkeit von Ergebnissen unterlaufen werden. Es wäre sinnvoll, möglichst detaillierten Vorgaben – sofern man sie für zweckmäßig hält – für einzelne Wissenschaftsdisziplinen zu formulieren, fortlaufend zu überprüfen und bei Bedarf zu aktualisieren. Dabei sollten die verschiedenen Möglichkeiten, wie TKI im Forschungsprozess genutzt werden kann, und die verschiedenen Textgattungen mitgedacht werden. Denn sie beeinflussen, welche Informationen für Forschende der jeweiligen Disziplin unerlässlich sind, um Ergebnisse evaluieren zu können. Im Folgenden werden beispielhaft einige offene Fragen aufgegriffen und aufgezeigt, dass für ihre Beantwortung

disziplinspezifische Regelungen notwendig sind.

Zum Ort der Angaben und allgemeine Hinweise zur Nutzung

Da TKI ein Werkzeug darstellt, scheint es angebracht zu sein, Angaben zur Nutzung, wie bei anderer Software auch, in der Sektion der Methoden auszuführen. Denkbar wäre, dass die notwendigen Angaben einen Hinweis auf den Anbieter, das Modell, die Version, den Zeitpunkt der Nutzung und die nutzende Person umfassen. Sollte die Angabe eines genauen Zeitpunktes nicht möglich sein, wäre es etwa auch denkbar, hier einen Zeitraum anzugeben. Die erwähnten funktionellen Angaben sollten ferner um eine allgemeine Angabe darüber, wofür die TKI eingesetzt wurde, ergänzt werden. Aus GWP-Sicht ist insbesondere die Benennung der nutzenden Person sinnvoll, um aufzuschlüsseln, wer Verantwortung für die Steuerung der TKI trug – wobei davon unabhängig auch die Frage zu betrachten ist, in wessen Namen der Einsatz von TKI erfolgte. So wäre etwa zu klären, ob bei gemeinschaftlichen Veröffentlichungen angenommen werden kann, dass alle Autor*innen generell dem Einsatz von TKI zustimmten. Daran anzuschließen wäre die Frage, wie die nutzende Person im CRediT-System anzugeben wäre. Decken die bestehenden Kategorien diese Tätigkeit ab oder ist die Steuerung von TKI vielmehr als eigenständige (neue) Kategorie zu betrachten?

Bei der Frage bezüglich des Ortes der Angaben wird allerdings auch offenbar, dass die Antwort von der Textart abhängig ist: In vielen Veröffentlichungsformaten wie etwa bei geisteswissenschaftlichen Beiträgen ist kein Methodenteil vorgesehen. Wo hier genau die entsprechende Angabe gemacht werden sollte – ob im Vorwort, in der Einleitung, in einer Fußnote oder der Bibliografie – gilt es in disziplinspezifischen bzw. auf verschiedene Textgattungen zugeschnittenen Regelungen zu erläutern. Ob die Sektion der Danksagung ein möglicher Ort für eine Erwähnung wäre, könnte ebenso in Betracht gezogen werden. In diesem Teil einer wissenschaftlichen Veröffentlichung wird insbesondere das Mitwirken von Forschenden offengelegt, welche zum Inhalt einer Veröffentlichung beigetragen haben, deren Beitrag aber keine Autorschaft begründet. Dies ermöglicht einerseits die Zuschreibung von Anerkennung, andererseits von Verantwortung. Die Angabe in dieser Sektion wird durchaus von verschiedenen Seiten vorgeschlagen.²⁷ (Fachspezifische) Diskussionen sollten klären, ob die Erwähnung einer Software in dieser Sektion im Einklang mit den Regeln der GWP steht.

Dokumentation der prompts

Um sich hierbei einer Lösung anzunähern, ist vielleicht grundsätzlich die Unterscheidung von zwei verschiedenen Szenarien sinnvoll: Erstens die Nutzung von TKI für große Teile einer Veröffentlichung durch itera-

tive Einbindung in diverse Schritte des Schreibprozesses; und zweitens eine eher punktuelle, sehr eingegrenzte Nutzung von TKI. Das erstgenannte Szenario soll hier zunächst im Fokus stehen, während auf das zweitgenannte Szenario im nächsten Abschnitt näher eingangen werden soll.

Die Problematik um die Nachvollziehbarkeit von Ideen- und Gedankenbildung/-schärfung oder des Textschreibens ist dabei nicht neu. So wies Katharina Beier darauf hin, dass bislang auch nicht offengelegt wird, wann Forschende welche Fachartikel lesen oder wann sich in Kolloquien Gedankenanstöße ergaben. Letztlich seien diese Dinge auch schwer nachvollziehbar oder abbildbar. Durch die Verwobenheit des »Austauschs« zwischen Forschenden und TKI sind in einem solchen Kontext Angaben von *prompts* im Text in Kombination mit der Zitation der ausgegebenen Antwort schwer vorstellbar. Bei dieser Art der Nutzung könnte es sich vielmehr anbieten, den Ablauf von *prompts* in gesammelter Form abzulegen und somit der erforderlichen Transparenz nachzukommen. Dabei kann hier zwischen einer intern einsehbaren Dokumentation und einer extern einsehbaren Dokumentation unterschieden werden. Forschende forderten etwa in einem Journal der American Chemical Society: »(i) [...] Clearly indicate which parts of the manuscript used the output of the language bot, and provide the prompts and questions, and/or transcript in the Supporting Information«.²⁸

