HANDOUT: ETHIK DER KI-NUTZUNG IM HOCHSCHULKONTEXT

Künstliche Intelligenz (KI) hält zunehmend Einzug in den Hochschulbereich und bietet vielfältige Möglichkeiten, Lehre, Forschung und Verwaltung zu verbessern und neue Chancen für Effizienz, Personalisierung und Innovation zu schaffen. Von personalisierten Lernplattformen bis hin zu automatisierten Prozessen verändert KI die Art und Weise, wie Hochschulen arbeiten.

Gleichzeitig wirft der Einsatz von KI jedoch wichtige ethische Fragen auf: Wie können Datenschutz und Chancengleichheit gewährleistet werden? Welche Verantwortung tragen Hochschulen bei der Nutzung und Entwicklung von KI-Systemen? Und wie können Studierende und Lehrende für einen kritischen Umgang mit dieser Technologie sensibilisiert werden?

Die Auseinandersetzung mit ethischen Aspekten ist essenziell, um sicherzustellen, dass KI im Hochschulkontext verantwortungsvoll und im Einklang mit gesellschaftlichen Werten eingesetzt wird. Zudem soll hier auch eine kurze Einsicht in das dazu notwendige ethische Handwerk gegeben.

WAS IST ‚ETHIK'?

„Ethik [ist] die Lehre vom richtigen, gelingenden, guten Handeln", die als „angewandte Ethik [..] den Versuch [bildet], mit den Mitteln der Ethik [d.h. ihren Begründungsmustern und Prinzipien] Menschen dabei zu helfen, sich in bestimmten Situationen moralisch richtig zu verhalten, in denen Unklarheit oder Unsicherheit herrscht, was in dieser Situation richtig wäre" (Stoecker, et al., 2011/ 2023 S. 6,7).

Ethik bietet somit einen Weg menschliche Handlungen auf Grund von nachvollziehbaren Begründungen mit Wertungen wie ‚moralisch gut', moralisch neutral' oder ‚moralisch schlecht' zu versehen. Damit können einerseits Handlungsempfehlungen, aber auch zunächst Argumente für mögliche Handlungsempfehlungen (zur Frage ‚Was soll ich tun?') aufgestellt werden.

In der Philosophie gibt es eine Reihe von Theorien in der normativen Ethik. Dabei sind drei Gruppen moralphilosophischer Theorien besonders prävalent, die die Spannweite ethischer Begründungsmuster recht gut abbilden:

1. In der Tugendethik, die bereits im antiken Griechenland (durch Socrates, Aristoteles und Platon) vertreten wurde, wird das ‚Gute' einer Handlung durch vorliegende Tugenden, d.h. in Charaktereigenschaften und Absichten, bestimmt. Es gibt aber auch moderne Verfechter:innen (z.B. Nussbaum und MacIntyre).
2. In deontologischen Theorien wird eine Handlung nach dem durch die Handlung verkörperten Handlungscharakter bestimmt. Das heißt, es wird geprüft, ob eine Handlung mit Pflichten einhergeht oder mit Rechten kollidiert. Hier sind insbesondere die kategorischen Imperative Kants oder moderne Vertragstheorie (z.B. nach Rousseau oder Rawls) verortet.
3. In teleologischen Theorien werden Handlungen primär nach den aus der Handlung resultierenden Folgen bestimmt. Hier findet sich insbesondere der Konsequentialismus, der z.B. in Form des Utilitarismus (geprägt durch Bentham und Mill) die Folgen von Handlungen nach Ihrem in ihrem Nutzen oder Schaden bewertet.

Dabei ist es wichtig zu beachten, dass sich die aus diesen Theorien abgeleiteten moralischen Urteile im Ergebnis je nach dem Kontext bestimmter Situationen unterscheiden können, aber sich nicht zwingend unterscheiden müssen.

WAS IST ‚KÜNSTLICHE INTELLIGENZ'?

