Katharina Miller, Milena Valeva, Julia Prieß-Buchheit (Hrsg.)

Verlässliche Wissenschaft

Dieses Projekt wurde finanziert im Rahmen des Horizon 2020 Forschungs- und Innovationprogramms der Europäischen Union unter der Finanzhilfevereinbarung Nr. 824488.

Katharina Miller, Milena Valeva, Julia Prieß-Buchheit (Hrsg.)

Verlässliche Wissenschaft

Bedingungen, Analysen, Reflexionen

wbg Academic

Der folgende Text ist ein vorabgedruckter Auszug aus dem Band „Verlässliche Wissenschaft – Bedingungen, Analysen, Reflexionen“ von Katharina Miller, Milena Valeva und Julia Prieß-Buchheit (Hrsg.), der im Programm der wbg erscheinen wird (ISBN 978-3-534-40607-4).

Forschungsintegrität und Künstliche Intelligenz mit Fokus auf den wissenschaftlichen Schreibprozess

Traditionelle Werte auf dem Prüfstand für eine neue Ära

Nicolaus Wilder, Doris Weßels, Johanna Gröpler, Andrea Klein & Margret Mundorf

Einleitung

Dass Künstliche Intelligenz (KI) die traditionelle Art und Weise, wie Wissenschaft praktiziert wird, in ihren Grundfesten verändern wird, ist nicht erst seit der Lösung des seit etwa 50 Jahren im Zentrum biochemischen Rätselns stehenden Problems der Proteinfaltung durch DeepMinds Künstliche Intelligenz AlphaFold 2.0 offenkundig, der zugesprochen wird, die Biologie von Grund auf zu transformieren (Callaway, 2020). Die Diskussion um die Bedeutung, die KI für wissenschaftliche Erkenntnisprozesse hat, hat eine lange Tradition (The Royal Society, 2017; The Royal Society & The Alan Turing Institute, 2019). Was sich aber besonders in den letzten Jahren verändert hat, ist die Präsenz KI-basierter Anwendungen (Tahiru, 2021), die sich in den (akademischen) Alltag nahezu unbemerkt eingeschlichen hat, sowie deren Qualität, die die Ergebnisse mitunter nicht mehr als Erzeugnisse einer Künstlichen Intelligenz erkennen lässt. Besonders deutlich wird dies am Beispiel des aktuellen und hoch kontroversen Diskurses um GPT-3 von OpenAI, einem KI-basierten Sprachprozessor, der mühelos Texte generiert, die die MIT Technology Review als „shockingly good – and completely mindless" (Heaven, 2020) bezeichnet. Mithilfe solcher Software-Tools reicht ein Klick, um Texte umschreiben, paraphrasieren, zusammenfassen oder gänzlich neu generieren zu lassen. Und diese Möglichkeit hat unzweifelhaft Auswirkungen auf die fundamentalen Praktiken wissenschaftlichen Arbeitens sowie das Erlernen dieser Praktiken. Die klassische Publikation – als das Prädikat wissenschaftlicher Leistungen von Individuen oder Forschungskollektiven seit 2500 Jahren – steht wohlmöglich an einem Scheidepunkt, was z. B. die vollständig KI-generierte Publikation „Lithium-Ion Batteries – A Machine-Generated Summary of Current Research" (Beta Writer, 2019) beim Springer-Verlag verdeutlicht.

Damit steht schlussendlich das wissenschaftliche Selbstverständnis zumindest auf dem Prüfstand – unweigerlich verbunden mit der Frage, welche Werte und welches

Selbstverständnis von Wissenschaft dem akademischen Nachwuchs mit auf den Weg gegeben werden soll, um diesen auf die komplexen Herausforderungen der Zukunft vorzubereiten. Können uns die traditionellen Werte zur Forschungsintegrität auch in Zeiten KI-gestützter Wissenschaft die nötige Orientierung bieten oder verlieren sie ihre Klarheit und Anwendbarkeit? Wie verhält es sich mit Transparenz und Nachvollziehbarkeit beim Einsatz von Anwendungen, die sich dadurch auszeichnen, bei jeder Verwendung genuin Neues und somit Nichtreproduzierbares zu erzeugen? Reicht es gegebenenfalls, Modifikationen in der Deutung der Werte vorzunehmen, oder müssen sie aufgegeben und durch neue ersetzt werden? Diese Fragen sind dabei nicht nur theoretisch reizvoll, sondern auch praktisch dringlich, besteht doch einerseits sowohl eine bereits stark ausgeprägte faktische Präsenz von KI im Bildungsbereich (Chen et al., 2020) als auch ein immer stärker werdender politischer Wille, KI fundamental in Wissenschaft und Forschung zu integrieren (GWK, 2020), und andererseits der Grundgedanke, dass für den Einsatz und die Ergebnisse von KI letztlich immer der Mensch in der Verantwortung steht (Université de Montréal, 2018).

Um sich diesen Fragen zu nähern, wird zunächst exemplarisch vergleichend ein Kanon an Grundwerten wissenschaftlichen Arbeitens über verschiedene Ebenen – europäisch, national, organisations- und disziplinspezifisch – aufgezeigt. In einem zweiten Schritt wird ein Einblick gegeben in das, was bereits jetzt KI-gestützt im akademischen Alltag möglich ist, am Beispiel der wissenschaftlichen Textproduktion, um dann in einer den Beitrag abschließenden Synthese zu diskutieren, inwieweit die Grundwerte zur Forschungsintegrität anwendbar sind auf eben jene Praktiken. Der hier vorliegende Artikel versteht sich dabei als Eröffnung eines Diskurses über die Adäquatheit der zur Diskussion stehenden Grundwerte.

Grundwerte zur Forschungsintegrität

Der Anfang wissenschaftlichen Denkens – also des Logos – kann als Reaktion auf den „Zerfall des Mythos“ (Deppert, 2019, 72) im antiken Griechenland gedeutet werden und ist somit der Versuch, ein durch diesen Zerfall entstandenes, für gesellschaftliches Zusammenleben jedoch notwendiges Orientierungsvakuum zu füllen. Diese Idee von Wissenschaft als Orientierung – die weit über das bloße Ansammeln deskriptiven Wissens hinausgeht – hat bis in die Gegenwart Bestand (Mittelstraß, 2019). Doch will Wissenschaft Orientierung stiften, bedarf sie selbst orientierender Prinzipien, eben Grundwerten zur Forschungsintegrität, da sie sich ansonsten in performativen Selbstwidersprüchen (Kranz, 2017) verstrickt und damit ihre Glaubwürdigkeit aufs Spiel setzt. Den Entwurf dieser neuen Orientierungsprinzipien sieht Deppert in der Philosophie Sokrates’ zu einem ersten Höhepunkt kulminiert mit dem Vermeiden innerer Widersprüche (einer Art moralisches Konsistenzkriterium), der Idee der Selbstverantwortlichkeit, dem „Vertrauen in die Verläßlichkeit des eigenen Denkens“ (Deppert, 2019, 79f.) sowie dem Gebrauch der Vernunft, verstanden als Streben nach dem Rechten und

Guten, also dem, was Mensch und Welt dienlich ist und beiden keinen Schaden zufügt (ebd., 127). Letztlich sind es im Kern diese 2500 Jahre alten Überlegungen, die ideengeschichtlich den Diskurs um Forschungsintegrität bis in die Gegenwart bestimmen, wo jedoch nicht mehr der Kampf gegen den Mythos im Zentrum steht, sondern vielmehr gegen die immer häufiger an die Öffentlichkeit gelangenden Fälle absichtlichen Fehlverhaltens in der Wissenschaft (Resnik & Shamoo, 2011; Fanelli, 2009; OECD, 2007).[1]

