

Dirk Heckmann und Sarah Rachut

Rechtssichere Hochschulprüfungen mit und trotz generativer KI

I. Einleitung

1. Generative KI „überrollt“ universitäres Lernen, Lehren und Prüfen

- a) Perspektive der Studierenden
- b) Perspektive der Hochschule
- c) Perspektive der Lehrenden
- d) Zwischenfazit

2. Einordnung generativer KI im Prüfungswesen

- a) Anknüpfungspunkt „zugelassene Hilfsmittel“
- b) Anknüpfungspunkt „eigenständige Prüfungsleistung“
- c) Fazit

3. Rechtliche Herausforderungen

- a) *de lege lata*
- b) *de lege ferenda*

II. Rechtliche Eckpunkte

1. Relevante rechtliche Anforderungen

- a) Grundrechte (Art. 12 Abs. 1, Art. 3 Abs. 1, Art. 5 Abs. 3 GG)
- b) Rechtsstaatliche Anforderungen, insb. Gesetzesvorbehalt
- c) Einfachgesetzliches Prüfungsrecht
- d) Satzungsrecht (Prüfungsordnungen)
- e) Datenschutzrecht

2. Perspektivisch: „KI-Recht“

III. Eckpunkte zur Systematik des Einsatzes generativer KI

1. Begriffliche Annäherung

- a) Keine Definition von Künstlicher Intelligenz
- b) Zulassung, Einschränkung oder Verbot „technischer Hilfsmittel“ bei der Vorbereitung, Durchführung und (Täuschungs-) Kontrolle von Prüfungen

2. Übersicht der Anwendungsszenarien I – Perspektive der Studierenden

- a) Vorbemerkung: Unterscheidung von beaufsichtigten Prüfungen (z.B. Klausuren) und unbeaufsichtigten schriftlichen Arbeiten (Bachelorarbeiten etc.)
- b) Digitale Anwendungen für „unterstütztes Lernen“
- c) Generative KI als Recherche-Tool zu Prüfungsaufgaben
- d) Generative KI zur formalen Verbesserung eigenständig erstellter Texte
- e) Generative KI als Instrument zu Inspiration und Vorentwurf einer Aufgabenlösung
- f) Generative KI als Co-Erstellerin einer Prüfungsleistung

3. Übersicht der Anwendungsszenarien II – Perspektive der Hochschule

a) Einsatz von KI-Detektoren zur Täuschungskontrolle und Aufdeckung des KI-Einsatzes durch Studierende

b) Exkurs: Einsatz von KI zur Klausuraufsicht

c) Exkurs: Einsatz von KI zum Lehr-/Lern-Monitoring

d) Exkurs: Einsatz generativer KI in der Hochschulverwaltung

4. Übersicht der Anwendungsszenarien III – Perspektive der Lehrenden

a) Einsatz von generativer KI zur Unterstützung der Lehre

b) Einsatz von generativer KI zur Erstellung von Prüfungsaufgaben

c) Einsatz von generativer KI als Bestandteil innovativer Lehr- und Prüfungsformate

IV. Leitgedanken zum rechtssicheren Einsatz generativer KI im Prüfungskontext

1. Disclaimer („work in progress“)

2. Regulierungsbedarf

3. Verhältnismäßiger Ausgleich von Lehr- und Lernfreiheit mit Chancengerechtigkeit

V. Fazit

I. Einleitung

1. Generative KI „überrollt“ universitäres Lernen, Lehren und Prüfen

Seit das US-amerikanische Unternehmen OpenAI im November 2022 seine KI-Anwendung ChatGPT¹ nicht nur einer breiten (weltweiten) Öffentlichkeit vorstellte, sondern auch zur kostenlosen Nutzung freigab, löste dies einen regelrechten Hype aus, der spätestens seit dem Beginn des Jahres 2023 auch Deutschland erreichte. Damit erlebte Künstliche Intelligenz (KI) ihren „iPhone-Moment“, weil diese technische Innovation ähnlich wie Smartphones seit 2007 für viele Menschen intuitiv bedienbar wird. ChatGPT ist ein fortschrittlicher, natur-sprachbasierter Chatbot und basiert auf einer Klasse von künstlichen neuronalen Netzwerken (sog. Transformer-Architektur), die die Sprachverarbeitung (NLP – Natural Language Processing) in den letzten Jahren auf ein Neues Effizienz- und damit auch Qualitätsniveau gehoben haben. KI-gestützte Chatbots wie Chat-GPT haben eine Vielzahl von Fähigkeiten, die es ihnen ermöglichen, auf

¹ Disclaimer: Dieser Aufsatz verwendet ChatGPT des US-amerikanischen Unternehmens OpenAI pars pro toto für die neuere Generation von Large Language Models, die insbesondere der automatisierten Generierung von Texten dienen. Es gibt mehr

als dieses eine Produkt (z.B. Jasper, Google Bard oder Perplexity), aber ChatGPT hat den größten Bekanntheitsgrad, kommt wahrscheinlich auch am häufigsten zum Einsatz und kann damit die hier aufgeworfenen Themen plakativ adressieren.

menschliche Sprache zu reagieren und eine Konversation aufrechtzuerhalten. So kann ChatGPT Fragen in verschiedenen Sprachen (auch Deutsch) beantworten und seine Antworten in der folgenden Interaktion vertiefen. Die Anwendung kann auch aus dem Kontext einer Konversation, also aus deren bisherigem Verlauf, lernen und ist damit in der Lage, künftige Interaktionen innerhalb derselben Konversation zu verbessern. Dabei wirkt es wie ein zwar sehr talentiertes, über großes Wissen verfügendes, aber doch vielfach auch unerfahrenes, ungebildetes und lernbedürftiges (digitales) Assistenzsystem, dessen beeindruckendes Sprachmodell eine massive Wissensbasis zuweilen auch dazu verwendet, plausible Formulierungen für sachlich falsche Aussagen zu formulieren. Damit sind zugleich die größten Vorteile und Chancen und die größten Nachteile und Risiken beschrieben: Bei richtigem, reflektiertem und verantwortungsbewusstem Einsatz bietet ChatGPT eine große Hilfe bei textbasierten Arbeitsprozessen, die schneller, genauer und effizienter werden können – ein eher unkritischer Umgang mit solchen KI-Textgeneratoren bis hin zu ungeprüfter Verwendung seiner Ergebnisse kann jedoch je nach Tragweite der fehlerhaften Textinhalte fatale Folgen haben. Vergleichbares gilt für multimodale Anwendungen, also Bilder, Bewegtbild (Videos) und weiteren Content. Diese technische Entwicklung hat auch für Hochschulen allergrößte Bedeutung: Zum einen, weil Studierende ChatGPT wie selbstverständlich nutzen (können), zum anderen, weil es die verantwortungsvolle Aufgabe von Wissenschaft ist, die mit dieser Entwicklung einhergehenden Disruptionen zu erkennen, einzuordnen und die richtigen Schlussfolgerungen für Lehre und Studium, Prüfung und Arbeitswelt zu ziehen. Dabei sind unterschiedliche Perspektiven einzunehmen.

a) Perspektive der Studierenden

Aus Sicht der Studierenden gilt, dass ein Tool wie ChatGPT so oder so – zumindest von einem Teil – eingesetzt wird, ob nun zur Arbeitserleichterung, mit Täuschungsabsicht oder auch einfach deshalb, weil generative KI schon bald ein selbstverständlicher Bestandteil großer Teile des Arbeitslebens und seiner Wertschöpfungsketten sein wird: Kann und darf man das im Studium ignorieren? Haben Studierende nicht einen Anspruch darauf, dass die Hochschulen sie auf ihr künftiges Berufsleben angemessen vorbereiten – und zwar unter den künftigen Bedingungen, nicht nach Maßgabe überkommener Curricula?

b) Perspektive der Hochschule

Für die Hochschulen stellen sich – vice versa – grundlegende Fragen, wie sie die durch generative KI noch einmal beschleunigte digitale Transformation insbesondere der Arbeitswelt in ihren Studien- und Prüfungsstrukturen berücksichtigen können. Ganz kurzfristig müssen sie aber auch darauf reagieren, dass Prüfungsergebnisse durch den Einsatz generativer KI verfälscht sein können. Können oder müssen Hochschulen solche innovativen Anwendungen am Ende sogar zumindest im Prüfungsgeschehen verbieten? Wie aber kontrolliert man dies?

c) Perspektive der Lehrenden

Die Lehrenden wiederum sind jene Akteure, die einerseits den Bedürfnissen der Studierenden Rechnung tragen müssen, andererseits Vorgaben der Hochschulleitung umsetzen sollen, soweit dies etwa der Unterbindung von Täuschungshandlungen bzw. der Herstellung von Chancengerechtigkeit dient. Bei alledem muss ihr eigener Gestaltungsspielraum in fachlicher und didaktischer Hinsicht erhalten bleiben. Dies bietet ein breites Spektrum an Reaktionsmöglichkeiten: von der proaktiven eigenen Nutzung generativer KI bis zur Entwicklung von Bewertungsmaßstäben in Bezug auf die neue Mensch-Maschine-Interaktion.

d) Zwischenfazit

Die mit dem Einsatz generativer KI verbundenen hochschulrelevanten Fragen sind vielschichtig. Auch wenn sich etwa die Ebenen notwendiger Studienreformen, Anpassungen der Prüfungsordnungen, Kontrollfragen und strategischen Ausrichtungen zum Teil überschneiden, soll doch der Schwerpunkt dieses Aufsatzes auf der Regulierung in einer allgemeinen Prüfungsordnung liegen.

2. Einordnung generativer KI im Prüfungswesen

a) Anknüpfungspunkt „zugelassene Hilfsmittel“

Generative KI kann als ein (technisches) Hilfsmittel angesehen werden, welches die Prüfungsteilnehmenden bei der Erstellung einer (schriftlichen) Prüfungsleistung auf unterschiedliche Arten unterstützen kann, je nachdem wie und bis zu welchem Punkt man es einsetzt (hierzu näher unten III.2). Die mögliche „Unterstützungsleistung“ reicht von einfachen Erklärungen einer Prüfungsaufgabe über Recherchen und stilistische Textvorschläge bis zur Erstellung eines kompletten Textentwurfs.

