

Tiefgreifende Umbrüche

KI im Lehr- und Prüfungsbetrieb

| ALEXANDER PRETSCHNER | DIRK HECKMANN | Die

fortschreitende Entwicklung von Künstlicher Intelligenz (KI) verändert auch die Hochschullehre und -prüfung. Welche Auswirkungen zeigen sich hier? Vier Dimensionen im Fokus.

Software eats the world“ – und Künstliche Intelligenz scheint noch deutlich hungriger zu sein. Ihr Appetit macht auch vor der Hochschulbildung nicht Halt. Die Auswirkungen der fortschreitenden Entwicklung von KI auf den Lehr- und Prüfungsbetrieb an Hochschulen werden in vier Dimensionen betrachtet: 1. KI als Lehrinhalt in beliebigen Fächern, 2. Auswirkungen der KI auf Lerninhalte, 3. KI als Unterstützung für den Lern- und Lehrbetrieb und 4. KI als Herausforderung und Unterstützung für das Prüfungswesen.

KI als Lehrinhalt

Zu dieser ersten Dimension an dieser Stelle nur eine Bemerkung: KI ist nur ein softwaretechnisches Werkzeug unter vielen. Die Debatte sollte daher nicht allein auf KI oder auf den Begriff des computational thinking verengt werden. Vielversprechender erscheint es uns, außerdem das Denken in kybernetischen Systemen zu vermitteln – insbesondere auch jenseits der Informatik. Dazu gehören das Zerlegen komplexer Systeme in Subsysteme mit klar definierten Schnittstellen, ein grundlegendes Verständnis von Daten und Datenmodellierung, von Algorithmen und maschinellem Lernen sowie – wo relevant – von Rückkopplungsprozessen zwischen Maschinen sowie zwischen Mensch und Maschine.

KI und Lerninhalte

Die Auswirkungen von KI auf Lerninhalte außerhalb der KI sind naturgemäß fachspezifisch. Zugleich werfen sie grundlegende Fragen für die Hochschullehre insgesamt auf: Sollen primär Fähigkeiten oder Fertigkeiten vermittelt werden? Welchen Stellenwert hat die Wissensvermittlung in einem kompetenzorientierten Lehrrahmen, wenn Wissen zunehmend verlässlich durch generative KI bereitgestellt werden kann? Und in welchem Maß lässt sich im Studium eine solide Grundlage für das Urteilsvermögen schaffen? Solche Fragen lassen sich am besten anhand konkreter Beispiele diskutieren, etwa die Programmierausbildung im Informatikstudium: Wenn Maschinen künftig immer besser

programmieren können, müssen Studierende dann noch Programmieren lernen? Unsere Antwort: Ja, das bleibt notwendig. Denn bei nichttrivialen und potentiell kritischen Anwendungen wird es immer einen Menschen geben müssen, der die Verantwortung für den erzeugten Code trägt – und der dafür in der Lage sein muss, die Funktionalität und verschiedene Qualitätsmerkmale zu beurteilen. Das ist nur möglich, wenn er selbst programmieren kann. Die Einsicht, dass nur beurteilen kann, wer auch selbst gestalten kann, lässt sich allerdings nicht ohne Weiteres auf alle Bereiche übertragen. Sie gilt beispielsweise weniger für nichtkritischen Code – etwa einfache Webseiten oder Datenbankabfragen, deren Funktionalität und Angemessenheit sich relativ leicht überprüfen lassen.

Eine ähnliche Überlegung lässt sich auch im Jurastudium anstellen. So müssten juristische Methodik, logisches und kritisches Denken sowie der Umgang mit Datenbanken und anderen Wissensquellen im ersten „KI-freien“ Studienjahr intensiviert werden, um alsdann KI als Werkzeug so selbstverständlich in die Fallbearbeitung zu integrieren, wie das bislang bei Gesetzestexten und Kommentaren der Fall war. Ein stets präsenten Basiswissen ist unverzichtbar, Spezialwissen hingegen generierbar und in seiner fachlichen Qualität kontrollierbar: Juristen als Qualitätsmanager und Verantwortungsträger in rechtlichen Kontexten, die Scheinlösungen entlarven.