Als Zäsur für die aktuelle Auseinandersetzung um die Bestimmung gemeinsamer orientierender Werte in der Wissenschaft bietet sich das „Singapore Statement on Research Integrity" (World Conference on Research Integrity, 2010) an, das im Rahmen der „2. World Conference on Research Integrity" formuliert wurde und erstmalig auf transkontinentaler und disziplinübergreifender Ebene einen Konsens über Prinzipien und Verantwortungen für Wissenschaftler:innen abbildet. Auf der Grundlage von vier Grundprinzipien – *Ehrlichkeit*, *Verantwortlichkeit*, *Professionelles Verhalten und Fairness* sowie *klare Organisation* – werden 14 konkrete Verantwortungen für die Anwendung in unterschiedlichen wissenschaftlichen Handlungsfeldern bestimmt. Andere Leitfäden auf globaler Ebene versuchen der Komplexität des Diskurses durch differenziertere, dafür weniger trennscharfe Grundwerte gerecht zu werden (IAC & IAP, 2012), die aber letztlich in späteren Dokumenten wieder zu vier Grundwerten zusammengeführt werden, wie im „Statement of Principles for Research Integrity" (Global Research Council, 2013) als *Ehrlichkeit*, *Verantwortlichkeit*, *Fairness* und *Rechenschaftspflicht*. Dieses orientiert sich begrifflich sehr dicht am Singapore Statement, lediglich die spezielle Bezeichnung der klaren Organisation wird im weiteren Begriff der Rechenschaftspflicht aufgehoben.

Auf europäischer Ebene wurde mit der „Europäischen Charta für Forscher" bereits 2005 ein Kodex mit Rechten und Pflichten für Wissenschaftler:innen formuliert, in dem jedoch keine Grundwerte genannt werden, sondern die Einhaltung von lokal oder disziplinspezifisch geltenden Grundwerten eingefordert wird (Europäische Kommission, 2005). Das ändert sich mit der Veröffentlichung des „Europäischen Verhaltenskodex für Integrität in der Forschung" (ALLEA, 2011), der in der ersten Fassung zunächst noch acht Grundwerte benennt, diese jedoch in der aktuellen überarbeiteten Fassung (ALLEA, 2018) – orientiert am „Statement of Principles for Research Integrity" – zu vier Grundwerten zusammenfasst. Diese sind: *Zuverlässigkeit*, *Ehrlichkeit*, *Respekt* und *Rechenschaftspflicht*. Verweisen Zuverlässigkeit und

[1] Aufgrund der inhaltlichen Ausrichtung des vorliegenden Beitrags müssen eigentlich zu klärende Aspekte vernachlässigt bleiben. Dies betrifft insbesondere Begriffsarbeit sowohl an dem sich noch in der Aushandlung befindenden Begriff der Forschungsintegrität selbst (Shaw, 2019) als auch an einer näheren Bestimmung dessen, was damit gemeint ist: Werte, Prinzipien, Normen etc. (Peels et al., 2019). Was im Folgenden versucht wird, ist eine Annäherung an die inhaltliche Bestimmung im Diskurs ausgehandelter orientierender Grundwerte – hier schlicht verstanden als die Bevorzugung bestimmter Handlungen vor anderen (Schwemmer, 2013) – in der Wissenschaft in dem Maße, wie sie eine Übertragung auf den Gegenstandsbereich der KI sinnvoll ermöglichen.

Verantwortlichkeit in den unterschiedlichen Kodizes auf sehr Ähnliches, nämlich eine verantwortungsbewusste Durchführung von Forschung, und zielen somit auf eine Qualitätssicherung methodischen Vorgehens ab – letztlich herrscht Einigkeit in Bezug auf den Gegenstand und Uneinigkeit in Bezug auf den adäquaten Begriff –, ist der Unterschied zwischen Fairness und Respekt ein qualitativer in Bezug auf den Gegenstand. Während sich Fairness im Wesentlichen auf den Umgang mit anderen Personen bezieht, bezieht sich Respekt gleichermaßen auf Mensch und Welt, ist also als Erweiterung zu verstehen.

Auf nationaler Ebene gelten für Deutschland seit 1998 und aktuell in der dritten Fassung vorliegend die „Leitlinien zur Sicherung guter wissenschaftlicher Praxis" (DFG, 2019) als Referenzdokument für Forschungsintegrität. Auch dort werden zunächst allgemeine abstrakte Prinzipien festgelegt, um anschließend in Handlungsaufforderungen für gute wissenschaftliche Praxis in Bezug auf den Forschungsprozess konkretisiert zu werden. Anders jedoch als bei den globalen und europäischen Kodizes wird hier nicht zunächst ein verbindlicher Wertekanon festgeschrieben, sondern es geht vielmehr um die (Selbst-)Verpflichtung von Organisationen und Wissenschaftler:innen zu den in ihrer Disziplin und Organisation geltenden Standards sowie um Organisationsprinzipien der Forschungslandschaft. In den weiteren Ausführungen werden dann aber doch eben jene Werte genannt, wie *Verantwortung* in Bezug auf die Qualitätssicherung des Forschungsprozesses sowie gegenüber der Gesellschaft, *Respekt* im Umgang mit Mensch, Natur und Kultur, die Forderung nach einem begründeten, nachvollziehbaren und dokumentierten Vorgehen – was der Idee der *Rechenschaftspflicht* entspricht – sowie *Ehrlichkeit*. Auf organisationsspezifischer Ebene handelt es sich dann in der Regel um eine Umsetzung und Konkretisierung der DFG-Leitlinien (siehe z. B. CAU, 2017; TH Wildau, 2002).[2]

Ein abschließender Blick auf disziplinspezifische Leitlinien (z. B. VDI, 2002) lässt ebenfalls keine nennenswerten Abweichungen in Bezug auf die Grundwerte zur Forschungsintegrität erkennen. So ist beispielsweise auch im Ethik-Kodex der Deutschen Gesellschaft für Erziehungswissenschaft (DGfE, 2016) die Rede von *Lauterkeit* in Bezug auf den wissenschaftlichen Arbeitsprozess, *Fairness* in Bezug auf den Umgang mit beteiligten Personen sowie *Verantwortung* im Hinblick auf den Einsatz von Ressourcen.

Da die im Europäischen Verhaltenskodex für Integrität in der Forschung (ALLEA, 2018) formulierten vier Grundwerte – Zuverlässigkeit, Ehrlichkeit, Respekt, Rechenschaftspflicht – allem Anschein nach so etwas wie einen gegenwärtigen Konsens der scientific community über die Grundwerte zur Forschungsintegrität abbilden und die Werte dort zudem sehr weit gefasst werden, bietet es sich an, zunächst diese Werte im Hinblick auf ihre Anwendbarkeit

[2] „Mit Inkrafttreten des Kodex zum 01.08.2019 müssen alle Hochschulen und außerhochschulischen Forschungseinrichtungen die 19 Leitlinien und ihre Erläuterungen rechtsverbindlich umsetzen, um Fördermittel durch die DFG erhalten zu können." Abgerufen am 21.06.2021 von https://www.dfg.de/foerderung/grundlagen_rahmenbedingungen/gwp/index.html.

und Sinnhaftigkeit in Bezug auf KI-gestützte wissenschaftliche (Schreib-)Praktiken hin zu betrachten. Die größte Herausforderung bei diesem Unterfangen besteht darin, dass in keinem der genannten Dokumente eine intensionale Bestimmung der Grundwerte vorgenommen wird. Der über ein alltagssprachliches Verständnis hinausgehenden Bedeutung der Werte kann man sich letztlich nur nähern, indem man versucht, diese aus der Extension – in diesem Fall der sich daraus ergebenden Rechte und Pflichten für bestimmte wissenschaftliche Handlungsfelder – zu rekonstruieren. Doch bevor dies geschieht, soll zunächst ein exemplarischer Einblick in das gegeben werden, was bereits mit einfachen Mitteln an KI-gestützten wissenschaftlichen Praktiken möglich ist, um so eine fundierte Diskussion der Grundwerte zu erlauben.