Um ethische Problemstrukturen zu beleuchten, gegebenenfalls auch begründete Urteile fällen und in entsprechende Handlungen umzusetzen zu können, benötigt es zunächst eine gewisse Klarheit über das Objekt einer ethischen Untersuchung.

Im Fall von ‚KI' ist es wie mit vielen Objekten der angewandten ethischen Untersuchung: Eindeutige, vollständige und universell akzeptierte Definitionen gibt es selten. Genauso gibt es zu ‚KI' viele, mögliche Definition (Sheikh, et al., 2023 S. 15).

Als Arbeitsdefinition bietet sich z.B. der Definitionsvorschlag der Expertengruppe der Europäischen Kommission zu KI an: KI-Systeme bezeichnen menschengemachte Software- (und ggf. auch Hardware-) Systeme, die bei Vorgabe komplexer Ziele, in physischen oder digitalen Handlungen ausführen, in dem sie ihre Umgebung durch

Datenerfassung wahrnehmen, die gesammelten strukturierten und unstrukturierten Daten interpretieren, auf Grundlage der aus diesen Daten abgeleiteten Informationen Schlussfolgerungen ziehen und daraus die besten Maßnahmen ableiten, um die vorgegebenen Ziele zu erreichen (High-Level Expert Group on Artificial Intelligence, 2019 S. 6).

Im Kontext der Hochschule spielt gerade der Einsatz von generativer KI (generative AI; wie z.B. Le Chat, Gemini oder Chat GPT) eine besondere Rolle und sollte deshalb von der obigen, sehr breiten Definition abgegrenzt werden: Als generative KI werden jene KI-Systeme bezeichnet, die auf Grundlage von Eingaben (‚prompts'), Trainingsdaten und mit Hilfe neuronaler Netzwerke gestütztem maschinellen Lernens (‚deep learning') menschenähnliche Inhalte (z.B. in Text und Bild) generieren (Michel-Villareal, et al., 2023 S. 1).

Eine Auswahl ethischer Herausforderungen in der Nutzung von KI an Hochschulen

KI kann Lehrende im Hochschulkontext unterstützen, doch ist auch Vorsicht geboten, da Verzerrungen, Halluzinationen, Eingriffe in Persönlichkeitsrechte und Menschenwürde und Gefahren für die Demokratie auftreten können:

KI-generierten Inhalte können Verzerrungen (‚biases') aufweisen. Zum Beispiel können KI-Inhalten begrenzte und ggf. nicht aktuelle Datensätze (Michel-Villareal, et al., 2023 S. 2), aber auch Datensätze mit geringer fachlicher Qualität oder gar Falschbehauptungen zugrunde liegen. Weiterhin können KI-Resultate durch Datensätze oder Eingriffe, die versteckte normative (d.h., wertende) Annahmen und diskriminierende Vorurteile beinhalten, beeinflusst sein (Sheikh, et al., 2023 S. 147).

Auch unabhängig vom Einfluss der Trainingsdatensätze oder Eingriffen in die Programmierung von KI kann es vorkommen, dass KI scheinbare Zusammenhänge herbei konstruiert (also ‚halluziniert') und zum Beispiel nicht-belegte oder belegbare Schlussfolgerungen als belegt darstellt. Auch für KI gilt: ‚Eine bestehende Korrelation muss keine dementsprechende Kausalität implizieren.' oder anders gesagt: ‚Nur, weil eines oft auf das andere (zeitlich oder räumlich) folgt, muss das nicht heißen, dass das eine das andere auch verursacht.'

Solche Verzerrungen und Halluzinationen sind für Nutzende meist nicht oder nur unvollständig nachvollziehbar. Dies gilt gerade dann, wenn KI-Anbieter und -Entwickler die verwendeten Trainingsdatensätze, Funktionsweise der eingesetzten Algorithmen des maschinellen Lernens oder Begründungen für generierte Schlussfolgerungen, nicht offen legen wollen oder können (Sheikh, et al., 2023 S. 75).