Damit vereint ein Tool wie ChatGPT letztlich die Funktionalität von Suchmaschinen, Datenbanken, Schreibprogrammen und Unterstützungsleistungen, die bisher von anderen Personen vorgenommen wurden (Korrekturlesen, Sprachoptimierung, „Sparringspartner“). Genau hier ist anzusetzen, wenn man die Grenzen des Einsatzes generativer KI im Prüfungskontext bestimmen und regulatorisch festhalten will. Anknüpfungspunkt ist insofern das in der jeweiligen Hochschulprüfung „zugelassene Hilfsmittel“. Auch wenn die Prüfungsordnungen der Hochschulen regelmäßig jede Nutzung eines nicht ausdrücklich zugelassenen Hilfsmittels als Täuschungshandlung bewerten, ist hierbei zu differenzieren. Gerade bei unbeaufsichtigten Prüfungsformaten, wie Seminar- oder Hausarbeiten, werden nicht die zugelassenen Hilfsmittel (positiv) benannt, sondern vielmehr unerlaubte Handlungsweisen, wie die Bearbeitung durch eine andere Person (negativ) ausgeschlossen. Nachdem Systeme wie ChatGPT viele der bisher zugelassenen oder zumindest tolerierten Hilfsmittel (Suchmaschinen, Übersetzungsleistungen o.ä.) umfasst, ist nicht generell davon auszugehen, dass die Nutzung von ChatGPT per se die eines unerlaubten Hilfsmittels ist.² Auf welcher Normebene eine Einschränkung von ChatGPT im Rahmen der zugelassenen Hilfsmittel überhaupt reguliert werden kann oder soll und wie sich dies wiederum zum Gestaltungsspielraum der Lehrenden verhält, bleibt zu untersuchen (hierzu unten II.1 b und IV.3).

b) Anknüpfungspunkt „eigenständige Prüfungsleistung“

Neben den „zugelassenen Hilfsmitteln“ stellt die Eigenständigkeit der Prüfungsleistung einen weiteren Anknüpfungspunkt für eine mögliche Regulierung dar. Dies ergibt sich daraus, dass generative KI kein herkömmliches Hilfsmittel wie Fachliteratur oder ein Rechenprogramm darstellt, dessen Einsatz die Eigenständigkeit der eigentlichen Prüfungsleistung nicht in Frage stellen würde. Vielmehr liegen die generativer KI zugrundeliegende Methode des maschinellen Lernens und der daraus entstehende Leistungsumfang im Rahmen dessen, was zumindest bei herkömmlichen Prüfungen der zu bewertenden Leistung des Kandidaten entspricht. Je nachdem, wie man die zu erbringende Prüfungsleistung genau bestimmt bzw. vorgibt, kann aus dem Einsatz generativer KI die Schlussfolgerung gezo-

gen werden, dass die Prüfungsleistung nicht mehr oder noch eigenständig von dem Prüfungskandidaten erbracht wurde.

c) Fazit

Der Einsatz generativer KI durch Prüfungsteilnehmende entwertet die Prüfungsleistung nicht a priori, er ändert jedoch den tatsächlichen Kontext des Prüfungsgeschehens. Es kommt daher entscheidend darauf an, wie in der jeweiligen Prüfung die „zugelassenen Hilfsmittel“ bestimmt wurden und was genau die zu erbringende (eigenständige) Prüfungsleistung darstellen soll. Genau hier kann und sollte die Regulierung seitens der Hochschule ansetzen, dies auch in Verbindung mit den entsprechenden Vorgaben der Lehrenden, was die (eigenständige) Prüfungsleistung ist und welche (technischen) Hilfsmittel zugelassen werden.

3. Rechtliche Herausforderungen

a) de lege lata

Der Einsatz bzw. die Nutzung generativer KI wirft vielfältige Rechtsfragen auf und berührt dabei zahlreiche relevanten Rechtsnormen des aktuell geltenden Rechts („de lege lata“). Im Prüfungskontext umfasst dies besonders die von den Hochschulen zu beachtenden Grundrechte aller Akteure, aber auch die allgemeinen Prüfungsgrundsätze, die zum Teil in Prüfungsordnungen und darüber hinaus im Hochschulrecht verankert sind. Weil es hier explizit auch um die Verarbeitung personenbezogener Daten geht, spielt das Datenschutzrecht, allen voran die Datenschutzgrundverordnung (DSGVO) eine wichtige Rolle (hierzu näher II.)

b) de lege ferenda

Es empfiehlt sich, darüber hinaus auch jene Rechtsvorschriften in den Blick zu nehmen, die derzeit noch keine Geltungskraft haben, thematisch aber durchaus einschlägig sind. Das betrifft besonders die sog. KI-Verordnung (engl. Artificial Intelligence Act) der EU. Zwar wird diese („de lege ferenda“) voraussichtlich erst im Juni 2024 in Kraft treten und dann aufgrund einer ausdrücklich normierten Übergangsphase erst sukzessive Geltung erlangen. Es wäre aber fahrlässig, dies gar nicht erst in die Überlegungen einzubeziehen, weil man sonst eine Prüfungsordnung und eine neue Prüfungspraxis schafft, die kurze Zeit später wieder zu ändern wäre.

² Nicht überzeugend hingegen *Birnbaum*, NVwZ 2023, 1127, 1128: „Dass KI aus sich heraus kein erlaubtes Hilfsmittel ist, versteht sich, ohne dass es dafür einer gesonderten Regelung in einer Prüfungsordnung bedarf. Jegliche Hilfsmittel sind verboten, die die notwendige Eigenleistung des Prüflings substituieren.“ Diese

Begründung setzt voraus, was erst begründet werden müsste: „Verboten ist, was verboten sein sollte“. Sie ist damit nicht tragfähig, auch weil pauschal von „KI“ gesprochen wird, obwohl es zahlreiche Arten und Anwendungen von Künstlicher Intelligenz gibt.

II. Rechtliche Eckpunkte

Beim Umgang mit generativer KI im Prüfungskontext sind sowohl verfassungsrechtliche als auch einfachgesetzliche und untergesetzliche Normen zu beachten. In diesem Beitrag werden diese in ihrer Bedeutung und in ihrem einschlägigen Inhalt dargestellt. Eine vertiefte Analyse bleibt weiteren Beiträgen vorbehalten, nicht zuletzt in einer Anschlussforschung, die der vorliegende Beitrag anstoßen möchte.

1. Relevante rechtliche Anforderungen

a) Grundrechte (Art. 12 Abs. 1, Art. 3 Abs. 1, Art. 5 Abs. 3 GG)

Von besonderer Bedeutung sind die einschlägigen Grundrechte, an die zumindest staatliche Hochschulen gebunden sind (Art. 1 Abs. 3 GG).

So folgt aus Art. 12 Abs. 1 GG (Berufsfreiheit, Ausbildungsfreiheit), dass den Studierenden, die in einem bestimmten Studiengang eingeschrieben sind, ein berufsqualifizierender Abschluss mit entsprechenden Prüfungen nicht vorenthalten werden darf. Das hatte gerade während der Pandemie eine besondere Bedeutung, weil dieser Prüfungsanspruch nur durch das Angebot elektronischer Fernprüfungen aufrechterhalten werden konnte.³ Darüber hinaus ist aber zu bedenken, dass sowohl die Hochschulen und ihre Fakultäten als auch die Lehrenden selbst innerhalb hochschulgesetzlicher Vorgaben einen Gestaltungsspielraum bei der Ausgestaltung der Prüfungsformate haben. Pointiert gesagt: Die Studierenden haben keinen grundrechtlich fundierten Anspruch auf innovative Lehrinhalte und Prüfungsbedingungen, solange Lehre und Prüfung nicht komplett an der Lebenswirklichkeit vorbeigehen. Insofern könnte die Billigung des Einsatzes generativer KI bei Hochschulprüfungen zwar bildungs- und arbeitsmarktpolitisch wünschenswert sein – dies nicht zu billigen wäre aber noch keine Verletzung von Art. 12 Abs. 1 GG.

Im Mittelpunkt der Grundrechtsbindung von Hochschulen in Bezug auf das Prüfungsgeschehen steht der Gleichheitssatz (Art. 3 Abs. 1 GG), hier verstanden als die Gewährleistung von Chancengleichheit. Der Grundrechtsschutz der Prüfungsteilnehmenden besteht besonders darin, dass eine faire, für alle gleiche Prüfungssituation herzustellen ist. Danach soll niemand einen unberechtigten Vorteil etwa durch Täuschungshandlungen

erlangen, weil ansonsten Ungleiches (unterschiedliche Leistungsstärken) gleichbehandelt würde (im Sinne einer vergleichbar guten Note, die nur durch Nutzung nicht zugelassener Hilfsmittel erlangt wurde), ohne dass ein sachlicher Grund dies rechtfertigt. Auch hier muss man allerdings berücksichtigen, dass eine „echte Chancengleichheit“ ohnehin nicht herstellbar ist – zu groß sind schon die faktischen Unterschiede der Teilnehmenden, etwa in Bezug auf deren finanzielle Möglichkeiten, Lernverhältnisse oder bestimmte kognitive Fähigkeiten. So versteht sich die grundrechtliche Perspektive auf Chancengleichheit auch eher im Sinne einer zumutbaren Optimierung der Prüfungsumstände⁴ und einer Vermeidung von Willkür.⁵

Zu beachten ist schließlich auch die Lehrfreiheit (als Ausdruck der Wissenschaftsfreiheit, Art. 5 Abs. 3 GG), die den Lehrenden ein Minimum an prüfungsdidaktischen Entscheidungen gewährt. Vor diesem Hintergrund darf die Hochschule Prüfungsinhalte und Prüfungsformate nur allgemein vorgeben, ohne den Grundsatz „wer lehrt, der prüft“ auszuhebeln. Insbesondere muss es bei entsprechenden Vorgaben einen sachlichen Grund geben, dies zentral zu regeln.

Die genannten Grundrechte stehen zuweilen auch in einem Spannungsverhältnis. So könnte zur Optimierung der Chancengleichheit auf eine Prüfungsform und eine Prüfungsumgebung hingewirkt werden, die sich durch möglichst wenige Variablen und damit Möglichkeiten für Unterschiede zwischen den Studierenden auszeichnet. Dies wäre z.B. bei einer Prüfung ohne zugelassene Hilfsmittel und ausschließlich geschlossene Fragestellungen, die nur eine korrekte Lösung zulassen, der Fall. Gleichwohl können anders ausgestaltete Prüfungen, die auch den Umgang mit in der Praxis sowieso verfügbaren Hilfsmitteln einbeziehen, dem Berufsqualifizierungsziel besser gerecht werden.

b) Rechtsstaatliche Anforderungen, insb. Gesetzesvorbehalt

Die Ausgestaltung einer Hochschulprüfung ist nicht alleine Sache der Hochschulen oder der Lehrenden. Vielmehr bedarf es gesetzlicher Vorgaben, soweit etwa Grundrechte wesentlich betroffen sind (Wesentlichkeitstheorie des BVerfG). Das bedeutet: gerade eine Regelung des Einsatzes generativer KI berührt sowohl Art. 12 Abs. 1 GG als auch Art. 3 Abs. 1 GG oder Art. 5 Abs. 3 GG. Sowohl eine pauschale Erlaubnis als

³ Hierzu Heckmann/Rachut, E-Klausur und Elektronische Fernprüfung, 2023, S. 150 ff.