Gegenwärtige Möglichkeiten KI-gestützter wissenschaftlicher Textgenerierung

IT-gestützte Technologien für Lernen, Lehre und Forschung gewinnen nicht erst durch die COVID-19-Pandemie an Bedeutung. Auffällig ist dabei, dass KI-basierte Tools bisher, zumindest in der Wahrnehmung derjenigen, die sich aus professioneller Sicht mit Lernen und Lehren beschäftigen, stark unterrepräsentiert bis nicht existent sind (Hart, 2020). Während klassische Tools für die Unterstützung des Schreibprozesses, wenn es um Grammatik oder Rechtschreibung geht, vielfältige Hilfestellungen bieten, scheitern diese jedoch bei anspruchsvolleren Aufgaben, die zur Verbesserung der semantischen Textqualität führen sollen. Das bedeutet, dass z. B. der Aufbau einer Argumentationskette oder auch die Stringenz der Argumentation vom Menschen selbst übernommen werden muss (Strobl et al., 2019). Im Rahmen des Schreibprozesses werden von Studierenden zudem Online-Paraphrasierungs-Tools eingesetzt (Prentice & Kinden, 2018). Aber auch hier kann die Qualität der Ergebnisse leiden, wenn Studierende die zugrunde liegenden Texte nicht verstanden haben und durch das unbedarfte Paraphrasieren fehlerhafte Aussagen und semantische Brüche entstehen. Darüber hinaus birgt dies Plagiatsrisiken in Bezug auf die Originalität der Ergebnisse (Rogerson & McCarthy, 2017).

Im Hinblick auf die Generierung und Bearbeitung von Texten wird rückblickend vermutlich 2019 als das Jahr für den Durchbruch der Künstlichen Intelligenz in Form des Natural Language Processing (NLP) bewertet werden. Das Sprachmodell GPT-2 der früheren Non-Profit-Organisation OpenAI aus San Francisco steht für diesen Durchbruch. Analog zu den bekannten Möglichkeiten der Autovervollständigung bei Messenger-Diensten sind derartige statistische Sprachmodelle darauf trainiert, jeweils schrittweise das nächste Wort auf Basis der vorherigen Textsequenzen als das wahrscheinlichste vorherzusagen. Außerdem können Tools, die auf diesen Sprachmodellen basieren, sich an dem Sprachstil des Inputs orientieren und in diesem Stil auch Textsequenzen generieren und fortführen (Radford, Wu, Amodei et al., 2019).

Das Akronym GPT-2 steht für ein künstliches neuronales Netz mit dem Namen Generative Pretrained Transformer 2. Es handelt sich um ein Deep Learning System (Jones, 2014), das auf der Transformer-Architektur von Google beruht (Vaswani et al., 2017). Die Veröffentlichung der ersten Version dieses Algorithmus wurde begleitet von der Befürchtung, dass durch diese selbstständig textgenerierende Software vermehrt Fake News produziert und verbreitet würden. Daher wurde sie zunächst nur in Auszügen veröffentlicht (Kremp, 2019). Die Vollversion des Modells wurde erst im November 2019 publiziert. Die Datengrundlage des 1,5 Milliarden Parameter umfassenden neuronalen Netzes besteht aus mehr als 8 Millionen Dokumenten, die ca. 40 Gigabyte Text entsprechen (Radford, Wu, Child et al., 2019).

Bereits im März 2020 wurde der Nachfolger GPT-3 veröffentlicht, der seit Juni 2020 das erste kommerzielle Produkt von OpenAI darstellt. Microsoft hat eine Exklusivlizenz und Zugang zum Quellcode des Sprachmodells. Eine Integration in die Azure-Dienste sowie ein Öffnen explizit auch für die Wissenschaft sind geplant (Scott, 2020). Diese Version arbeitet bereits mit 175 Milliarden Parametern basierend auf einer Datengrundlage von 45 Terrabyte, d. h. der mehr als zehnfachen Menge im Vergleich zu den bisherigen KI-Sprachmodellen (Brown et al., 2020). Erste Anwendungsbeispiele zeigen, dass neben der automatischen Generierung von Texten (Moorstedt, 2020) auch vielfältige andere Einsatzgebiete abgedeckt werden (Chojecki, 2020).

Forschende und Studierende können sich mit Software-Tools, die auf solchen Modellen basieren, auf immer einfachere und effizientere Art und Weise schriftliche Texte bis hin zu Abschlussarbeiten generieren lassen. Darüber hinaus enthalten Übersetzungslösungen, wie z. B. DeepL, bereits leistungsstarke KI-Funktionalitäten (DeepL GmbH, 2020), die häufig nicht als solche wahrgenommen werden. Werden derartige Übersetzungslösungen in Verbindung mit bereits vorhandenen, digital verfügbaren Dokumenten (z. B. Studienarbeiten anderer Studierender) und Rewriting-Tools, wie z. B. Quillbot, in geschickter Weise kombiniert, entstehen quasi auf Knopfdruck vermeintlich neue Texte (Weßels, 2020). Auf diese Art generierte Texte können nicht mehr als Plagiate identifiziert werden, weisen doch aktuelle Plagiatserkennungs-Softwarelösungen selbst bei nicht durch KI erstellten Texten nur unzureichende Ergebnisse auf (Foltýnek et al., 2020). Daher werden bereits erste Forderungen nach einer

Kennzeichnungspflicht für KI-generierte Texte laut (Meier, 2020). Neben dieser Diskussion, wie Wissenschaft angemessen auf die neuen Möglichkeiten reagieren soll, beginnt parallel der Diskurs darüber, ob und inwieweit auch Forschung oder Teile des Forschungsprozesses, darunter auch das wissenschaftliche Schreiben, durch KI ersetzt werden (Voshmgir, 2020). Dabei scheint sich ein neues Paradigma zu formen, das KI als eine Art Teammitglied betrachtet, mit dem man gemeinsam eine kollaborative Intelligenz bildet, um so zukünftige Probleme zu lösen (Kankanhalli, 2020).

Anwendbarkeit der Grundwerte im Zeitalter Künstlicher Intelligenz

Der Einsatz von KI zum Zweck wissenschaftlicher Textproduktion in Forschung und Lehre birgt also gleichermaßen Möglichkeiten wie Risiken (Meyer & Weßels, 2020; Strobl et al., 2019; Schmohl et al., 2019). Will Wissenschaft die bisher weder faktisch noch theoretisch absehbaren Möglichkeiten zum Guten einsetzen, sie also in den Dienst der Gemeinschaft stellen, bedarf es orientierender Prinzipien. Während in den zuvor diskutierten Kodizes zur Forschungsintegrität eine Vielzahl von Handlungsfeldern explizit thematisiert werden, bleibt auf dieser allgemeinen Ebene (bisher) eine Orientierung für den Einsatz von KI in der Wissenschaft aus. Daneben gibt es einen umfangreichen Diskurs über orientierende Prinzipien für KI (Bartneck et al., 2021; Europäische Kommission, 2019; Floridi, 2019; Université de Montréal, 2018). Diese betreffen jedoch im Wesentlichen die Entwicklung. Was bisher unterbeleuchtet bleibt, ist die Frage, wie Wissenschaftler:innen KI als Instrument zu Forschungszwecken einsetzen sollen (Hwang et al., 2020). Daher gilt es im Folgenden zu diskutieren, ob die traditionellen Grundwerte dazu geeignet sind, eine ausreichende Orientierung für den Umgang mit KI zu leisten.