Weiterhin werden viele KI-Tools nicht nur mit Nutzereingaben (‚prompts') bedient, sondern diese Nutzereingaben können ggf. auch als Trainingsdaten weiterverwendet werden. Sollten Eingaben persönliche Daten der Nutzenden oder anderer unbeteiligter Personen enthalten, kann dies einerseits gegen den Datenschutz verstoßen. Andererseits können auch weitere ethische Probleme auftreten, wenn diese Daten auf die oben genannten möglichen Verzerrungen durch wertende Annahmen oder sogar diskriminierende Vorurteile treffen und diese dann weiterverbreitet werden (Sheikh, et al., 2023 S. 50).

Wenn KI-Ergebnisse ohne weitere Prüfung übernommen werden, können generierte Fehlinformationen, Verfälschungen und Verzerrungen verbreitet werden. Diese Einflüsse können freie und offene Deliberation (d.i. die gesamtgesellschaftliche Willensbildung) untergraben, die grundsätzlich für eine Demokratie ist (Cohen, 2007 S. 291ff.). Nutzende können sich nicht einfach darauf verlassen, dass KI-Systeme demokratische Grundrechte, wie die Unantastbarkeit der Menschenwürde, universelle Menschenrechte und Grundrechte, sowie Prinzipien der Rechtsstaatlichkeit jederzeit achten.

Wenn KI herangezogen wird, um Entscheidungen zu fällen, besteht das Risiko, dass einerseits mögliche Betroffene objektifiziert werden. Das heißt, dass Betroffene nicht mehr als Subjekt, welche unveräußerliche Rechte, wie gewisse Abwehrrechte gegen Eingriffe des Staates und anderer Personen, innehaben, behandelt werden. Objektifizierte Betroffene werden eben nur als Objekt verstanden, denen keine Rechte auf Mitbestimmung und keine Pflicht gegenüber einzuräumen sind.

Dies wäre nach der Kant'schen zweiten Formulierung des kategorischen Imperatives, nach der Menschen niemals nur als Mittel, sondern auch immer auch als Zweck „an sich selbst behandel[t]" werden sollte (Kant, 1910-1983 S. AA IV, 429, 433), nicht erlaubt. Das hieße, mögliche Betroffene können zwar in der Erreichung eines Zwecks als Mittel eingesetzt werden können, um Objektifizierung zu vermeiden, müssten sich die zu erreichenden Zwecke auch auf die Zwecke der möglichen Betroffenen ausdehnen; die zu erreichenden Zwecke müssen zu einem gewissen Teil auch ‚im Sinne' der Betroffenen sein. Wenn Menschen nicht in sie betreffende Entscheidungsprozess miteinbezogen und deren Ergebnisse für sie prinzipiell unbegründet bleiben, bleiben sie als Selbstzweck außer Acht und ihre Menschenwürde ist damit möglicherweise verletzt.

Sobald Verantwortungen an KI abgegeben werden, können die Verantwortlichkeiten für die von KI eigenständig getroffenen Entscheidungen ggf. nicht mehr klar zugeordnet werden. Darüber hinaus werden Fragen nach der Haftbarkeit zwischen menschlichen und künstlichen Akteuren aufgeworfen. Dies wird in der Philosophie als Problem der Verantwortlichkeitslücke (‚responsibility gap') (Matthias, 2004 S. 176) thematisiert.

Grundsätzlich ist es wichtig zu beachten, dass diese Aufzählung möglicher ethischer Herausforderungen nicht vollständig ist. Es muss auch weiterhin auf zukünftige Weiterentwicklungen von KI reagiert und die ethische Auseinandersetzung dahingehend erweitert werden. Doch schon angesichts dieser Risiken, ist der Erwerb von KI-Kompetenz und eine gewisse Grundkenntnis in Ethik für Lehrende und Studierende unerlässlich (Holmes, et al., 2022 S. 543ff.) & (UNESCO, 2021 S. 23, 35).

