

Workshop: KI-gestützte Erstellung von Prüfungsleistungen

Bastian Küppers¹, Svenja Bedenlier², Stefanie Gerl³, Matthias Bandtel⁴

Abstract: ChatGPT ist dieser Tage in aller Munde. Auf den ersten Blick ist ChatGPT eine Bedrohung für Prüfungsleistungen, speziell für asynchrone Prüfungsleistungen wie Hausarbeiten, da der anforderbare Output von ChatGPT sehr geeignet für die Lösung entsprechender Aufgabenstellungen erscheint. Doch ein näherer Blick zeigt: Chat GPT ist ein Sprachmodell und hat entsprechende Limitierungen. Der Workshop „KI-gestützte Erstellung von Prüfungsleistungen“ hat mit einem praxisorientierten Fokus Expertise aus der Praxis der aktuellen Prüfungslandschaft an verschiedenen Hochschulen gebündelt und Probleme und Chancen von generativer KI betrachtet.

Keywords: Künstliche Intelligenz, Elektronische Klausuren, Klausuren, Prüfungen

1 Einleitung

ChatGPT bzw. generative KI auf Basis von Sprachmodellen ist dieser Tage in aller Munde. Auf den ersten Blick stellen solche Technologien eine Bedrohung für Prüfungsleistungen, speziell für asynchrone Prüfungsleistungen wie Hausarbeiten dar, da der anforderbare Output sehr geeignet für die Lösung entsprechender Aufgabenstellungen erscheint. Doch ein näherer Blick zeigt: Generative KI kann nur auf den Informationen arbeiten, mit denen sie trainiert wurde, und hat entsprechende Limitierungen. Dementsprechend lohnt sich ein näherer Blick, um Risiken und Chancen sinnvoll bewerten zu können. Dabei hat der Workshop einen praxisorientierten Fokus und soll Expertise aus der Praxis und den aktuellen Prüfungslandschaften an verschiedenen Hochschulen bündeln.

Im Workshop wurden dazu drei Leitfragen jeweils in verschiedenen Arbeitsmodi bearbeitet:

¹ RWTH Aachen University IT Center / Learning Technologies Research Group, Seffenter Weg 23, D-52074 Aachen, kueppers@itc.rwth-aachen.de

² FAU Erlangen-Nürnberg, Institut für Lern-Innovation, Dr.-Mack-Straße 77, D-90762 Fürth, svenja.bedenlier@ili.fau.de

³ FAU Erlangen-Nürnberg, Institut für Lern-Innovation, Dr.-Mack-Straße 77, D-90762 Fürth, stefanie.gerl@ili.fau.de

⁴ KIT Karlsruhe, Karl-Friedrich-Straße 17, D-76133 Karlsruhe, matthias.bandtel@kit.edu

1. Wie sehr bedroht generative KI die aktuell etablierten Prüfungsformate an Hochschulen tatsächlich?
2. Wie können aktuell etablierte Prüfungsformate gegebenenfalls angepasst werden, damit generative KI nicht als Tool zur (teilweisen) Lösung von Prüfungen verwendet werden kann?
3. Welche Chancen birgt generative KI möglicherweise im Hinblick auf (elektronische) Prüfungen?

2 Beschreibung der Arbeitsmodi

2.1 Frage 1: Wie sehr bedroht generative KI die aktuell etablierten Prüfungsformate an Hochschulen tatsächlich?

Anhand einer leicht abgewandelten 6-Hüte-Methode⁵ sollte die tatsächliche „Bedrohungslage“ von Prüfungen im Hochschulbereich durch generative KI erörtert werden. Die Methode wurde dabei gewählt, um eine möglichst umfassende Betrachtung zu ermöglichen, wurde jedoch leicht abgewandelt, um der Fragestellung und dem Kontext gerecht zu werden. Die gewählten Hüte sollten jeweils die folgenden Aspekte berücksichtigen:

- Weißer Hut: Analytisches Denken
- Roter Hut: Emotionales Denken
- Schwarzer Hut: Kritisches Denken
- Gelber Hut: Optimistisches Denken

Beim analytischen Denken sollte dabei eine möglichst reelle Einschätzung der Bedrohungslage erarbeitet werden, indem sowohl Prüfungslandschaft als auch generative KI analysiert werden und darauf aufbauend die tatsächliche Bedrohungslage abgeleitet werden.