Zuverlässigkeit

Zuverlässigkeit kann wohl als der fundamentalste wissenschaftliche Wert gesehen werden, ist er doch vielmehr Definitionskriterium von Wissenschaft, deren Funktion es ist, Wissen zu generieren, das systematischer, nachvollziehbarer und begründeter – eben zuverlässiger – ist als individuell situatives bzw. lebensweltliches Wissen. Ein Verzicht auf Zuverlässigkeit hätte unweigerlich einen gesellschaftlichen Vertrauensverlust zur Folge und würde das gesamte tradierte Konzept Wissenschaft infrage stellen. Wie aber lässt sich Zuverlässigkeit im Zeitalter Künstlicher Intelligenz konkretisieren? Hierzu bedarf es zunächst einer Rollenklärung im Zusammenspiel von Mensch und Maschine. Der Mensch in Wissenschaft und Forschung kann in diesen drei primären Rollen auftreten:

1. Mensch als „Creator“ der KI (Programmierung von Algorithmen, Modellierung von Modellen, Auswahl und Bereitstellung der Datenbasis, Testen der Software, Überwachung des Systemverhaltens usw.)
2. Mensch als „Tool-Expert“ für die zielkonforme Auswahl oder auch Konfiguration einer KI-Anwendung
3. Mensch als „User“ von bereitgestellten und (gesellschaftlich akzeptierten) KI-Anwendungen

Eine Beurteilung der Zuverlässigkeit kann nun spezifisch für die obigen Rollen erfolgen. In den diskutierten Kodizes wird sichtbar, dass diese (noch) nicht differenziert werden. Es wird die Position vertreten, dass die jeweiligen Forschenden, die User also, dafür verantwortlich sind, die Zuverlässigkeit der im Forschungsprozess eingesetzten Verfahren und Instrumente sicherzustellen. Dieses Vorgehen ist weder leistbar noch sinnvoll vor dem Hintergrund der oben differenzierten Rollen.

Um Zuverlässigkeit im Folgenden zu analysieren, wird eine systemische Perspektive gewählt und die soziotechnische Zusammenarbeit von Mensch und Maschine betrachtet, letztere verstanden als KI-basierte Anwendung des maschinellen Lernens. Hierfür können als Grundlage die folgenden Kriterien für Zuverlässigkeit im Sinne der Vertrauenswürdigkeit KI-gestützter Systeme (Huchler et al., 2020) angeführt werden:

- Qualität der verfügbaren Daten
- Transparenz, Erklärbarkeit und Widerspruchsfreiheit
- Verantwortung, Haftung und Systemvertrauen

Das Vertrauen der Gesellschaft und des Individuums (hier als Tool-Expert und User) in KI-basierte Anwendungen erfordert von den Akteur:innen in der Rolle des Creators im Hinblick auf die obigen drei Gruppen von Kriterien ein hohes Maß an Transparenz, Erklärbarkeit und Überprüfbarkeit. Um dieser Zielsetzung gerecht zu werden, benötigt eine gute wissenschaftliche Praxis im KI-Zeitalter neue und vor allem differenzierte Orientierungen, die die User im Hinblick auf die Sicherstellung der Zuverlässigkeit entlasten. Hierzu bedarf es eines neuen Regelungssystems der KI-Governance mit klar definierten Strukturen und Prozessen, die sowohl als Definitionsrahmen (für das WAS) wie auch als Handlungsrahmen (für das WIE) konzipiert werden müssen. Eine KI-Governance könnte zu einer Kennzeichnungspflicht der von Wissenschaftler:innen verwendeten KI-Systeme in ihren Publikationen führen und im Bereich NLP letztlich zu qualitätsgeprüften KI-Sprachmodellen und darauf basierenden Werkzeugen. Der TÜV-Verband (2020) hat in seinem Positionspapier zur Sicherheit KI-gestützter Anwendungen seine Mitwirkung signalisiert und die Bedeutung unabhängiger Prüforganisationen betont.

Ehrlichkeit

„Ehrlich" sind im allgemeinen Sprachgebrauch Personen, die die Wahrheit sagen, zuverlässig und rechtschaffen sind. In der Wissenschaft ist Ehrlichkeit „bei der Entwicklung, Durchführung, Überprüfung, Berichterstattung und Kommunikation von Forschungsarbeiten in transparenter, fairer, vollständiger und unvoreingenommener Weise" (ALLEA, 2018, 4) essenziell. Jeder Schritt von Antragstellung bis zur Veröffentlichung der Ergebnisse soll im Sinne einer nahtlosen Qualitätssicherung überprüfbar sein, um Datenmanipulationen oder erfundene Ergebnisse gar nicht erst zu ermöglichen bzw. frühzeitig zu erkennen und zu ahnden sowie Methodenfehler aufzudecken (DFG, 2019, 14f.). Transparenz im Umgang mit neuen Methoden, wozu auch die Einbeziehung von KI in den Forschungsprozess gezählt werden kann, ist laut Leitlinie 7 der DFG besonders wichtig. Das Prinzip 5 der „Montréal Declaration for a Responsible Development of Artificial Intelligence" zielt im Umgang mit KI genau auf diesen Aspekt der Transparenz ab: „AIS must meet intelligibility, justifiability, and accessibility criteria, and must be subjected to democratic scrutiny, debate, and control" (Université de Montréal, 2018, 12). Die Funktionsweise und der Quellcode des KI-gestützten Systems sollen demnach offengelegt und jegliche unerwünschten Verhaltensweisen dokumentiert und gemeldet werden. Die Verantwortung soll immer bei dem Creator oder User liegen. In einem Co-Autor:innen-Gespann aus Mensch und KI-Tool würde der/die menschliche Autor:in demnach immer auch für die KI verantwortlich sein. Wie kann ein:e Wissenschaftler:in aber einem System Ehrlichkeit auferlegen, das im Sinne der Reproduzierbarkeit nie zweimal dieselben Ergebnisse liefern wird, weil es programmiert wurde, um „mitzudenken" und so neue Ideen anzustoßen? Oder noch fundamentaler formuliert: Das Konzept der Ehrlichkeit ist schlicht nicht anwendbar auf NLP-Systeme, deren Funktion darin besteht, eingehende Informationen nach einem bestimmten Algorithmus zu verarbeiten, unabhängig davon, ob der Input wahr oder falsch ist.

Die Möglichkeiten der KI-gestützten Systeme einzuschränken, um sie unter Kontrolle zu behalten, widerspricht wiederum der Idee von KI, die laut Definition der EU-Kommission „Systeme mit einem ‚intelligenten' Verhalten, die ihre Umgebung analysieren und mit einem gewissen Grad an Autonomie handeln, um bestimmte Ziele zu erreichen", umfasst (Europäische Kommission, 2018, 1). Es bedarf einer modifizierten Definition des Wertes Ehrlichkeit, wenn Forschende zwar ihren Anteil am Text sowie die verwendeten KI-Tools offenlegen können, aber die Funktionsweise der Letzteren selbst nicht vollständig durchdrungen haben. Können Ehrlichkeit und Transparenz nur so weit erwartet werden, wie der/die menschliche Partner:in ausschließlich KI-generierte Textpassagen oder Ideen und Übersetzungen verwendet, die er/sie selbst nachvollziehen und erläutern kann? Dies bedeutete allerdings, das Potenzial von KI-gestützter Textproduktion nicht auszuschöpfen und intelligenten Systemen grundsätzlich zu misstrauen. Wird Ehrlichkeit darüber hinaus herangezogen, um das künst

-liche Manipulieren oder Generieren von Daten zu unterbinden, so werden die Grenzen der Anwendung dieses Wertes auf den Einsatz von KI-Systemen deutlich, ist doch genau das deren Funktion.