⁴ S. hierzu etwa zu den aus dem Gleichheitsgrundsatz folgenden Vorgaben für die Gestaltung von Prüfungen Heckmann/Rachut,

E-Klausur und Elektronische Fernprüfung, 2023, S. 166 ff.

⁵ Vgl. Jeremias, in: Fischer/Jeremias/Dieterich, Prüfungsrecht, 8. Aufl. 2022, Rn. 402 f.

auch ein pauschales Verbot müssten entweder im Hochschulgesetz verankert werden oder es müsste eine Ermächtigung an den Verordnungsgeber bzw. Satzungsgeber erteilt werden, dies nach Maßgabe bestimmter gesetzlicher Anforderungen zu regeln. In ähnlicher Weise hatte etwa der bayerische Gesetzgeber 2020 eine Verordnungsermächtigung zur Einführung elektronischer Fernprüfungen geschaffen, nachdem die Pandemie Präsenzprüfungen wesentlich erschwerte. Auf dieser Grundlage entstand wiederum die Bayerische Fernprüfungserprobungsverordnung, die zum Vorbild für das Fernprüfungsrecht in ganz Deutschland wurde.⁶

Ob etwas Ähnliches bei der Regulierung der prüfungsrechtlichen Anforderungen an den Einsatz generativer KI gelingt, ist fraglich, weil der Handlungsdruck kaum gesehen wird bzw. zu spüren ist. Auch wenn eine gesetzliche Rahmensetzung unter dem Aspekt des Gesetzesvorbehalts notwendig erscheint, wäre schon die Regelung in einer untergesetzlichen Norm, etwa der „Allgemeinen Prüfungsordnung“ auf Hochschulebene ein rechtsstaatliches „Plus“ gegenüber dem jetzigen Rechtszustand.

c) Einfachgesetzliches Prüfungsrecht

Auf der Ebene des einfachen Gesetzesrechts finden sich Ansätze für das Prüfungsrecht insbesondere in den Landeshochschulgesetzen. Die dort zu finden Vorschriften zu Prüfungen und Prüfungsordnungen⁷ regeln insoweit nur allgemeine Grundsätze wie die Art und die Bewertung von Prüfungsleistungen – auf Besonderheiten einer Mensch-Maschine-Interaktion wie beim Einsatz generativer KI in Prüfungen gehen diese Vorschriften aber (nachvollziehbar) nicht ein.

Wenn man insbesondere die Ermächtigung zur Regelung in Prüfungsordnungen der Hochschule⁸ sehr großzügig auslegt, könnte man zumindest einstweilen auf die Satzungsebene ausweichen, um den Rahmen für den Einsatz generativer KI zu regeln.

d) Satzungsrecht (Prüfungsordnungen)

Auch in den bestehenden Rahmenprüfungsordnungen der Hochschulen für den Bachelor-Studiengang bzw. für den Master-Studiengang finden sich bislang keine expli-

ziten Regelungen zum Umgang mit (generativer) KI im Prüfungskontext. Dies wäre allerdings der richtige Ort für die eine oder andere Klarstellung. Dies betrifft insbesondere (am Beispiel einer Bachelor-Ordnung der Universität Freiburg⁹):

- § 6: Fachspezifische Bestimmungen
- § 12: Bachelorprüfung
- § 13: Studienleistungen
- § 14: Studienbegleitende Prüfungsleistungen
- § 17: Studienbegleitende schriftliche Prüfungsleistungen
- § 18: Online-Prüfungen
- § 18a: Datenverarbeitung bei Online-Prüfungen
- § 21: Bachelorarbeit
- § 23: Rücktritt, Täuschung, Ordnungsverstoß

Dort könnten Einzelheiten zu den zugelassenen Hilfsmitteln, zur Eigenständigkeit der Leistung im Rahmen einer Mensch-Maschine-Interaktion, zu den Bewertungsmaßstäben, zu den Grenzen der KI-Nutzung u.a.m. geregelt werden.

e) Datenschutzrecht

Weil der Einsatz generativer KI immer auch mit der Verarbeitung personenbezogener Daten einhergeht, sind die Anforderungen des Datenschutzrechts, insbesondere der DSGVO zu beachten. Hierzu zählen etwa die allgemeinen Grundsätze (Art. 5 DSGVO), die Notwendigkeit einer Rechtsgrundlage (Art. 6 DSGVO), die Gestaltung eines Einwilligungsprozesses (Art. 7 ff. DSGVO), allgemeine Transparenzanforderungen (Art. 12 DSGVO) und Informationspflichten (Art. 13 ff. DSGVO) sowie das Verbot einer automatisierten Prüfungsentscheidung (Art. 22 DSGVO). Hinzu kommt ggf. das Erfordernis einer Datenschutzfolgenabschätzung (Art. 35 DSGVO). Welche genauen Anforderungen gelten, hängt stark von dem Einsatzszenario ab: Geht es um die Nutzung der KI durch die Studierenden im Rahmen registrierter Nutzerkonten oder um den KI-Einsatz durch die Hochschule im Rahmen von Lehrveranstaltungen oder zur Täuschungskontrolle?

In erster Linie hat der Anbieter von Systemen generativer KI (also etwa das Unternehmen OpenAI im Fall von ChatGPT) die Anforderungen der DSGVO (sowie

⁶ Grundlegend hierzu Heckmann/Rachut, E-Klausur und Elektronische Fernprüfung, 2023. Vgl. auch Rachut, E-Klausur und elektronische Fernprüfung: Technologischer Fortschritt und Prüfungskulturwandel im Spiegel des Rechts - Ein Werkstattbericht, ODW 2023, S. 89 ff.

⁷ Vgl. etwa §§ 32, 32a HochSchG BW („Die Prüfungsordnungen enthalten Regelungen zum Prüfungsverfahren und den Prüfungsanforderungen, insbesondere über ... 3. die Bewertung von Prüfungsleistungen.“); Art. 84 Abs. 3 BayHIG („Die Prüfungsordnung regelt die wesentlichen Fragen im Hinblick auf Prüfungsan-

forderungen und Prüfungsverfahren, insbesondere 1. den Zweck der Prüfung, die Gegenstände der Prüfung und die Anforderungen in der Prüfung, ... 10. die Grundsätze für die Bewertung der einzelnen Prüfungsleistungen“).

⁸ Zum Beispiel § 32 Abs. 3 HochSchG BW, Art. 84 Abs. 3 Satz 1 BayHIG.

⁹ https://www.jsl.uni-frei-burg.de/informationen_fuer_studierende_web/pruefungsordnungen/bachelor_of_science/b_sc__pruefungsordnung_aktuell_rahmenordnung.pdf.

künftig der KI-Verordnung) einzuhalten. Weil dies zumindest in Frage gestellt werden kann, hatte etwa die Datenschutzaufsichtsbehörden in Italien ChatGPT zwischenzeitlich verboten; die deutsche Datenschutzkonferenz hat OpenAI einen umfangreichen Fragenkatalog zur Beantwortung übermittelt, um den Sachverhalt zu klären, bevor etwaige Aufsichtsmaßnahmen ergriffen werden. Ob Hochschulen für den Fall, dass sie solche Systeme generativer KI gleichsam „offiziell“ in ihr Prüfungsgeschehen (bzw. auch übergeordnet in die Lehre) einbinden, Verantwortliche i.S.d. DSGVO werden, bleibt zu klären. Insoweit wären die Maßstäbe zu prüfen, die der EuGH in seiner Entscheidung zu den Facebook-Fanpages vorgegeben hat¹⁰. Zweifellos datenschutzrechtlich relevant wäre demgegenüber der Einsatz von sog. KI-Detektoren, weil bei der Kontrolle von Prüfungsarbeiten personenbezogene Daten der Studierenden verarbeitet werden.¹¹ Hierzu bedürfte es einer ausdrücklichen gesetzlichen Ermächtigungsgrundlage (die derzeit fehlt), weil die Einwilligung in solchen Fällen nicht freiwillig erteilt werden kann.¹² Außerdem sind die strengen Anforderungen des Art. 22 DSGVO¹³ zu beachten, wenn und soweit ein negatives Prüfungsergebnis unmittelbar oder mittelbar auf das Ergebnis des Prüfungsvorgangs des KI-Detektors gestützt wird.

2. Perspektivisch: „KI-Recht“

Derzeit ist der KI-Einsatz in öffentlich-rechtlichen Kontexten (wie eben innerhalb einer staatlichen Hochschule) kaum reguliert – am ehesten greifen, wie gesehen, noch datenschutzrechtliche Anforderungen. Dies wird sich ändern, sobald die europäische KI-Verordnung (KI-VO), die gerade das Trilogverfahren durchlaufen hat, in Kraft getreten ist und Geltung beansprucht. Welche Auswirkungen dies für den Einsatz generativer KI in Studium, Lehre und Prüfung haben wird, hängt aufgrund des risikobasierten Regulierungsansatzes wiederum von dem jeweiligen Einsatzszenario ab. Es ist empfehlenswert, diese Entwicklung bei der eigenen Regulierung im Auge zu behalten. Im Rahmen der Trilogverhandlungen kam es zuletzt nochmals zu umfassenden Änderungen,

insbesondere hinsichtlich der der KI-VO zugrundeliegenden Begriffsdefinition von KI-Systemen, sowie der Regulierung generativer KI.¹⁴ Die in der bisher verfügbaren englischen Version als „general purpose AI models“ bezeichneten KI-Systeme werden nochmals einer eigenen Risikobewertung unterzogen und „general purpose AI models with systemic risk“ umfassender geregelt. Die Vorgaben der KI-VO richten sich hierbei vordergründig an die Anbieter der KI-Systeme, jedoch ebenso an die Betreiber, die Einführer, Händler und die Endnutzenden.¹⁵ Somit wirkt sich die Nutzung einer möglicherweise europarechtswidrigen Plattform auch für diejenigen aus, die eine solche Nutzung dulden, erlauben oder gar ausdrücklich in ihrem Prüfungssystem verankern. Neben der Frage, wie generative KI durch die KI-VO reguliert ist, steht im jeweiligen Einsatzszenario die Zuordnung zu einer der durch die KI-VO vorgesehenen Risikogruppen im Vordergrund. Gerade die sog. Hochrisiko-KI-Systeme, zu denen gem. Art. 6 Abs. 2 KI-VO i.V.m. Anhang III Nr. 3 ebenso KI-Systeme im Kontext von allgemeiner und beruflicher Bildung zählen, werden durch die Verordnung umfassend reguliert. Allemal ist schon jetzt anzuraten, Formulierungen zu wählen, die der dynamischen Entwicklung in diesem Sektor Rechnung tragen.

