

Steffen Albrecht

# ChatGPT als doppelte Herausforderung für die Wissenschaft

Eine Reflexion aus der Perspektive der Technikfolgenabschätzung

**Abstract:** The public release of ChatGPT in November 2022 has shaken up the academic world. Many see AI-based text generators as a potential threat to the integrity and competences of scientists, but there is also a huge potential to improve research not only in the quantitative sciences, but also in the arts and in qualitative social sciences. This contribution presents the results of an investigation into the opportunities and limitations of ChatGPT, and its implications for different areas of society, from a technology assessment perspective. Furthermore, it discusses the challenges of using systems like ChatGPT in scientific policy advice and other academic fields.

## 1 Einleitung

Als ChatGPT im November 2022 veröffentlicht wurde, waren Wissenschaftler\*innen nicht nur unter den ersten, die davon Notiz nahmen, auch die Auswirkungen und das disruptive Potenzial dieses so gänzlich neuen KI-basierten Chatbots wurden in der Wissenschaft besonders früh spürbar. Studierende nutzten das System, um Unterstützung bei der Abfassung ihrer Hausarbeiten zu bekommen oder diese gleich in Gänze schreiben zu lassen,<sup>1</sup> Forschende überließen dem Chatbot die Autorschaft über Abschnitte ihrer Publikationen<sup>2</sup> und es zeigte sich, dass ChatGPT verschiedene standardisierte Prüfungen an Hochschulen mit respektablen Ergebnissen absolvieren kann.<sup>3</sup>

---

1 Stephen Marche, „The College Essay Is Dead“, *The Atlantic*, 2022, in [theatlantic.com/technology/archive/2022/12/chatgpt-ai-writing-college-student-essays/672371/] (Zugriff: 15.10.2023).

2 Debby R. E. Cotton/Peter A. Cotton/J. Reuben Shipway, „Chatting and cheating: Ensuring academic integrity in the era of ChatGPT“, *Innovations in Education and Teaching International*, 2023 (DOI: 10.1080/14703297.2023.2190148).

3 Jonathan H. Choi et al., „ChatGPT Goes to Law School“, *Minnesota Legal Studies Research Paper*, Bd. 23(3), 2023 (DOI: 10.2139/ssrn.4335905); Tiffany H. Kung et al., „Performance of ChatGPT on USMLE. Potential for AI-Assisted Medical Education Using Large Language Models“, *PLOS Digital Health*, 2023 (DOI: 10.1371/journal.pdig.0000198).

ChatGPT bildet, wie auch andere sogenannte „foundation models“, die Grundlage für eine Reihe von Anwendungen, die auf den Leistungen der KI-Textgeneratoren aufbauen, wie etwa Übersetzungsdienste, Programme zur Lernunterstützung oder für die wissenschaftliche Recherche. Die Modelle sind aber auch in dem Sinne grundlegend, dass sie Auswirkungen auf die menschliche Kommunikation und damit die Grundlage des gesellschaftlichen Zusammenlebens haben. Schon bald begann daher in Wissenschaft und Öffentlichkeit eine Debatte darüber, welche Folgen KI-Textgeneratoren direkt oder indirekt und auch in längerfristiger Perspektive haben werden.

Mit diesen Fragen ist auch die Technikfolgenabschätzung (TA) auf den Plan gerufen, die Potenziale und mögliche Risiken neuer wissenschaftlicher und technologischer Entwicklungen mit wissenschaftlichen Methoden untersucht. Wie viele wissenschaftliche Disziplinen, die sich mit KI-Textgeneratoren beschäftigen, ist TA in doppelter Weise durch die Entwicklungen herausgefordert. Zum einen bietet sich ihr ein neuer Untersuchungsgegenstand, eine Technologie mit potenziell weitreichenden gesellschaftlichen Folgen. Zum anderen wirkt sich diese Technologie aber auch auf die wissenschaftliche Arbeit der TA und die Rahmenbedingungen von TA aus. Dieser doppelten Herausforderung geht der vorliegende Beitrag nach.

Einleitend soll zunächst die besondere Perspektive der TA vorgestellt werden. TA bildet zwar keine eigenständige Disziplin im engeren Sinn, ist aber als reflexive wissenschaftliche Praxis etabliert und in Deutschland wie auch in vielen anderen Ländern institutionalisiert. Es gibt mehrere Einrichtungen, die sich schwerpunktmäßig mit TA beschäftigen, darunter sowohl Hochschulinstitute als auch Abteilungen in Unternehmen und Beratungseinrichtungen, etwa bei Parlamenten (wie das Büro für Technikfolgen-Abschätzung beim Deutschen Bundestag [TAB]). Außerdem besteht ein Kanon von Methoden und theoretischen Grundlagen, der nicht zuletzt in spezialisierten Vereinigungen, Tagungen und Zeitschriften diskutiert und weiterentwickelt wird.<sup>4</sup>

Als reflexiv kann TA insofern gelten, als sie neue wissenschaftliche und/oder technologische Entwicklungen zum Gegenstand hat und deren Auswirkungen auf die Gesellschaft wissenschaftlich untersucht – welchen Nutzen sie versprechen, welche Risiken möglicherweise mit ihnen verbunden sind, und wie sich die weitere Entwicklung beeinflussen lässt. TA hat dabei den Anspruch, möglichst unvoreingenommen auf sowohl Chancen wie auch Risiken zu schauen und durch die Überparteilichkeit des „ehrlichen Maklers“<sup>5</sup> auch ein Gegengewicht zu Hype, übertriebenen Erwartungen und einseitiger Vereinnahmung zu bilden.

