

FAQ ZUR KI-KORREKTUR

1. Warum probieren wir eine KI-Korrektur aus?	1
2. Was ist bisher passiert, was läuft gerade und was planen wir?	1
3. Ist eine vollständige KI-Umstellung schon beschlossene Sache? Ab wann werden Klausuren durch KI korrigiert?	2
4. Liefere ich durch meine Bearbeitung der KI Datenpunkte, die mich letztlich als Jurist/in überflüssig machen wird?	2
5. Warum suchen wir nicht bessere Korrekturkräfte?	3
6. Geht es nur ums Sparen? Wie schlecht bezahlt werden Korrekturkräfte?	3
7. Muss die KI so geschwätzig sein? 17 Seiten Anmerkungen zu einer 10-seitigen Klausur?.....	4
8. Kann die KI mit Abweichungen von der Musterlösung umgehen?	4
9. Muss die KI so penibel sein und jeden Tippfehler monieren?	5
10. Wie soll eine KI wie ein Mensch bewerten?.....	5

1. WARUM PROBIEREN WIR EINE KI-KORREKTUR AUS?

Im Klausurenkurs fallen jede Woche im Schnitt 200 Klausuren an. Dies gilt für rund 52 Wochen im Jahr – die fehlenden Wochen werden durch das zweimal jährliche schriftliche Probeexamen, wo jeweils 6 Klausuren am Stück geschrieben werden, mehr als kompensiert. Diese sollen zeitnah korrigiert und zurückgegeben werden. Bereits seit mehreren Semestern häufen sich Beschwerden von Studierenden, dass es vielfach nur wenige und ggf. formalistische/oberflächliche Anmerkungen der Korrektorinnen und Korrektoren gab und die Noten deshalb kaum nachvollziehbar waren.

Gleichzeitig haben in den vergangenen Jahren generative KI-Systeme enorme Fortschritte gemacht. Die gerade in den Anfangsjahren beobachteten Halluzinationen und irritierenden Zufallsartefakte haben erheblich abgenommen. Gleichzeitig können KI-Systeme mit größeren Kontextfenstern umgehen, d.h. mehr aktuelle Daten zur konkreten Anfrage verarbeiten. Im Rahmen des Projekts DeepWrite wurde wesentliche Grundlagenforschung für das automatisierte Auswerten von Klausuren geschaffen, das nunmehr nutzbar gemacht werden soll.

Wichtig: Es geht weder aktuell noch langfristig um den Einsatz von KI-Systemen für echte Prüfungsleistungen (Zwischenprüfung, Hauptstudium, Schwerpunktbereich, Examen) – hier handelt es sich um den Bereich von „Hochrisiko-KI“, die besonders strengen rechtlichen Anforderungen unterliegt. Auch die diesbezüglichen datenschutzrechtlichen Anforderungen (Art. 22 DS-GVO) würden die aktuell getesteten Systeme nicht erfüllen. Vielmehr sollen die KI-Systeme beim (besseren) Klausurschreiben unterstützen und umfangreiches, individuelles Feedback liefern.

2. WAS IST BISHER PASSIERT, WAS LÄUFT GERADE UND WAS PLANEN WIR?

Wir haben einen ersten Pilotversuch mit drei bereits auf dem Markt verfügbaren KI-Systemen im Rahmen der Klausur im Zivilrecht vom 27.06.2026 bei Prof. Harnos durchgeführt. Die Kosten hierfür (rund 1.500 € - ca. 500 € pro KI-System) wurden nicht aus Fakultätsmitteln, sondern durch private Spenden aufgebracht. Insgesamt haben sich 145 Studierende (von 166 Studierenden) bereiterklärt, ihre Klausuren zusätzlich durch die drei KI-Systeme korrigieren zu lassen. Die Klausur wurde besprochen und im Anschluss insbesondere über die Qualität und Sinnhaftigkeit der KI-Korrektur diskutiert.