III. Eckpunkte zur Systematik des Einsatzes generativer KI

1. Begriffliche Annäherung

a) Keine Definition von Künstlicher Intelligenz

Auch wenn sich die empfohlene Regulierung des Einsatzes generativer KI (wie etwa ChatGPT) zumindest auf der Ebene der Prüfungsordnungen mit einer Software befasst, die (schon vom Namen her) mit „Künstlicher Intelligenz“ zu tun hat, ist es weder möglich noch ratsam, eine Legaldefinition von KI aufzunehmen. Eine solche gleichsam abschließende Definition scheitert schon daran, dass es weder in der Wissenschaft noch in der Rechtspraxis einen Konsens darüber gibt, was KI genau

¹⁰ Hierzu auch *Paschke*, Social Media-Nutzung von Hochschulen vor dem Aus? Verfassungsrechtliche Analyse der Untersagungsverfügung des BfDI gegen das BPA vom 17.02.2023, ODW 2023, S. 165 ff.

¹¹ Vgl. zu den rechtlichen Bedenken gegenüber KI-Systemen, die vorgeben, Abweichungen vom Standardverhalten als Indiz für einen Täuschungsversuch in Prüfungen zu messen *Rachut/Besner*, Künstliche Intelligenz und Proctoring Software - Einsatzfelder und rechtlicher Rahmen im Kontext von elektronischen Fernprüfungen an Hochschulen. MMR 2021, S. 851 ff.

¹² S. u.a. *Nehlsen/Fleck*, DSRITB 2023, 15, 20 f.; sowie allgemein für die Tauglichkeit der Einwilligung im Verhältnis Hochschule und Studierende im Prüfungskontext *Heckmann/Rachut*, E-Klausur

und Elektronische Fernprüfung, 2023, S. 181 f.

¹³ Hier sind auch die Einschränkungen für automatisierte Entscheidungen zu berücksichtigen, die der EuGH kürzlich in seinem Urteil v. 7.12.2023 (C-634-21, „Schufa“) CR 2024, 29 ff. mit Anm. *Kerkemeyer*, vorgenommen hat.

¹⁴ Die Regulierung von generativer KI wurde zwischenzeitlich in Art. 28a und Art. 28b KI-VO-E vorgesehen, wobei die sog. „Basismodelle“ (engl. foundation models) im Mittelpunkt standen. Die nunmehr im Trilog abgestimmte Fassung sieht hingegen Regelungen in den Art. 52a ff. KI-VO für „general purpose AI models“ vor und führt hierbei eine eigenen Risikostufe („general purpose AI models with systemic risk“) ein.

¹⁵ *Bomhard/Siglmüller*, RD 2024, 45, 46.

ist und wie man es so definiert, dass es alles Wesentliche umfasst und sich gegenüber verwandten IT-Technologien trennscharf abgrenzen lässt.

Dies gelingt nicht einmal der KI-Verordnung, deren Art. 3 Nr. 1 einer der zentralen Diskussionspunkte im Trilogverfahren war¹⁶ und nunmehr lautet:

„AI system‘ is a machine-based system designed to operate with varying levels of autonomy and that may exhibit adaptiveness after deployment and that, for explicit or implicit objectives, infers, from the input it receives, how to generate outputs such as predictions, content, recommendations, or decisions that can influence physical or virtual environments“.

Die Datenethikkommission der Bundesregierung verzichtete in ihrem Abschlussbericht 2019 auf eine Definition von KI und lieferte stattdessen folgende Beschreibung

„Sammelbegriff für diejenigen Technologien und ihre Anwendungen, die durch digitale Methoden auf der Grundlage potenziell sehr großer und heterogener Datensätze in einem komplexen und die menschliche Intelligenz gleichsam nachahmenden maschinellen Verarbeitungsprozess ein Ergebnis ermitteln, das ggf. automatisiert zur Anwendung gebracht wird.“¹⁷

Schon hier wird deutlich, dass es zwar bestimmte prägende Merkmale gibt, über die man sich verständigen kann – ebenso aber viele offene Flanken, die einen Interpretationsspielraum lassen (und angesichts der Dynamik der technischen Entwicklung vielleicht auch lassen müssen). So gesehen geht es bei KI annäherungsweise um maschinelles Lernen auf der Grundlage großer Datensätze, woraus automatisiert Ergebnisse ermittelt werden, die nicht unbedingt bereits Bestandteil einer menschlichen Eingabe in den algorithmischen Prozess entsprechen. Damit ist gleichsam ein quasi schöpferischer informationstechnischer Prozess umschrieben, der an menschliche Kreativität erinnert. Genau diese Nähe innovativer Computerleistungen zu menschlicher Leistungsbeschreibung ist es, die das Prüfungswesen herausfordert. Was ist eine geistige menschliche Leistung noch wert, wenn sie eine Maschine ebenso erbringen kann?

b) Zulassung, Einschränkung oder Verbot „technischer Hilfsmittel“ bei der Vorbereitung, Durchführung und (Täuschungs-) Kontrolle von Prüfungen

Statt allgemein und etwas diffus mit dem (Rechts-) Begriff der KI zu operieren, bietet es sich an, an den Begriff der „technischen Hilfsmittel“ anzuknüpfen. Bis-

lang ist die Zulassung von Hilfsmitteln Sache der Prüfenden in ihrem jeweiligen Verantwortungsbereich. So können etwa in Klausuren der Mathematik, Informatik oder ähnlichen technischen Disziplinen Taschenrechner, Formelsammlungen oder auch bestimmte Datenbanken oder Programme genutzt werden. Noch weitergehend öffnet man bei sog. Open Book-Klausuren die Möglichkeit, auf Teile oder das gesamte Lernmaterial zurückzugreifen oder sogar darüberhinausgehende Recherchen während der Prüfungszeit zu erlauben. Welche Hilfsmittel zugelassen werden, hängt auch vom jeweiligen Prüfungsfach und dessen Anforderungen ab. Je mehr es etwa auf die Beherrschung bestimmter Methoden zur Anwendung von Wissen ankommt, um so eher mag es vertretbar sein, das Wissen nicht als Präsenzwissen zu verlangen und (mit) abzufragen, sondern stattdessen die Verhältnisse abzubilden, die die Absolventen in der späteren Arbeitspraxis vorfinden. Dort aber kann man vielfach auf schnell erreichbares Wissen (etwa in Datenbanken) zurückgreifen, es wird also nicht verlangt, dass man dies gleichsam „auswendig“ aufsagen könnte. Auch vor diesem Hintergrund wäre es aber verfehlt, auf jegliches Lernen von Fakten zu verzichten.

Ohne ein stabiles Fundament an präsentem Wissen wird man auch die Systeme zum Wissensabruf nur unvollkommen bedienen können. So kann etwa ein Gesetzeskommentar hilfreich sein, um einen komplexen juristischen Fall zu lösen. Innerhalb kurzer Zeit wird man damit aber wenig anfangen können, wenn man nicht genau weiß, wonach man sucht.

Nimmt man dies als Ausgangslage, kann und muss bei der Frage, inwieweit so etwas wie generative KI auch im Kontext von Hochschulprüfungen zugelassen werden kann, mehrfach unterschieden werden:

Eine erste Unterscheidung betrifft die Art der Prüfung. Bereits aus prüfungsdidaktischen Gründen wird man zunächst zwischen Klausuren unter Aufsicht und schriftlichen Arbeiten wie Seminarreferaten, Bachelor- oder Masterarbeiten unterscheiden müssen. Schon weil etwaige Kontrollmaßnahmen im letztgenannten Fall ungleich schwerer umzusetzen wären, soll dieser im Mittelpunkt dieses Aufsatzes stehen. Inwieweit man die daraus resultierenden Empfehlungen dann auf Aufsichtsklausuren übertragen könnte, wäre später zu diskutieren.

Eine zweite Unterscheidung betrifft den Zeitpunkt des generativen KI-Einsatzes. So kann man zwischen der Vorbereitung/Einarbeitung (Vorfeld), der Texterstellung (Prüfungsleistung) und der späteren Bewertung (Kontrolle) unterscheiden.

¹⁶ Zu den verschiedenen begrifflichen Annäherungen s. z.B. Bronner, jurisPR-ITR 15/2023 Anm. 2.

¹⁷ Gutachten der Datenethikkommission, 2019, S. 34.

- Was die Vorbereitung der Prüfungsleistung betrifft (das Einarbeiten in eine Aufgabenstellung, Recherchen, das Erschließen von Wissen und Methodik etc.), kann es kaum eine Einschränkung geben. So zählt es zur (auch grundrechtlich geschützten) „Studierfreiheit“ (Art. 12 Abs. 1 i.V.m. Art. 5 Abs. 3 GG), dass Studierende weitgehend selbst entscheiden können, auf welche Weise sie sich den Prüfungsstoff erschließen (Besuch von Vorlesungen/Übungen, Selbststudium, privater Repetitor, Lektüre von wissenschaftlichen Lehrbüchern oder Skripten, Nutzung von Lernsoftware etc.). Insofern hat es im Laufe der letzten Jahrzehnte ohnehin schon einen Wandel gegeben: Während es früher unentbehrlich war, Literaturrecherchen für wissenschaftliche Arbeiten in Bibliotheken vorzunehmen, lässt sich dies heute bequem von einem Laptop mit Internetzugang über zahlreiche Datenbanken erledigen. Schon die dort integrierten Suchmaschinen nehmen einen Teil der Rechercheskills ab, die man sich früher durch Grundkurse und stetige Übung erst angeeignet hat. Sollte das früher ein Bestandteil von wissenschaftlicher Leistungserbringung gewesen sein (was man auch bestreiten könnte), hätte sich dieser stillschweigend verabschiedet, wäre durch die technische Entwicklung überholt worden, ohne dass dies durch eine Prüfungsreform normativ abgesegnet worden wäre. So lautet das Zwischenergebnis: Die Nutzung generativer KI zu Zwecken der Recherche, zur weiteren Erschließung von relevantem Wissen oder ggf. auch zur Hilfestellung bei Konzeption, Gliederung oder ersten Entwürfen kann und sollte kaum verboten werden. Aus Gründen der Rechtssicherheit kann man dies dann aber auch ausdrück-

lich zulassen – nicht zuletzt, um Chancengleichheit zwischen jenen, die ohne eine solche Zulassung bei der Nutzung zögern, und jenen, die insofern keine Scheu haben, herzustellen. Wenn dem wiederum so ist, wird man natürlich all das, was mehr oder weniger von der Software erledigt wird oder werden kann, später nicht in die Bewertung einfließen lassen. Es wird als selbstverständlich angesehen.¹⁸