Gerade in Fällen stark iterativer Nutzung mag es jedoch nicht sinnvoll sein, alle *prompts* zugänglich für Dritte offenzulegen. Katharina Beier verwies in diesem Kontext auf die unterschiedliche »intellektuelle Flughöhe der *prompts*«. Womöglich wäre es eine sinnvolle Lösung, nur entscheidende *prompts* in Veröffentlichungen offenzulegen, aber intern eine ausführliche Dokumentation aller *prompts* anzulegen. Guido Juckeland verdeutlichte in diesem Zusammenhang, dass genauere Angaben von Input/Output im Nachhinein eine Überprüfung, beispielsweise mit Blick auf Plagiate, erleichtern könnten. Hjördis Czesnick warf die Frage auf, ob *prompts* als Rohdaten anzusehen seien. Folgt man diesem Gedanken, dann wäre es sinnvoll, über eine Aufbewahrungsfrist nachzudenken. Es kann aber auch sein, dass TKI nur für sehr bestimmte Schritte eingesetzt wurde oder die *prompts* weniger explorativen Charakter haben, als vielmehr eine sehr strukturierte Abfolge von Schritten darstellen, ähnlich wie eine Befehlsabfolge bei Software Codes. In diesem Kontext könnte es wiederum sinnvoll erscheinen, gar als notwendig erachtet werden, alle *prompts* offenzulegen. Hier zeigt sich abermals, dass das angemessene Maß an Dokumentation abhängig ist von den spezifischen Modi der Wissensgenerierung einzelner Fachdisziplinen und Forschungsprojekte. Es sollte auch die Frage geklärt werden, wo die verwendeten *prompts* zu notieren sind. Für viele Textarten mögen sich hierfür die sogenannten Supplements anbieten,

doch, ähnlich zu der Problematik bezüglich des Methodenteils, sind Supplements bzw. Anhänge nicht in allen Textformaten üblich.

In-Text-Zitationen

Neben dieser eher auf kollaboratives Mensch-Maschine-Arbeiten am Text ausgelegten Nutzung ist noch eine weitere Form denkbar. Diese ist gekennzeichnet durch wörtliche Übernahmen aus einer Antwort, welche aus einem oder nur sehr wenigen *prompts* heraus generiert wurde. In diesem Fall kommt zu der Frage nach der Dokumentation der *prompts* die Frage nach den In-Text-Zitationen hinzu, ebenso wie die Frage, welche Angaben im Literaturverzeichnis gemacht werden sollten.

Eine Möglichkeit wäre es, wie die APA-Guidelines aufzeigen, im Text den genauen *prompt* und die gegebene Antwort wiederzugeben unter Verweis auf den Anbieter der Software und das Jahr.²⁹ In der Referenzliste sollte dann zusätzlich der Hinweis auf den Namen der Software, die Version und die Art der Software (large language model) inklusive Link erfolgen, wie das von APA angeführte Beispiel aufzeigt: »OpenAI. (2023). ChatGPT (Mar 14 version) [Large language model]. <https://chat.openai.com/chat>«.³⁰ Die Forschenden Mohammad Hosseini, Lisa M. Rasmussen und David B. Resnick wiesen insbesondere darauf hin, dass in der Liste der Referenzen auch angegeben werden sollte, wer den *prompt* gab. Ein Beispiel sähe folgendermaßen aus: »OpenAI chatbot. 2022b. Response to Query Made by David B Resnik, December 11, 2022, 10:48pm EST. [Google Scholar]«.³¹

Demgegenüber halten die Empfehlungen des ICMJE fest: »Authors should not list AI and AI-assisted technologies as an author or co-author, nor cite AI as an author«.³² Dies verdeutlicht den noch fehlenden Konsens in dieser Frage sowie den Bedarf an disziplinspezifischen Lösungen. Erschwert wird die Frage nach der Zitation TKI-generierter Inhalte durch die stochastiche Funktionsweise von TKI. Die Reproduktion der Ergebnisse ist aufgrund der nach probabilistischen Mustern erzeugten Texte ausgeschlossen, selbst wenn die genauen *prompts* bekannt sind. Folglich sind Texte von TKI im strengen Sinne auch nicht zitierfähig, wie Debora Weber-Wulff hervorhob.

GWP-Abweichungen bzw. Täuschung

Für Hannes Bajohr ist Täuschung integraler Bestandteil von TKI: »Zugespitzt gesagt: *Das Wesen von KI ist es, artifizielle Texte als natürliche Texte auszugeben.*«³³ Das funktioniere überhaupt nur aufgrund der Standarderwartung an unbekannte Texte, die stets menschliche Urheberschaft annehme. »Künstliche Intelligenz basiert also von Anfang an auf dem Prinzip der Täuschung – und sie muss es«; denn die fragliche Intelligenz werde am subjektiven Eindruck nicht an objektiven Eigen-