---

4 Armin Grunwald, *Technikfolgenabschätzung*, Baden-Baden <sup>3</sup>2022, 70 ff.

5 Roger A. Pielke, *The Honest Broker. Making Sense of Science in Policy and Politics*, Cambridge 2007.

Ein wichtiges Merkmal der TA ist die interdisziplinäre Perspektive. Die Technologien werden nicht nur aus einer technischen, sondern auch zum Beispiel aus gesellschafts- oder umweltwissenschaftlicher Perspektive betrachtet, um ein möglichst umfassendes Bild zu gewinnen – ähnlich der in diesem Tagungsband vertretenen Vielfalt in Bezug auf KI-Textgeneratoren. Neben der Berücksichtigung wissenschaftlicher Disziplinen werden auch Positionen von Betroffenen und Stakeholdern außerhalb der Wissenschaft berücksichtigt, da es bei Fragen zu Auswirkungen von Technologien auch um solche der Akzeptanz und der praktischen Nutzung im Alltag geht. In diesem Sinne ist die TA auch transdisziplinär – es geht häufig, insbesondere in der sogenannten konstruktiven TA, darum, gemeinsam mit Akteuren der Praxis eine neue Technologie bzw. die Rahmenbedingungen ihrer Entwicklung und Anwendung zu gestalten.

Da die allermeisten Quellen in TA-Untersuchungen wissenschaftliche Texte sind, wirken sich KI-Textgeneratoren zumindest potenziell sehr weitreichend auf die TA aus. Während es im Oktober 2022 noch absurd geklungen haben mag, einen Chatbot für wissenschaftliche Studien zu nutzen (außer als Gegenstand derselben), gibt die Wissenschaftszeitschrift *Nature* nur ein Jahr später bereits Tipps, welcher Chatbot sich am besten für welche Art von Forschungsaufgabe eignet.<sup>6</sup> Die Arbeitsschritte bei TA-Untersuchungen bestehen insbesondere in der Erfassung, Bewertung und zielgruppenorientierten Umformulierung von meist in Textform vorliegenden Erkenntnissen und liegen damit im Funktionsbereich der KI-Textgeneratoren. Daher sollen in diesem Beitrag neben den gesellschaftlichen Auswirkungen von KI-Textgeneratoren (Abschnitt 2) auch ihr Einfluss auf die Arbeit und die Rahmenbedingungen von entsprechenden Studien reflektiert werden (Abschnitt 3), um besser zu verstehen, welche Auswirkungen sich für wissenschaftliche Diskurse und die wissenschaftliche (Text-)Arbeit aus einer zunehmenden Nutzung von KI-Textgeneratoren ergeben.

## 2 KI-Textgeneratoren als Gegenstand der Technikfolgenabschätzung

Als Gegenstand von TA lassen sich KI-Technologien wie die ChatGPT zugrunde liegenden Textgeneratoren (auch als Large Language Models bezeichnet) mit dem bewährten Instrumentarium der TA analysieren. Dabei werden anhand der technischen Grundlagen und möglicher Anwendungsszenarien der Technologie poten-

---

<sup>6</sup> Elizabeth M. Humphries et al., „What’s the best chatbot for me? Researchers put LLMs through their paces“, *Nature Careers*, 2023, in [nature.com/articles/d41586-023-03023-4] (Zugriff: 15.10.2023).

zielle Implikationen untersucht, meist spielen dabei ethische, rechtliche und soziale Aspekte eine Rolle, außerdem der Blick auf ökologische Auswirkungen. Mit einer solchen Untersuchung zu ChatGPT wurde das TAB zu Beginn des Jahres 2023 vom Deutschen Bundestag beauftragt. Der Fokus der Untersuchung lag zu diesem frühen Zeitpunkt weniger auf der Identifizierung von Handlungsoptionen als vielmehr darauf, eine Übersicht der verfügbaren Erkenntnisse aus möglichst unterschiedlichen Perspektiven zu schaffen. Mit der Veröffentlichung eines Hintergrundpapiers,<sup>7</sup> dessen Ergebnisse im Folgenden zusammengefasst werden, und der Veranstaltung eines öffentlichen Fachgesprächs leistete der Ausschuss für Bildung, Forschung und Technikfolgenabschätzung des Deutschen Bundestages auch einen Beitrag zur Weiterentwicklung der öffentlichen Debatte über ChatGPT.

## 2.1 Technologische Grundlagen

In technischer Hinsicht stellt ChatGPT<sup>8</sup> einerseits eine Fortsetzung der Entwicklung bei sprachverarbeitenden KI-Systemen dar, andererseits auch einen technologischen Sprung, von dem selbst Expert\*innen überrascht waren. Die Architektur von ChatGPT geht auf das sogenannten Transformermodell zurück, das bereits 2017 von einem Forschungsteam von Google vorgestellt wurde. Transformer zeichnen sich dadurch aus, dass sie Texte nicht allein sequenziell abarbeiten, sondern auch Verbindungen zwischen weit entfernt stehenden Wörtern erkennen und dadurch sprachliche Kontexte berücksichtigen können. Zudem können für das Training parallel arbeitende Grafikprozessoren eingesetzt werden, so dass sehr große Modelle mit einer Vielzahl von Parametern und einer Unmenge von Trainingsdaten technisch möglich werden – im Fall von GPT 3.5 175 Mrd. Parameter und 300 Mrd. Textbestandteile Trainingsmaterial.

Als vierter Aspekt neben der Architektur, den umfangreichen Hardwareressourcen und den Datenmengen ist bei ChatGPT das Feintuning des Systems auf Konversationsfähigkeit hin zu beachten, es gilt als „Erfolgsrezept“ für die große Verbreitung des Systems.<sup>9</sup> Nach einem ersten, weitgehend automatisierten Training

---

7 Steffen Albrecht, *ChatGPT und andere Computermodele zur Sprachverarbeitung – Grundlagen, Anwendungspotenziale und mögliche Auswirkungen. Hintergrundpapier Nr. 26*, Berlin 2023.

8 Hier beziehe ich mich auf die zunächst veröffentlichte GPT-Version 3.5, die – im Unterschied zur kostenpflichtigen Variante – weiterhin ohne größere Zugangsbeschränkungen nutzbar ist, allerdings ebenfalls weiterentwickelt wurde.