- Entscheidend ist bei einer schriftlichen Prüfungsaufgabe, die ohne Aufsicht über einen gewissen Zeitraum in eigener Verantwortung der zu Prüfenden erbracht wird, die Erstellung des Textes, der letztlich von den Prüfern bewertet werden soll. Hier besteht der eigentliche Regulierungsbedarf. Ob und inwieweit man die Nutzung generativer KI auch hier zulassen oder ihre Zulassung eher einschränken möchte, hängt stark davon ab, was als Prüfungsleistung gelten soll. Hier zeichnet sich ein Paradigmenwechsel ab. Während man früher davon ausgegangen ist, dass das „Schreiben“ (Formulieren) einer Arbeit (Referat, Bachelorarbeit etc.) durch die Studierenden höchstpersönlich erfolgt – was dann auch Gegenstand der „Eigenständigkeitserklärung“¹⁹ war –, muss heute erwogen werden, ob man nicht Entwürfe durch generative KI zulassen will und den Schwerpunkt der Bewertung nunmehr auf die inhaltliche Qualität legen möchte, die Prüfungsleistung damit schwerpunktmäßig in der Qualitätsprüfung sowie den individuellen Ergänzungen läge.²⁰

2. Übersicht der Anwendungsszenarien I – Perspektive der Studierenden

Für die weitere Diskussion kann es nützlich sein, sich die wesentlichen Anwendungsszenarien vor Augen zu füh-

¹⁸ An dieser Stelle könnte die Frage aufgeworfen werden, ob es mit dem Grundsatz der Chancengerechtigkeit vereinbar ist, wenn man die Nutzung generativer KI zumindest im Vorfeld der Prüfungsleistung zulässt (was bekanntlich manches vereinfacht), obwohl vielleicht nicht jeder Zugang zu der entsprechenden Software hat. Dem lässt sich entgegen, dass es eine absolute Gleichheit zwischen den Studierenden nicht geben kann. Vielmehr ist durch entsprechende Verfahrensvorgaben und die jeweils angelegten Bewertungskriterien dafür Sorge zu tragen, dass alle über möglichst vergleichbare Prüfungsbedingungen verfügen. Nicht auszuschließen und auch hinzunehmen ist dabei, dass bestimmte Studierende durch ihre individuellen Umstände z.B. über einen besseren oder schnelleren Computer verfügen, sich Literatur kaufen können, statt diese (möglicherweise mit Wartezeit) aus der Bibliothek ausleihen zu müssen oder sich schlicht bestimmte Fähigkeiten technisch angeeignet haben, die ihnen gegenüber anderen einen Effizienzvorteil bringen. Inzwischen gibt es viele generative KI-Systeme, auf die Studierende kostenlos oder gegen die Zahlung eines kleinen einmaligen oder monatlichen Betrags zugreifen können. Ein ausdrücklicher Hinweis auf die gestattete Nutzung kann daher vielmehr zu einer Angleichung der tatsächlichen Prüfungsbedingungen führen. Weiterführend

zum Grundsatz der Chancengerechtigkeit im Hochschul- und Prüfungskontext *Heckmann/Rachut*, E-Klausur und Elektronische Fernprüfung, 2023, S. 89 ff., 166 ff.

¹⁹ Beispiel: „Hiermit erkläre ich an Eides statt, dass ich die vorliegende schriftliche Hausarbeit XXX an der Ludwig-Maximilians-Universität München im WS 2023/24 selbstständig verfasst und keine anderen als die von mir im Literaturverzeichnis angegebenen Werke / Hilfsmittel benutzt habe. Die Stellen der Arbeit, die anderen Werken dem Wortlaut oder dem Sinne nach entnommen sind, wurden in jedem Fall unter Angabe der Quellen (einschließlich des World Wide Web und anderer elektronischer Text- und Datensammlungen) kenntlich gemacht. Dies gilt auch für beigegebene Zeichnungen, bildliche Darstellungen, Skizzen und dergleichen sowie muttersprachliches Gegenlesen der ohne fremde Hilfe verfassten Arbeit.“

²⁰ Aktuell hängt die Qualität des KI-generierten Textes noch sehr stark von der Qualität der Prompts ab, sodass auch darin ein gewisse, durchaus anspruchsvolle Teilleistung gesehen werden kann. Indes ist davon auszugehen, dass sich die Systeme stetig verbessern und in naher Zukunft auch schwächere menschliche Eingaben zu qualitativ hochwertigen Ergebnissen führen können.

ren, damit diese Diskussion nicht pauschal und damit verzerrt geführt wird. Dabei gilt es nach den verschiedenen Perspektiven (Studierende, Hochschule, Lehrende) zu differenzieren.

a) Vorbemerkung: Unterscheidung von beaufsichtigten Prüfungen (z.B. Klausuren) und unbeaufsichtigten schriftlichen Arbeiten (Bachelorarbeiten etc.)

- Bei unbeaufsichtigten schriftlichen Arbeiten sollte man generative KI bis zu einem bestimmten Nutzungsgrad (siehe die nachfolgenden Anwendungsszenarien) zulassen: Zum einen dürfte die Kontrolle eines Verbots rechtskonform nur schwer möglich sein. Zum anderen gibt es gute Gründe, die Nutzung generativer KI zum Gegenstand eines praxisorientierten Studiums mit Blick auf Bedürfnisse des Arbeitsmarktes zu machen. Ein solcher proaktiver Umgang mit einer innovativen Technologie stellt deren Chancen und nicht einseitig die Risiken in den Mittelpunkt.
- Für beaufsichtigte Prüfungen bietet es sich demgegenüber – zumindest als Regelfall – an, den Einsatz generativer KI zu untersagen. Zum einen lässt sich dies bei Aufsichtsprüfungen leichter kontrollieren. Zum anderen hätte man eine größere Bandbreite an Prüfungsformaten, weil bestimmte Lerninhalte und Methoden doch die höchstpersönliche Erbringung der Prüfungsleistung im gesamten Umfang erfordern.

Bei den unbeaufsichtigten schriftlichen Arbeiten muss wiederum unterschieden werden, ob das „technische Hilfsmittel“ im Vorfeld der eigentlichen Prüfungsleistung steht oder zu dessen Bestandteil wird.

b) Digitale Anwendungen für „unterstütztes Lernen“

Sowohl rechtlich als auch praktisch problemlos ist die Nutzung von digitalen Anwendungen für „unterstütztes Lernen“. Damit sind sämtliche Software-Anwendungen gemeint, die von den Studierenden eingesetzt werden, um den Lernprozess zu unterstützen: Das reicht von eBooks oder Podcasts über fachbezogene Datenbanken und spezifischer Lernsoftware bis zu generativer KI wie ChatGPT, die man als digitalen Assistenten zu Lernzwecken einsetzen kann. Dies gilt nicht nur außerhalb der

Prüfungen, sondern auch, wenn man die Anwendungen nutzt, um sich den (möglicherweise bislang nicht oder nur unvollkommen erschlossenen) Prüfungsstoff zu erschließen. Auch wenn man sich bei der Nutzung bereits in der Prüfungsphase befindet, findet die konkrete Nutzung im Sinne eines „unterstützenden Lernens“ noch im Vorfeld der eigentlichen Prüfungsleistung statt.

c) Generative KI als Recherche-Tool zu Prüfungsaufgaben

Einen Schritt weiter geht man, wenn generative KI konkret bezogen auf eine bestimmte Prüfungsaufgabe eingesetzt wird, insbesondere zu Recherchen mit Blick auf die konkrete Fragestellung. Damit befindet man sich in einer Phase, die nach herkömmlichem Verständnis bereits zur Leistung gezählt wird, die in der Prüfung zu erbringen ist (gerade bei wissenschaftlichen Hausarbeiten, Bachelor- oder Masterarbeiten). Nach dem vorstehend skizzierten Paradigmenwechsel könnte man diese Recherchearbeit allerdings auch aus der Beschreibung der bewerteten Prüfungsleistung herausnehmen und sich der modernen Möglichkeiten innovativer Anwendungen der generativen KI bedienen. Die eigentliche Prüfungsleistung würde dann erst daran anschließen.²¹

d) Generative KI zur formalen Verbesserung eigenständig erstellter Texte

Auf einer gedachten „Eigenständigkeits-Skala“ der Mensch-KI-Interaktion lässt sich die nächsthöhere Stufe dort einordnen, wo die KI bereits am Prüfungstext selbst eingesetzt wird, dies aber nur zur formalen Verbesserung des ansonsten eigenständig erstellten Textes. Ein solcher Einsatz generativer KI scheint schon deshalb vertretbar, weil er gar nicht weit weg von dem herkömmlichen Vorgehen einzustufen ist, wonach die eigene Hausarbeit durch andere Personen redigiert oder in ein „Schreibbüro“ gegeben wird, wo sie auf Rechtschreibung, Grammatik und Stil überprüft wird.²²

e) Generative KI als Instrument zu Inspiration und Vorentwurf einer Aufgabenlösung

In eine „Grauzone“ kommt man dann, wenn generative KI als Instrument zu Inspiration und Vorentwurf einer Aufgabenlösung eingesetzt wird: Einerseits bekommt

²¹ Birnbaum, NVwZ 2023, 1127, 1128 lässt eine andere Ansicht erkennen und will bereits das Aufbauen auf einer Gliederung oder einem Entwurf als Täuschung qualifizieren: „Die Täuschung im Rechtssinne ist eher technischer Natur: Der als Prüfungsleistung hereingereichte Text darf weder unmittelbar noch mittelbar eine „textliche Übernahme“ der KI-Ausarbeitung sein.“ Damit wird letztlich ein strengerer Maßstab angelegt, als dies bislang bei Hausarbeiten der Fall war.