Respekt

Unter Respekt wird gemeinhin eine Haltung der Achtung, Wertschätzung und Anerkennung verstanden. Dies bezieht sich im Rahmen von Forschung auf „Kollegen, Forschungsteilnehmer, die Gesellschaft, Ökosysteme, das kulturelle Erbe und die Umwelt" (ALLEA, 2018, 4). Wie aber lässt sich Respekt im Hinblick auf KI-gestützte Textgenerierung deuten?

Wer den Kolleg:innen Respekt zollt, achtet sowohl sie als Person als auch ihre Leistungen für das Feld. Im Wissenschaftsbetrieb wird Anerkennung derzeit in hohem Maße über die Bezugnahme auf Texte ausgedrückt. Wenn künftig vermehrt KI-Tools zum Einsatz kommen, ist der Mehrwert dieser Praxis des Zitierens fraglich, da das Erschließen, Überblicken und Zitieren bestehender Gedanken KI-Tools schneller erledigen als der Mensch. Letztlich wären es also die Maschinen, die den Respekt zollen würden, womit die Praktik obsolet wäre. Vielmehr könnte es darauf ankommen, den eigenen Beitrag deutlich zu benennen, um den erzielten Fortschritt zu kennzeichnen. Dies würde inhaltlich der Forschung und auf der persönlichen Ebene dem Vorankommen des/der Autor:in dienen.

Respekt gegenüber Kolleg:innen zeigt sich auch durch Transparenz bezüglich aller Beitragenden zu einer Publikation. Gemäß ALLEA (2018, 7) sollen „wichtige Arbeiten und intellektuelle Beiträge anderer, einschließlich von Kooperationspartnern, Assistenten und Förderern, die die veröffentlichten Forschungsarbeiten in geeigneter Form beeinflusst haben", gewürdigt und zitiert werden, wobei sich schon jetzt unterschiedliche Praktiken bei der (Nicht-)Nennung von Personen zeigen, die z. B. durch Feedback den Forschungsprozess prägen (Osborne & Holland, 2009, 3). Da bei KI-gestützten Schreibprozessen aufgrund der Verschmelzung von Mensch und Maschine die jeweilige Leistung nicht abgegrenzt werden kann, ist diese geforderte Transparenz nicht zu gewährleisten.

Unter Respekt für Gesellschaft und Umwelt sind das Nichtschadensprinzip, die Kultursensibilität und das Einbeziehen der Öffentlichkeit zu verstehen. Forschende sollen potenzielle Schäden, aber auch Risiken, die mit ihrer Tätigkeit einhergehen, „erkennen und bewältigen" (ALLEA, 2018, 6). Angesichts der rasant fortschreitenden Entwicklung technischer Möglichkeiten, insbesondere in Bezug auf KI, vermag jedoch niemand die Folgen sicher abzuschätzen. Bisherige Orientierungen greifen hier somit nicht. Besser steuerbar erscheint hingegen die Bewusstheit über den kulturellen Kontext, in dem Forschung stattfindet und den sie möglicherweise verändert. Durch den Einsatz von KI, insbesondere des NLP, besteht die Gefahr, dass Forschung einseitig wird und stetig den Mainstream inklusive seiner Fehler reproduziert (Hao, 2020). Auf Respekt im Sinne von Kultursensibilität von Forschung ist demnach

verstärkt zu achten. Das Einbeziehen der Gesellschaft in Forschungsprojekte (Citizen Science) könnte durch KI-Tools gefördert, der Zugang zu wissenschaftlichem Wissen und somit Partizipation erleichtert werden. Im Dienste von Wissenschaftskommunikation bzw. Public Understanding of Science können KI-Tools den Forschenden die Aufgabe abnehmen, ihre Ergebnisse in gut verständlicher Form aufzubereiten, sodass die Öffentlichkeit besser informiert wird.

Rechenschaftspflicht

Der Europäische Verhaltenskodex für Integrität in der Forschung bezieht Rechenschaftspflicht umfänglich auf den gesamten Forschungsprozess „von der Idee bis zur Veröffentlichung, für deren Verwaltung und Organisation, für Ausbildung, Aufsicht und Betreuung und für ihre weiteren Auswirkungen" (ALLEA, 2018, 4). Im Hinblick auf den Einsatz KI-basierter Systeme ist die Frage, wer wofür und wem gegenüber rechenschaftspflichtig ist, bislang ungeklärt. So stellt z. B. die Montréal Declaration einerseits eindringlich klar, der Einsatz von KI dürfe nicht die Verantwortung des Menschen bei Entscheidungen mindern (Université de Montréal, 2018, 16). Andererseits will sie bei Schäden die Schuld und Verantwortung von Entwickler:innen und Nutzer:innen der KI beschränkt wissen, solange die KI zuverlässig und bestimmungsgemäß eingesetzt wurde (ebd.).

Dies zieht auch für den KI-gestützten wissenschaftlichen Schreibprozess eine Reihe ungeklärter Fragen nach sich: Können wissenschaftlich Schreibende als bloße User (siehe oben) überhaupt ihrer Rechenschaftspflicht Genüge tun, wenn sie KI einsetzen? Was bedeutet Rechenschaftspflicht, wenn sie entsprechende Werkzeuge nutzen, ohne sich über die Prozesse der Textentstehung (geschweige denn die performative Wirkung des Schreibproduktes) im Klaren zu sein? Die „Ethik-Leitlinien für eine vertrauenswürdige KI" der Europäischen Kommission (2019) nennen Rechenschaftspflicht in Beziehung mit sechs weiteren Anforderungen. Konkretisierend werden beispielhaft „Nachprüfbarkeit, Minimierung und Meldung von negativen Auswirkungen, Kompromisse und Rechtsbehelfe" aufgezählt (ebd., 18). Prüft man die genannten Kategorien der Rechenschaftspflicht im Hinblick auf KI-gestützte wissenschaftliche Schreibprozesse, so lässt sich feststellen, dass sie kaum individuell einlösbar sind.

Mit dem Kriterium der Nachprüfbarkeit verbindet die hochrangige Expertengruppe für künstliche Intelligenz, „dass Algorithmen, Daten und das Entwurfsverfahren einer Bewertung unterzogen werden können" (ebd., 24). Interne und externe Prüfung sowie Bewertungsberichte sollen das Vertrauen in die Technik fördern. Den einzelnen wissenschaftlich Schreibenden als User von KI-gestützter Software bleibt lediglich, möglichst auf entsprechend geprüfte und bewertete Produkte zurückzugreifen und diese auszuweisen. In Gänze nachprüfbar ist jedoch weder der wissenschaftliche Schreibprozess mithilfe von KI-Tools noch der fertige Text.

Beim Kriterium der Minimierung und Meldung von negativen Auswirkungen geht es um Abwendung von Risiken, negativen Folgen und den Schutz von Personen und Personengruppen. Hier bleibt ein Spielraum der Abwägung und des Ermessens – und somit der Unsicherheit, denn auch dieses Kriterium bleibt an die fehlende Nachprüfbarkeit geknüpft.

Bei Interessen- und Wertekonflikten verlangen die Ethik-Leitlinien nachdrücklich ethisch vertretbare Kompromisse, die mit einer Begründungs- und Dokumentationspflicht einhergehen und kontinuierlich zu prüfen sind. Kommt ein Kompromiss nicht zustande, so ist eine andere Nutzung oder Entwicklung von KI-Systemen gefordert. Als letzte Maßnahme bleiben Vorkehrungen für Rechtsschutz und Rechtsmittel.