Fünf Prinzipien Im Umgang mit KI

Aus den vorgestellten aber auch weiteren moralphilosophischen Theorien können konkretere ethische Prinzipien für KI in der Anwendung abgeleitet werden.

Hier sollen beispielhaft fünf übersichtliche Prinzipien nach (Floridi, et al., 2019 S. 5-8) vorgestellt werden:

1. Wohltätigkeit (‚beneficence'): Der Einsatz von KI sollte ausschließlich in einer Weise erfolgen, die menschliches Wohl fördert, menschliche Würde bewahrt und nachhaltig ist.
2. Nichtschädigung (‚non-maleficence'): Während ‚Nichtschädigung' zunächst wie äquivalent zu ‚Wohltätigkeit' erscheint, ist dem nicht zwingend so. Denn, gerade wenn Risiken nicht vollständig vermieden werden können oder Ungewissheit über das Eintreten möglicher Risiken besteht, ist ein Prinzip das zur Schadensvermeidung und Schadensbegrenzung aufruft wichtig und nicht allein durch ein Wohltätigkeitsprinzip vollständig aufgefangen.
3. Autonomie (‚autonomy'): Das Prinzip der Autonomie ruft Nutzende auf die Übertragung von Entscheidungs- und Handlungsgewalt an KI und die Bewahrung von eigener und fremder Handlungs- und Entscheidungsfreiheit gegeneinander abzuwägen. Dabei sind alle möglichen Beteiligten in diese Abwägung miteinzubeziehen und der Wahrung menschlicher Handlungs- und Entscheidungsfreiheit und ihrer Würde besondere Priorität einzuräumen.
4. Gerechtigkeit (‚justice'): Die Nutzung von KI sollte der Förderung des Wohlergehens jedes Menschen, der Erhaltung von Solidarität, und der Vermeidung von Ungerechtigkeiten dienen. Im Umgang mit KI sollte das mögliche Auftreten von Diskriminierung (z.B. in den bereitgestellten Schlussfolgerungen oder in den zugrundeliegenden Datensätzen, die eine KI verwendet) berücksichtigt, ausgeglichen oder zumindest abgedämpft werden. Dabei ist auch auf Gerechtigkeit im Diskurs über Herausforderungen und Lösungsansätze in der Nutzung von KI zu achten.
5. Erklärbarkeit (‚explicability'): Die Nutzung von KI verpflichtet—so weit wie möglich—die eigenen Nutzungsabsichten verständlich zu machen und sich der eigenen Rechenschaftspflicht über die entstehenden Nutzungsfolgen anzunehmen. Darüber hinaus sind diese Aspekte auch in passender Weise transparent zu machen, sodass Nutzen, Risiken und Schäden von allen Beteiligten (und sogar gesamtgesellschaftlich) diskutiert werden können.

Bei diesen fünf Prinzipien nach Floridi, et al., 2019 handelt es sich aber nur um eine Möglichkeit; andere Philosoph:innen haben weitere Alternativvorschläge anzubieten. Ein recht verbreiteter Vorschlag ist das ART-Modell, dessen drei namensgebende Prinzipien (Rechenschaft (‚accountability'), Verantwortung (‚responsibility') und Transparenz (‚transparency')) (Ifenthaler, 2023 S. 79f.) sich auch teils in den obigen Prinzipien wiederfinden lassen.

Mittlerweile gibt es geradezu ein Sammelsurium vorgeschlagener Prinzipien. Auch das bleibt nicht ohne Folgen: Es birgt die Gefahr der Versuchung, dass KI-Nutzer (und Hersteller) sich in der Menge der verfügbaren Prinzipien (‚principle proliferation') (Floridi, et al., 2019 S. 2) nur solche herauszusuchen, die zum eigenen Vorteil sind oder sogar Schädigungen anderer billigend in Kauf nehmen. Diese ‚einfache' Route ist zu vermeiden und bleibt in sich eine ethische Herausforderung.