Die Qualität der Antwort einer generativen KI hängt unmittelbar von der Qualität der Eingabe ab – also des Prompts. Ein guter Prompt zeichnet sich unter anderem durch Präzision, durch die Verwendung fachlich fundierter – im Studium vermittelter – Terminologie

AUTOREN



Foto: TU München

Alexander Pretschner ist Professor für Software & Systems Engineering an der TU München und Gründungsdirektor des Bayerischen Forschungsinstituts für Digitale Transformation (bidt).



Foto: TU München

Dirk Heckmann ist Professor für Recht und Sicherheit der Digitalisierung an der TU München und Direktor des Bayerischen Forschungsinstituts für Digitale Transformation (bidt).

Forschung & Lehre

Foto: xxxxxxxx

und Konzepte sowie durch eine hinreichend vollständige Beschreibung des Problems aus. Gerade bei komplexeren Fragestellungen erfordert dies zunächst die Fähigkeit, ein Problem in sinnvolle Teilprobleme zu zerlegen. Im Sinne von Kleists „allmählicher Verfertigung der Gedanken beim Sprechen“ lässt sich zudem selten bereits zu Beginn die nötige Vollständigkeit des Prompts erreichen. Das Prompten ist deshalb ein iterativer Prozess, bei dem Zwischenergebnisse geprüft, angepasst und schließlich zusammengeführt werden müssen, was wiederum analytische Fähigkeiten und ein gutes Urteilsvermögen voraussetzt. Das heißt nicht, dass der Einsatz generativer KI nicht zu erheblichen Produktivitätsgewinnen führen kann – ganz im Gegenteil. Es bedeutet jedoch, dass im Studium wieder verstärkt Grundlagen gelegt werden müssen für präzises Denken und klaren Ausdruck sowie für die Fähigkeit, Wesentliches von Unwesentlichem zu unterscheiden. Solche Fähigkeiten erwerben Studierende (wie auch Schülerinnen und Schüler) durch Exzerpieren, durch das Verfassen von Protokollen und allgemein durch gezielte Übung. Vor diesem Hintergrund erscheint es uns problematisch, wenn im Zeichen vermeintlicher Technikoffenheit etwa empfohlen wird, generative KI für Recherche, Konzeption und Verfassen jeder Seminararbeit zu nutzen, um den Einsatz anschließend zu reflektieren. Denn der eigentliche Zweck dieses Formats liegt doch nicht im fertigen Produkt, sondern im Prozess seiner eigenständigen Erarbeitung.

KI für den Lehr- und Lernbetrieb

Verschiedene Studien zeigen übereinstimmend, dass inzwischen über 90 Prozent der Studierenden generative KI im Studium nutzen. Das ist nicht aufzuhalten. Diskussionen über Verbote sind unseres Erachtens daher realitätsfern und rückwärtsgewandt. Im Sinne der eingangs

»Ein wesentlicher Sinn von Hausaufgaben liegt gerade darin, dass sie anspruchsvoll und im besten Sinne anstrengend sind.«

verhännten Kompetenzorientierung halten wir allerdings zu viel Unterstützung bei der Bearbeitung von Hausaufgaben für problematisch. Bei allem Verständnis für die Diskussion um kognitive Überforderung oder Frustration im Studium gilt: Ein wesentlicher Sinn von Hausaufgaben liegt gerade darin, dass sie anspruchsvoll und im besten Sinne anstrengend sind. Wenn eine typische Hausaufgabe im Informatikstudium etwa zehn Stunden in Anspruch nimmt und davon sieben Stunden auf das Verfolgen letztlich unbrauchbarer Lösungswege entfallen, dann liegt der Lernerfolg oft genau in diesen sieben Stunden. Die Studierenden lernen, welche Strategien nicht funktionieren, entwickeln Ausdauer und vertiefen ihr Verständnis zentraler Konzepte. Aus diesem Grund sollten KI-gestützte Hilfsmittel, einschließlich sogenannter sokratischer Tutoren, bei der Bearbeitung von Hausaufgaben nur gezielt und maßvoll eingesetzt werden. Diese Sichtweise kollidiert allerdings mit einer stark effizienzorientierten Haltung mancher Studierender, die Hausaufgaben

vor allem als Mittel zur Notengewinnung betrachten – und im weiteren Sinne das Studium allein als Vorbereitung auf Prüfungen. Ein möglicher Ausweg könnte in regelmäßigeren Prüfungsformaten in Präsenz bestehen, bei denen der Einsatz von KI ausdrücklich ausgeschlossen ist – wie es derzeit etwa in der Programmierausbildung an der TU München praktiziert wird. Wir halten gezielte, moderat eingesetzte KI-Interventionen für zielführender als Hausaufgabenhilfen. Ein Beispiel dafür ist das ursprünglich an der TU München entwickelte One-Tutor-System. Es ermöglicht