Letzten Endes zielen die Ethik-Leitlinien auf „Vorkehrungen […], die die Verantwortlichkeit und Rechenschaftspflicht für KI-Systeme und deren Ergebnisse vor und nach deren Umsetzung gewährleisten" (ebd., 24). Diese Forderung greift für die wissenschaftlich schreibenden User jedoch nur beschränkt. Sie können begründend, beschreibend und dokumentierend transparent und nachprüfbar machen, mit welchen – womöglich zertifizierten – Tools sie wozu und in welcher Weise arbeiten. Darüber hinaus können sie als einzelne User keine zuverlässigen Angaben machen. Einer Idee von Rechenschaftspflicht, bei der die/der Einzelne nicht für jedes Detail verantwortlich sein muss, sehr wohl aber für das Ganze der Arbeit (Shamoo & Resnik, 2009, 101f.), entspricht dies nicht.

Ausblick

Die hier geführte Diskussion offenbart sowohl Notwendigkeit als auch Schwierigkeit der Reflexion traditioneller Werte guter wissenschaftlicher Praxis im Hinblick darauf, sie an einen durch KI-Technologien ausgelösten disruptiven Wandel wissenschaftlicher Praktiken am Beispiel der Textgenerierung anzupassen. Die vier Grundwerte Zuverlässigkeit, Ehrlichkeit, Respekt und Rechenschaftspflicht müssen sich nun nicht mehr nur auf rein menschliche Autor:innenschaft beziehen lassen, sondern auf ein kollaboratives Konstrukt aus Mensch und Maschine, das beim heute schon absehbaren Einsatz von Schreibbots eine symbiotische Struktur aufweist und genau darin einen großen gesellschaftlichen Nutzen entfalten kann. Wenn Mensch mit Maschine gemeinsam Texte beliebig konstruiert, modifiziert und veröffentlicht, ergeben sich in allen Lebenszyklen der wissenschaftlichen Textarbeit vielfältige neue Herausforderungen, zu denen sich Wissenschaft positionieren muss. Zumindest prima facie scheint eine einfache Adaption des Grundwertekanons keine ausreichende Orientierung für den Umgang mit diesen Herausforderungen zu bieten. Insbesondere die traditionelle Zuweisung der Verantwortung für Forschungsintegrität an die jeweils forschenden Individuen und Kollaborationen ist von den Einzelnen nicht mehr zu leisten. Sie bedarf daher einer alternativen Verteilung.

Die sich andeutende Komplexität und Radikalität des hier skizzierten Problemraums muss es darüber hinaus erlauben, sich zunächst einmal von dem Faktischen, Altbekannten zu lösen, um im Raum des Möglichen nach Lösungen zu suchen (Kralemann, 2011). Dies kann durch einen radikalen Perspektivenwechsel geschehen, der als Pendant genauso disruptiv wirkt wie die auslösende KI-Disruption im Problemraum. Eine solche, den Diskurs eröffnende Möglichkeit, könnte wie folgt aussehen:

Während wir heute in der wissenschaftlichen Praxis eine – nebenbei bemerkt ressourcenverschlingende – Kennzeichnungspflicht für die Gedanken anderer verlangen, könnten wir im Zeitalter Künstlicher Intelligenzen und der Kollaboration von Mensch und Maschine eine Kennzeichnung primär für die eigenen Textpassagen verlangen, die entweder direkt von dem/der menschlichen Autor:in stammen oder unter seiner/ihrer Leitung als machine leader in der Kollaboration von Mensch und Maschine entstanden sind. Das wiederum würde bedeuten, dass wir die Qualität einer wissenschaftlichen Arbeit vorrangig an der intellektuellen und originären Eigenleistung der eigenen Textpassagen festmachen würden, und hätte zudem den angenehmen Nebeneffekt, Wissenschaftler:innen wieder mehr Zeit für ihre eigentlichen denkenden und forschenden Tätigkeiten zu eröffnen.[3] Die in der Praxis dominierende Literaturarbeit rückt damit in den Hintergrund und verliert drastisch an Bedeutung. Und es stellt sich die Frage, ob wir die Quellenverwendungsnachweise zukünftig überhaupt noch als erforderlich erachten müssen oder auch darauf verzichten könnten, denn sie stehen dann ohnehin der wissenschaftlichen Community als ubiquitäres Gut im Sinne der Zielsetzung von Open Science (European Commission, 2019) zur Verfügung.

Literatur

ALLEA – All European Academies (Hrsg.). (2011). The European Code of Conduct for Research Integrity. Abgerufen am 22.06.2021 http://archives.esf.org/fileadmin/Public_documents/Publications/Code_Conduct_ResearchIntegrity.pdf.

ALLEA – All European Academies (Hrsg.). (2018). Europäischer Verhaltenskodex für Integrität in der Forschung. Abgerufen am 22.06.2021 von http://www.allea.org/wp-content/uploads/2018/06/ALLEA-European-Code-of-Conduct-for-Research-Integrity-2017-Digital_DE_FINAL.pdf.

[3] Dieser Gedanke ist strenggenommen weder besonders radikal noch innovativ, ist doch die gegenwärtige Belegpraxis als grundlegendes Qualitätsmerkmal wissenschaftlicher Publikationen eine Entwicklung erst des mittleren 20. Jahrhunderts. So heißt es z. B. noch bei Carnap (1934, VI): „An manchen Stellen im Text werden Hinweise auf die wichtigste Literatur gegeben. Vollständigkeit ist dabei nicht angestrebt worden."

Bartneck, C., Lütge, C., Wagner, A., & al (2021). An Introduction to Ethics in Robotics and AI. Cham: Springer International Publishing. Abgerufen am 22.06.2021 von https://doi.org/10.1007/978-3-030-51110-4.

Beta Writer (2019). Lithium-Ion Batteries: A Machine-Generated Summary of Current Research. Springer Nature. Abgerufen am 22.06.2021 von https://doi.org/10.1007/978-3-030-16800-1.

Brown, T. B., Mann, B., Ryder, N., & al. (2020). Language Models are Few-Shot Learners. arXiv:2005.14165 [cs]. Abgerufen am 22.06.2021 von http://arxiv.org/abs/2005.14165.

Callaway, E. (2020). 'It will change everything': DeepMind's AI makes gigantic leap in solving protein structures. *Nature*, 588(7837), 203–204. Abgerufen am 22.06.2021 von https://doi.org/10.1038/d41586-020-03348-4.

Carnap, R. (1934). Logische Syntax der Sprache. Berlin, Heidelberg: Springer. Abgerufen am 22.06.2021 von https://doi.org/10.1007/978-3-662-25375-5.

CAU – Christian-Albrechts-Universität zu Kiel (Hrsg.). (2017). Richtlinien der CAU zur Sicherung guter wissenschaftlicher Praxis. Abgerufen am 22.06.2021 von https://www.uni-kiel.de/fileadmin/user_upload/forschung/integritaet-ethik/downloads/CAU-Richtlinie_Gute_wissenschaftliche_Praxis.pdf.

Chen, X., Xie, H., & Hwang, G.-J. (2020). A multi-perspective study on Artificial Intelligence in Education: Grants, conferences, journals, software tools, institutions, and researchers. *Computers and Education: Artificial Intelligence*, 1, 100005. Abgerufen am 22.06.2021 von https://doi.org/10.1016/j.caeai.2020.100005.

Chojecki, P. (2020). Crazy GPT-3 Use Cases. Discover how powerful GPT-3 from OpenAI really is. Abgerufen am 22.06.2021 von https://medium.com/towards-artificial-intelligence/crazy-gpt-3-use-cases-232c22142044.

DeepL GmbH (2020). Erneuter Durchbruch bei der KI-Übersetzungsqualität. Abgerufen am 22.06.2021 von https://www.deepl.com/blog/20200206.html.

Deppert, W. (2019). Theorie der Wissenschaft. Band 2: Das Werden der Wissenschaft. Wiesbaden: Springer. Abgerufen am 22.06.2021 von https://doi.org/10.1007/978-3-658-14043-4.