schaften des Systems festgemacht.³⁴ Die transparente Offenlegung der Nutzung von TKI und das Verständnis von Abweichungen von der GWP bedingen also einander unmittelbar: Die unzureichend gekennzeichnete Verwendung von TKI verstößt gegen die GWP. Denn der Einsatz von TKI ohne hinreichende Kenntlichmachung stellt eine Täuschung der Rezipient*innen dar. *Der Europäische Verhaltenskodex für Integrität in der Forschung* stuft in seiner neuen Fassung vom Juni 2023 entsprechend »[h]iding the use of AI or automated tools in the creation of content or drafting of publications«³⁵ zwar nicht als wissenschaftliches Fehlverhalten im engeren Sinne (FFP) ein, aber führt sie dennoch als nicht akzeptable Forschungspraxis an. Es wäre zu überprüfen, ob neben Auslegungen von bestehenden GWP-Regeln auch zusätzlich Ergänzungen zur Definition und zum Verständnis von wissenschaftlichem Fehlverhalten bzw. fragwürdiger Forschungspraxis notwendig sind, also nicht nur der zulässige Einsatz von TKI definiert werden sollte, sondern ebenso wann der Einsatz unzulässig ist. Auch in diesem Kontext könnten ggf. Auslegungen der bestehenden Definitionen helfen. Die Schwere der Täuschung und somit des Regelverstoßes kann dabei sicherlich unterschiedlich ausfallen, je nachdem, ob etwa einzelne Angaben fehlen, in grob fahrlässiger Weise fehlerhafte Inhalte in Umlauf gebracht werden oder Kompetenzen, Leistungen oder Produktivität vorgetäuscht werden. Diskussionen sind hier erforderlich, um zu erörtern, was etwa als geringfügige Täuschung, was als besonders schwerwiegende Täuschung angesehen werden sollte. Dabei sollte auch der entstandene Schaden und die jeweiligen Adressat*innen der Täuschung beachtet werden: Lesende, Reviewer*innen, Forschungsförderer, weitere Akteur*innen aus der Gesellschaft, Prüfungskommissionen.

Unklar ist auch, wie eine unzureichende Kenntlichmachung nachgewiesen werden sollte. Ein gesicherter Nachweis von Regelverstößen ist die Grundlage für Sanktionierung. Der Einsatz von TKI ist kaum nachweisbar, wie etwa Debora Weber-Wulff betont.³⁶ OpenAI etwa hatte längere Zeit ein Detektionstool beworben, dieses jedoch im Juli 2023 aus Gründen der Zuverlässigkeit vorerst zurückgezogen.³⁷

Neben den Fragen zu GWP-Abweichungen im Kontext von Autorschaft und mangelnder Offenlegung der Nutzung gibt es noch einen ganz anderen Bereich der GWP, in dem die Nutzung von TKI sehr kritisch zu sehen ist: das Formulieren von Gutachten etwa im Zuge der Peer Review. Eingereichte Manuskripte und Förderanträge etwa enthalten vertrauliche Informationen. Es ist sehr fraglich, ob der Schutz dieser Daten bei der Nutzung von TKI gewährleistet werden kann. Das National Institute of Health in den USA etwa stellt fest, dass die Nutzung von TKI eine Verletzung der von den Regeln der GWP geforderten Vertraulichkeit darstellt, und hat den Einsatz von TKI beim Begutachten von

Förderanträgen entsprechend untersagt.³⁸ Es bleibt zu klären, ob auch in Deutschland der Einsatz von TKI für Gutachten unterbunden werden sollte bzw. in welchen Fällen er ggf. für zulässig erachtet wird.³⁹

Die hohen Verlässlichkeitsstandards der Wissenschaft werden also durch die Anwendung von TKI herausgefordert. Steffen Albrecht notiert in seiner Studie für das TAB zutreffend, die Anwendung sei »nur unter strikter – und aufwendiger – Kontrolle möglich«.⁴⁰ Die TKI erfindet manchmal Tatsachen oder Quellen, macht unpräzise Angaben und liefert aufgrund ihrer Trainingsdaten verzerrte, vorurteilsbeladene Ergebnisse. Genau hier schließt sich die Verantwortung von Autor*innen wissenschaftlicher Veröffentlichungen an. Die Verantwortung für den Inhalt können Forschende nicht an eine TKI auslagern. Es ist umso wichtiger, dass Forschende darüber informiert sind, dass TKI nicht faktenfrei arbeitet, sondern Texte durch stochastische Prozesse generiert. Neuere Entwicklungen im Bereich KI könnten hier auch andere Fragen mit sich bringen, nämlich wenn TKI kombiniert wird mit weiteren Tools zum Faktencheck, etwa durch Schnittstellen zu Literaturdatenbanken. Bislang ist aber festzuhalten: Werden von einer TKI angegebene Quellen, Zitate, Fakten oder Wissensstände nicht überprüft, entspricht dieses Verhalten nicht den Regeln der GWP, ist also als ein Verstoß gegen eben diese zu werten. Die Rolle des Steuerns zu übernehmen, heißt auch, diese Dinge zu überprüfen.

Albrecht betont neben diesen Problemen drei Risiken speziell für den Wissenschaftsbetrieb. So nennt er erstens die »fehlende Zuordnung von Informationen zu einer Quelle«.⁴¹ Die Urheberschaft bleibe unklar (Wer hat einen Text bzw. die Texte geschrieben, auf welche die TKI zurückgreift?) und es finde keine Kontextualisierung im Forschungsdiskurs statt (Wo ist ein Text, ein Gedanke einzuordnen?). Fraglich ist auch, ob Anwender*innen TKI mit Sachverstand und Expertise verwenden und die eingangs geforderte steuernde Rolle einnehmen, oder aber TKI gerade dazu nutzen, um über fehlende Expertise hinwegzutäuschen und der eigentlich erforderlichen Einarbeitung in Themen nicht nachkommen. Wenn TKI nicht adäquat gesteuert wird, kann sich das in qualitativ minderwertigen Artikeln niederschlagen.