Beim emotionalen Denken sollten hauptsächlich Bedenken hinsichtlich der Bedrohungslage gesammelt werden. Dabei sollte zunächst nicht unterschieden werden, ob eine Bedrohung ein realistisches Szenario oder nur ein befürchtetes Szenario darstellt.

Beim kritischen Denken sollten realistische Bedrohungen erarbeitet werden, dabei sollte jedoch ein kritischer Grundgedanke im Vordergrund stehen, sodass auch mögliche, jedoch sehr unwahrscheinliche, Bedrohungen gesammelt werden. Im Gegensatz dazu sollten beim optimistischen Denken zwar ebenfalls realistische Bedrohungen erarbeitet werden,

⁵ De Bono E. (1985). Six thinking hats. ISBN: 0-14-013784-X

jedoch mit einem optimistischen Fokus, sodass die Menge der gesammelten Bedrohungen am Ende nur wahrscheinliche Szenarien enthält.

2.2 Frage 2: Wie können aktuell etablierte Prüfungsformate gegebenenfalls angepasst werden, damit generative KI nicht als Tool zur (teilweisen) Lösung von Prüfungen verwendet werden kann?

Mit dem methodischen Ansatz des „World Café“⁶ sollten Möglichkeiten erarbeitet werden, wie aktuell verwendete Prüfungsformate angepasst werden könnten, damit generative KI keine Bedrohung mehr für die Reliabilität der Prüfungsleistungen darstellt. Dabei gab es Tischen zu folgenden Themen:

- Präsentationen
- Portfolios
- Klausuren
- E-Klausuren
- Hausarbeiten

In drei direkt aufeinanderfolgenden Runden sollten die Teilnehmenden des Workshops selbstgewählt jeweils an einem anderen Tisch pro Runde die Entwicklungsmöglichkeiten der jeweils am Tisch repräsentierten Prüfungsform erörtern. Die Themenpaten an den Tischen haben ab der zweiten Runde jeweils kurz die Ergebnisse der vorherigen Runden vorgestellt, sodass die Runden thematisch aufeinander aufgebaut waren.

2.3 Frage 3: Welche Chancen birgt generative KI möglicherweise im Hinblick auf (elektronische) Prüfungen?

Die dritte Frage wurde in zwei Gruppen diskutiert, wobei eine Gruppe einen Fokus auf E-Prüfungen hatte und die andere Gruppe sich mit Papierprüfungen beschäftigt hat. Dazu muss erwähnt werden, dass sich generative KI bei Papierprüfungen nicht ohne Weiteres während der Prüfungsdurchführung einsetzen lässt, sodass der Fokus in dieser Gruppe auf der Vor- und Nachbereitung von Prüfungen lag, während der Fokus der ersten Gruppe deutlich mehr auf die Prüfungsdurchführung ausgerichtet war. Beiden Gruppen war während des Workshops freigestellt, den Arbeitsmodus selbst zu wählen.

⁶ Brown, J.J., & Isaacs, D. (2005). *The World Café: Shaping Our Futures Through Conversations That Matter*. ISBN: 978-1-57675-258-6

3 Ergebnisse und Diskussion

Die Ergebnisse der einzelnen Sessions wurden jeweils nach den Sessions im Plenum vorgestellt und kurz diskutiert. Die jeweils erzielten Ergebnisse werden in den folgenden Abschnitten dargestellt und kurz diskutiert.

3.1 Ergebnisse zu Frage 1

Die Gruppen mit den jeweiligen Hüten haben während der Sessions folgende Punkte erarbeitet:

Weißer Hut – Analytisches Denken

Die einzelnen Komponenten einer Prüfung müssen in der Bewertung anders gewichtet werden, d.h. der Fokus muss auf Punkte wie Argumentationen und Schlussfolgerungen gelegt werden. Zudem müssen sich die geprüften Kompetenzen verändern: Weg von der Reproduktion von Wissen hin zur Anwendung des Wissens.