Für die obige Auswahl möglicher Herausforderungen *könnten* auf Grundlage der vorgestellten Prinzipien folgende Handlungsanweisungen formuliert werden: Nutzende sind in der Pflicht sich zu fragen und haben (ggf. mit möglichen Betroffenen) zu untersuchen, ob generierte KI-Ergebnisse einer eigenen Prüfung standhalten und ob die mögliche Weiterverbreitung der Ergebnisse die Chancengleichheit und Würde aller Menschen achtet.

Nutzende stehen weiterhin in der Pflicht sich zu fragen und haben (ggf. mit möglichen Betroffenen) zu untersuchen, ob eine Eingabe persönlicher Daten ihnen oder anderen schaden könnte, ob die Daten nicht doch anonymisiert werden können oder ganz darauf verzichtet werden könnte.

Hochschulen sind in der Pflicht Richtlinien zum Umgang mit persönlichen Daten zu entwickeln, bereitzustellen, deren Einhaltung mit passenden Überprüfungsinstrumenten (z.B. durch verpflichtende ethische Begutachtung) abzusichern.

Nutzende sollten – zumindest solange, wie Fragen nach Verantwortung und Haftung nicht grundsätzlich abschließend beantwortet sind – abschließende Entscheidungen selbst treffen und damit Verantwortung für das eigene Handeln zu übernehmen. Hochschulen sind in der Pflicht klare Verantwortlichkeiten in allen Bereichen des Hochschulbetriebs zu definieren.

Ethik bietet Begründungsmuster – keine Begründungsvorgaben

Grundsätzlich ist festzuhalten, dass Ethik sich *nur bedingt* eignet um präskriptive Vorgaben zu machen.Noch weniger kann Ethik Vorgaben liefern, die ohne eigene Auseinandersetzung, einfach befolgt oder übernommen werden könnten. Eine ethische Betrachtung ist eine ständige kritische Auseinandersetzung und Abwägung möglicher, aber gegebenenfalls kollidierender Begründungen menschlichen Handelns. Die Auseinandersetzung mit diesen Begründungen ist als gesamtgesellschaftliche Herausforderung eine Aufgabe für alle Menschen gleichermaßen.

Das schließt nicht aus, dass Universitäten und Hochschulen als (Bildungs-)Institutionen und Arbeitgeber in einer Pflicht gegenüber ihren Studierenden und Lehrenden, aber auch der gesamten Gesellschaft sind, bindende Regeln für die Nutzung, aber auch Entwicklung von KI-Systemen zu entwickeln. Wie solche Regeln zum Umgang mit KI in der Nutzung, Forschung und Entwicklung aussehen können, zeigt sich am Beispiel der Richtlinien zur Nutzung von System der Künstlichen Intelligenz an der Technischen Informationsbibliothek (TIB), in der acht Grundsätze festgeschrieben sind und im Annex System nach Anwendungsfall und in Risikoklassen eingeordnet werden (TIB - Leibniz-Informationszentrum Technik und Naturwissenschaften (TIB), 2024).

Weiterhin bleiben Entwickelnde von KI-Systemen, sowohl in Form von Unternehmen, öffentlichen Institutionen und Forschungsgruppen, als auch individuelle Personen in der Pflicht besondere Vorsicht walten zu lassen. Sie haben eine besondere Verantwortung sich allen ethischen Herausforderungen anzunehmen, alle möglichen Folgen ernst zu nehmen und aktiv in den gesamtgesellschaftlichen Diskurs um KI, ihre Nutzung und Entwicklung einzutreten.

Andererseits obliegt es aber weiterhin jedem Menschen sich für die möglichen Risiken und mögliche Schäden, die durch die Nutzung von KI für sie selbst und andere entstehen können, zu sensibilisieren. Ethik erlaubt uns den eigenen Umgang mit KI strukturiert und begründet zu reflektieren, konstruktiv an der gesamtgesellschaftlichen Aushandlung ethischer Problemstellungen teilzunehmen und unsere Verantwortung für das eigene Handeln anzunehmen.