Zudem führt Albrecht die Beschleunigung des Publikationsprozesses zu Ungunsten sorgfältiger Prüfung an; dadurch könnte TKI die Aufmerksamkeit auf schon bestehende Unzulänglichkeiten im Betrieb lenken, doch abseits dieser Hoffnung werde wohl ein negativer Effekt eintreten, nämlich die Verschärfung von Arbeitslast und Konkurrenzdruck durch den verbreiteten Einsatz von KI, wodurch die Nutzung von KI attraktiver werde (und damit ein sich selbst verstärkender Prozess eintritt). Weitere risikobehaftete Veränderungen für das wissenschaftliche Publikationswesen sind denkbar: Ba-johr vermutet eine baldige »Flut artifizieller Texte«.⁴²

Auf die Wissenschaft gemünzt dürfte eine enorme Zunahme an Textproduktion diverse Akteure überfordern: beispielsweise hätten Fachzeitschriften und Verlage mehr Einreichungen zu sichtern, bei Forschungsförderorganisationen müsste eine größere Anzahl an Anträgen begutachtet werden und Paper Mills könnten aufgrund der Leistungsfähigkeit aktueller TKI mehr glaubwürdig anmutende Texte produzieren. Dies bedeutet im Umkehrschluss auch, dass Forschende sich mit größerer Achtsamkeit durch eine größere Anzahl an Forschungsliteratur arbeiten müssten. Schon jetzt heißt es oft, dass mehr geschrieben wird, als gelesen werden kann. Mit Hilfe von TKI kann sich dieses Ungleichgewicht weiter vergrößern, wie auch in dem oben genannten Workshop thematisiert wurde. Der Produktivitätsgewinn durch TKI würde dadurch geschrägt, oder es käme zwangsläufig auch zum Einsatz von KI zur Sichtung von Literatur, um der »Flut« gerecht zu werden. Dies könnte weitere Probleme mit sich bringen, wenn beispielsweise nicht nachvollziehbar ist, wie der Selektionsprozess der eingesetzten KI funktioniert. Welche Literatur rezipiert wird, wird dann teilweise von den Algorithmen der KIs abhängig. Wie weit Forschende diese Entscheidungsprozesse aus der Hand geben wollen und dabei gleichzeitig GWP-konform handeln, muss zukünftig ausgehandelt werden.

Ferner bemängelt Albrecht die »fehlende Faktenorientierung der KI-Modelle«,⁴³ die auch risikobehaftet für das Verhältnis von Wissenschaft und Öffentlichkeit sei; wissenschaftlich klingende Texte könnten die Desinformation vorantreiben und Wissenschaftler*innen selbst könnten plausibel klingenden, aber unwissenschaftlichen Texten aufsitzen. Letztlich spricht er hier eins der größten Risiken an, welches die Nutzung von TKI mit sich bringt: den Verlust des Vertrauens in wissenschaftliche Ergebnisse und den wissenschaftlichen Prozess. Vertrauen ist eines der höchsten Güter in der Wissenschaft. Ohne die Grundannahme von Vertrauen funktioniert Wissenschaft nicht. Dieser besonderen Verantwortung müssen sich Forschende bewusst sein. Gleichzeitig spielt auch das Vertrauen der Gesellschaft in die Wissenschaft hier eine Rolle. Neben der Sicherstellung, dass der Umgang mit TKI GWP-konform ist, sollte aber auch der Aspekt der Fairness in Betracht gezogen werden. Blickt man mit diesem Filter auf TKI in der Forschung ergeben sich Fragen, wer welchen Zugang zu TKI (in Zukunft) hat und wer am meisten von ihrem Einsatz profitiert.

TKI – Zugang und Infrastrukturen

Manchmal wird in Programmen wie ChatGPT die Chance gesehen, dass TKI Forschenden überall zur Verfügung steht und damit gar eine Art Demokratisierung der Wissenschaft bewirken könnte. Dabei ist fraglich, ob TKI auf lange Sicht wirklich disruptiv wirken kann oder sich vielmehr in bestehende Infrastrukturen und Mecha-

nismen integriert. Denn Zugangsfragen und Infrastrukturen sind seit Langem ein kontroverses Thema in der Wissenschaft, dem regelmäßig Probleme und Handlungsbedarf nachgesagt wird, Verbesserungen im Vergleich zur Wichtigkeit aber nur langsam implementiert werden. Dies betrifft insbesondere das wissenschaftliche Publikationssystem, in dem sich wissenschaftliche Werte u. a. mit einigen sehr großen und profitorientierten Unternehmen diametral gegenüberstehen. Obwohl viele Forschende dem Geschäftsmodell und den Praktiken der großen Verlage (Elsevier, Springer und Wiley beherrschen zusammen 50 % des Marktes) kritisch gegenüberstehen, können sie sich dem System nicht gänzlich entziehen, zu groß ist die Sorge, dadurch im wissenschaftlichen Wettbewerb benachteiligt zu sein. Open Access galt einst als großer disruptiver Hoffnungsträger gegen die Verlagsmacht, wurde aber mittlerweile von den großen Verlagen vereinnahmt und nach eigenen Regeln umgesetzt. Hinzu kommen die von den Verlagen erstellten Plattformen, die den Zugang zu Artikeln, die Einreichung von Manuskripten sowie die Verknüpfung mit anderen Research Tools ermöglichen. Das macht es für Forschende noch schwerer, die Nutzung zu verweigern, obgleich hinlänglich bekannt ist, dass die großen Verlage diese Plattformen für die Sammlung von Nutzerdaten verwenden, ein Geschäft mit dem zum Beispiel Elseviers Mutterunternehmen RELX seinen Hauptumsatz macht. Der Einsatz und die Nutzung von TKI durch Forschende geschieht also im Kontext bereits bestehender Strukturen: Sie stellt einerseits einen neuen Faktor dar, deren Gebrauch aber andererseits nicht unabhängig von etablierten Mechanismen betrachtet werden kann.