Neben der mündlichen Aussprache und persönlichen Gesprächen wurde eine anonyme Evaluation durchgeführt, an der sich 50 der 145 Studierenden, deren Klausur durch die KI bewertet wurde, sowie 31 weitere Personen beteiligt haben. Die Ergebnisse wurden bereits in einer separaten Datei veröffentlicht – sie

enthalten berechtigte Kritik an den eingesetzten KI-Systemen, die als Verbesserungsvorschläge aufgegriffen wurden.

Auf Grundlage des so erlangten Feedbacks wird in einem zweiten Versuch die Klausur im Zivilrecht vom 11. Juli 2026 (Dr. Maus) ebenfalls zusätzlich durch ein auf dieser Grundlage entwickeltes KI-System korrigiert, die entstehenden Kosten werden ebenfalls durch private Spende gedeckt. Auch hier soll sich eine Diskussion und Evaluation anschließen.

Ein dritter Versuch soll im Rahmen des Probeexamens bei Prof. Kuhn erfolgen. Auch hier soll das System weiterentwickelt sein und die bis dahin gesammelte Kritik berücksichtigt werden.

Neben der Beurteilung durch die Studierenden ist auch der kritische Blick der beteiligten Aufgabenstellerinnen und Aufgabensteller auf die KI-Korrektur gerichtet. Neben der Reaktion auf Rückfragen in der Besprechung (die es auch bei den menschlichen Korrekturen stets gab) stehen dabei vor allem die Klausuren im Vordergrund, bei denen die Note der menschlichen Bewertung deutlich von der maschinellen Bewertung abweicht. Darüber hinaus sollen – wiederum unter Einsatz von KI – diejenigen Klausuren ausgefiltert werden, die in besonderem Maße von der Lösungsskizze / dem Rohpunkteschema und/oder den Arbeiten der anderen Teilnehmenden abweichen.

Parallel dazu wird versucht, die Bedienbarkeit auch für technisch nicht versierte Aufgabenstellerinnen und Aufgabensteller zu gewährleisten. Dabei soll die KI insbesondere Vorschläge zur Verbesserung von Lösungsskizzen und zur Entwicklung von Rohpunkteschemata machen.

Es handelt sich nicht um ein abstraktes Forschungsprojekt, sondern es sollen zeitnah mit überschaubarem Aufwand Vorteile für die Studierenden (bessere Klausurgutachten) erzielt werden. Bei jedem Versuchsdurchgang wird es (zum Teil fundamentale) Veränderungen geben und wir werden jedes Mal Feedback einholen und Verbesserungen suchen. Voraussichtlich werden sich weitere Testläufe anschließen. Wir verbinden damit aber die Hoffnung, dass wir selbst und unsere Studierenden die KI-Korrektur letztlich für ihre Examensvorbereitung sinnvoller finden als das, was wir ihnen derzeit durch Korrekturassistentinnen und -assistenten bieten können. Das muss die KI aber beweisen – wir geben hier nicht wegen berechtigter Kritik sofort vollständig auf, sondern greifen diese als Anregung für Verbesserungen auf.

3. IST EINE VOLLSTÄNDIGE KI-UMSTELLUNG SCHON BESCHLOSSENE SACHE? AB WANN WERDEN KLAUSUREN DURCH KI KORRIGIERT?

Die Einführung einer KI-Korrektur hängt nicht nur von der Zustimmung der Studierenden (die unverzichtbar ist) ab, sondern auch von organisatorischen Fragen: Neben der Klärung der technischen Voraussetzungen gilt es auch haushalts- und vergaberechtliche Aspekte zu berücksichtigen, aber auch Fragen des Datenschutzrechts. Hinzu kommt, dass eine entsprechende Einführung auch die Lehrfreiheit der einzelnen Aufgabenstellerinnen und Aufgabensteller berührt, so dass auch die Professorinnen und Professoren hier überzeugt sein müssen. Insoweit gilt es noch einige Hürden vor der flächendeckenden KI-Korrektur zu überwinden. Rein praktisch kann es schon angesichts der teilweise vollständigen Ablehnung nicht auf eine Einstimmigkeitsentscheidung hinauslaufen, aber auch eine einfache Mehrheit gäbe keinen Anlass für einen Systemwechsel.