Studierenden, sich anonym mit einem Chatbot über vorlesungsspezifische Inhalte in der Terminologie dieser Vorlesung auszutauschen. Zusätzlich können sie sich mithilfe von KI-generierten Multiple-Choice- und Freitext-Fragen selbst testen. Unsere bisherigen Erfahrungen zeigen, dass sowohl die generierten Fragen als auch die generierten Antworten derzeit noch durch Dozierende kuratiert werden müssen. Diese wiederum haben Einblick in die anonym gestellten Fragen und die aggregierten Quiz-Ergebnisse – was ihnen ermöglicht, gezielt auf noch unklare Inhalte einzugehen. So entsteht auch in großen Lehrveranstaltungen eine Rückkopplung zwischen Lehrenden und Lernenden. Die Möglichkeiten, die sich darüber hinaus aus einer rechtssicheren Gestaltung individualisierter Lernpfade ergeben, sind beträchtlich. Unserer Überzeugung nach lassen sich Vorlesungen durch solche Systeme effizienter gestalten: virtualisiert, individualisiert und in stärkerem Maße bidirektional. Dennoch glauben wir nicht, dass es sinnvoll ist, Präsenzformate wie

Vorlesungen oder Übungsstunden durch digitale Angebote zu ersetzen. Ziel muss es sein, diese sinnvoll zu ergänzen. Die Idee einer vollständigen Virtualisierung hat bereits im Zuge der Begeisterung für Massive-Open-Online-Courses letztlich kaum nachhaltigen Einfluss auf die Hochschullehre ausgeübt. Ein möglicher Grund dafür liegt darin, dass ein Studium nicht nur vom Wissen der Lehrenden lebt, sondern mindestens ebenso vom Austausch mit anderen Studierenden – fachlich, persönlich und darüber hinaus. Gerade an exzellenten Hochschulen spielen auch entstehende Netzwerke eine zentrale Rolle für die spätere Entwicklung. Dem steht allerdings eine seit dem Ende der Coronapandemie zu beobachtende Campusflucht entgegen. Mögliche Ursachen sind die leichte Verfügbarkeit von Videomaterial, ein wachsendes Anspruchsdenken in Bezug auf Flexibilität, deutlich gestiegene Lebenshaltungskosten – aber womöglich auch ein grundsätzlicher Wandel in der studentischen Haltung gegenüber dem Studium. Der zunehmende und in unseren Augen unvermeidbare Einsatz von KI-basierten Assistenzsystemen wie OneTutor kann diese Entwicklung noch verstärken. Es liegt daher an den Hochschulen, das Campuserlebnis so attraktiv zu gestalten, dass Studierende gerne zurückkehren. An der TU München planen wir deswegen im Wintersemester – unter den dortigen besonderen Skalierungsbedingungen – die Erprobung von Präsenztutorien, die durch eine KI moderiert werden. Es ist zudem gut vorstellbar, dass sich die Rolle der Lehrenden wandeln wird, Persönlichkeitsmerkmale wie Motivationsfähigkeit und Kommunikationsstärke könnten künftig stärker an Bedeutung gewinnen. Auch die Diskussion darüber, in welcher Form es etwa dreißigmal eine sehr ähnliche Einführungsveranstaltung „Informatik 1“ an bayerischen Hochschulen geben muss, dürfte neuen Schwung erhalten. Als Überleitung zu unserer vierten Dimension sei noch erwähnt, dass KI nicht nur den Lernprozess der Studierenden unterstützt, sondern auch Lehrende bei der Erstellung von Lehrmaterialien. Dazu zählen etwa Übungsbeispiele, die Generierung von Skripten aus Vorlesungsmitschnitten sowie die Erstellung und Vorkorrektur von Haus- und Prüfungsaufgaben.