TKI ist damit ein Werkzeug, bei dem nicht nur der Einsatz geregelt werden muss, sondern auch der Zugang. Schon jetzt existiert neben der kostenfreien Version von ChatGPT eine Bezahlvariante, die den Abonennt*innen zeitlich uneingeschränkten Zugang und bestimmte Features bietet. Mit der Einführung von GPT-4 wandte sich die Entwicklerfirma außerdem von ihrem Open-Source-Ansatz ab, der damit nur noch im Namen OpenAI existiert. Die Zugänglichkeit kann sich also schnell ändern, wenn sie vollständig in der Hand von kommerziellen Firmen liegt.⁴⁴ Zugangsbeschränkungen konterkarieren die Idee, TKI trage zu einer Demokratisierung in der Wissenschaft bei. Denn schon jetzt ist klar, dass die (Weiter-)Entwicklung potenter Sprachmodelle den Einsatz enormer Ressourcen bedarf, der nur von wenigen Akteuren geleistet werden kann.⁴⁵ Es stellt sich die Frage, ob eine vergleichbare TKI, deren Entwicklung und Betrieb in öffentlicher Hand liegt, möglich ist.

Auch wenn der jetzige Preis der Bezahlvariante von monatlich 20 US-Dollar für viele erschwinglich ist, lässt sich mutmaßen, dass sich – analog zum wissenschaftlichen Publikationssystem – eine Struktur entwickelt, bei der das Budget von Institutionen darüber entschei-

det, wer welchen Zugang zu welchen TKI-Werkzeugen bekommt. Anstatt Ungleichheiten abzubauen, würde TKI sie weiter vergrößern. Dies beträfe dann sowohl Forschende verschiedener Forschungseinrichtungen innerhalb eines Landes, aber auch global gesehen verschiedene Länder untereinander. Länder des Globalen Südens, die generell geringere Forschungsbudgets haben, vulnerablere Energieinfrastrukturen und weniger in die Entwicklung von TKIs einbezogen sind, könnten im internationalen Wettbewerb dadurch größere Nachteile erleiden – insbesondere je größer der positive Einfluss von TKI auf Produktivität und Output im Globalen Norden wird. Der oft genannte Vorteil von TKI für Nichtmuttersprachler*innen, ihre englischsprachigen Manuskripte publikationsreif zu bekommen, kann nicht eingelöst werden, wenn der Zugang zur TKI nicht gegeben ist.

Manche zukünftigen Entwicklungen lassen sich also antizipieren, das heißt jedoch nicht, dass jetzt schon abzusehen ist, wie die besten Gegenstrategien aussehen könnten. Es ist zu erwarten, dass Universitäten und Forschungseinrichtungen versuchen werden, ihren Forschenden den Zugang zu den besten TKIs zur Verfügung zu stellen, auch weil es in ihrem eigenen Interesse ist. Denkbar ist, dass sich in Zukunft ähnliche Machtkämpfe entspinnen, wie es diese zurzeit zwischen Bibliotheken und Universitäten auf der einen und den großen Verlagskonzernen auf der anderen Seite gibt. Dieser Fall tritt insbesondere dann ein, wenn sich auf dem Feld der KI ebenso einige wenige große Player bilden, die aufgrund ihrer Marktmacht die Preise bestimmen können. Vor Kurzem hat Elsevier seine erste eigene KI, ScopusAI, angekündigt. Da diese KI höchstwahrscheinlich in Zukunft Zugriff auf die gesamte bei Elsevier veröffentlichte Forschungsliteratur haben wird, dürfte sie für Forschende nochmal eine andere Relevanz haben als derzeit ChatGPT. KIs könnten damit Teil der Verlagsangebote werden und somit für die Verlage lukrativer und für Universitäten begehrlicher. Dies könnte es für Universitäten umso schwieriger machen, keine Verträge mit den großen Verlagen abzuschließen. Mit der starken Zunahme von Open-Access-Publikationen und Preprints wäre eine offene TKI-Alternative denkbar, aber trotzdem, wie bereits angesprochen, teuer in der Umsetzung. Zumal der Vorsprung, den kommerzielle Anbieter bereits haben, schwer einzuholen wäre.

Neben strukturellen Strategien, kann an das individuelle Verhalten von Forschenden appelliert werden, vergleichbar mit dem Elsevier-Bojkott. Das Abwälzen der Verantwortung auf Forschende als Individuen ist jedoch problematisch, wenn die Rahmenbedingungen dieses Verhalten nicht belohnen oder es gar bestrafen. Dabei ist im Kontext von KI gar nicht die Nutzung per se kritisch zu bewerten, sondern die Konsequenzen ungleich verteilter Zugänge. Strukturell müsste also sichergestellt werden, dass die Forschenden, die TKI nutzen

wollen, diese auch in dem Umfang nutzen können, wie es ihr Forschungsvorhaben bedarf. Ansonsten wird es zu Ausschlüssen kommen, die möglicherweise stellenweise durch Schattenangebote überbrückt werden (wie das heutzutage mit Sci-Hub und ähnlichen Schattenbibliotheken der Fall ist). Das Unwohlsein, dass Forschende proprietäre Software nutzen (müssen), weil es keine offenen Alternativen gibt, und dies somit zu den Konditionen der Firmen tun müssen, spiegelt sich in den Datensammelpraktiken der großen Verlage. Dass die Macht der großen Verlage so schwer zu brechen ist, liegt auch daran, dass sie so eng mit dem wissenschaftlichen System verwoben sind, sodass Metriken, die ursprünglich zur Auswahl von Journals entwickelt wurden, mittlerweile zwar verhasste, aber trotzdem etablierte Indikatoren für wissenschaftlichen Erfolg geworden sind. Wie schwer es ist, sich aller Kritik zum Trotz von diesem System zu lösen, lässt sich mit Blick auf die langanhaltende Debatte gut nachvollziehen. Es wäre wünschenswert, früh genug über Fragen von Zugang und Infrastruktur in Bezug auf KI nachzudenken, um ähnliche Abhängigkeiten in Zukunft zu vermeiden.

