

THEMENSCHWERPUNKT

„ERZIEHUNGSWISSENSCHAFT NACH CHATGPT“

Die (produktive) Rolle von generativer KI im erziehungswissenschaftlichen Diskurs

Bardo Herzig

Abstract

The (Productive) Role of Generative AI in Educational Science Discourse

To determine the role of generative AI in educational science discourse, the focus first shifts to knowledge relevant to the field of educational science. This is followed by an examination of the discursive functions of media, before addressing the specific characteristics of generative AI, its applications, and its implications.

Erziehungswissenschaftlicher Diskurs

Disziplinäre Besonderheiten erziehungswissenschaftlichen Wissens sind neben der Pluralität von Paradigmen, Begriffsverständnissen und Theorien u. a. die explizite Auseinandersetzung mit Normen und Zielen oder die Relationierung von Theorie und Praxis und die (nicht nur) damit verbundene Reflexionsarbeit. Eine erziehungswissenschaftliche Theorie lässt sich als eine explizite und stimmig verknüpfte Aufstellung von Begriffen und Aussagen zu einem bestimmten Gegenstandsbereich verstehen, die in intersubjektiv nachvollziehbarer Weise begründet ist. Sie kann normative, empirisch-deskriptive, empirisch-hypothetische und logische Aussagen sowie anwendungsbezogene Sätze umfassen und auf das Verstehen, das Erklären, die Prognose oder die Gestaltung von Erziehungs- und Bildungsphänomenen gerichtet sein (vgl. Kommesser/Büttemeyer 2020, S. 143f.).

Erziehungswissenschaftlich relevantes Wissen wird in einem wissenschaftlichen Diskurs verhandelt, der in der Regel in (zeichen-)sprachlicher Form geführt wird. Institutionelle Orte der Dokumentation eines solchen Diskurses sind z. B. Fachzeitschriften, Bücher oder kuratierte Foren in analoger oder digitaler Form. Der Diskurs lebt von seiner Dynamik, er spiegelt verschiedene Perspektiven wider, verändert sich und kann beispielsweise durch eine Diskursanalyse in seinen ggf. widersprüchlichen Perspektiven, Ordnungen und Praktiken rekonstruiert werden (vgl. z. B. Fegter et al. 2015). Die Teilnahme

am Diskurs wird über die Scientific Community geregelt, z. B. über Review-Verfahren, in denen Beiträge auf die Einhaltung wissenschaftlicher Standards und Gütekriterien geprüft werden.

Je nach Perspektive könnte hier eingewendet werden, dass es dabei nicht um Wissen geht, sondern um Mitteilungen bzw. Informationen, die erst über die Verarbeitung durch ein Individuum oder kognitives System zu Wissen werden (vgl. Herzig 2012, S. 221ff.). Der Einfachheit halber wird im Folgenden von Wissen bzw. potenziellem Wissen gesprochen.

Die Rolle von Medien und generativer KI im erziehungswissenschaftlichen Diskurs

Die Rolle der (technischen) Medien bei der (Weiter-)Entwicklung und Dissemination von Wissen lässt sich (vereinfachend und historisch betrachtet) in drei Phasen skizzieren: die Wissensrepräsentation, die Wissensordnung und die Wissensproduktion. In der ersten Phase dienen die analogen Medien vor allem als Zeichenträger zur Repräsentation und zum Zeichentransport, etwa beim Buchdruck. Dies beeinflusste insbesondere die Verfügbarkeit potenziellen Wissens als Grundlage eines Diskurses.

Mit der Digitalisierung – in der zweiten Phase – und der Möglichkeit, Zeichen maschinell zu verarbeiten, werden Informationen als potenzielle Wissensbestände analysiert, systematisiert, geordnet, gerant und vernetzt. Für eine Suchmaschine bedeutet dies beispielsweise festzulegen, in welcher Form welche Datenbestände durchsucht und anhand von Schlüsselwörtern (Metatags) indiziert, wie eine Anfrage zerlegt und wie im Vergleich von Anfrage und Suchindex Dokumente gerant werden sollen (vgl. Herzig 2012, S. 233ff.). Damit wird kein originäres Wissen erzeugt, sondern vorhandenes Wissen – nach bestimmten kombinatorischen Regeln – auffindbar gemacht. Auch mithilfe von symbolischer, d. h. regelbasierter KI lassen sich Daten(-mengen) mit statistischen Verfahren auf Muster und Strukturen hin untersuchen, die mit menschlicher Intelligenz nicht oder nur unter erheblichem Aufwand erkennbar wären. In diesem Fall werden in vorhandenen Daten Muster aufgedeckt, deren Interpretation zu neuem Wissen führen kann. Die maschinelle Herstellung solcher Wissensordnungen und informationeller (Hyper-)Strukturen hat den disziplinären Diskurs nicht nur beschleunigt, sondern auch qualitativ beeinflusst.