9 Will D. Heaven, „The inside story of how ChatGPT was built from the people who made it“, *MIT Technology Review*, 2023, in [technologyreview.com/2023/03/03/1069311/inside-story-oral-history-how-chatgpt-built-openai/] (Zugriff: 15.10.2023). Bei keinem Onlinedienst war bisher ein vergleichbares

mit großen Mengen von Texten aller möglichen Sprachen und Genres wurde das Modell weiter justiert, um möglichst generisch und flexibel auf ganz unterschiedliche sprachliche Eingaben (die sogenannten Prompts) reagieren zu können. Dazu wurden die Antworten des Systems mithilfe von menschlichem Feedback<sup>10</sup> so angepasst, dass unerwünschte, etwa rassistische, sexistische oder illegale Aktivitäten befördernde Äußerungen unterdrückt und möglichst passende Reaktionen auf die Prompts generiert werden.

## 2.2 Möglichkeiten...

So beeindruckend die Leistung von ChatGPT und vergleichbaren Systemen auch ist, sie stellen nur eine Form von KI-Systemen dar, neben anderen auf Deep Learning oder auch symbolischer künstlicher Intelligenz beruhenden Systemen, die alle konstruktionsbedingte Möglichkeiten und Grenzen haben. Hinsichtlich ihrer Möglichkeiten zeichnen sich KI-Textgeneratoren durch die Breite ihrer Fähigkeiten aus. KI-Textgeneratoren stellen „powerful generalist models“ für ganz unterschiedliche sprachliche Aufgaben dar.<sup>11</sup> Dies umfasst die Erzeugung von Texten mit beachtlicher Länge<sup>12</sup> zu einem bestimmten Thema in ganz unterschiedlichen Sprachen,

---

Wachstum beobachtet worden: Innerhalb von fünf Tagen nach der Veröffentlichung am 30. November 2022 meldeten sich 1 Mio. Nutzende an, am 2. Februar 2023 wurden 100 Mio. aktive Nutzende/Monat gemeldet, vergleiche dazu: [www.tagesschau.de/wirtschaft/digitales/chatgpt-wachstum-bezahlangebot-abo-101.html] (Zugriff: 15.10.2023). Bei TikTok beispielsweise wurde die Schwelle von 100 Mio. aktiven Nutzenden nach etwa neun Monaten erreicht, bei Instagram nach zweieinhalb Jahren (Oliver Bünte (dpa), „ChatGPT: Rekord-Wachstum und Abo-Modell in den USA“, *heise online*, 2023, in [heise.de/news/ChatGPT-startet-Abo-Modell-in-den-USA-7480052.html] (Zugriff: 15.10.2023). Mittlerweile scheint allerdings eine Sättigung erreicht, zumindest die Nutzung von ChatGPT über das Webinterface ging laut Zahlen der Analysefirma Similarweb im Sommer 2023 zurück, vgl. David F. Carr, „ChatGPT Drops About 10 % in Traffic as the Novelty Wears Off“, *similarweb Blog*, in [www.similarweb.com/blog/insights/ai-news/chatgpt-traffic-drops/] (Zugriff: 15.10.2023).

**10** Offenbar erfolgte diese Arbeit in Niedriglohnländern unter prekären Arbeitsbedingungen, vgl. hierzu Reed Albergotti/Louise Matsakis, „OpenAI has hired an army of contractors to make basic coding obsolete“, *Semafor*, 2023, in [www.semafor.com/article/01/27/2023/openai-has-hired-an-army-of-contractors-to-make-basic-coding-obsolete] (Zugriff: 15.10.2023); Billy Perrigo, „Exclusive: OpenAI Used Kenyan Workers on Less Than \$2 Per Hour to Make ChatGPT Less Toxic“, *TIME*, 2023, in [https://time.com/6247678/openai-chatgpt-kenya-workers] (Zugriff: 15.10.2023).

**11** Chengwei Qin et al., *Is ChatGPT a General-Purpose Natural Language Processing Task Solver?*, arXiv, 2023, 11.

**12** Aktuelle KI-Textgeneratoren sind in der Lage bis zu 75.000 Wörter Text als Eingabe zu verarbeiten und mehrere Tausend Wörter Text auszugeben, in [heise.de/news/Claude-2-ist-da-Anthropiveroeffentlicht-neue-Version-des-Chatbots-und-der-API-9214163.html] (Zugriff: 15.10.2023).

Stilen und Niveaus, ebenso die Übersetzung zwischen diesen und andere Formen der Bearbeitung wie die Analyse, Zusammenfassung oder Korrektur gegebener Texte. Durch die Transformer-Architektur können die Systeme den Kontext von Ausdrücken berücksichtigen und sie beispielsweise einem übergeordneten Thema zuzuordnen. Die Fähigkeit, Texte über unterschiedliche Sprachen hinweg zu verarbeiten, erstreckt sich auch auf andere regelhaft verwendete Codes wie Programmiersprachen – die Erzeugung von Programmcode hat sich als eine der vielversprechendsten Anwendungsmöglichkeiten von KI-Textgeneratoren herausgestellt. Neuere Versionen sind in der Lage, auch mit gesprochener Sprache und multimodal zu arbeiten, d. h. Inhalte in Text- oder aber Bildform zu verarbeiten bzw. auszugeben.

### 2.3 ...und Grenzen

Trotz ihrer generativen Fähigkeiten sind KI-Textgeneratoren, wie andere KI-Systeme, durch ihr Training auf bestimmte Aufgaben ausgerichtet, in diesem Fall die Verarbeitung sprachlicher Texte im Rahmen des Materials, mit dem sie trainiert (und feinjustiert) wurden. Auch wenn sie Konversationsfähigkeit simulieren können, haben sie kein Verständnis der verarbeiteten Texte,<sup>13</sup> keinen Bezug zu der Welt, um die es in den Texten geht, und Schwierigkeiten mit Abstraktion und logischen Problemstellungen. Es handelt sich um Sprachmodelle, keine Wissensmodelle.<sup>14</sup> Sie antworten in der Regel auch dann, wenn sie zu einer Frage oder Aufforderung keine passende Antwort haben, und produzieren überzeugend formulierte, aber frei erfundene Inhalte ohne verlässlichen Bezug zur Realität. Dies schließt auch Quellenangaben in wissenschaftlichen Texten ein, die sich bei einer kritischen Überprüfung mitunter als fiktiv herausstellen. Reproduzierbarkeit, wie sie für wissenschaftliche Zwecke erwartet wird, ist bei KI-Textgeneratoren nicht gegeben, sie können auf ein und denselben Prompt unterschiedliche Antworten liefern. Bei längeren Konversationen tendieren die Modelle dazu, den Fokus zu

---

13 Emily M. Bender/Alexander Koller, „Climbing towards NLU: On Meaning, Form, and Understanding in the Age of Data“, *Association for Computational Linguistics*, Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, July, 2020, 5185–5198; Celeste Biver, „The easy intelligence tests that AI chatbots fail“, *Nature*, Bd. 619(7971), 2023, 686–689.