Anmerkungen

- 1 Die Verfasser*innen nutzen in diesem Artikel die Abkürzung TKI für textgenerierende künstliche Intelligenz, da sie sich auf die Technologie der großen Sprachmodelle (large language models) konzentrieren und insbesondere deren Funktion, auf Grundlage von sogenannten *prompts* Text zu erstellen. Andere Technologien, die unter den Begriff der künstlichen Intelligenz gefasst werden und in der Wissenschaft zum Einsatz kommen (können), sollen – auch wenn hierzu unbestritten großer Diskussionsbedarf besteht – nicht im Fokus stehen.
- 2 Madiega, T. *Artificial Intelligence Act*, 2023.
- 3 Albrecht, S. *ChatGPT und andere Computermodelle zur Sprachverarbeitung – Grundlagen, Anwendungspotenziale und mögliche Auswirkungen*. Büro für Technikfolgen-Abschätzung beim Deutschen Bundestag (TAB), 2023.
- 4 Leschke, J. und P. Salden. *Didaktische und rechtliche Perspektiven auf KI-gestütztes Schreiben in der Hochschulbildung*. Ruhr-Universität Bochum, 2023.
- 5 Siehe beispielsweise: Nature. Tools such as ChatGPT threaten transparent science; here are our ground rules for their use [online]. *Nature*, 2023, 613(7945), 612. Verfügbar unter: doi:10.1038/d41586-023-00191-1
- 6 Das zeigt z.B. die Diskussion ab Minute 30:34 der online abrufbaren Aufzeichnung: <https://youtu.be/kACFINu110Y?t=1834>
- 7 Ein Kurzbericht zur Veranstaltung mit weiteren Informationen ist einsehbar unter: <https://ombudsman-fuer-die-wissenschaft.de/11596/workshop-textgenerierende-ki-und-gwp/>. Die Verfasser*innen danken den geladenen Expert*innen Katharina Beier, Iryna Gurevych, Guido Juckeland, Kirsten Hüttemann, Nadine Lordick, Doris Weßels und Debora Weber-Wulff für die sehr rege Diskussion, welche dazu beigetragen hat, Gedanken zu schärfen, und den vorliegenden Beitrag mitgeprägt hat. Einige Ideen und Thesen von ihnen flossen auch dezidiert in den Text ein. In den betreffenden Fällen erfolgt eine Namensnennung im Text, jedoch keine weitere Referenz.