In der dritten Phase hat sich mit dem Aufkommen generativer KI die Situation noch einmal grundlegend geändert. Generative KI kann eigenständig potenzielles Wissen produzieren und wird vielfach mit einem eigenen Akteursstatus versehen. Sie zeichnet sich dadurch aus, dass sie auf der Basis einer Eingabe solche Datensätze erzeugt, die bestehenden Daten sehr ähnlich sind – im Falle eines Sprachmodells wie ChatGPT oder DeepSeek also solche Ausgaben, die von menschlicher Sprache nicht oder nur schwer zu unterscheiden sind. Mithilfe

von Verfahren der Repräsentation bedeutungsrelevanter Aspekte von Wörtern in hochdimensionalen Vektorräumen oder der Erfassung von Zusammenhängen durch Aufmerksamkeitsmechanismen wird sprachliches Wissen auf der Basis von Wahrscheinlichkeiten mithilfe neuronaler Netze erzeugt.

Während in den ersten beiden Phasen vorhandenes Wissen kopiert und disseminiert bzw. analysiert, systematisiert und vernetzt wird, geht es in der dritten Phase bei generativer KI darum, aus vorhandenem Wissen ein stochastisches Modell zu entwickeln, mit dem neues Wissen generiert werden kann – mit Folgen für den Diskurs.

Konsequenzen für den erziehungswissenschaftlichen Diskurs

Bei KI-generiertem Wissen handelt es sich nicht um originäre Beiträge im klassischen Sinne, sondern um Muster von Wortfolgen, die Zusammenhangswahrscheinlichkeiten in Trainingsdaten widerspiegeln. Damit können häufig vorkommende Muster – abhängig z. B. von Publikationsorganen, -sprachen oder -metriken – die Oberhand gewinnen, und die Breite des erziehungswissenschaftlichen Diskurses im Sinne von Diversität wird nicht abgebildet. Die Systemlogiken von Wissenschaft und von Sprachmodellen bzw. damit verbundenen Informatiksystemen sind mit Wahrheit und Faktizität auf der einen und Wahrscheinlichkeiten auf der anderen Seite zudem nicht kommensurabel. Theorien beispielsweise werden damit zwar als plausible und wahrscheinliche, aber eben nicht notwendigerweise auch präzise, kohärente und sinnstiftende Aussagensysteme ausgegeben. Zudem besteht die Gefahr, statistische (korrelative) Zusammenhänge als Grundlage erziehungswissenschaftlichen Denkens und Handelns zu priorisieren und der Theoriearbeit keine oder weniger Bedeutung beizumessen (vgl. Thompson 2023). Nicht zuletzt kommt den nicht offenliegenden Trainingsdaten eine hohe Bedeutung zu, weil sie Verzerrungen (*biases*) enthalten können, die durch Sprachmodelle dann noch verstärkt werden (Matthäus-Effekt). Damit geht auch die für wissenschaftliches Arbeiten bedeutsame Referenzierung von Wissen auf Quellen verloren. In Bezug auf Verzerrungen muss allerdings konstatiert werden, dass diese auch heute schon bestehen, z. B. weil bestimmte Methoden als Goldstandards gelten wie im Fall der empirisch-quantitativen Bildungsforschung (vgl. z. B. Münch 2022) oder weil nicht das gesamte erziehungswissenschaftliche Wissen digitalisiert ist.

Formal kann die Qualität eines Sprachmodells u. a. an den Kriterien von Präzision und Diversität festgemacht werden (vgl. Chang et al. 2024, S. 26ff.). Die Präzision gibt an, wie genau das Modell arbeitet, d. h. inwieweit sich die Ausgaben des Modells auch als Aussagen in den jeweiligen Gegenstandsbereichen finden, insbesondere also keine Halluzinationen oder Falschaussagen produziert werden. Die Diversität kennzeichnet den Grad der Breite, mit der die Modellausgaben die in den Gegenstandsbereichen zu findenden Aussagen

und Positionen abdecken, d. h. hierbei geht es vor allem um mögliche Verzerrungen (s. o.). Dabei zeigt sich allerdings, dass letztlich immer ein Kompromiss zwischen der Präzision und der Diversität gemacht werden muss. Das Feintuning eines Modells kann seine Präzision erhöhen, jedoch zulasten der Diversität. Umgekehrt bedeutet eine hohe Diversität eine geringere Präzision (vgl. Le Bronnec et al. 2024, S. 2). Die Qualität eines Sprachmodells wird auch dann beeinflusst, wenn KI-generiertes Wissen als Trainingsmaterial verwendet wird. Werden Sprachmodelle mit ihrem eigenen Output trainiert, laufen sie Gefahr, in autophagen Schleifen zu geraten und damit Präzision oder Diversität auf Dauer deutlich zu verschlechtern (vgl. Alemohammad et al. 2023). Dies lässt sich nur dann verhindern, wenn hinreichend „neue“ Daten – in diesem Fall originäres, nicht synthetisches erziehungswissenschaftliches Wissen – als Trainingsdaten verwendet werden.