KI und Prüfungen

Prüfungen sind Grundrechtseingriffe, dienen aber auch der Grundrechtserhaltung. Die Existenz und Nutzung generativer KI bewirken einen Paradigmenwechsel, der zur Neugestaltung von Prüfungsformen und -ordnungen zwingt. Weder pauschale Verbote noch

»Generative KI bewirkt einen Paradigmenwechsel, der zur Neugestaltung von Prüfungsformen und -ordnungen zwingt.«

die pauschale Erlaubnis des KI-Einsatzes und schon gar nicht ein (gleichgültiges) stilles Dulden genügen verfassungsrechtlichen Anforderungen eines chancengerechten Prüfungssystems. Ein „Weiter-so“ ist schon deshalb ausgeschlossen, weil Sinn und Zweck von Prüfungen, nämlich die Feststellung eigenständiger Leistungen und eines entsprechenden Studienfortschritts, in Zeiten unkontrollierbarer und unvermeidbarer Co-Creation mit KI-Systemen verfehlt werden. In der Neugestaltung von (Hochschul-)Prüfungen sind die folgenden Fragen zu beantworten: 1. Welche Fähigkeiten und Fertigkeiten werden unter Berücksichtigung der zunehmenden Leistungsfähigkeit von KI-Systemen in spezifischen Arbeitskontexten künftig verlangt und wie können Hochschulen adäquat darauf vorbereiten? 2. Was bedeutet das insbesondere für die Kompetenz der Studierenden, KI sinnvoll einzusetzen und deren Ergebnisse kritisch reflektierend zu kontrollieren? 3. Welche Prüfungsinhalte und -formate eignen sich (überhaupt noch) zur Leistungsfeststellung? 4. Wie kann man den auch durch Art. 5 Abs. 3 Grundgesetz geschützten Gestaltungsspielraum der Lehrenden mit den Bedürfnissen von Studierenden und der Arbeitswelt in Einklang bringen? 5. Was ist auch aus verfassungsrechtlichen Gründen (Gesetzesvorbehalt) auf Gesetzesebene zu regeln, was könnte Gegenstand einer Rechtsverordnung sein (wie etwa bei der Bayerischen Fernprüfungserprobungsverordnung – BayFEV – für elektronische Fernprüfungen) und was gehört in allgemeine und fachspezifische Studien- und Prüfungsordnungen?

Aktuell gibt es vereinzelte Initiativen, aber kein abgestimmtes Vorgehen an den deutschen Hochschulen. KI-Strategien sind sinnvoll, bleiben als Grundlage für eine operative Umsetzung aber meist zu vage. Allemaal werden sie der Heterogenität der Fallgestaltungen in einzelnen Studiengängen kaum gerecht.

Die damit einhergehende Rechtsunsicherheit ist nicht nur ärgerlich, sondern führt in vielen Einzelfällen auch zu verzerrten Prüfungssituationen oder gar einer rechtswidrigen Prüfungspraxis, die weitgehend hingenommen wird, weil die Kontroll-

maßstäbe unklar sind. Deshalb entwickelt das Forschungsprojekt AFFAIRE am Bayerischen Forschungsinstitut für Digitale Transformation (bidt) bis 2027 Musterprüfungsordnungen und stellt diese im Rahmen einer Roadshow an deutschen Hochschulen zur Diskussion.

Druck durch KI-Realität

Wie stark KI am Ende die Hochschulbildung tatsächlich verändern wird, lässt sich derzeit nicht genau sagen. Es steht außer Frage, dass tiefgreifende Umbrüche bevorstehen und dass sich die Hochschulen dem Druck der KI-Realität auf Dauer nicht entziehen können. Zugleich wissen wir heute nicht, wie effektiv KI-gestützte Lehrformate wirklich sind. Deshalb untersuchen wir im Sinn einer Evidenzbasierung am bidt seit diesem Sommersemester im Rahmen eines dreijährigen Projekts, wie effektiv und in welchen Dimensionen der Effektivität das OneTutor-System den Lernprozess unterstützt. Diese Studie zur Evidenzbasierung läuft an zehn bayerischen Universitäten, Hochschulen für angewandte Wissenschaften und der Virtuellen Hochschule Bayern – und zwar in ganz unterschiedlichen Fachrichtungen.

Wir sind fest davon überzeugt, dass es in unserem ureigensten Interesse liegt, frühzeitig Erfahrungen mit solchen Systemen zu sammeln und ihre Wirkung gezielt zu analysieren. Mit einigem Erstaunen stellen wir dabei fest, dass es nicht wenige Hochschulen gibt, die neue Technologien in der Lehre eher aus äußerem Druck als aus innerer Überzeugung einführen – und dass angesichts der anstehenden Umwälzungen die Offenheit für neue Technologien nicht bei allen Kolleginnen und Kollegen überall gleich ausgeprägt ist.