²² A.A. wohl Birnbaum, NVwZ 2023, 1127, 1128, der eine relevante Täuschungshandlung im Rahmen von nicht beaufsichtigten Prüfungen bereits dann bejaht, wenn nicht mehr ausschließlich die eigene Intelligenz des Prüflings bei der Prüfungsbearbeitung zum Einsatz kommt, sodass jeglicher Einsatz einer „künstlichen“ sowie „nicht-künstlichen Intelligenz“ eine Täuschungshandlung darstellen soll.

die KI nun eine „Rolle“, die bereits sehr nahe an der regelmäßigen Prüfungsleistung einzuordnen ist. Andererseits wird es nun schwer, die „Anteile“ von Mensch und KI hinreichend abzugrenzen. Wenn man dies zulassen wollte, könnte man auch gleich auf die nächste Stufe gehen.

f) Generative KI als Co-Erstellerin einer Prüfungsleistung

Der „mutigste“, vielleicht aber auch der rechtlich am besten umsetzbare und prüfungsdidaktisch innovativste Schritt wäre es, generative KI als Co-Erstellerin einer Prüfungsleistung zuzulassen – auch um keinen Zweifel daran zu lassen, dass damit ein neues Leitbild der „Eigständigkeit“ der (unbeaufsichtigten, schriftlichen) Prüfungsleistung geschaffen wird, das der innovativen Entwicklung Rechnung trägt und den gewandelten Praxisanforderungen genügen will: Wenn nämlich künftig in der Berufspraxis ohnehin generative KI eingesetzt wird, sollten Hochschulen dies weder verbieten noch ignorieren, sondern proaktiv einsetzen.

Eine Abgrenzung der eigentlichen Prüfungsleistung wäre ebenso möglich: Diese bestünde in der wissenschaftlichen Überarbeitung des KI-Entwurfs, besonders mit Blick auf Quellen, die die KI nicht berücksichtigt hat, aber auch um Fehler („Halluzinationen“) zu verbessern und jenen wissenschaftlichen Ansprüchen zu genügen, die mit der Aufgabenstellung explizit genannt und gefordert werden. Je größer der errechnete/geschätzte „Workload“ der KI ist, um so mehr kann von den Studierenden an Qualitätsmanagement und Optimierung verlangt werden. Eingereichte Arbeiten, die genau dies nicht erkennen lassen, werden den Anforderungen nicht genügen.

3. Übersicht der Anwendungsszenarien II – Perspektive der Hochschule

a) Einsatz von KI-Detektoren zur Täuschungskontrolle und Aufdeckung des KI-Einsatzes durch Studierende

Lässt eine Hochschule bzw. lassen Prüfende den KI-Einsatz nicht wie im vorgestellten Sinne zu, so stellt sich die Frage, wie ein unerlaubter KI-Einsatz der Studierenden aufgedeckt und prüfungsrechtlich nachgewiesen werden kann. Hierbei liegt zunächst der Einsatz von KI-Detek-

toren zur Täuschungskontrolle nahe. Solch einem „KI-Einsatz zur Aufdeckung eines KI-Einsatzes“ stehen indes gravierende rechtsstaatliche Bedenken entgegen. Sog. KI-Detektoren werben u.a. damit, dass sie die Erstellung eines Textes durch KI-Systeme mit einer gewissen Wahrscheinlichkeit beziffern und demnach als Indiz für eine KI-gestützte Bearbeitung herangezogen werden können. Dabei arbeiten diese Programme wiederum anhand intransparenter, mit menschlichen Sinnen nicht nachvollziehbaren Parametern. Stützt sich eine Hochschule bei der Annahme eines Täuschungsversuchs auf das ausgegebene Ergebnis, so genügt dies nicht den aus Art. 20 Abs. 3 GG folgenden rechtsstaatlichen Grundsätzen an Transparenz, Nachvollziehbarkeit und Vorhersehbarkeit staatlicher Entscheidungen. Die KI-Software zur Aufdeckung des vorherigen KI-Einsatzes stellt wiederum eine Blackbox dar, deren Ergebnis nicht ohne weiteres darzulegen, nachzuvollziehen und damit überprüfbar ist. Weiter muss beachtet werden, dass eine solch grundrechtsinvasive Maßnahme der Hochschule nach der Wesentlichkeitstheorie des BVerfG²³ einer entsprechenden Rechtsgrundlage bedarf, die die wesentlichen Vorgaben für den Einsatz einer solchen Software regelt.²⁴ Aus datenschutzrechtlicher Sicht ist im Kontext des Einsatzes von KI-Detektoren zudem Art. 22 DSGVO zu beachten. Nach der jüngsten EuGH-Rechtsprechung wurde das Vorliegen einer „automatisierten Entscheidung im Einzelfall“ i.S.v. Art. 22 Abs. 1 DSGVO für den SCHUFA-Score bejaht.²⁵ Begründet wurde dies damit, dass der auf personenbezogene Daten einer Person gestützter Wahrscheinlichkeitswert maßgeblich für die Entscheidung Dritter sei. Scoring ist auch nach dieser Rechtsprechung weiterhin zulässig, setzt jedoch voraus, dass Betroffene über die involvierte Logik des Algorithmus informiert werden müssen.²⁶ Dies muss auch für den Fall der KI-Detektoren gelten, wenn diese die prüfungsrechtliche Entscheidung bezüglich eines unzulässigen KI-Einsatzes und die daran knüpfende Folge des Nichtbestehens maßgeblich beeinflussen. Bestreitet man hingegen, dass ein solcher Score Einfluss auf die Prüfungsentscheidung hätte, so muss von dem Einsatz bereits aus Gründen der Verhältnismäßigkeit von vornherein abgesehen werden.

Ungeachtet dessen ist zu fragen, wie ein unerlaubter KI-Einsatz grundsätzlich aufgedeckt und nachgewiesen werden kann. Im prüfungsrechtlichen Verfahren gilt der Grundsatz, dass eine Täuschungshandlung durch die

²³ S. bereits z. I. b.; im Detail zur Bedeutung der Wesentlichkeitstheorie im Kontext elektronischer (Fern-)Prüfungen s. *Heckmann/Rachut*, E-Klausur und Elektronische Fernprüfung, 2023, S. 145 ff.

²⁴ Vgl. hierzu v.a. im Hinblick auf den Einsatz von sog. Plagiatssoft-

ware, *Nehlsen/Fleck*, DSRITB 2023, 15, 20 ff.

²⁵ EuGH, Urt. 7.12.2023 – C-634/21 – SCHUFA Holding u.a. (Scoring).

²⁶ *Radtke*, MMR 2024, 153, 157.

Hochschulen nachzuweisen ist.²⁷ Dies umfasst sowohl die objektive Täuschungshandlung als auch den subjektiven Willen, sich einen unzulässigen Vorteil zu verschaffen, eine fahrlässige Täuschungshandlung kommt daher ebenfalls nicht in Betracht.²⁸ Reichen die Beweismittel für die Feststellung bzw. die hinreichende Gewissheit einer Täuschung nicht aus, so muss die Leistung nach den üblichen Maßstäben bewertet werden.²⁹ Bei unbeaufsichtigten Prüfungsarbeiten kann die Täuschungshandlung indes nicht durch etwaiges Aufsichtspersonal beobachtet werden, sondern muss im Nachhinein festgestellt werden. Hierfür kommen auch im Prüfungsrecht die Grundsätze des Beweises des ersten Anscheins zur Anwendung. Demnach kann aufgrund allgemeiner Erfahrungswerte ein Rückschluss auf eine fragliche Tatsache (hier: Täuschungshandlung) gezogen werden, wenn es sich um einen typischen Sachverhalt handelt. Der Prüfling wiederum kann dies entkräften, indem er die ernsthafte Möglichkeit eines atypischen Geschehensablaufs darlegt.³⁰ Bereits auf der ersten Stufe – dem typischen Sachverhalt – stellt sich das Problem, dass es schlicht an vergleichbaren „typischen“ Geschehensabläufen und damit auch entsprechenden Erfahrungswerten mangelt. Während bisher insbesondere aufgrund identischer Abgaben – inklusive der (Tipp-) Fehler – auf eine unerlaubte Zusammenarbeit geschlossen werden konnte, ändern sich die Umstände durch die elektronische Bearbeitung, neue Prüfungsmodi und nunmehr den möglichen KI-Einsatz.³¹ Ein KI-System wird hierbei bei identischer Eingabe unterschiedliche Ausgaben hervorbringen, sodass ein KI-generiertes Ergebnis nicht direkt reproduzierbar ist. Etwaige weitere denkbare Parameter, wie ein bestimmter Schreibstil, Fehler oder ebenso die Fehlerlosigkeit eines Textes sind

ebenfalls nicht geeignet, hinreichende Indizien für einen KI-Einsatz darzustellen.³² Gerade das Abstellen auf eine besonders gute Bearbeitung würde den Prüfungsgedanken ad absurdum führen: Bestraft würden hierbei gerade besonders gute Studierende. Auch ein Vergleich mit den bisherigen Leistungen eines Studierenden kann, bei einer nunmehr deutlich besseren Leistung, unter dem Grundsatz der Chancengleichheit nicht als Indiz für eine Täuschungshandlung herhalten.³³ Bei der Heranziehung von durch Software ermittelter Wahrscheinlichkeitswerte ergibt sich wiederum das Problem der Intransparenz und damit ein rechtswidriger Ausschluss der Berücksichtigung atypischer Geschehensabläufe. Der Umstand, dass die bisherigen Mechanismen zur nachträglichen Täuschungsaufdeckung bei KI-generierten Texten an ihre Grenzen stoßen, mag im Einzelfall zu unbefriedigenden Ergebnissen führen, darf jedoch nicht zu einem Absenken der Anforderungen an den Anscheinsbeweis führen. Eine solche Verschiebung der Maßstäbe tastet die geltenden Beweislastregeln an und würde letztlich zum Nachteil aller Prüfungsteilnehmenden führen, die im Zweifel ihre (gute) Leistung zu begründen hätten. Gelingt es ihnen nicht, einen atypischen Verlauf darzulegen, würde eine Leistungssteigerung dann zu ihrem Nachteil gereichen.³⁴ Dies erscheint insbesondere vor dem Hintergrund, dass die Hochschulen die Prüfungen gestalten – und damit über das Verfahren, die konkrete Prüfungsaufgabe und die zugelassenen Hilfsmittel bestimmen – ungerecht. Denn sie haben es mit ihrem Gestaltungsspielraum in der Hand, faire Prüfungsbedingungen zu schaffen und damit Aufgaben zu stellen, die sich im Zweifel nicht ohne Weiteres durch KI lösen lassen oder die Bewertungsmaßstäbe an eine erlaubte Co-Kreation anzupassen.

²⁷ Jeremias, in: Fischer/Jeremias/Dieterich, Prüfungsrecht, 8. Aufl. 2022, Rn. 236, 869.

²⁸ Jeremias, in: Fischer/Jeremias/Dieterich, Prüfungsrecht, 8. Aufl. 2022, Rn. 237.