Roter Hut – Emotionales Denken

Es wird eine Ungerechtigkeit in Prüfungen befürchtet, da generative KI nicht von allen Studierenden gleich gut genutzt werden kann, da auch dies bestimmte Kompetenzen erfordert. Hinsichtlich der Vor- und Nachbereitungen von Prüfungen wurden zudem Vor- und Nachteile erarbeitet. Es wird als positiv gesehen, dass generative KI-Routineaufgaben übernehmen könnte und der Umgang damit sogar Spaß machen kann bzw. eine gewisse Faszination mit sich bringt. Auf der anderen Seite wird befürchtet, dass manche Aufgaben und damit langfristig Arbeitsplätze ganz von einer KI übernommen werden könnten und so zudem die Wertschätzung für geleistete Arbeit abnimmt. Darüber hinaus wird auf die Möglichkeit technischer Probleme bzw. generell auf neue Herausforderungen im Arbeitsalltag bei der Verwendung generativer KI im Arbeitsalltag hingewiesen.

Schwarzer Hut – Kritisches Denken

Bei nicht in Präsenz abgelegten Prüfungen wird es schwierig die abgegebene Prüfungsleistung gerecht zu bewerten. Darüber hinaus sind aktuelle Prüfungsordnungen nicht auf Prüfungsszenarien in denen generative KI – erlaubt oder unerlaubt – eingesetzt wird ausgelegt. Daher wird rechtliche Unsicherheit bei der Durchführung von Prüfungen befürchtet.

Gelber Hut – Optimistisches Denken

Generative KI kann als unerlaubtes Hilfsmittel, während einer nicht in Präsenz abgelegten Prüfung verwendet werden, allerdings birgt diese Prüfungsmodus immer ein erhöhtes Risiko, dass die Prüfungsabgabe unter Zuhilfenahme unerlaubter Werkzeuge bzw. unerlaubter Hilfe anderer Menschen angefertigt worden ist. Daher wird die tatsächliche Bedrohung durch generative KI zwar als existent anerkannt, aber nicht als grundsätzlich

neuartiges Problem eingeordnet. Darüber hinaus werden auch sehr positive Aspekte generativer KI gesehen, da diese Technologie es beispielsweise ermöglicht Prüfungsaufgaben dynamisch generieren zu lassen und so mehrere gleichwertige Varianten einer Aufgabe in einer Prüfung einzusetzen. Somit wird einfaches Abschreiben bei Präsenzprüfungen deutlich erschwert, was die Reliabilität einer Prüfung am Ende sogar verstärkt. Zudem kann über automatisch generierte Aufgaben den Studierenden zur Vorbereitung auf eine Prüfung ein wirkungsvolles Hilfsmittel zum Lernen an die Hand gegeben werden. Dies gilt insbesondere, wenn die generierten Aufgaben auch mit einer Musterlösung versehen sind. Sofern zu einer bestimmten Aufgabe eine Musterlösung mittels generativer KI erzeugt werden kann, könnte dies darüber hinaus auch bei der Korrektur einer Prüfung helfen und so den Arbeitsaufwand für die Korrektur senken.

3.2 Ergebnisse zu Frage 2

In den drei Runden des World Cafés wurden während des Workshops folgende Punkte an den einzelnen Thementischen erarbeitet:

Präsentationen

Generative KI kann bei einer Präsentation nur in der Vorbereitung eingesetzt werden, beispielsweise indem eine Gliederung oder ganze Folien generiert werden. Bei der Präsentation selbst und einer anschließenden Diskussion sind die Prüflinge auf sich allein gestellt. Daher sollte der Fokus auf diesen beiden Elementen der Prüfung liegen und die Folien als solche nur sekundäres Kriterium zur Bewertung sein. Zudem wird die Nutzbarkeit von automatisch generierten Folieninhalten insgesamt in Frage gestellt, da die Folien immer auf die Zuhörenden und den Kontext der Präsentation (Seminar, Kolloquium, ...) abgestimmt sein. Zudem ist die Bewertung und Spezifika einer Präsentation sowie der Gestaltung zugehöriger Folien sehr von der Fachdisziplin abhängig, sodass eine generative KI diesen Kontext kennen müsste, um verwertbare Ergebnisse zu liefern. Zusätzlich zu den vorstehend genannten Punkten ist die Generierung bzw. Einbindung von nicht-textuellen Inhalten wie Bildern oder Tabellen zumindest im Moment noch nicht möglich.