Bis dahin werden wir weitere Testversuche fahren und Feedback einholen.

4. LIEFERE ICH DURCH MEINE BEARBEITUNG DER KI DATENPUNKTE, DIE MICH LETZTLICH ALS JURIST/IN ÜBERFLÜSSIG MACHEN WERDEN?

Dass juristische Arbeit teilweise durch KI erledigt werden kann, ist kein Geheimnis und wird durch Kanzleien, Unternehmen und sogar den Staat mit großem Aufwand vorangetrieben. Der Einsatz der Korrekturkapazitäten wird daran nichts ändern.

Es werden vielmehr schon keine neuen Datenpunkte geschaffen, anhand derer die KI-Systeme als solche „dazulernen könnten“: Die verwendeten KI-Modelle sind für kommerzielle Zwecke konzipiert und werden u.a. auch von Unternehmen genutzt. Sie sind daher gerade für vertrauliche und datenschutzrechtlich

sensible Informationen gedacht. Entsprechend regeln die Nutzungsbedingungen explizit, dass eine Weiternutzung zum Training, eine Weitergabe an Dritte und auch nur eine längerfristige Aufbewahrung ausgeschlossen sind. Die betreffenden Unternehmen verpflichten sich zudem nach den Standardvertragsklauseln (SCC) die Vorgaben der DS-GVO auch bei Verarbeitung im Ausland einzuhalten.

Die übermittelten Daten werden so weit wie möglich pseudonymisiert – hierzu werden alle PDF-Dateien vor der Übermittlung serverseitig noch einmal neu gespeichert, um Meta-Daten zu entfernen; die Dateien werden mit einem nur für die konkrete Klausur gültigen einmaligen Code versehen und lassen sich so auch anhand von Matrikelnummern nicht zuordnen. Da die Klausuren getippt sind und alle Bearbeitungen sich mit dem gleichen Sachverhalt befassen, ist auch so eine Individualisierung nicht möglich. Durch die Verpflichtung Altdaten zu löschen kann auch kein Profil geschaffen werden, das ggf. einzelne Klausuren anhand des Schreibstils identifizieren könnte – wenn der Stil bei formalisierten juristischen Arbeit überhaupt hinreichend individuell ist. Da die nächste Klausur einen anderen Sachverhalt (und in der Regel sogar ein anderes Rechtsgebiet) behandelt, können auch Argumentationsmuster, Fehlerprofile und Formulierungen nicht für die Zuordnung herangezogen werden.

Zudem sollte man sich fragen, ob eine Klausurlösung – mit allen darin enthaltenen Fehlern - wirklich das geeignete Trainingsmaterial für eine juristische KI ist und nicht vielleicht die der Korrektur zugrundeliegende Lösung (die ggf. schon durch KI korrigiert wurde) diesen Effekt schon vorweggenommen hat.

Auch eine Abhängigkeit von US-Anbietern droht nicht. Für den Notfall haben wir bereits lokale KI-Systeme getestet, welche mit geeigneten Prompts und Kontextinformationen vergleichbare Korrekturen liefern würden (aber ein deutliches Mehr an lokalen Speicher- und Rechenressourcen sowie Wartungsaufwand benötigen würden).