14 Reinhard Heil, „Einige ethische Implikationen großer Sprachmodelle“, *KIT Scientific Working Papers* 221, 2023, 1–2 (DOI: 10.5445/IR/1000158914).

verlieren und immer stärker vom erwünschten Kommunikationsverhalten abzuweichen.<sup>15</sup>

Eine Abhängigkeit vom Training bzw. den dabei verwendeten Daten zeigt sich auch darin, dass die Systeme solche Sprachen besonders gut beherrschen, die besonders häufig im Trainingsmaterial vorkamen. Auch die Inhalte, die von KI-Textgeneratoren erzeugt werden, spiegeln z.T. deutlich das Trainingsmaterial wider, was sich etwa in verzerrten Repräsentationen bestimmter sozialer Gruppen zeigt<sup>16</sup> sowie in einer generell konservativen Tendenz der Systeme, bestehende, in den Trainingsdaten aufgefundene Muster zu wiederholen und zu verstärken.<sup>17</sup> Auch wenn im Training gezielt Vorkehrungen gegen einen solchen Bias getroffen werden, zeigt sich, dass diese Sicherheitsvorkehrungen von den Nutzenden umgangen werden können oder aber die Leistungsfähigkeit des Systems einschränken.

Weitere Grenzen, die sich aus der Konstruktion der Systeme ergeben, betreffen deren Transparenz und Nachhaltigkeit. So ist in Bezug auf den Schutz persönlicher Daten kaum transparent, welche Daten Eingang in die Modelle gefunden haben und wie sie in ihnen repräsentiert sind.<sup>18</sup> Außerdem wird davor gewarnt, ChatGPT sensible Daten im Zuge der Nutzung zu übermitteln, da nicht klar ist, in welcher Form diese durch das System bzw. seine Betreiber verwertet werden.<sup>19</sup> KI-Textgeneratoren benötigen aufgrund ihrer hohen Hardwareansprüche beim Training große Mengen an Ressourcen, was unter anderem zu hohen klimaschädlichen Emissionen führt.<sup>20</sup> Auch wenn Rechenzentren an Strategien für effizientere Pro-

---

15 Kevin Roose, „Bing’s A.I. Chat: ‚I Want to Be Alive‘“, *New York Times*, 2023, in [nytimes.com/2023/02/16/technology/bing-chatbot-transcript.html] (Zugriff: 15.10.2023).

16 Davey Alba, „OpenAI Chatbot Spits Out Biased Musings, Despite Guardrails“, *Bloomberg*, 2022, in [www.bloomberg.com/news/newsletters/2022-12-08/chatgpt-open-ai-s-chatbot-is-spitting-out-biased-sexist-results] (Zugriff: 15.10.2023).

17 Cory Doctorow, „Our Neophobic, Conservative AI Overlords Want Everything to Stay the Same“, *Blog/LA Review of Books*, 2020, in [https://blog.lareviewofbooks.org/provocations/neophobic-conservative-ai-overlords-want-everything-stay/] (Zugriff: 15.10.2023); Paola Lopez, „ChatGPT und der Unterschied zwischen Form und Inhalt“, *Merkur*, Bd. 77(891), 2023, 21–22.

18 Uri Gal, „ChatGPT is a data privacy nightmare. If you’ve ever posted online, you ought to be concerned“, *The Conversation*, 2023, in [https://theconversation.com/chatgpt-is-a-data-privacy-nightmare-if-youve-ever-posted-online-you-ought-to-be-concerned-199283] (Zugriff: 15.10.2023).

19 Melissa Heikkilä, „Could ChatGPT do my job?“, *MIT Technology Review*, 2023, in [www.technologyreview.com/2023/01/31/1067436/could-chatgpt-do-my-job/] (Zugriff: 15.10.2023).

20 Emily M. Bender et al., „On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?“, *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, 2021, 610–623; Pengfei Li et al., *Making AI Less ‚Thirsty‘: Uncovering and Addressing the Secret Water Footprint of AI Models*, arXiv, 2023.

zesse arbeiten,<sup>21</sup> stellt das derzeitige Wachstum sowohl der Größe der Systeme als auch ihrer Nutzung in ökologischer Hinsicht ein Problem dar.

## 2.4 Auf die Anwendung kommt es an!

Was technisch möglich oder nicht möglich ist, steckt nur den Rahmen ab, in dem das Potenzial von KI-Textgeneratoren durch die konkrete Nutzung ausgeschöpft werden kann oder in dem sich möglicherweise unerwünschte Effekte ergeben. Die bisherigen innovativen Ideen für die Nutzung von KI-Textgeneratoren zeigen, dass dieser Rahmen sehr weit gesteckt ist. Neben Programmierung, kreativen Tätigkeiten und Anwendungen in Unternehmen werden im TAB-Hintergrundpapier Anwendungen in den Bereichen Gesundheit, Information und Öffentlichkeit, Rechtswesen, öffentlicher Verwaltung sowie Bildung und Forschung dargestellt und ihre möglichen Auswirkungen untersucht. Da sich die Entwicklung noch immer rasant vollzieht und immer neue Anwendungsmöglichkeiten erdacht oder realisiert werden, lässt sich nur skizzieren, welche Implikationen mit der Nutzung von KI-Textgeneratoren in den jeweiligen Bereichen verbunden sind. Dies soll im Folgenden exemplarisch anhand der Bereiche Gesundheitsversorgung, öffentliche Kommunikation und Forschung gezeigt werden.