Portfolio

Portfolios sind generell als kompetenzorientierte Prüfungsform anzusehen. Daher ist die reine Reproduktion von Informationen durch eine generative KI ohnehin nicht sonderlich hilfreich. Hinzu kommt, dass jedes Artefakt des Portfolios einzeln im Kontext des Portfolios abgewogen werden muss, um sinnvoll zum Gesamtwerk beizutragen. Diese kontextabhängige Abwägung und Auswahl von Artefakten kann durch eine generative KI nicht in dieser Form vorgenommen werden.

Klausur

Bei Präsenzklausuren wird der Einsatz von generativer KI als unerlaubtes Hilfsmittel als nahezu unmöglich angesehen. Anders sieht es bei Distanzklausuren aus, wo es stark von den genauen Umständen abhängt, ob generative KI unbemerkt eingesetzt werden kann. Zeitbegrenzung und notwendiges Spezialwissen aus der zugehörigen Lehrveranstaltung lassen aber Zweifel daran aufkommen, ob generative KI bei einer Klausur gewinnbringend als unerlaubtes Hilfsmittel eingesetzt werden kann. Anders sieht es aus, wenn generative KI als explizit erlaubtes Hilfsmittel eingesetzt werden soll, um die Kompetenz im Umgang mit diesem Werkzeug unter Beweis zu stellen. In diesem Fall muss der verwendete Prompt mit abgegeben werden, um diesen in die Bewertung einzubeziehen.

E-Klausuren

Bei elektronischen Klausuren hängt es sehr von der Organisation der Prüfung ab, ob der Zugang zu generativer KI gesperrt werden kann, oder nicht. In einem PC-Pool der Hochschule ist dies möglich, bei einer BYOD-Prüfung eher nicht. Daher sollten die Vorteile elektronischer Prüfungen ausgespielt werden, sodass generative KI den Aufgabenstellungen nicht mehr gerecht werden kann. Dies kann zum Beispiel durch multimediale Aufgaben geschehen, oder durch Bezug zu konkreten Fallbeispielen oder Projekten. Zudem sollte ein stärkerer Fokus auf den Lösungsweg, nicht nur auf die Lösung selbst gelegt werden.

Hausarbeiten

Da Hausarbeiten in der Regel ohne direkte Aufsicht und über einen längeren Zeitraum angefertigt werden, scheint die Möglichkeit zur unerlaubten Nutzung generativer KI hier besonders realistisch. Um dem zuvorzukommen sollte bei der Bewertung ein Fokus auf den Arbeitsprozess, nicht nur auf die finale Abgabe gelegt werden. Der Arbeitsprozess kann im besten Fall auch durch regelmäßige Treffen zwischen Lehrenden und Studierenden dokumentiert werden. Zudem kann die Hausarbeit zusätzlich zur Anfertigung der eigentlichen Arbeit mit einer Präsentation abgeschlossen werden, sodass potenziell ein Korrektiv zur Verfügung steht, wenn etwas mit der Hausarbeit selbst nicht stimmen sollte. Außerdem kann eine zusätzliche Fokussierung auf das Fazit sowie weiterführende Arbeiten die eigene Leistung der Studierenden deutlich hervorheben, da diese Punkte nicht ohne eigene Beschäftigung mit dem Thema ausformuliert werden können.