5. WARUM SUCHEN WIR NICHT BESSERE KORREKTURKRÄFTE?

Früher wurde ein großer Teil der Korrekturen durch wissenschaftliche Mitarbeiterinnen und Mitarbeiter der Fakultät erbracht – allerdings nicht im Rahmen ihrer normalen Arbeitstätigkeit, sondern aufgrund eines zusätzlich abgeschlossenen Werkvertrags (gegen Einzelvergütung). Aufgrund eines Schreibens aus dem Kreis der Sozialversicherung wurde u.a. die Universität Passau darüber in Kenntnis gesetzt, dass in Arbeitsverhältnissen (§§ 611, 611a BGB) stehende Beschäftigte nicht durch Werkvertrag mit Aufgaben betraut werden dürfen, die auch im Rahmen des dienstlichen Weisungsrechts (§ 106 GewO) angeordnet werden könnten. Da wiss. Beschäftigte ausdrücklich für wissenschaftliche Dienstleistungen angestellt sind (Art. 72 Abs. 2 S. 1 BayHIG), die neben der Forschung auch die Lehre umfasst (vgl. nur Art. 72 Abs. 3 S. 3 BayHIG), ist die Übertragung von Klausurkorrekturen, insbesondere auch von Probeklausuren im Examenskurs, eine solche Tätigkeit, die grundsätzlich in der Arbeitszeit für den Arbeitslohn zu erbringen ist. Bei Arbeitslohn sind insbesondere Sozialversicherungsbeiträge abzuführen, was wiederum eine entsprechende Tätigkeit im Vergleich zur Tätigkeit auf Werklohnbasis unattraktiv macht. Damit stehen wiss. Beschäftigte praktisch nicht mehr für Korrekturen zur Verfügung.

Die Suche nach Personen, die zeitlich in der Lage sind, Pakete von 20 oder gar 50 5-Stunden Klausuren innerhalb von vier Wochen zu korrigieren (und das im Idealfall jede Woche) auf dem freien Markt gestaltet sich schwierig: Gute Juristinnen und Juristen sind oft in Berufen, die ihre Arbeitszeit voll ausfüllen (Richterinnen/Richter, Anwältinnen/Anwälte, etc.) – nebenbei in der Freizeit Klausuren zu korrigieren ist wenig attraktiv und muss ohnehin meist schon für die Staatsprüfungen (wo oft 100 oder mehr Klausuren anstehen) erledigt werden. Referendarinnen und Referendare sind zwar in der Anfangsphase ihrer Ausbildung noch zeitlich verfügbar – aber in späteren Phasen sind sie durch ihre Tätigkeit in der Anwaltskanzlei oder die Examensvorbereitung gebunden.

6. GEHT ES NUR UMS SPAREN? WIE SCHLECHT BEZAHLT WERDEN KORREKTURKRÄFTE?

Die Juristische Fakultät der Universität Passau zahlt pro korrigierter Klausurenkursklausur 12,80 € (https://www.uni-passau.de/fileadmin/dokumente/beschaefigte/Personal/Formulare_Internetseite/Korrekturassistenten/JUR_Korrekturvorlage_mit_Formularfeldern.pdf). Zum Vergleich: Nach der einschlägigen Bekanntmachung des Bayerischen Staatsministeriums der Justiz (<https://www.gesetze->

bayern.de/Content/Document/BayVV_2038_3_3_1_J_14610/true) erhalten die Prüferinnen und Prüfer für die Korrektur einer (echten) Klausur in der Ersten Juristischen Staatsprüfung eine Vergütung von 13,84 € - diese wurde am 01.09.2024 erhöht, bis dahin betrug die Vergütung 12,58 €.

Im Vergleich zu anderen Universitäten bewegt sich dieser Betrag im Mittelfeld – die Universität zu Köln zahlt etwa 11 € pro Klausur (<https://klausurenkurs.uni-koeln.de/ueber-die-einrichtung/stellenangebote>), die Goethe Universität Frankfurt a.M. sogar nur 10 € pro Klausur (https://www.jura.uni-frankfurt.de/131749086/Rechnung_Korrekturleistung_interaktiv_202210.pdf), an der Universität Hamburg hingegen 3,5*5,72 € = 20,02 € (<https://www.jura.uni-hamburg.de/service/stellenangebote/korrekturassistenten.html>) und an der Bucerius Law School „nur“ 19,50 € pro Klausur (<https://www.law-school.de/services/stellenangebote/video-korrektorinnen-m/w/d>).