- 8 Stokel-Walker, C. ChatGPT listed as author on research papers: many scientists disapprove [online]. *Nature*. Verfügbar unter: https://www.nature.com/articles/d41586-023-00107-z?utm_medium=Social&utm_campaign=nature&utm_source=Twitter#Echobox=1674075390
- 9 Beta Writer. *Lithium-ion batteries. A machine-generated summary of current research*. Cham: Springer, 2019. Einen interessanten Einblick in die Genese dieser Meta-Studie stellt Jan Schwenkenbecher in seinem Beitrag »Die Schreib-Maschine. KI als Wissenschaftsautor« zusammen (In: *Forschung Frankfurt*, 2020, (1). Verfügbar unter: <https://www.forschung-frankfurt.uni-frankfurt.de/89891513.pdf>).
- 10 Fecher, B., M. Hebing, M. Laufer, J. Pohle und F. Sofsky. *Delphi Study: Exploring the Implications of Large Language Models on the Science System*. Zenodo, 2023.
- 11 Deutscher Ethikrat. *Mensch und Maschine – Herausforderungen durch Künstliche Intelligenz*, 2023.
- 12 Hannes Bajohr etwa hält hierzu fest: »Obwohl sie noch immer nicht intelligent sind – sie verstehen nicht tatsächlich, was sie tun –, vermitteln die allerneuesten KI-Modelle mehr denn je den Eindruck von Intelligenz. Dieses Erscheinen-Als betrifft die Art und Weise, wie Beobachterinnen Ausgaben interpretieren, wie sie ihnen gegenüberstehen und von ihnen auf das dahinterstehende System zurückschließen.« (Bajohr, H. Artifizielle und postartifizielle Texte. Über die Auswirkungen Künstlicher Intelligenz auf die Erwartungen an literarisches und nichtliterarisches Schreiben. *Sprache im technischen Zeitalter*, 2023, 61(245)).
- 13 Hosseini, M., D. B. Resnik und K. Holmes. The ethics of disclosing the use of artificial intelligence tools in writing scholarly manuscripts [online]. *Research Ethics*, 2023. Verfügbar unter: doi:10.1177/17470161231180449
- 14 COPE: Committee on Publication Ethics. *Authorship and AI tools* [online]. 1 August 2023 [Zugriff am: 1. August 2023]. Verfügbar unter: <https://publicationethics.org/cope-position-statements/ai-author>
- 15 Hoeren, T. Rechtsgutachten zum Umgang mit KI-Software im Hochschulkontext. In: *Didaktische und rechtliche Perspektiven auf KI-gestütztes Schreiben in der Hochschulbildung*. Ruhr-Universität Bochum, 2023, S. 26–27.
- 16 Salden, P., N. Lordick und M. Wiethoff. KI-basierte Schreibwerkzeuge in der Hochschule: Eine Einführung. In: *Didaktische und rechtliche Perspektiven auf KI-gestütztes Schreiben in der Hochschulbildung*. Ruhr-Universität Bochum, 2023, S. 10–11.
- 17 Barlösius, E., F. Knoke und M. Pook-Kolb. Was sind wissenschaftliche Eigenleistungen – während des Forschungsprozesses und nach der Publikation der Ergebnisse? In: L. Akremi, N. Baur, H. Knoblauch und B. Traue, Hg. *Handbuch Interpretativ forschen*. Weinheim: Beltz, 2018, S. 133–160.
- 18 Weßels, D. und O. Gottschalk. *Hochschullehre unter dem Einfluss des KI-gestützten Schreibens* [online], 2022. 15 Juli 2022 [Zugriff am: 7. September 2023]. Verfügbar unter: [https://wame.org/page3.php?id=106](https://hochschulforumdigitalisierung.de/de/blog/Hochschullehre-KI-gestuetztes-Schreiben. Eine abweichende Position wird von Jenkins und Lin (2023) vertreten; sie fordern, dass large language models, wenn sie wesentliche Beiträge zu einer Publikation leisten, als Autoren genannt werden sollten, da die Anerkennung sonst in unangemessener Weise auf die menschlichen Autor*innen entfiel (vgl. Hosseini 2023).</p>
<p>19 Bezogen auf die Hochschullehre gibt es noch viel Unklarheit zu dieser Frage: Die Forderungen reichen vom generellen Ausschluss von KI über eine Nutzung unter bestimmten Bedingungen bis hin zu der aktiven Förderung der Kompetenz, wie man KI zielgerichtet nutzt und welche Gefahren sie bergen kann.</p>
<p>20 Kritisch ist jedoch der Einsatz von KI in Peer-Review-Prozessen oder bei der Erstellung von Gutachten bei der Fördermittelvergabe zu sehen – etwa wegen des Aspektes der Vertraulichkeit der Informationen in den eingereichten Schriftstücken.</p>
<p>21 Albrecht, S. <i>ChatGPT und andere Computermodelle zur Sprachverarbeitung – Grundlagen, Anwendungspotenziale und mögliche Auswirkungen</i>. Büro für Technikfolgen-Abschätzung beim Deutschen Bundestag (TAB), 2023. S. 79.</p>
<p>22 Deutsche Forschungsgemeinschaft. <i>Leitlinien zur Sicherung guter wissenschaftlicher Praxis</i>. Bonn, 2019.</p>
<p>23 Ebd. S. 19.</p>
<p>24 Ebd. S. 19.</p>
<p>25 Ebd. S. 17.</p>
<p>26 Ebd. S. 17–18. Übrigens hält hierzu auch der europäische Verhaltenskodex der All European Academies fest, dass Forschende die Nutzung von KI offenlegen sollen, erkennt aber auch an, dass dies unter Berücksichtigung fachspezifischer Normen erfolgen müsse: »Researchers report their results and methods, including the use of external services or AI and automated tools, in a way that is compatible with the accepted norms of the discipline and facilitates verification or replication, where applicable.« ALLEA – All European Academies. <i>The European Code of Conduct for Research Integrity</i>: ALLEA – All European Academies, 2023, S. 7.</p>
<p>27 Siehe zum Beispiel: Nature. Tools such as ChatGPT threaten transparent science; here are our ground rules for their use [online]. <i>Nature</i>, 2023, 613(7945), 612. Verfügbar unter: doi:10.1038/d41586-023-00191-1#, oder WAME. <i>Chatbots, Generative AI, and Scholarly Manuscripts II</i> WAME [online]. 1 August 2023 [Zugriff am: 1. August 2023]. Verfügbar unter: <a href=)
- 28 Buriak, J. M., D. Akinwande, N. Artzi, C. J. Brinker, C. Burrows, W. C. W. Chan, C. Chen, X. Chen, M. Chhowalla, L. Chi, W. Chueh, C. M. Cradden, D. Di Carlo, S. C. Glotzer, M. C. Hersam, D. Ho, T. Y. Hu, J. Huang, A. Javey, P. V. Kamat, I.-D. Kim, N. A. Kotov, T. R. Lee, Y. H. Lee, Y. Li, L. M. Liz-Marzán, P. Mulvaney, P. Narang, P. Nordlander, R. Oklu, W. J. Parak, A. L. Rogach, M. Salanne, P. Samori, R.E. Schaak, K. S. Schanze, T. Sekitani, S. Skrabalak, A. K. Sood, I. K. Voets, S. Wang, S. Wang, A. T. S. Wee und J. Ye. Best Practices for Using AI When Writing Scientific Manuscripts [online]. *ACS Nano*, 2023, 17(5), 4091–4093. Verfügbar unter: doi:10.1021/acsnano.3c01544, S. 4092. Ganz ähnlich heißt es in den Empfehlungen der World Association of Medical Editors: »[...] all prompts used to generate new text, or to convert text or text prompts into tables or illustrations, should be specified« (WAME. *Chatbots, Generative AI, and Scholarly Manuscripts II* WAME [online]. 1 August 2023 [Zugriff am: 1. August 2023]. Verfügbar unter: <https://wame.org/page3.php?id=106>).
- 29 Siehe APA Guidelines; verfügbar unter: <https://apastyle.apa.org/blog/how-to-cite-chatgpt>
- 30 Ebd.
- 31 Hosseini, M., L. M. Rasmussen und D. B. Resnik. Using AI to write scholarly publications [online]. *Accountability in Research*, 2023, 1–9. Verfügbar unter: doi:10.1080/08989621.2023.2168535
- 32 ICMJE. *Defining the Role of Authors and Contributors* [online]. Verfügbar unter: <https://www.icmje.org/recommendations/browse/roles-and-responsibilities/defining-the-role-of-authors-and-contributors.html>
- 33 Bajohr, H. Artifizielle und postartifizielle Texte. Über die Auswirkungen Künstlicher Intelligenz auf die Erwartungen an literarisches und nichtliterarisches Schreiben. *Sprache im technischen Zeitalter*, 2023, 61(245). S. 42; Hervorhebung im Original.