Inwieweit sich die genannten Eigenschaften von Sprachmodellen tatsächlich auf den Diskurs auswirken, hängt nicht zuletzt davon ab, in welcher Konstellation generative KI in der Wissenschaft genutzt wird und in welchem Umfang synthetische Daten in den Diskurs einfließen. Generell lässt sich die Nutzung als eine Form hybrider Mensch-Maschine-Interaktion beschreiben, in der die Rolle der KI sehr unterschiedlich ausfallen kann. Als substitutive Funktion könnte das alleinige Generieren von Wissensprodukten (z. B. Erstellung eines Papers), als assistive Funktion die Bearbeitung von Teilaufgaben in einem Forschungsprozess (z. B. Datenanalyse oder -auswertung) und als kooperative Funktion die ko-produktive Erstellung von Wissensprodukten charakterisiert werden. Für alle Funktionen gibt es inzwischen hinreichend viele Beispiele. Generative KI erzeugt dabei jeweils synthetische Daten, deren Güte zunächst nur durch die Qualität des Trainingsdatenmaterials und des Sprachmodells sowie ggf. der Prompts bestimmt wird. Die Einschätzung, ob und inwieweit diese Daten geeignet sind, an verschiedenen Stellen des Forschungs- oder Erkenntnisprozesses genutzt zu werden und schließlich in den erziehungswissenschaftlichen Diskurs einzufließen, obliegt den einzelnen Wissenschaftler*innen. Mit anderen Worten, der Umgang mit synthetischen Daten vor dem Hintergrund wissenschaftlicher Standards entscheidet letztlich darüber, ob Wissenschaft (weiterhin) erkenntnis-, theorie- und erklärungsorientiert bleibt und divers über Ziele, Mittel und Methoden streitet oder ob Anwendungs- und Nutzerorientierung in den Vordergrund treten (vgl. Schmidt 2022, S. 81).

Ausblick

Die zukünftige Rolle, die generative KI in der erziehungswissenschaftlichen Diskussion und damit in Bezug auf Vorstellungen von Erziehung und Bildung haben wird, ist schwer abzuschätzen. Die hoch dynamische Weiterentwicklung von Sprachmodellen – z. B. durch Verbindung mit externen Datenquellen

zur Erhöhung der Präzision und Aktualität (Retrieval Augmented Generation, vgl. Gao et al. 2024) oder durch neue Trainingsmethoden zur Vermeidung von autophagen Schleifen (vgl. Alemohammad et al. 2024) – lässt weitere Qualitätsverbesserungen erwarten, ebenso wie Ansätze, Qualitätsstandards und normative Vorgaben selbst zum Gegenstand von KI-Modellen zu machen und damit synthetische Wissensprodukte zu prüfen (vgl. Bai et al. 2022). Bei aller dienlichen und unterstützenden Funktion von generativer KI im Forschungs- und Publikationsprozess bleibt aber die Verantwortung für erziehungswissenschaftliches Wissen letztlich bei den Wissenschaftler:innen und darf nicht Algorithmen oder Plattformlogiken überlassen werden. Vom Verhalten der Einzelnen wird es abhängen, ob einer *bot delusion* durch kritische Prüfung von Ergebnissen und das eigene Publikations- und Feedbackverhalten vorgebeugt werden kann (vgl. Wiczorek et al. 2023). Die Scientific Community kann neben der Aufrechterhaltung von Standards auch durch die Kuratierung von Trainingsdaten z. B. in Open-Access-Formaten dazu beitragen.

Dies bedeutet aber auch, dass die (selbst-)reflexive forschende Auseinandersetzung mit der eigenen Wissensproduktion und -dissemination eine zunehmend bedeutsame Aufgabe werden muss – sowohl in Bezug auf die Rolle und Nutzung generativer KI oder weiterführender Technologien als auch in Bezug auf die diesbezügliche Veränderung von Erziehungs- und Bildungszielen und damit verbundener Praktiken.

Bardo Herzig, Prof. Dr., ist Professor für Allgemeine Didaktik und Schulpädagogik unter Berücksichtigung der Medienpädagogik an der Universität Paderborn.