KI-Anwendungen für die Verbesserung der *Gesundheitsversorgung* werden schon seit längerem erforscht, bisher konnte in einzelnen Bereichen wie etwa der Radiologie ein Nutzen nachgewiesen werden.<sup>22</sup> In Bezug auf KI-Textgeneratoren werden hohe Erwartungen formuliert, etwa was die Unterstützung ärztlicher Diagnosen oder den Zugang zu therapeutischen Gesprächen bei psychischen Problemen angeht.<sup>23</sup> Bisherige KI-gestützte Assistenzsysteme haben die Erwartungen jedoch nicht erfüllen können,<sup>24</sup> die mangelnde Genauigkeit von KI-Textgeneratoren in Bezug auf Fakten wirft auch für die Zukunft Zweifel auf. Der Einsatz von KI-basierten Chatbots zur Diagnose und Therapieunterstützung bei psychischen Pro-

---

21 Reinhard Grünwald/Claudio Caviezel, *Energieverbrauch der IKT-Infrastruktur*, Berlin 2022.

22 Deutscher Ethikrat, *Mensch und Maschine – Herausforderungen durch Künstliche Intelligenz. Stellungnahme*, Berlin 2023, 140 ff.; Katrin Gerlinger, *Data-Mining – gesellschaftspolitische und rechtliche Herausforderungen*, Berlin 2022.

23 Rob Toews, „A Wave Of Billion-Dollar Language AI Startups Is Coming“, *Forbes*, 2022, in [forbes.com/sites/robtoews/2022/03/27/a-wave-of-billion-dollar-language-ai-startups-is-coming/] (Zugriff: 15.10.2023); Ron Li/Andre Kumar/Jonathan H. Chen, „How Chatbots and Large Language Model Artificial Intelligence Systems Will Reshape Modern Medicine. Fountain of Creativity or Pandora’s Box?“, *JAMA Internal Medicine*, Bd. 183(6), 2023, 596–597.

24 Gerlinger 2022, 191.



blemen bietet zwar durchaus Potenziale, etwa in Fällen, in denen keine Therapie verfügbar ist oder wenn Betroffene es bevorzugen, sich einer Maschine anstelle eines Menschen anzuvertrauen.<sup>25</sup> Allerdings ist die Qualitätskontrolle bislang mangelhaft, die Funktionsweise und der Umgang mit persönlichen Daten intransparent und es fehlt an einer Anbindung an das soziale Umfeld der Nutzenden, um im Falle einer Krise reagieren zu können.<sup>26</sup> Eine zunehmende Verbreitung KI-basierter Chatbots könnte zudem einen Anstieg der Zahl psychischer Erkrankungen mit sich bringen, etwa durch eine übermäßige Abhängigkeit von den als Kommunikationspartner angesehenen Systemen.

In Hinblick auf *öffentliche Kommunikation* wird unter anderem das Szenario diskutiert, dass automatisiert erstellte Beiträge den öffentlichen Diskurs zunehmend beeinflussen. Solche Beiträge spielen bereits seit einigen Jahren eine Rolle im Journalismus, könnten jedoch durch KI-Textgeneratoren eine neue Dimension erreichen. Erste Medienhäuser haben ChatGPT bereits eingesetzt (und Journalist\*innen ersetzt), die dabei auftretenden Falschinformationen und Plagiate haben aber zu negativer Resonanz geführt.<sup>27</sup> In einem Experiment des Bayerischen Rundfunks mit einem Vorgängermodell von ChatGPT erwiesen sich die Texte als so überarbeitungsbedürftig, dass der Einsatz nicht effizient erschien.<sup>28</sup>

Die Verlockung, die journalistische Arbeit etwa bei datenintensiven Recherchen zu erleichtern bzw. Nischenthemen erschließen zu können, dürfte allerdings anhalten. Und da ChatGPT allgemein zugänglich ist, wird erwartet, dass öffentliche Kommunikation in traditionellen wie auch sozialen Medien in zunehmendem Maß durch KI-Textgeneratoren bestimmt wird. Damit verbunden ist das Risiko einer Zunahme von gezielt oder unbewusst verbreiteter Desinformation, manipulativen Inhalten oder Deepfakes,<sup>29</sup> was wiederum zu einem Vertrauensverlust in die öf-

---

<sup>25</sup> Albrecht 2023, 56.

<sup>26</sup> Deutscher Ethikrat 2023, 158.

<sup>27</sup> Matthias Bastian, „CNET investigation shows lots of flaws in AI-written articles“, *the decoder*, 2023, in [<https://the-decoder.com/cnet-investigation-shows-lots-of-flaws-in-ai-written-articles/>] (Zugriff: 15.10.2023); Jon Christian, „CNET’s AI Journalist Appears to Have Committed Extensive Plagiarism“, *Futurism*, 2023, in [<https://futurism.com/cnet-ai-plagiarism>] (Zugriff: 15.10.2023); Jon Christian, „Magazine Publishes Serious Errors in First AI-Generated Health Article“, *Neoscope*, 2023, in [<https://futurism.com/neoscope/magazine-mens-journal-errors-ai-health-article>] (Zugriff: 15.10.2023).

<sup>28</sup> Marco Lehner, „Sind Künstliche Intelligenzen die besseren Journalist:innen?“, *Medium*, 2021, in [<https://medium.com/br-next/sind-künstliche-intelligenzen-die-besseren-journalist-innen-53e6a4e70d>] (Zugriff: 15.10.2023).

