

AI Literacy als kritische Praxis

Ein Plädoyer für einen bewussteren KI-Einsatz an Hochschulen

von Johanna Leifeld



Bereits Mitte der 1990er Jahre betonte Alison Adam, dass Technik nicht im luftleeren Raum entsteht, sondern geprägt wird – ähnlich wie ein Kind, das erzogen wird. Sie sprach damals zugespitzt von „Babyrobotern“, deren „Erziehung“ auch feministische Perspektiven bereichern könnte. Diese Metapher wirkt heute erstaunlich aktuell: Betrachtet man heutige Entwicklungen rund um generative KI, steckt sie tatsächlich noch in den sprichwörtlichen Kinderschuhen. Die Frage ist daher nicht nur, wie leistungsfähig sie wird, sondern auch, mit welchen Werten und gesellschaftlichen Perspektiven sie aufwächst. Eine reflektierte, sozialverträgliche Weiterentwicklung und Nutzung ist entscheidend, denn KI ist Ausdruck von Macht und diese ist ungleich verteilt.

Zahlen und Daten wirken auf den ersten Blick objektiv, doch sie sind es nicht. Viele KI-Systeme sind „Black Boxes“: Es bleibt unklar, welche Daten verwendet und wie

sie gewichtet werden. Zudem bedeutet Quantität nicht gleich Repräsentativität. Studien zeigen etwa, dass Online-Bilddaten dreimal häufiger Männer als Frauen darstellen (Guilbeault et al. 2024), während Personen aus dem Globalen Norden deutlich überrepräsentiert sind (Shankar et al. 2017). Ähnliches lässt sich bei textverarbeitenden KI-Systemen beobachten, die mit Inhalten aus dem Internet trainiert werden. Da diese Inhalte vor allem von jungen, männlichen und ökonomisch privilegierten Nutzern stammen, sind auch hier deren Perspektiven überproportional präsent (Gengler 2024).

Solche Verzerrungen in den Trainingsdaten führen dazu, dass KI-Systeme bestehende Stereotype und Machtungleichgewichte widerspiegeln und verstärken. In der Bildgenerierung etwa können geschlechter-, alters- oder kulturell geprägte Klischees reproduziert werden. Diese Darstellungen prägen gesellschaftliche Wahrnehmungen und

Sie wollen KI fairer einsetzen, wissen aber nicht, wo Sie anfangen sollen?

Die folgenden Ressourcen bieten einen guten Einstieg:

Ein Leitfaden von **Eva Gengler**, ursprünglich für Unternehmen entwickelt, lässt sich gut auf den Hochschulkontext übertragen. Das darin vorgestellte **KI-FAIRNESS-Framework** bietet wertvolle Impulse für faires Prompten.

Auch **Nicola Marsden** hat mit dem KI-Campus umfassende **Online-Kurse** veröffentlicht, in denen Methoden wie Re-Prompting vorgestellt werden – Techniken, mit denen stereotype KI-Ausgaben im Prompting-Prozess erkannt und vermieden werden können.



tragen zur Normalisierung und Reproduktion von Vorurteilen bei. Auch bei Sprachmodellen zeigen sich problematische Tendenzen: Eine aktuelle Studie belegt, dass Frauen bei Gehaltsfragen systematisch niedrigere Empfehlungen erhalten als Männer, und dass auch Migrationshintergrund oder ethnische Zuschreibungen Einfluss auf die Empfehlungen haben (Sorokovikova et al. 2025). Es droht somit eine algorithmische Verfestigung sozialer Ungleichheit.

KI als Werkzeug für eine gerechtere Zukunft?

Verzerrungen in KI sind kein individuelles Versagen, sondern strukturell bedingt – vor allem durch die Auswahl der Trainingsdaten und technologische Designentscheidungen. Beides entzieht sich (aktuell noch) dem direkten Einfluss von Hochschulen. Dennoch sind Hochschulen zentrale Akteur:innen, denn sie bilden diejenigen aus, die KI-Systeme nutzen und zukünftige KI gestalten werden – sei es in der Informatik, Psychologie oder Soziologie. Diese drei Ansatzpunkte bieten Hochschulen bereits heute die Möglichkeit aktiv Verantwortung zu übernehmen: **(1) AI Literacy für alle fördern:** Studierende wie auch Lehrende benötigen ein solides Verständnis davon, wie KI funktioniert – und welche Verzerrungen ihr zugrunde liegen können. **(2) Gesellschaftliche Dimensionen von KI sichtbar machen:** AI Literacy umfasst mehr als technisches Wissen – sie schließt auch ein kritisches Bewusstsein für Machtverhältnisse und soziale Auswirkungen ein. **(3) Reflektierte KI-Nutzung im Hochschulalltag stärken:** Lehrende und Studierende sollten darin unterstützt werden, ihr eigenes Nutzungsverhalten zu hinterfragen und verantwortungsvoll zu gestalten. Hierzu sind bereits hilfreiche Materialien und didaktische Ansätze entwickelt worden. Zwei Beispiele finden sich links in der Infobox.

Im Hinblick auf die „Erziehung der Babyroboter“ und die Frage, welche gesellschaftliche Wirkung generative KI künftig entfalten wird, ist eine fächerübergreifende Sensibilisierung für Daten, gerechte Entwicklung und verantwortungsvolle Nutzung entscheidend. Gelingt es, generative KI so zu gestalten, dass sie Vorurteile nicht verstärkt, sondern sichtbar macht, kann sie zu einem Werkzeug für eine gerechtere Gesellschaft werden. // Layout: Theresa Sommer



Johanna Leifeld

ist Projektmanagerin im Hochschulforum Digitalisierung für das CHE Centrum für Hochschulentwicklung. Sie arbeitet in der Strategieentwicklung und verantwortet dort die Peer-to-Peer-Fachbereichsberatung. Außerdem beschäftigt sie sich mit der Frage, wie der Einsatz künstlicher Intelligenz unsere Gesellschaft verändert und wie diese Entwicklung gerechter gestaltet werden kann. Sie gehört zudem zum Organisationsteam des University:Future Festival.

Quellen

- Adam, A. (1995). Artificial intelligence and women's knowledge: What can feminist epistemologies tell us. *Women's Studies International Forum*.
- Gengler, E. J. (2024). Sexism, Racism, and Classism: Social Biases in Text-to-Image Generative AI in the Context of Power, Success, and Beauty. *Wirtschaftsinformatik 2024 Proceedings*, 48.
- Guilbeault, D. R., Delecourt, S., Hull, T., Desikan, B.S., Chu, M. & Nadler, E. (2024) Online images amplify gender bias, *Nature*, 626(8001), pp. 1049–1055.
- Shankar, S., Halpern, Y., Breck, E., Atwood, J., Wilson, J. & Sculley, D. (2017). No classification without representation: Assessing geodiversity issues in open data sets for the developing world.
- Sorokovikova, A., Chizhov, P., Eremenko, I. & Yamshchikov, I.P. (2025). Surface Fairness, Deep Bias: A Comparative Study of Bias in Language Models. In *Proceedings of the 6th Workshop on Gender Bias in Natural Language Processing (GeBNLP)*, pp. 206–227, Vienna, Austria.