Bei 200 Klausuren über 50 Wochen (= 10.000 Klausuren) kostet dies die Fakultät 10.000 * 12,80 € = 128.000 €, die überwiegend aus Stundienzuschüssen, zum Teil aber auch aus zentralen Mitteln der Fakultät bezahlt werden. Würde man die Vergütung auf 13,84 € erhöhen, betrügen die Kosten 138.400 € - also 10.400 € zusätzlich. Diese müssten anderweitig eingespart werden – etwa durch die Streichung von Einzelcoachings, des Zusatztrainings für diejenigen, welche die Erste Juristische Staatsprüfung im ersten Versuch nicht bestanden haben oder der sonstigen Mittel des Instituts für Rechtsdidaktik. Aber auch bei 13,84 € ist der Anreiz für den einzelnen Korrektor / die einzelne Korrektorin kaum höher – bei 40 Klausuren läge der Mehrverdienst bei nur 41,60 €. Als Kompensation für den Aufwand ändert dies kaum etwas an der Angemessenheit.

Engagierte und motivierte Korrektorinnen und Korrektoren übernehmen die Klausuren oft auch nicht für das (überschaubare) Honorar, sondern um sich selbst besser auf die Prüfungen (etwa in der Zweiten Juristischen Staatsprüfung) vorzubereiten, aus Verbundenheit zur Universität Passau (etwa als Dank für die gute Betreuung im Examenskurs) oder aus didaktischem oder persönlichem Engagement (etwa um künftigen Generationen etwas Gutes zu tun). In der Gesamtbetrachtung bilden diese Korrektorinnen und Korrektoren aber keineswegs die Mehrheit.

Vielmehr haben einige Korrektorinnen und Korrektoren in großem Umfang Klausuren nicht nur aus Passau, sondern auch von anderen Universitäten übernommen und korrigieren hauptberuflich. Um hier einen angemessenen Stundenlohn zu erhalten (der gesetzliche Mindestlohn beträgt derzeit 13,90 €), muss man dann schon mehrere Klausuren pro Stunde erledigen – ab drei Klausuren dürfte man sogar profitabel arbeiten. Wenn man aber für eine fünfstündige Klausur nur 20 Minuten zur Korrektur investiert, kann keine tiefgreifende Analyse erwartet werden. Entsprechend schlecht fallen dann auch die Korrekturen aus. Dass zum Teil nicht mehr Zeit in eine Korrektur investiert wurde, lässt sich mitunter anhand der Meta-Daten der Dateien der Korrekturassistenten nachvollziehen.

7. MUSS DIE KI SO GESCHWÄTZIG SEIN? 17 SEITEN ANMERKUNGEN ZU EINER 10-SEITIGEN KLAUSUR?

Wie schon die drei im ersten Testlauf eingesetzten Systeme gezeigt haben, ist die Ausgabe stark anpassungsfähig. Die Kritik, dass die Gutachten vielfach zu lang und unübersichtlich sind und Randbemerkungen zu geschwätzig daherkommen, haben wir ernst genommen.

Das derzeit im Test befindliche KI-Prompt generiert kurze Abschlussvoten (ca. 2 Seiten + Hinweise für künftige Klausuren) und knappe Randbemerkungen (wenige Worte). Hier ist sicher noch viel Feintuning erforderlich; aber schon in kurzer Zeit haben wir eine Qualität erreicht, welche die meisten Kritikpunkte ausräumen dürfte.

8. KANN DIE KI MIT ABWEICHUNGEN VON DER MUSTERLÖSUNG UMGEHEN?

Neben der Korrektur als solcher ist ein Hauptfokus der aktuellen Tests die Verbesserung von Lösungen und Rohpunkteschemata durch KI-Werkzeuge. Eine Lösungsskizze, die auch auf andere Lösungswege eingeht und diese möglichst weitgehend ausformuliert, bildet eine sehr viel bessere Grundlage für eine spätere Korrektur. Gleichzeitig wird versucht, neben der Lösungsskizze noch weiteren Kontext einzubeziehen, um auch abweichende Gedankengänge abzubilden.