DFG – Deutsche Forschungsgemeinschaft (Hrsg.). (2019). Leitlinien zur Sicherung guter wissenschaftlicher Praxis. Abgerufen am 22.06.2021 von https://doi.org/10.5281/zenodo.3923602.

DGfE – Deutsche Gesellschaft für Erziehungswissenschaft (Hrsg.). (2016). Ethik-Kodex der Deutschen Gesellschaft für Erziehungswissenschaft. Abgerufen am 22.06.2021 von https://www.dgfe.de/fileadmin/OrdnerRedakteure/Satzung_etc/Ethikkodex_2016.pdf.

Europäische Kommission (Hrsg.). (2005). Europäische Charta für Forscher: Verhaltenskodex für die Einstellung von Forschern. Office for Official Publications of the European Communities. Abgerufen am 22.06.2021 von https://cdn2.euraxess.org/sites/default/files/brochures/eur_21620_de-en.pdf.

Europäische Kommission (Hrsg.). (2018). Künstliche Intelligenz in Europa. Abgerufen am 22.06.2021 von https://eur-lex.europa.eu/legal-content/DE/TXT/PDF/?uri=CELEX:52018DC0237&from=CS.

Europäische Kommission (Hrsg.). (2019). Ethik-Leitlinien für eine vertrauenswürdige KI. Abgerufen am 22.06.2021 von https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai.

European Commission (Hrsg.). (2019). Open Science. Abgerufen am 22.06.2021 von https://ec.europa.eu/info/sites/info/files/research_and_innovation/knowledge_publications_tools_and_data/documents/ec_rtd_factsheet-open-science_2019.pdf.

Fanelli, D. (2009). How Many Scientists Fabricate and Falsify Research? A Systematic Review and Meta-Analysis of Survey Data. PLOS ONE, 4(5): e5738. Abgerufen am 22.06.2021 von https://doi.org/10.1371/journal.pone.0005738.

Floridi, L. (2019). Establishing the rules for building trustworthy AI. *Nature Machine Intelligence*, 1(6), 261–262. Abgerufen am 22.06.2021 von https://doi.org/10.1038/s42256-019-0055-y.

Foltýnek, T., Dlabolová, D., Anohina-Naumeca, & al. (2020). Testing of Support Tools for Plagiarism Detection. arXiv:2002.04279 [cs]. Abgerufen am 22.06.2021 von http://arxiv.org/abs/2002.04279.

Global Research Council (Hrsg.). (2013). Statement of Principles for Research Integrity. Abgerufen am 22.06.2021 von https://www.globalresearchcouncil.org/fileadmin//documents/GRC_Publications/grc_statement_principles_research_integrity_FINAL.pdf.

GWK – Gemeinsame Wissenschaftskonferenz (Hrsg.). (2020). Bund-Länder-Vereinbarung gemäß Artikel 91b Absatz 1 des Grundgesetzes über die Förderinitiative „Künstliche Intelligenz in der Hochschulbildung“ vom 10. Dezember 2020. Abgerufen am 22.06.2021 von https://www.gwk-bonn.de/fileadmin/Redaktion/Dokumente/Papers/BLV_KI_in_der_Hochschulbildung.pdf.

Hao, K. (2020). The True Dangers of AI Are Closer Than We Think. *MIT Technology Review*, 123(6), 38–39.

Hart, J. (2020). Top Tools for Learning 2020: Results of the 14th Annual Survey. Abgerufen am 22.06.2021 von https://www.toptools4learning.com/.

Heaven, W. D. (2020). OpenAI's new language generator GPT-3 is shockingly good–and completely mindless. *MIT Technology Review*. Abgerufen am 22.06.2021 von https://www.technologyreview.com/2020/07/20/1005454/openai-machine-learning-language-generator-gpt-3-nlp/.

Huchler, N., Adolph, L., André, E., Bauer, & al. (2020). Kriterien für die Mensch-Maschine-Interaktion bei KI. Ansätze für die menschengerechte Gestaltung in der Arbeitswelt. Abgerufen am 22.06.2021 von https://www.plattform-lernende-systeme.de/files/Downloads/Publikationen/AG2_Whitepaper2_220620.pdf.

Hwang, G.-J., Xie, H., Wah, & al. (2020). Vision, challenges, roles and research issues of Artificial Intelligence in Education. *Computers and Education: Artificial Intelligence*, 1, 100001. Abgerufen am 22.06.2021 von https://doi.org/10.1016/j.caeai.2020.100001.

IAC & IAP – InterAcademy Council, & InterAcademy Partnership (Hrsg.). (2012). Responsible Conduct in the Global Research Enterprise. A Policy Report. Abgerufen am 22.06.2021 von http://nas-sites.org/responsiblescience/files/2015/03/IAP2012_Responsible-conduct.pdf.

Jones, N. (2014). Computer science: The learning machines. *Nature*, 505(7482), 146–148. Abgerufen am 22.06.2021 von https://doi.org/10.1038/505146a.

Kankanhalli, A. (2020). Artificial intelligence and the role of researchers: Can it replace us? *Drying Technology*, 38(12), 1539–1541. Abgerufen am 22.06.2021 von https://doi.org/10.1080/07373937.2020.1801562.

Kralemann, B. (2011). Παιδεία – Entdecke die Möglichkeiten! In Prieß, W. (Hrsg.). Wirtschaftspädagogik zwischen Erkenntnis und Erfahrung – Strukturelle Einsichten zur Gestaltung von Prozessen (87–119). Norderstedt: Books on Demand.

Kranz, M. (2017). Widerspruch, performativer; Widerspruch, pragmatischer. In Historisches Wörterbuch der Philosophie online. Basel: Schwabe Verlag. Abgerufen am 22.06.2021 von https://doi.org/10.24894/HWPh.4839.

Kremp, M. (2019). „Talk to Transformer": Diese künstliche Intelligenz schreibt beängstigend gut. Abgerufen am 22.06.2021 von https://www.spiegel.de/netzwelt/web/talk-to-transformer-kuenstliche-intelligenz-schreibt-texte-fertig-a-1295116.html.

Meier, C. J. (2020). Wie gefährlich ist ein Algorithmus, der schreibt wie ein Mensch? Abgerufen am 22.06.2021 von https://www.riffreporter.de/ki-fuer-alle/sprachsoftware-schreibt-wie-ein-mensch/.

Meyer, E., & Weßels, D. (2020). Original oder Plagiat? Das neue Kontinuum wissenschaftlicher Arbeiten. In Nees, F., Stengel, I., Meister, V.G., Barton, T., Herrmann, F., Müller, C. & Wolf, M. R. (Hrsg.). Angewandte Forschung in der Wirtschaftsinformatik 2020 (53–60). Heide: mana-Buch. Abgerufen am 22.06.2021 von http://akwi.de/documents/AKWI_Tagungsband2020.pdf.

Mittelstraß, J. (2019). Bildung in einer Wissensgesellschaft. *heiEDUCATION Journal. Transdisziplinäre Studien zur Lehrerbildung*, 3, 21–36. Abgerufen am 22.06.2021 von https://doi.org/10.17885/HEIUP.HEIED.2019.3.23942.

Moorstedt, M. (2020). Künstliche Intelligenz. Technologie oder Magie? Abgerufen am 22.06.2021 von https://www.sueddeutsche.de/kultur/kuenstliche-intelligenz-verfasst-texte-teilweise-auch-diesen-federhalter-1.4989758.

OECD – Organisation for Economic Co-operation and Development (Hrsg.). (2007). Best Practices for Ensuring Scientific Integrity and Preventing Misconduct. Abgerufen am 22.06.2021 von http://www.oecd.org/science/inno/40188303.pdf.