<sup>29</sup> Anil Ananthaswamy, „In AI, is bigger better?“, *Nature*, Bd. 615(7951), 2023, 202–205, 203; Michael Atleson, „Chatbots, deepfakes, and voice clones: AI deception for sale“, *Federal Trade Commission*, 2023, in [[www.ftc.gov/business-guidance/blog/2023/03/chatbots-deepfakes-voice-clones-ai-deception](http://www.ftc.gov/business-guidance/blog/2023/03/chatbots-deepfakes-voice-clones-ai-deception)]

fentliche Kommunikation führen kann.<sup>30</sup> Als positive Vision lässt sich darauf verweisen, dass die Anwendung von KI-Textgeneratoren als Übersetzungshilfe die Kommunikation zwischen unterschiedlichen Sprachgemeinschaften ebenso wie zwischen Menschen unterschiedlicher Sprachkompetenz erleichtern und zu einem stärkeren Austausch zwischen diesen Gruppen beitragen kann.

Im Bereich der *Forschung* wird schließlich die Technologie der KI-Textgeneratoren erprobt, um Experimente zu designen und zu automatisieren<sup>31</sup> und um Strukturen und Muster in Daten, etwa der biomedizinischen Forschung, zu identifizieren.<sup>32</sup> Während die Systeme in diesen Fällen Erkenntnis schaffen sollen, werden sie in vielen Fällen bereits als Hilfsmittel genutzt.<sup>33</sup> Sie können Übersetzungen und sprachliche Verbesserungen von Texten liefern oder die Recherche erleichtern, indem sie Zusammenfassungen erstellen und Suchfunktionen ergänzen. Sie können Code für wissenschaftliche Software generieren oder Anregungen für Fragestellungen geben, Forschungslücken identifizieren und Antragstexte verbessern. Auch die Unterstützung von Reviewprozessen oder die Auswertung und Verschriftlichung von Forschungsdaten werden diskutiert.<sup>34</sup> Eine Autorschaft von KI-Textgeneratoren wird bislang abgelehnt mit der Begründung, nur Menschen könnten die Verantwortung für die veröffentlichten Ergebnisse übernehmen.<sup>35</sup>

---

sale] (Zugriff: 15.10.2023); Europol, *ChatGPT. The impact of Large Language Models on Law Enforcement*, 2023, 11.

30 Deutscher Ethikrat 2023, 216.

31 Daniil A. Boiko/Robert MacKnight/Gabe Gomes, *Emergent autonomous scientific research capabilities of large language models*, arXiv, 2023.

32 Hugo Dalla-Torre et al., „The Nucleotide Transformer: Building and Evaluating Robust Foundation Models for Human Genomics“, *bioRxiv*, 2023 (DOI: 10.1101/2023.01.11.523679); Zeming Lin et al., „Evolutionary-scale prediction of atomic-level protein structure with a language model“, *Science*, Bd. 379(6637), 2023, 1123–1130.

33 In den textbasierten Wissenschaften können die beiden Funktionen zusammenfallen, da beispielsweise eine Übersetzung bereits ein Teil des Forschungsprozesses ist. Einen Überblick über Nutzungsmöglichkeiten als Hilfsmittel geben Benedikt Fecher et al., „Friend or Foe? Exploring the Implications of Large Language Models on the Science System“, *HIIG Preprint*, 2023, 4–5 (DOI: 10.5281/zenodo.8009429).

34 Matthew Salganik, „Can ChatGPT—and its successors—go from cool to tool?“, *Freedom to tinker*, 2023, in [https://freedom-to-tinker.com/2023/03/08/can-chatgpt-and-its-successors-go-from-cool-to-tool/] (Zugriff 15.10.2023); Gemma Conroy, „Scientists used ChatGPT to generate a whole paper from data“, *Nature*, Bd. 619(7970), 2023, 443.

35 Nature Editorial, „Tools such as ChatGPT threaten transparent science; here are our ground rules for their use“, *Nature*, Bd. 613(7945), 2023, 612; H. Holden Thorp, „ChatGPT is fun, but not an author“, *Science*, Bd. 379(6630), 2023, 313.

Viele der bisher genannten problematischen Aspekte einer Nutzung von ChatGPT betreffen auch die Forschung,<sup>36</sup> zudem könnten sich bestehende Probleme der Wissenschaft wie der Publikationsdruck, die Arbeitslast, der Konkurrenzdruck und die Schwierigkeit, hochwertige Informationen von minderwertigen zu unterscheiden, noch verstärken.<sup>37</sup> Eine offene Frage ist, wie sich die Nutzung von KI-Textgeneratoren auf die Ungleichheit im Wissenschaftssystem auswirkt. Zum einen könnte etwa die sprachliche Benachteiligung nicht-englischsprachiger Forschender gemildert werden, zum anderen könnten sich hinsichtlich des Zugangs zu leistungsfähigen KI-Systemen sowie der für ihren effektiven Einsatz benötigten Kompetenzen neue Benachteiligungen einstellen (bzw. alte verstärkt werden).

Inwieweit sich die hier skizzierten gesellschaftlichen Implikationen einer breiteren Anwendung von KI-Textgeneratoren realisieren, ist auch ein Jahr nach Veröffentlichung von ChatGPT noch nicht absehbar. Während die Technologie weiterentwickelt und dabei vielfältiger wurde, verlief diese Entwicklung evolutionär im Rahmen der Möglichkeiten und Grenzen der bisherigen Systeme. Die anfangs sehr intensive öffentliche Debatte hat sich ebenfalls ausdifferenziert, an die Stelle des massenmedialen Hypes sind verschiedene spezialisierte Diskussionen in den jeweils betroffenen gesellschaftlichen Teilbereichen getreten. Ein großes gesellschaftliches Interesse an der neuen Technologie und die Bereitschaft, sich damit auseinanderzusetzen, kann weiter angenommen werden. Dies gilt auch für den politischen Bereich. So wurden neben Bemühungen um eine gesetzliche Regulierung insbesondere auf Ebene der EU mehrere Ansätze zu einer globalen Beobachtung der Entwicklung gestartet, unter anderem unter dem Dach der Vereinten Nationen.<sup>38</sup>

---

**36** Einen Überblick geben Abeba Birhane et al., „Science in the age of large language models“, *Nature Review Physics*, Bd. 5, 2023, 277–280.