- 34 Ebd. Der hier zitierte Text Bajohrs befasst sich mit literarischen Texten, doch enthält einige Überlegungen, die ebenfalls auf wissenschaftliche Texte anzuwenden wären. Ein Beispiel ist der folgende, zentrale Gedanke Bajohrs: »Die Standarderwartung an Texte wird sich also auf kurz oder lang verschieben – von der Überzeugung, ein Mensch stehe dahinter, zum Zweifel, ob es nicht doch eine Maschine sein könnte. Damit aber wird auch die Unterscheidung zwischen natürlichen und artifiziellen Texten zusehends hinfällig. Wir würden dann womöglich in eine Phase postartifizieller Texte übergehen« (ebd., S. 50). Ob eine solche Phase auch in der Wissenschaft eintreten könnte und was sie zur Folge hätte, scheint doch überaus bedenkenswert.
- 35 ALLEA – All European Academies. *The European Code of Conduct for Research Integrity*: ALLEA – All European Academies, 2023, S. 10.
- 36 In einem Preprint hat sie gemeinsam mit anderen Forschenden untersucht, wie zuverlässig der Einsatz von TKI durch Detektionssoftware nachweisbar ist: Weber-Wulff, D., A. Anohina-Naumeca, S. Bjelobaba, T. Foltynek, J. Guerrero-Dib, O. Popoola, P. Šigut und L. Waddington. *Testing of Detection Tools for AI-Generated Text*, 21. Juni 2023.
- 37 Die entsprechende Mitteilung ist hier einsehbar: <https://openai.com/blog/new-ai-classifier-for-indicating-ai-written-text>
- 38 Lauer, M., S. Constant und A. Wernimont. *Using AI in Peer Review Is a Breach of Confidentiality* [online]. 20 September 2023 [Zugriff am: 20. September 2023]. Verfügbar unter: <https://nexus.od.nih.gov/all/2023/06/23/using-ai-in-peer-review-is-a-breach-of-confidentiality/>
- 39 Die DFG als wichtigste Forschungsförderin in Deutschland sieht hierzu auch Klärungsbedarf und erarbeitet zurzeit ein Positionspapier.
- 40 Albrecht, S. *ChatGPT und andere Computermodelle zur Sprachverarbeitung – Grundlagen, Anwendungspotenziale und mögliche Auswirkungen*. Büro für Technikfolgen-Abschätzung beim Deutschen Bundestag (TAB), 2023, S. 80.
- 41 Albrecht, S. *ChatGPT und andere Computermodelle zur Sprachverarbeitung – Grundlagen, Anwendungspotenziale und mögliche Auswirkungen*. Büro für Technikfolgen-Abschätzung beim Deutschen Bundestag (TAB), 2023. Hervorhebung im Original.
- 42 Bajohr, H. Artifizielle und postartifizielle Texte. Über die Auswirkungen Künstlicher Intelligenz auf die Erwartungen an literarisches und nichtliterarisches Schreiben. *Sprache im technischen Zeitalter*, 2023, 61(245), S. 47. Hervorhebung im Original.
- 43 Albrecht, S. *ChatGPT und andere Computermodelle zur Sprachverarbeitung – Grundlagen, Anwendungspotenziale und mögliche Auswirkungen*. Büro für Technikfolgen-Abschätzung beim Deutschen Bundestag (TAB), 2023. Hervorhebung im Original.
- 44 OpenAI begann als Non-Profit Organisation, die sich der freien Kollaborationen mit anderen verschrieb. Auf der Website heißt es unter »Introducing OpenAI«: »OpenAI is a non-profit artificial intelligence research company. Our goal is to advance digital intelligence in the way that is most likely to benefit humanity as a whole, unconstrained by a need to generate financial return«. Verfügbar unter: <https://openai.com/blog/introducing-openai>. Mittlerweile gab es eine Abkehr vom Non-Profit, was zeitgleich mit einer Zusammenarbeit mit Microsoft einherging.
- 45 Auch die Nutzung von KI verbraucht gewaltige Mengen an Ressourcen, sodass selbst beim Einsatz der kostenfreien Varianten die Frage aufkommt, wer die Kosten eigentlich trägt (Strubell, E., A. Ganesh und A. McCallum. *Energy and Policy Considerations for Deep Learning in NLP*, 5. Juni 2019).

Verfasser*innen



Dr. Katrin Frisch, Wissenschaftliche Referentin
»Dialogforen zur Sicherung wissenschaftlicher Integrität«, Schwerpunkt: Forschungsdaten, Ombudsman für die Wissenschaft, Jägerstraße 22–23, 10117 Berlin, Telefon +49 30 20370259, katrin.frisch@ofdw.de
Foto: privat



Dr. Felix Hagenström, Referendar an der Staatsbibliothek zu Berlin – Preußischer Kulturbesitz, Potsdamer Straße 33, 10785 Berlin, felix.hagenstroem@sbb.spk-berlin.de
Telefon +49 30 266 432-648

Foto: privat



Dr. Nele Reeg, Wissenschaftliche Referentin
»Dialogforen zur Stärkung wissenschaftlicher Integrität«, Schwerpunkt: Autorschaften, Ombudsman für die Wissenschaft, Jägerstraße 22–23, 10117 Berlin, Telefon +49 20370259, nele.reeg@ofdw.de
Foto: MK Photo Studio