Osborne, J. W. & Holland, A. (2009). What is authorship, and what should it be? A survey of prominent guidelines for determining authorship in scientific publications. *Practical Assessment, Research, and Evaluation*, 14, Artikel 15. Abgerufen am 22.06.2021 von https://doi.org/10.7275/25pe-ba85.

Peels, R., de Ridder, J., Haven, T., & al. (2019). Value pluralism in research integrity. *Research Integrity and Peer Review*, 4(1), Artikel 18. Abgerufen am 22.06.2021 von https://doi.org/10.1186/s41073-019-0076-4.

Prentice, F. M., & Kinden, C. E. (2018). Paraphrasing tools, language translation tools and plagiarism: An exploratory study. *International Journal for Educational Integrity*, 14(1), 11. Abgerufen am 22.06.2021 von https://doi.org/10.1007/s40979-018-0036-7.

Radford, A., Wu, J., Amodei, D., & al. (2019). Better Language Models and Their Implications. Abgerufen am 22.06.2021 von https://openai.com/blog/better-language-models/.

Radford, A., Wu, J., Child, R., & al. (2019). Language models are unsupervised multitask learners. Technical report. Abgerufen am 22.06.2021 von https://cdn.openai.com/better-language-models/language_models_are_unsupervised_multitask_learners.pdf.

Resnik, D. B., & Shamoo, A. E. (2011). The Singapore Statement on Research Integrity. *Accountability in Research*, 18(2), 71–75. Abgerufen am 22.06.2021 von https://doi.org/10.1080/08989621.2011.557296.

Rogerson, A. M., & McCarthy, G. (2017). Using Internet based paraphrasing tools: Original work, patchwriting or facilitated plagiarism? *International Journal for Educational Integrity*, 13(1), 1–15. Abgerufen am 22.06.2021 von https://doi.org/10.1007/s40979-016-0013-y.

Schmohl, T., Löffl, J., & Falkemeier, G. (2019). Künstliche Intelligenz in der Hochschullehre. In Schmohl, T. & Scheffer, D. (Hrsg.), Lehrexperimente der Hochschulbildung. Didaktische Innovationen aus den Fachdisziplinen (Bd. 2, 117–122). Bielefeld: wbv. Abgerufen am 22.06.2021 von https://www.pedocs.de/volltexte/2020/18564/.

Schwemmer, O. (2013). Wert (moralisch). In: Mittelstraß, J. (Hrsg.). Enzyklopädie Philosophie und Wissenschaftstheorie. Bd. 5 (604–606). Stuttgart, Weimar: J. B. Metzler.

Scott, K. (2020). Microsoft teams up with OpenAI to exclusively license GPT-3 language model. Abgerufen am 22.06.2021 von https://blogs.microsoft.com/blog/2020/09/22/microsoft-teams-up-with-openai-to-exclusively-license-gpt-3-language-model/.

Shamoo, A. E. & Resnik, D. B. (2009). Responsible conduct of research. 2nd ed. Oxford Univ. Press.

Shaw, D. (2019). The Quest for Clarity in Research Integrity: A Conceptual Schema. *Science and Engineering Ethics*, 25(4), 1085–1093. Abgerufen am 22.06.2021 von https://doi.org/10.1007/s11948-018-0052-2.

Strobl, C., Ailhaud, E., Benetos, K., & al. (2019). Digital support for academic writing: A review of technologies and pedagogies. *Computers & Educatio*n, 131, 33–48. Abgerufen am 22.06.2021 von https://doi.org/10.1016/j.compedu.2018.12.005.

Tahiru, F. (2021). AI in Education: A Systematic Literature Review. *Journal of Cases on Information Technology*, 23(1), 1–20. Abgerufen am 22.06.2021 von https://doi.org/10.4018/JCIT.2021010101.

TH Wildau – Technische Fachhochschule Wildau (Hrsg.). (2002). Ordnung zur Sicherung guter wissenschaftlicher Praxis an der Technischen Fachhochschule Wildau. Entwickelt nach den Empfehlungen der Deutschen Forschungsgemeinschaft. Abgerufen am 22.06.2021 von https://www.th-wildau.de/files/2_Dokumente/Amtliche_Mitteilungen/13_2002_ordnung_sicherung_wissenschaftlicher_praxis.pdf.

The Royal Society (Hrsg.). (2017). Machine learning: The power and promise of computers that learn by example. Abgerufen am 22.06.2021 von https://royalsociety.org/-/media/policy/projects/machine-learning/publications/machine-learning-report.pdf.

The Royal Society, & The Alan Turing Institute (Hrsg.). (2019). The AI revolution in scientific research. Abgerufen am 22.06.2021 von https://royalsociety.org/-/media/policy/projects/ai-and-society/AI-revolution-in-science.pdf?la=en-GB&hash=5240F21B56364A00053538A0BC29FF5F.

TÜV-Verband (Hrsg.). (2020). Zur Sicherheit KI-gestützter Anwendungen. Positionspapier. Abgerufen am 22.06.2021 von https://www.vdtuev.de/positionspapiere/tuev-verband-zur-sicherheit-ki.

Université de Montréal (Hrsg.). (2018). Montréal Declaration for a Responsible Development of Artificial Intelligence. Abgerufen am 22.06.2021 von https://5dcfa4bd-f73a-4de5-94d8-c010ee777609.filesusr.com/ugd/ebc3a3_506ea08298cd4f8196635545a16b071d.pdf.

Vaswani, A., Shazeer, N., Parmar, N., & al. (2017). Attention Is All You Need. arXiv:1706.03762 [cs]. Abgerufen am 22.06.2021 von http://arxiv.org/abs/1706.03762.

VDI – Verein Deutscher Ingenieure (Hrsg.). (2002). Ethische Grundsätze des Ingenieurberufs. Abgerufen am 22.06.2021 von https://www.vdi.de/fileadmin/pages/mein_vdi/redakteure/publikationen/VDI_Ethische_Grundsaetze.pdf.

Voshmgir, S. (2020). GPT3 & Parametric Academic Writing. Abgerufen am 22.06.2021 von https://sherminvoshmgir.medium.com/gtp3-the-future-of-parametric-academic-writing-a81f02e10426.

Weßels, D. (2020). Zwischen Original und Plagiat. Abgerufen am 22.06.2021 von https://www.forschung-und-lehre.de/management/zwischen-original-und-plagiat-2754/.

World Conference on Research Integrity (Hrsg.). (2010). Singapore Statement on Research Integrity. Abgerufen am 22.06.2021 von https://www.wcrif.org/documents/327-singapore-statement-a4size/file.

Autor*innen:

Nicolaus Wilder, Dipl. Päd., wissenschaftlicher Mitarbeiter am Institut für Pädagogik der Christian-Albrechts-Universität Kiel, Projektleiter im Horizon2020-Projekt Path2Integrity, wilder@paedagogik.uni-kiel.de

Doris Weßels, Dr. rer. pol., Professorin für Wirtschaftsinformatik an der Fachhochschule Kiel, Fachgruppenleiterin „KI und Academic Writing“ beim KI-ExpertLab Hochschullehre, doris.wessels@fh-kiel.de

Johanna Gröpler, M. A., wissenschaftliche Mitarbeiterin TH Wildau College/Schreibwerkstatt, Projektmitarbeiterin in der Hochschulbibliothek der Technischen Hochschule Wildau, johanna.groepler@th-wildau.de

Andrea Klein, Dr., selbstständige Dozentin, Coach, Autorin und Forscherin zum Thema Wissenschaftliches Arbeiten, Gründerin des Online-Kongresses Studienfeuer, www.wissenschaftliches-arbeiten-lehren.de

Margret Mundorf, M.A., freie Schreib- & Kommunikationstrainerin, Lehrbeauftragte, aktiv

in SIG Digitalität/SIG Schreibforschung der gefsus e. V., KI-ExpertLab Academic Writing, margret.mundorf@hs-kl.de