3.3 Ergebnisse zu Frage 3

In den beiden Themengruppe zu elektronischen Prüfungen und Papierprüfungen werden im Workshop die folgenden Ergebnisse erarbeitet:

Papierklausuren mit Fokus auf Prüfungsvor- und Nachbereitung

Generative KI macht es möglich Aufgaben dynamisch anzupassen, sowie den Kontext der Aufgaben zu personalisieren. Beispielsweise können Aufgaben zum gleichen Thema auf

verschiedenen Niveaustufen oder eingekleidete Aufgaben mit einem persönlichen Kontext erstellt werden. Zudem kann generative KI auch von Studierenden zur Vorbereitung auf eine Prüfung genutzt werden, beispielsweise indem Übungsaufgaben generiert werden oder Feedback zu Probeklausuren gegeben wird. Letztgenannter Mechanismus könnte ebenfalls bei der Korrektur der Prüfung genutzt werden, um Bewertungs- bzw. Notenvorschläge zu erzeugen. Dies deckt sich mit den Ergebnissen, die teilweise bereits zu Frage 1 erarbeitet worden sind.

Elektronische Klausuren mit Fokus auf Prüfungsdurchführung

Die gezielten Einsatzmöglichkeiten generativer KI während der Durchführung sind breit gefächert. Tools wie ChatGPT können beispielsweise in open-book Klausuren verwendet werden, um die dort abgefragten Inhalte für die Weiterbearbeitung durch die Studierenden vorzubereiten. Dies würde eine Abkehr vom Auswendiglernen bedeuten, was generell begrüßenswert erscheint. Allerdings muss auch der Einsatz generativer KI zunächst fachbezogen erlernt werden. Daher sollte neben den Lernzielen einzelner Module in einem Studium auch ein übergeordnetes Lernziel „KI-Kompetenz“ existieren, was beispielsweise in einem eigenen Soft-Skill-Modul, ähnlich dem vielerorts angebotenen Kurs „wissenschaftliches Arbeiten“, realisiert werden könnte.

Ähnlich wie die elektronische Klausur selbst, bietet auch generative KI viele Chancen sinnvoll in einer Prüfung eingesetzt zu werden. Dies sollte als Impuls aufgefasst werden, Prüfungen jetzt weiterzuentwickeln, um das Potenzial beider Punkte sinnvoll zu entfalten. Damit einher geht auch die Frage, was eine Modulprüfung wirklich prüfen soll: Nur das Endergebnis bzw. gemerktes Wissen, oder doch erworbene Kompetenzen bzw. die Entwicklung der Studierenden während des Lernens?

4 Diskussion der Ergebnisse und Fazit

Die während des Workshops unter Berücksichtigung der drei Fragen zusammengetragenen Ergebnisse zeichnen ein klares Bild: Generative KI wird nicht als ernstzunehmende Bedrohung des Prüfungswesens in den Hochschulen bewertet. An Stellen, wo generative KI als unerlaubtes Hilfsmittel eingesetzt werden könnte, gibt es hinreichend viele Ansätze, Prüfungen derart zu modifizieren, dass der Einsatz solcher Tools nicht mehr als gewinnbringend einzustufen ist.

Darüber hinaus wird aber das große Potenzial gesehen, das neue Technologien – und hier geht die Perspektive an mancher Stelle durchaus über die generative KI der aktuellen Generation hinaus – für Prüfungen im Hochschulwesen mitbringen. Dieses Potenzial zu nutzen und sinnvoll in das Prüfungswesen zu integrieren wird eine der großen Herausforderungen der nächsten Jahre sein. Gerade im Kontext elektronischer Prüfungen, die selbst oft noch nicht zur Gänze exploriert wurden bzw. im Einsatz sind, scheint generative KI als weiterer wertvoller Baustein, um diese Art der Prüfungen weiterzuentwickeln.

Sicher ist in jedem Falle Folgendes: Die technologische Entwicklung wird in rasantem Tempo weiter gehen. Diese einfach zu ignorieren und Prüfungen nicht daran anzupassen wird selbst mittelfristig nicht mehr funktionieren. Am Ende müssen die Prüfenden entscheiden, ob sie die Entwicklung der nächsten Jahre mitgestalten wollen, oder ob sie eines Tages von der Entwicklung überholt werden wollen.