**37** Tony Ho Tran, „A Doctor Published Several Research Papers With Breakneck Speed. ChatGPT Wrote Them All“, *Daily Beast*, 2023, in [<https://www.thedailybeast.com/how-this-doctor-wrote-dozens-of-science-papers-with-chatgpt>] (Zugriff: 15.10.2023).

**38** United Nations, *High-Level Advisory Body on Artificial Intelligence*, in [[www.un.org/techenvoy/ai-advisory-body](http://www.un.org/techenvoy/ai-advisory-body)] (Zugriff: 15.10.2023); Ian Bremmer/Mustafa Suleyman, „The AI Power Paradox Can States Learn to Govern Artificial Intelligence—Before It’s Too Late?“, *Foreign Affairs*, Bd. 102(5), 2023, 26–43; Lewis Ho et al., *International Institutions for Advanced AI*, arXiv, 2023.

### 3 Die Rolle von KI-basierten Textgeneratoren für die Praxis der Technikfolgenabschätzung

Außer als Gegenstand von TA-Untersuchungen können KI-Textgeneratoren die Arbeitsweise der TA auf zwei Ebenen beeinflussen. Zum einen können sie unmittelbar die Arbeit verändern, indem sie in zunehmendem Maß als (Hilfs-)Mittel genutzt werden. Da die Arbeitsschritte bei TA-Untersuchungen in erster Linie die Bearbeitung von Textmaterialien betreffen, kommt eine Nutzung in nahezu allen Bereichen in Betracht. Zum anderen kann eine zunehmende Verbreitung der Systeme in der Wissenschaft oder (im Fall von TA als Politikberatung) in der Politik die Rahmenbedingungen von TA-Untersuchungen verändern und sich indirekt auf die Arbeit auswirken.

Potenziale von KI-Textgeneratoren speziell im Bereich der wissenschaftlichen Politikberatung werden insbesondere bei der Recherche wissenschaftlicher Evidenz und der Erstellung von Berichten gesehen.<sup>39</sup> Im Fall von Recherchen könnten KI-Textgeneratoren Zeit sparen, indem sie große Mengen von wissenschaftlichen oder medialen Texten durchsuchen, auf die in ihnen verhandelten Fragestellungen, Methoden und Ergebnissen hin analysieren und so die Erstellung von Literaturübersichten zum Forschungsstand oder von Analysen des medialen Diskurses erleichtern oder dabei helfen, unberücksichtigte Perspektiven zu entdecken. Dabei könnte auch eine größere Zahl von Sprachen berücksichtigt werden, als es bei allein menschlicher Bearbeitung möglich ist. Berichte für Entscheidungsträger\*innen folgen meist einem inhaltlichen Muster, werden allerdings unter hohem Zeitdruck erstellt. Hier könnten KI-Textgeneratoren für erste Entwürfe einzelner Abschnitte (Management Summary, Zusammenfassung des Forschungsstands oder der geltenden Regulierung) eingesetzt werden und so die Erstellung einer endgültigen Fassung beschleunigen. Perspektivisch wird auch eine automatisierte individuelle Anpassung von Berichten an die Interessen der jeweiligen Lesenden diskutiert.<sup>40</sup>

Erste Experimente mit aktuell verfügbaren KI-Textgeneratoren zeigen allerdings die Probleme, die dabei entstehen.<sup>41</sup> So erfordert die Bewertung von wis-

---

39 Zum Folgenden: Chris Tyler et al., „AI tools as science policy advisers? The potential and the pitfalls“, *Nature*, Bd. 622(7981), 2023, 27–30; zu Teilaspekten auch Isabelle Borucki/Dennis Michels/Stefan Marschall, „Die Zukunft digitalisierter Demokratie – Perspektiven für die Forschung“, *Zeitschrift für Politikwissenschaft*, Bd. 30, 2020, 359–378 (zur Verarbeitung großer Textmengen); Pauline Riousset/Steffen Albrecht, „Zukünftige Technologien zur Untersuchung technologischer Zukünfte“, *TAB-Brief*, Nr. 51, Dezember 2020, 33–39 (zu Diskursanalysen).

40 Tyler et al. 2023, 29.

41 Ebd.

senschaftlicher Evidenz oder von politischen Optionen ein Urteilsvermögen, das sich aus Erfahrungen und Werten im Bezug zur realen Welt speist und keinen Verzerrungen durch die Orientierung an spezifisch ausgewählten Daten unterliegt. KI-generierte Texte werden zudem bislang als zu vorhersehbar und am Erwartbaren orientiert beurteilt, was sie für Politikberatung weniger geeignet macht. Bei Zusammenfassungen kann es zur Auslassung relevanter Aspekte kommen, wie generell die Zuverlässigkeit der KI-Textgeneratoren gering ist und aufwendige Prüfungen erforderlich macht. Und solange ein vertraulicher und datenschutzkonformer Umgang mit den eingegebenen Daten nicht sichergestellt ist, verbietet sich eine Nutzung im Zusammenhang mit sensiblen persönlichen oder anderweitig schützenswerten Informationen. Es bleibt angesichts dieser Eindrücke abzuwarten, inwieweit sich die in Studien berichteten Produktivitätssteigerungen<sup>42</sup> auch im Bereich der TA realisieren lassen.