²⁹ Jeremias, in: Fischer/Jeremias/Dieterich, Prüfungsrecht, 8. Aufl. 2022, Rn. 236.

³⁰ BVerwG, Beschluss vom 23.1.2018 – 6 B 67/17 = NJW 2018, 1896; Jeremias, in: Fischer/Jeremias/Dieterich, Prüfungsrecht, 8. Aufl. 2022, Rn. 237, 870 m.w.N.

³¹ Vgl. zur Übertragbarkeit der Grundsätze des ersten Anscheins auf geänderte Prüfungsformate Rachut, jurisPR-ITR 19/2023 Anm. 4; Morgenroth, OdW 2022, 273, 274.

³² Vgl. hierzu auch Tegethoff, jurisPR-BVerwG 12/2018 Anm. 3, wonach bei dem Vergleich mit einer Musterlösung nicht allein die inhaltliche bzw. qualitative Ähnlichkeit ausreichend ist, sondern sich darüber hinaus Übereinstimmungen in einzelnen Formulierungen sowie dem Aufbau und der Gedankenführung notwendig sind; A.A. VG München, Beschl. v. 28.11.2023 – M 3 E 23.4371, wobei das Gericht hierbei auf die durch die Bewertenden dargelegte besondere Strukturiertheit, Prägnanz, Inhaltsdichte, das Fehlen von gedanklichen Brüchen sowie eine fehlerfreie Orthographie

und Rechtsschreibung abstellt.

³³ Vgl. hierzu Jeremias, in: Fischer/Jeremias/Dieterich, Prüfungsrecht, 8. Aufl. 2022, Rn. 237, wonach bisher ausschließlich auf Prüfungsarbeiten anderer Prüfungsteilnehmenden sowie Übereinstimmung mit der Musterlösung oder ähnlichen Dokumenten abgestellt wird; ebenso konnte ein Student die Neubewertung seiner Prüfungsleistung nicht aufgrund der großen Diskrepanz der Bewertung zu seinen bisher (sehr) guten Prüfungsleistungen fordern, VG Ansbach, Urteil vom 14. April 2016 – AN 2 K 15.02220 – juris Rn. 33; nicht überzeugend hingegen VG München, Beschl. v. 28.11.2023 – M 3 E 23.4371 Rn. 40, das in seiner Entscheidung auf eine vergleichbare Essayaufgabe aus dem vorherigen Semester desselben Kandidaten abstellt und damit die Leistungssteigerung als rechtfertigungsbedürftig einstuft.

³⁴ Da die bloße Möglichkeit eines alternativen Geschehens nicht reicht, vgl. Jeremias, in: Fischer/Jeremias/Dieterich, Prüfungsrecht, 8. Aufl. 2022, Rn. 871, – mithin auch hier durchaus hohe Anforderungen an die Darlegung des atypischen Verlaufs gestellt werden – müssen diese Anforderungen erst recht bereits auf der ersten Stufe des Anscheinsbeweises durch die Hochschule erfüllt werden.

Es mag nicht auszuschließen sein, dass in ganz bestimmten Prüfungssituationen aufgrund der konkreten Aufgabenstellung und den Umständen des Einzelfalls dennoch aus den äußeren Faktoren einer Prüfungsleistung Rückschlüsse auf einen KI-Einsatz möglich sind. Diese Einzelfälle können jedoch nicht darüber hinwegtäuschen, dass – insbesondere ohne prüfungsdidaktische Anpassungen – der Einsatz von KI durch die Studierenden in der Regel nicht nachvollzogen und belegt werden kann. Ein etwaiges Verbot kann daher nicht durchgesetzt werden und begegnet schon deshalb rechtsstaatlichen Zweifeln.

Davon unbenommen bleibt die Möglichkeit, mittels (KI-)Software Verstöße gegen die wissenschaftliche Praxis, wie bspw. Plagiate, aufzudecken, da diese nachvollzogen werden können.³⁵

b) Exkurs: Einsatz von KI zur Klausuraufsicht

Nur kurz erwähnt sei die weitere Möglichkeit, KI einzusetzen, um bei beaufsichtigten Prüfungen Täuschungsversuche aufzudecken, insbesondere mit bestimmten Proctoring-Funktionen. Dies ist rechtlich sehr problematisch. Einerseits geht ein solcher Einsatz stets mit tiefgreifenden Grundrechtseingriffen einher, sodass hierfür zunächst eine entsprechende Rechtsgrundlage geschaffen werden müsste.³⁶ Andererseits stellt sich hierbei ebenfalls die Problematik der fehlenden Nachvollziehbarkeit solcher Systeme, wenn sie bspw. einen „Täuschungsscore“ ausgeben, sodass prüfungsrechtliche Entscheidungen hierauf nicht gestützt werden können. Insgesamt sind die rechtskonformen Einsatzszenarien daher sehr beschränkt.³⁷

c) Exkurs: Einsatz von KI zum Lehr-/Lern-Monitoring

Außerhalb des Prüfungsgeschehens kommt KI zuweilen auch dort zum Einsatz, wo individualisierte Lehrangebote auf den Einzelnen zugeschnitten werden sollen. Weil dies nicht ohne spezifische Profildaten umgesetzt werden kann, sind an dieser Stelle strenge datenschutzrechtliche Anforderungen einzuhalten und die Datenverarbeitung auf eine entsprechende Rechtsgrundlage zu stützen, wobei fraglich ist, ob dies im Kontext von staatlicher Hochschule und Studierenden allein durch eine Einwilligung zu erreichen wäre. Darüber hinaus sind insbeson-

dere die diesbezüglich in der KI-VO vorgesehenen Vorgaben für Hochrisikosysteme relevant.

d) Exkurs: Einsatz generativer KI in der Hochschulverwaltung

Ganz allgemein kann generative KI auch in der Hochschulverwaltung eingesetzt werden, wie dies derzeit auch in der allgemeinen Verwaltung³⁸, aber ebenso bei Krankenkassen oder Unternehmen diskutiert wird. Auch hierbei gelten die allgemeinen datenschutzrechtlichen Regelungen, sodass es unter anderem zu hinterfragen gilt, wie die allgemeinen Datenschutzgrundsätze des Art. 5 DSGVO (z.B. Zweckbindung und Datensparsamkeit) in diesem Zusammenhang umgesetzt werden können.

4. Übersicht der Anwendungsszenarien III – Perspektive der Lehrenden

a) Einsatz von generativer KI zur Unterstützung der Lehre

Generative KI kann nicht nur durch Studierende eingesetzt werden. Auch für Lehrende ergeben sich zahlreiche Vorteile. So können sie generative KI wie andere digitale Unterstützungssysteme zur Vorbereitung ihrer Lehre einsetzen, Textbausteine für die eigene Kommunikation entwerfen oder Checklisten anlegen lassen. Im Sinne eines digitalen Assistenzsystems sind die Einsatzmöglichkeiten auch im Bereich der Lehre vielfältig.

b) Einsatz von generativer KI zur Erstellung von Prüfungsaufgaben

Dies gilt ebenso für das Prüfungsgeschehen selbst. So ist denkbar, dass Lehrende entsprechende Anwendungen einsetzen, um Prüfungsaufgaben erstellen zu lassen. Von den Prüfenden erstellte Prüfungsaufgaben unterliegen grundsätzlich dem Urheberrechtsschutz. Tritt nun eine generative KI hinzu, so kann es an der notwendigen persönlichen geistigen Schöpfung i.S.d. § 2 Abs. 2 UrhG fehlen. Ähnlich der Eigenständigkeit der Prüfungsleistung müsste abgewogen werden, ob es sich hier noch – z.B. durch Anpassungen der Vorschläge – um eine eigene schöpferische Leistung des Prüfenden handelt oder eben nicht mehr. Die urheberrechtlichen Fragen von generativer KI sind aktuell noch Bestandteil des rechtswissenschaftlichen Diskurses³⁹ und können daher nicht

³⁵ S. hierzu *Nehlsen/Fleck*, DSRITB 2023, 15.

³⁶ Vgl. zum Bedürfnis einer Rechtsgrundlage *Heckmann/Rachut*, E-Klausur und Elektronische Fernprüfung, 2023, S. 144 ff.

³⁷ Vgl. hierzu *Rachut/Besner*, MMR 2021, 851 ff.; EuGH, Urteil v. 7.12.2023 (C-634-21, „Schufa“) CR 2024, 29.

³⁸ Zum KI-Einsatz in der öffentlichen Verwaltung vgl. den Überblick bei *Heckmann/Rachut*, Digitale Verwaltung, in: Heckmann/Paschke, juris Praxiskommentar Internetrecht, 8. Aufl. 2024, Kap. 5 Rn. 1166 ff.

³⁹ Vgl. *Schack*, NJW 2024, 113 ff.; *de la Durantaye*, ZUM 2023, 645 ff.; *Hoeren*, MMR 2023, 81 ff.

abschließend beantwortet werden. Folgen eines fehlenden urheberrechtlichen Schutzes wirken sich indes nicht auf das Prüfungsgeschehen an sich aus, sondern können nachgelagert, z.B. bei der Frage, ob Studierende Altprüfungen vervielfältigen dürfen, relevant werden. Auch wenn generative KI bei der Aufgabenstellung unterstützend eingesetzt wird, verbleibt es bei der Verantwortung des Prüfenden für eine den prüfungsrechtlichen Vorgaben entsprechende Aufgabenstellung und anschließende Bewertung.

c) Einsatz von generativer KI als Bestandteil innovativer Lehr- und Prüfungsformate

Schließlich kann eine Software wie ChatGPT auch selbst Bestandteil innovativer Prüfungsformate werden, womit sich der Kreis zur Perspektive der Studierenden schließt. Sie kann dabei gezielt in der Lehre eingesetzt werden, um den Umgang mit ihr zu schulen aber auch Schwächen und Risiken aufzuzeigen. Das sog. Prompten hat maßgeblichen Einfluss auf die durch ChatGPT ausgegebenen Ergebnisse und stellt damit eine Fähigkeit dar, die für eine effiziente Bewältigung, der sich im Berufsleben stellenden Aufgaben durchaus förderlich ist bzw. mitunter sogar erwartet wird. Auch das Prompten kann daher im Sinne einer innovativen und praxisorientierten Ausbildung Teil von Lehre und sogar Prüfungen sein. Denkbar ist etwa, dass Studierende bei der Erstellung einer Haus- oder Seminararbeit gezielt zum Einsatz einer generativen KI angehalten werden, dabei ihre Prompts und anschließenden Bearbeitungsschleifen dokumentieren müssen.