Literaturverzeichnis

Cohen, Joshua. 2007. Deliberative Democracy. [ed.] Shawn W. Rosenberg. *Deliberation, Participation and Democarcy.* London : Palgrave Macmillan, 2007, pp. 219-236. https://doi.org/10.1057/9780230591080_10.

Floridi, Luciano and Cowls, Josh. 2019. A Unified Framwork of Five Principles for AI in Society. *Harward Data Science Review.* 2019, Vol. 1, 1, pp. 1-14. https://doi.org/10.1162/99608f92.8cd550d1.

High-Level Expert Group on Artificial Intelligence. 2019. *A Definition of AI: Main Capabilities and Disciplines.* [Report for the European Commission] Brussels : European Commission, 2019. https://digital-strategy.ec.europa.eu/en/library/definition-artificial-intelligence-main-capabilities-and-scientific-disciplines.

Holmes, Wayne and Tuomi, Ilkka. 2022. State of the art and practice in AI in education. *European Jounral of Education.* 2022, Vol. 57, 4, pp. 542-570. https://doi.org/10.1111/ejed.12533.

Ifenthaler, Dirk. 2023. Ethische Perspektiven auf Künstliche Intelligenz im Kontext der Hochschule. [ed.] Tobias Schmohl, Alice Watanabe and Kathrin Schelling. *Künstliche Intelligenz inder Hochschulbildung: Chancen und*

Grenzendes KI-gestützten Lernens und Lehrens. Bielefeld : transcript, 2023, pp. 71-86. https://doi.org/10.25656/01:27831.

Kant, Immanuel. 1910-1983. *Gesammelte Schriften.* [ed.] Königlich Preußische Akademie der Wissenschaften. Berlin : W. De Gruyter, G. Reimer, 1910-1983. p. 655. https://gallica.bnf.fr/ark:/12148/bpt6k255406/f431.image. Vol. 4. https://gallica.bnf.fr/ark:/12148/bpt6k255406/f431.image.

Matthias, Andreas. 2004. The responsibility gap: Ascrbing responsbility for the actions of learning automata. [ed.] Kluwer Academic Publisher. *Ethics and Information Technology.* 2004, Vol. 6, pp. 175-183. https://doi.org/10.1007/s10676-004-3422-1.

Michel-Villareal, Rosario, et al. 2023. Challanges and Opportunites of Generative AI for Higher Education as Explained by Chat GPT. *education sciences.* 13, 2023, p. 856. https://doi.org/10.3390/educsci13090856.

Sheikh, Haroon, Prins, Corien and Schrijvers, Erik. 2023. *Mission AI: Research for Policy.* Cham : Springer, 2023. pp. 1-410. https://doi.org/10.1007/978-3-031-21448-6. https://doi.org/10.1007/978-3-031-21448-6.

Stoecker, Ralf, Neuhäuser, Christian and Raters, Marie-Luise. 2011/ 2023. Einführung und Überblick. *Handbuch Angewandte Ethik.* 2. Edition. Berlin : J.B.Metzler/ Springer Nature, 2011/ 2023, pp. 3-18. https://doi.org/10.1007/978-3-476-05869-0.

TIB - Leibniz-Informationszentrum Technik und Naturwissenschaften (TIB). 2024. Richtlinien zur Nutzung von Systemen der Künstlichen Intelligenz an der Technischen Informationsbibliothek (TIB). [Online] TIB - Leibniz-Informationszentrum Technik und Naturwissenschaften (TIB), November 19, 2024. [Cited: April 9., 2025.] https://www.tib.eu/de/die-tib/policies/ki-policy.

UNESCO. 2021. *AI and education: Guidance for policy-makers.* Paris : s.n., 2021. https://doi.org/10.54675/PCSP7350.