Neben diesen unmittelbar ersichtlichen Herausforderungen können auch indirekte Effekte der Nutzung sowohl im Zuge von TA-Untersuchungen als auch im wissenschaftlichen oder politischen Umfeld der TA auftreten. So kann sich die Qualität wissenschaftlicher Texte schleichend verschlechtern, falls Fehler oder Bias der KI-Systeme nicht entdeckt werden und Eingang in die Texte finden; langfristig kann dies einem Verlust des Vertrauens Vorschub leisten, zu dem auch eine fehlende Transparenz bezüglich der Nutzung von KI-Textgeneratoren bei der Erstellung bzw. Verbreitung von Publikationen beitragen kann. Wie Baird am Beispiel wissenschaftlicher Instrumente gezeigt hat, sind technische Geräte zudem nicht wertfrei, sondern durch Designentscheidungen und ihren Entstehungskontext geprägt.<sup>43</sup> Bei ChatGPT wurde ein tendenziell links-liberaler politischer Einschlag in den Texten identifiziert,<sup>44</sup> eine stärkere Nutzung dieses Systems in der wissenschaftlichen Politikberatung könnte daher dem Anspruch auf Überparteilichkeit entgegenstehen. Nicht zuletzt könnten aufgrund der Orientierung der KI-Systeme an bestehenden Mustern gerade solche gesellschaftlichen Perspektiven übersehen werden, die neu und innovativ und daher noch nicht im Trainingsmaterial der KI-Systeme repräsentiert sind. Solche politischen Risiken der Anwendung stellen sich zusätzlich zu den genannten praktischen Problemen und lassen eine Nutzung außerhalb von Experimenten gegenwärtig fraglich erscheinen.

---

42 Shakked Noy/Whitney Zhang, „Experimental evidence on the productivity effects of generative artificial intelligence“, *Science*, Bd. 381(6654), 2023, 187–192.

43 Davis Baird, *Thing Knowledge. A Philosophy of Scientific Instruments*, Berkeley 2004.

44 David Rozado, „The Political Bias of ChatGPT“, *Social Sciences*, Bd. 12(3), 2023 (DOI: 10.3390/socsci12030148), 148.

## 4 Offenheit und Skeptizismus als Leitlinien

Wissenschaft als Institution sollte grundsätzlich offen für Beiträge aus allen möglichen Quellen sein, so lässt sich, etwas verkürzt, die Norm des Universalismus in Mertons Konzeption von Wissenschaft verstehen.<sup>45</sup> Aus heutiger Perspektive kann das auch die Beiträge von KI-Textgeneratoren einschließen, auch wenn bisher Konsens herrscht, dass sie keine originäre Schöpfung darstellen, sondern nur im Verbund mit menschlichem Handeln zu sehen sind. Gerade diese Verbundenheit von nichtmenschlicher und menschlicher Produktion (nicht allein wissenschaftlicher) Texte wird in jüngerer Zeit theoretisch immer stärker in den Blick genommen. Im Poststrukturalismus wird die intertextuelle Qualität von Texten herausgestellt, bis hin zur plakativen Ausrufung des „Tods des Autors“. Ausgehend von der Wissenschaftssoziologie setzt sich die Anerkennung durch, dass Wissenschaft eine kollektive und sozial eingebundene Tätigkeit ist, worauf etwa Merton am Beispiel der metaphorischen Redewendung hinweist, wissenschaftlicher Fortschritt erfolge immer „von den Schultern von Riesen aus“.<sup>46</sup> Mit der Akteur-Netzwerk-Theorie<sup>47</sup> und der Systemtheorie<sup>48</sup> stehen gleich mehrere theoretische Ansätze bereit, mit denen sich das Verhältnis und das Zusammenwirken von menschlichen und nichtmenschlichen Aktanten in der wissenschaftlichen Praxis beschreiben und analysieren lässt.

Nicht nur im Sinn der wissenschaftlichen Offenheit und Neugier sind Erprobungen und Weiterentwicklungen des Verhältnisses zwischen Forschenden und KI-Textgeneratoren interessant und wünschenswert, auch angesichts der gesellschaftlichen Herausforderungen, mit denen sich die Technikfolgenabschätzung und die Wissenschaft insgesamt beschäftigen, ist technische Unterstützung grundsätzlich höchst willkommen (und kommt bereits vielfach zum Einsatz). Die vielen aufgeworfenen Fragen und Probleme machen aber deutlich, dass für das skizzierte Zusammenwirken von Mensch und KI-basierter Maschine neue Regelungen nötig sind, sowohl auf gesetzlicher Ebene als auch beispielsweise in Form von Standards, Tests bzw. Benchmarks und Richtlinien guter wissenschaftlicher Praxis. Merton hat der Norm des Universalismus die Norm des organisierten Skeptizismus an die Seite

---

45 Robert K. Merton, „The normative structure of science“, in *The Sociology of Science*, hg. von dems., Chicago/London 1973, 267–278.

46 Robert K. Merton, *On the Shoulders of Giants. A Shandean Postscript*, New York 1965.

47 Bruno Latour, „Zirkulierende Referenz. Bodenstichproben aus dem Urwald am Amazonas“, in *Die Hoffnung der Pandora. Untersuchungen zur Wirklichkeit der Wissenschaft*, hg. von dems., Frankfurt a. M. 2000, 36–95.

48 Elena Esposito, *Artificial Communication. How Algorithms Produce Social Intelligence*, Cambridge/London 2022.

gestellt, also des grundsätzlichen Infragestellens „jedes Aspekts von Natur und Gesellschaft“. <sup>49</sup> Ein solcher Skeptizismus erscheint auch gegenüber den Ergebnissen von KI-Textgeneratoren angebracht, die sich gerade nicht durch Unsicherheit in Bezug auf die eigenen Aussagen auszeichnen und daher Prinzipien der Wissenschaftlichkeit entgegenlaufen.

Die Rolle der TA kann dabei sein, über die Möglichkeiten und Grenzen von KI-Textgeneratoren im Rahmen wissenschaftlicher Praxis aufzuklären und dabei zu helfen, auch indirekte Effekte und Pfadabhängigkeiten zu berücksichtigen. Sie kann auch aufzeigen, welche Art von Weiterentwicklungen aus der Perspektive der Wissenschaft wünschenswert sind, um den jeweiligen Problemstellungen besser gerecht zu werden. Die auf diesen Erkenntnissen basierende wissenschaftsinterne Reflexion der technologischen Entwicklung kann dazu beitragen, die treibenden Kräfte und möglichen Folgen dieser Entwicklung besser zu verstehen und sie verantwortungsvoll auf die Realisierung der beachtlichen Potenziale auszurichten, insbesondere, wenn dabei vielseitige und ausgewogene Expertise zum Tragen kommt.

---

<sup>49</sup> Merton 1973, 277.