IV. Leitgedanken zum rechtssicheren Einsatz generativer KI im Prüfungskontext

1. Disclaimer („work in progress“)

Der vorliegende Beitrag dient zur ersten Orientierung für einen rechtssicheren Einsatz generativer KI im Prüfungskontext. Diese Thematik wirft zahlreiche, zum Teil ganz neue Rechtsfragen auf, die einer vertieften Behandlung bedürfen. Auch wenn man auf der Grundlage der hier formulierten Erkenntnisse und Empfehlungen Anpassungen in den jeweiligen Prüfungsordnungen vornimmt, gilt es doch, die technische Entwicklung und die weitere Klärung offener Rechtsfragen im Auge zu behalten und agil zu reagieren.

2. Regulierungsbedarf

Aktuell besteht eine erhebliche Rechtsunsicherheit im Hinblick auf den Umgang mit technischen Innovationen in Studium, Lehre und Prüfung. Es erscheint wenig hilf-

reich, sich von Seiten der Hochschule oder der Fakultäten in dieser Situation passiv abwartend zu verhalten: Zum einen entstünde so ein Flickenteppich an Einzellösungen nach Einschätzung der Lehrenden; zum anderen wären auch die Studierenden alleine gelassen in ihrer Entscheidung, bis zu welchem Grad sie generative KI bei der Anfertigung von Hausarbeiten, Seminarreferaten, Bachelor- oder Masterarbeiten nutzen dürfen oder sollen. Dies wirkt sich auch auf die Chancengleichheit aus, weil die studentischen Leistungen in solchen Prüfungen immer weniger vergleichbar werden.

Insbesondere genügt es nicht, ein pauschales Verbot der Nutzung generativer KI aus einer allgemeinen Studien- und Prüfungsordnung herzuleiten, wonach die Nutzung unzulässiger Hilfsmittel als Täuschungsversuch zu werten sei („es ist jedes Hilfsmittel verboten, das nicht ausdrücklich zugelassen ist“). Im Gegensatz zu Aufsichtsarbeiten gibt es bei häuslichen Arbeiten keine Positivliste zugelassener Hilfsmittel, so dass man davon auch nicht abweichen kann. Das zeigt schon die unbestrittene Zulässigkeit der Nutzung von Suchmaschinen oder Diktier- und Rechtschreibprogrammen, die ebenfalls KI-basiert sind.

Dabei schlägt auch eine Berufung auf Grundsätze „guter wissenschaftlicher Praxis“ fehl. Angesichts der Tatsache, dass allgemein zugängliche Anwendungen generativer KI erst seit kurzem existieren, kann sich eine solche wissenschaftliche Praxis noch gar nicht herausgebildet haben. Man müsste also mit Analogien arbeiten, etwa dem Plagiatsverbot oder der verbotenen Nutzung eines „Ghostwriters“. Dies beträfe aber nur die vollständige Übernahme fremder Texte bzw. deren fehlende Kennzeichnung. Was aber ist in dem Fall, dass ChatGPT nur als Chatpartner eingesetzt wird, so wie man schon immer (juristische) Hausarbeiten mit Kommilitonen diskutiert hat, gemeinsam Lösungsskizzen erstellt hat etc.? Allein dies zeigt bereits, dass die innovativen Möglichkeiten des Einsatzes generativer KI eine Neujustierung der Anforderungen an die Eigenständigkeit einer Leistung fordern, die nur der demokratisch legitimierte Gesetzgeber leisten kann.

Auch aus verfassungsrechtlicher Sicht besteht Regulierungsbedarf: Die pauschale oder einzelfallbezogene Erlaubnis oder Duldung bzw. das mehr oder weniger weitreichende Verbot der Nutzung generativer KI bei schriftlichen Prüfungsarbeiten berühren die Grundrechte aus Art. 12 Abs. 1, 5 Abs. 3, 3 Abs. 1 sowie Art. 2 Abs. 1 i.V.m. Art. 1 Abs. 1 GG. Sowohl ein aktives Tätigwerden im Rahmen der Prüfungsgestaltung als auch ein schlichtes Unterlassen sind damit grundrechtswesentlich, so

dass es rechtlicher Regelungen bedarf. Hier wäre eine entsprechende Ermächtigungsnorm in dem jeweiligen Hochschulgesetz geboten, um Details durch eine Rechtsverordnung oder zumindest auf Satzungsebene der Prüfungsordnungen zu regeln. Insofern sollten die Hochschulen aber nicht auf den Landesgesetzgeber warten, sondern selbst regulierend tätig werden, um ein gewisses Maß an Rechtssicherheit und Orientierung für Studierende und Lehrende zu bieten. Die Hochschulen haben insoweit eine eigene Verantwortung, einen verfassungskonformen Zustand herzustellen.

3. Verhältnismäßiger Ausgleich von Lehr- und Lernfreiheit mit Chancengerechtigkeit

Wenn man vor dem Hintergrund des sowohl verfassungsrechtlich als auch hochschul- und bildungspolitisch begründeten Regulierungsbedarfs Regeln für den Einsatz generativer KI in Prüfungen aufstellen will, gibt es nie „die“ eine richtige Lösung. Vielmehr muss die Regelung am Ende eines gestaltenden Abwägungsprozesses stehen, der sowohl die Lehrfreiheit der Lehrenden als auch die Lernfreiheit der Studierenden in den Blick nimmt und eine Verbesserung der Chancengerechtigkeit innerhalb der Prüfungskohorten anstrebt.

Dabei können die folgenden Eckpunkte handlungsleitend sein:

- » Generativer KI wird von großen Teilen in der Wissenschaft eine prägende Rolle in der gegenwärtigen digitalen Transformation von Staat, Wirtschaft und Gesellschaft zugeschrieben. Sie ist kein bloßer Hype, sondern ein Disruptionsfaktor.
- » Generative KI hat erheblichen Einfluss auf zahlreiche Berufsbilder. Sie anzuwenden und bestenfalls zu beherrschen (auch unter Berücksichtigung von Qualitätsdefiziten, Herausforderungen im sachgerechten Umfang, rechtlichen und ethischen Implikationen), hat eine relevante curriculare Bedeutung.
- » Ein Verbot der Nutzung generativer KI im Vorfeld der Erstellung von Prüfungstexten bei Hausarbeiten, Seminarreferaten, Bachelor- oder Masterarbeiten ist rechtskonform nicht überprüfbar und schon deshalb rechtswidrig.
- » Ein Verbot der Nutzung generativer KI bei der Erstellung entsprechender Prüfungstexte ist weder durch sog. KI-Detektoren noch durch weitere objektive Parameter, die einen ersten Anschein einer Täuschungshandlung begründen können, möglich, damit nicht überprüfbar und schon deshalb rechtswidrig.

- » Die konkrete Gestaltung von Prüfungsformaten unterliegt einem Gestaltungsspielraum der Lehrenden, der durch fachdidaktische und fachspezifische Besonderheiten geprägt ist. Bei der Bemessung von Schwierigkeitsgrad, Workload und Lernzielen einer Prüfungsaufgabe ist zu berücksichtigen, dass sowohl ein Verbot als auch eine stillschweigende Duldung der Nutzung generativer KI dazu führen wird, dass ein Teil der Studierenden solche Systeme gleichwohl nutzt, ein anderer Teil aber nicht. Es bedarf deshalb sachlicher Überlegungen dahingehend, wie in diesem Fall Chancengerechtigkeit hergestellt wird.
- » Will man die Nutzung generativer KI bei häuslichen schriftlichen Prüfungsaufgaben grundsätzlich zulassen, bedarf es einer differenzierten Darstellung der zugelassenen (technischen) Hilfsmittel und einer transparenten Information über die Bewertungsmaßstäbe im Hinblick auf die verbleibende Eigenständigkeit der erbrachten Prüfungsleistung. Die Prüfungsziele müssen den neuen technischen Möglichkeiten angepasst werden.
- » Im Zeitalter einer selbstverständlichen Nutzung generativer KI verlagert sich die eigenständige menschliche Leistung im Textstellungsprozess auf die qualitätsorientierte Prüfung maschinell erstellter Textentwürfe und die Ergänzung von Textteilen (einschließlich transparenter Quellenangaben), an deren Erstellung die KI derzeit noch scheitert. Mensch und Maschine avancieren zu Co-Produzenten. Lehre und Prüfung berücksichtigen diesen Paradigmenwechsel.

V. Fazit

Generative KI ist gekommen, um zu bleiben. Auch und gerade im Kontext von Wissenschaft, Studium und Lehre. Übereilte Verbote durch einzelne Universitäten, wie der SciencePo in Paris oder der Universität Tübingen, haben nur noch anekdotischen Wert. Längst haben die Hochschulen die große Chance erkannt, die in der Integration dieser Technologien in den Wissenschaftsbetrieb steckt – auch deshalb, weil Hochschulen die Verbindung zur (Berufs-) Praxis nicht verlieren dürfen, ohne ihren Bildungsauftrag zu vernachlässigen. Es geht aber weit über diese praktischen Überlegungen hinaus: Hochschulen waren schon immer ein Ort kritischer Reflektion, Innovationen in dem Kontext einer Werteordnung zu betrachten. Das ist beim Thema Künstliche Intelligenz dringender denn je. Denn den enormen Chancen

stehen auch Risiken gegenüber. Diese richtig einordnen zu können, bedarf es einer fundierten Beschäftigung und Auseinandersetzung in Ansehung konkreter Anwendungen und ihrer Folgen. Das gelingt nur, wenn man über und mit (generativer) KI forscht und sich die Einordnung des unmittelbar Erlebten zunutze macht. Die im vorliegenden Beitrag angestellten rechtlichen Überlegungen dienen dazu, dieser unverzichtbaren Befassung einen Ordnungsrahmen zu geben. Ordnung der Wissenschaft dient in Zeiten von Künstlicher Intelligenz immer auch der Bewahrung von Souveränität des Menschlichen im Technischen.

Prof. Dr. Dirk Heckmann ist Inhaber des Lehrstuhls für Recht und Sicherheit der Digitalisierung an der Technischen Universität München. Nebenamtlich wirkt er als Direktor am Bayerischen Forschungsinstitut für Digitale Transformation (www.bidt.digital) und als Verfassungsrichter am Bayerischen Verfassungsgerichtshof.

Sarah Rachut ist wissenschaftliche Mitarbeiterin am Lehrstuhl für Recht und Sicherheit der Digitalisierung (Prof. Dr. Dirk Heckmann) an der Technischen Universität München und Geschäftsführerin der Forschungsstelle TUM Center for Digital Public Services. Sie forscht und lehrt zu verfassungsrechtlichen Fragen der Digitalisierung, schwerpunktmäßig in den Bereichen E-Government, E-Health und E-Education.