Eine perfekte Berücksichtigung aller (auch eindeutig falscher) Lösungswege wird die KI aber voraussichtlich ebenso wenig leisten können wie ein menschlicher Korrektor – es kann nur eine deutlich höhere Kulanz als in den bisherigen Testläufen gewährleistet werden. Sie kann aber mehr Lösungswege spontan im Blick haben, stilistische Fehler eindeutig identifizieren, konsistentes, gezieltes, individuelles und sachliches Feedback dort liefern, wo Menschen mitunter durch Überlastung, schlechte Laune oder individuell fehlende Kenntnisse insbesondere in Randbereichen des Pflichtfachbereichs oberflächlich, knapp oder intolerant arbeiten. Sowohl Menschen als auch KI machen Fehler und sehen Vertretbares als falsch an – nur hat die KI hier ggf. eher die Neutralität, Unsicherheiten offenzulegen und ggf. den Aufgabensteller direkt zu fragen. Ein zentraler Aspekt im weiteren Projektverlauf wird die Prüfung sein, wie viele echte inhaltliche Fehler die KI-Systeme machen und hier nachzubessern.

9. MUSS DIE KI SO PENIBEL SEIN UND JEDEN TIPPFEHLER MONIEREN?

Nein, welche Ausgabe ein KI-System generiert ist abhängig von der konkreten Aufgabenstellung und dem gelieferten Kontext, insbesondere von Beispielen und Mustern. Durch die Möglichkeit, hier deutlich mehr Inhalt zu liefern als noch vor zwei Jahren, können die Antworten sehr viel stärker angepasst werden.

Das derzeit im Test befindliche KI-System ist schon sehr viel kulanter, aber bei schweren Fehlern auch streng. Auch hier wird weiteres Feintuning erforderlich sein, um eine ausgewogene Lösung zu finden.

10. WIE SOLL EINE KI WIE EIN MENSCH BEWERTEN?

Es ist eine irrige Vorstellung, dass „Menschen“ homogen bewerten. Die Personen, die in der Ersten Juristischen Staatsprüfung korrigieren sind überwiegend berufserfahrene Menschen aus der Praxis oder Professorinnen und Professoren. Im Klausurenkurs korrigieren dagegen entweder engagierte junge Freiwillige oder Personen, die hierin eine Einnahmemöglichkeit sehen und keine lukrativere Möglichkeit sehen, ihre Zeit einzusetzen. Die Korrektorengruppen sind also in ihren Erwartungen, Erfahrungen und ihrer Motivation durchaus unterschiedlich. Wer im Klausurenkurs regelmäßig 8 Punkte erzielt, wird das Examen wohl sicher bestehen – mit welcher Note ist aber völlig offen.

Die KI wird mit der Lösungsskizze und Begleitmaterial als Kontext an die Klausur gesetzt. Je breiter die dort abgedeckten Möglichkeiten, desto flexibler wird die Korrektur. Die KI ist soweit möglich anhand von Methodikunterlagen, Anpassungen durch erfahrene Prüfer:innen und sehr weitreichenden Kontext kalibriert und versucht maximale Notengerechtigkeit zu schaffen. Menschliche Willkür wird so zwar nicht simuliert, aber es wird eine begründete Meinung (konsequent) vertreten.

Die menschliche Kompensation fehlt auch bei einer KI-Korrektur nicht – neben der Möglichkeit, Fragen im Rahmen der Besprechung zu stellen ist bei jedem KI-System eine stetige menschliche Überwachung von Auffälligkeiten erforderlich. Klausuren, bei denen das KI-System Beurteilungsschwierigkeiten hat, werden so manuell (d.h. menschlich) überprüft und das KI-System verbessert.

Da es nicht um echte Prüfungsleistungen geht, ist eine menschliche Kontrolle auch nicht erforderlich. Auch die aktuellen Korrekturassistenten sind nur eine Annäherung an die Bewertung in der realen Prüfungssituation – welche ein KI-System nach unserer ersten Tests durchaus in vergleichbarer Weise mit entsprechenden Anpassungen sogar besser leisten könnte. Diese Hoffnung muss sich aber noch bestätigen, weshalb wir laufend weitere Versuche durchführen.