Literatur

- Alemohammad, Sina/Casco-Rodriguez, Josue/Luzi, Lorenzo/Humayun, Ahmed Imtiaz/Babaei, Hossein/LeJeune, Daniel/Siahkoohi, Ali/Baraniuk, Richard G. (2023): Self-Consuming Generative Models Go MAD. In: arXiv e-prints. <https://arxiv.org/pdf/2307.01850> [Zugriff: 18. Februar 2025]. <https://doi.org/10.48550/arXiv.2307.01850>.
- Alemohammad, Sina/Humayun, Ahmed Imtiaz/Agarwal, Shruti/Collomosse, John/Baraniuk, Richard G. (2024): Self-Improving Diffusion Models with Synthetic Data. In: arXiv e-prints. <https://arxiv.org/pdf/2408.16333> [Zugriff: 18. Februar 2025]. <https://doi.org/10.48550/arXiv.2408.16333>.
- Bai, Yuntao et al. (2022): Constitutional AI: Harmlessness from AI Feedback. In: arXiv e-prints. <https://arxiv.org/pdf/2212.08073> [Zugriff: 18. Februar 2025]. <https://doi.org/10.48550/arXiv.2212.08073>.
- Chang, Yupeng et al. (2024): A Survey on Evaluation of Large Language Models. In: ACM Transactions on Intelligent Systems and Technology 15, 3, S. 1-45. <https://doi.org/10.1145/3641289>.

- Fegter, Susann/Kessler, Fabian/Langer, Antje/Ott, Marion/Rothe, Daniela/Wrana, Daniel (2015): Erziehungswissenschaftliche Diskursforschung. Theorien, Methodologien, Gegenstandskonstruktionen. In: Fegter, Susann/Kessler, Fabian/Langer, Antje/Ott, Marion/Rothe, Daniela/Wrana, Daniela (Hrsg.): Erziehungswissenschaftliche Diskursforschung. Empirische Analysen zu Bildungs- und Erziehungsverhältnissen. Wiesbaden: Springer VS, S. 9-55. https://doi.org/10.1007/978-3-531-18738-9_1.
- Gao, Yunfan et al. (2024): Retrieval-Augmented Generation for Large Language Models: A Survey. In: ArXiv e-prints. <https://arxiv.org/pdf/2312.10997>. [Zugriff: 18. Februar 2025].
- Herzig, Bardo (2012): Medienbildung. Grundlagen und Anwendungen. München: kopaed.
- Kornmesser, Stephan/Büttemeyer, Wilhelm (2020): Wissenschaftstheorie. Eine Einführung. Stuttgart: J.B. Metzler. <https://doi.org/10.1007/978-3-476-04743-4>.
- Le Bronnec, Florian/Vérine, Alexandre/Negrevergne, Benjamin/Chevalere, Yann/Allauzen, Alexandre (2024): Exploring Precision and Recall to Assess the Quality and Diversity of LLMs. 62nd Annual Meeting of the Association for Computational Linguistics. Bangkok, Thailand. <https://hal.science/hal-04584210v1> [Zugriff: 18. Februar 2025]. <https://doi.org/10.18653/v1/2024.acl-long.616>.
- Münch, Richard (2022): Die Herrschaft der Inzidenzen und Evidenzen. Regieren in den Fallstricken des Szientismus. Frankfurt am Main, New York: Campus.
- Schmidt, Jan C. (2022): Wandel und Kontinuität von Wissenschaft durch KI. Zur aktuellen Veränderung des Wissenschafts- und Technikverständnisses. In: Gethmann, Carl Friedrich/Buxmann, Peter/Distelrath, Julia/Humm, Bernhard G./Lingner, Stephan/Nitsch, Verena/Schmidt, Jan C./Spiecker genannt Döhmann, Indra (Hrsg.): Künstliche Intelligenz in der Forschung (= Ethics of Science and Technology Assessment Band 48). Berlin: Springer, S. 79-125. https://doi.org/10.1007/978-3-662-63449-3_4.
- Thompson, Christiane (2023): Allgemeine Erziehungswissenschaft und die Problematisierung pädagogischen Wissens – in Zeiten von KI. In: Vierteljahresschrift für Wissenschaftliche Pädagogik 99, 4, S. 497-512. <https://doi.org/10.30965/25890581-09703117>.
- Wieczorek, Oliver/Steinhardt, Isabel/Schmidt, Rebecca/Mauermeister, Sylvi/Schneiderberg, Christian (2023): The Bot Delusion. Large Language Models and Anticipated Consequences for Academics' Publication and Citation Behavior. <https://ssrn.com/abstract=4779200> [Zugriff: 18. Februar 2025]. <https://doi.org/10.2139/ssrn.4779200>.