

Evaluation KI-Korrektur

(Klausur vom 27.06.2026)

1. Antworten der 50 Personen, die teilgenommen haben.....	2
a. Künftig soll die Korrektur erfolgen durch.....	2
b. Vergleich der Korrekturen	2
c. Negatives Feedback	2
d. Positives Feedback	15
e. Sonstige Anmerkungen	23
2. Antworten der 31 Personen die nicht teilgenommen haben.....	33
a. Künftig soll die Korrektur erfolgen durch.....	33
b. Vergleich der Korrekturen	33
c. Negatives Feedback	33
d. Positives Feedback	36
e. Sonstige Anmerkungen	38
f. Warum haben Sie nicht teilgenommen?	41

Die Ergebnisse der Evaluation geben Anlass zu erheblichen Anpassungen. Bemerkenswert sind dabei die Unterschiede zwischen denjenigen, die teilgenommen (also ihre Klausuren durch die KI bewerten ließen), und denjenigen, die nicht teilgenommen haben: In der ersten Gruppe würden 68 % eine reine KI-Korrektur bevorzugen und 16 % eine rein menschliche Korrektur; in der zweiten Gruppe hielten sich beide Präferenzen mit jeweils 41 % die Waage. In beiden Gruppen schnitten die KI-Korrekturen nach den vorgegebenen Kriterien deutlich besser ab als die menschlichen Korrekturen. Als größte Schwächen der KI wurden zu lange, repetitive und unübersichtliche Ausgaben genannt; zudem wurde bemängelt, dass die KI zu starr an den vorgegebenen Lösungswegen festhält und vertretbare alternative Lösungswege mitunter als „falsch“ bewertet. Die meisten wünschen sich – soweit möglich – eine Form des hybriden Modells. Vereinzelt kam generelle Systemkritik hinzu.

1. Antworten der 50 Personen, die teilgenommen haben

a. Künftig soll die Korrektur erfolgen durch...

8: Nur Mensch	8: Offen	34: Nur KI
---------------	----------	------------

b. Vergleich der Korrekturen

	Mensch	Klausuren-Kiste	DeepWrite	Institut für Rechtsdidaktik
Qualität und Quantität der Randbemerkungen	1.65 Sterne	3.62 Sterne	3.62 Sterne	4.02 Sterne
Übersichtlichkeit und Verständlichkeit des Abschlussvotums	1.92 Sterne	3.6 Sterne	3.79 Sterne	4.11 Sterne
Nachvollziehbarkeit der Note	2.33 Sterne	3.77 Sterne	3.87 Sterne	4.15 Sterne
Nützlichkeit von Arbeits-/ Klausurtipps	1.63 Sterne	3.38 Sterne	3.85 Sterne	4.13 Sterne
Gesamtbeurteilung	1.94 Sterne	3.57 Sterne	3.6 Sterne	4.13 Sterne

c. Negatives Feedback

Klausurenkiste: Sehr knappe Anmerkungen die sich nicht ganz mit der Gewichtung der Musterlösung decken; Sehr starre Anwendung von Regeln ohne positiv Bewertung von als gut betitelten Stellen

Eine **menschliche Korrektur** habe ich nicht erhalten

Menschliche Korrektur hat keine Randbemerkungen

Menschliche Korrektur stellt nur abstrakt auf Grundlagen des Rücktritts- und Arbeitsrechts ab, ohne konkret Fehler zu veranschaulichen

Es wird im Großen und Ganzen auf die Lösungsskizze verwiesen, die man auch ohne eine Korrektur angeschaut hätte.

Menschliche Korrektur spricht mehr von Motivation und verweist im letzten Satz darauf, dass man sich angewöhnen sollte Klausuren am PC zu schreiben, obwohl objektiv und neutral gesehen die Entscheidung weiterhin beim Verfasser liegt, solange das LJPA eine analoge Bearbeitung zulässt.

Menschliche Korrektur: kaum bis wenig Randbemerkungen, keine Erläuterungen zu den Fehler, keine Verbesserungsvorschläge, keine Anmerkungen hinsichtlich Sprache und Stil

Die **KI-Korrekturen** benutzen ähnliche Formulierungen, je nachdem, wie die Qualität der Bearbeitung ist. Das ist im Hinblick auf die Vergleichbarkeit positiv zu werten, trifft aber nicht den Punkt, wo sich die eigene Bearbeitung verbessern müsste.

Menschliche Korrektur: Ich habe nur 2 Randbemerkungen bekommen, jeweils nur ein Wort, nämlich vertretbar!

Das Abschlussvotum der **menschlichen Korrektur** ist sehr oberflächlich, dort heißt es lediglich eine bessere Struktur und tiefere Behandlung der Probleme wäre besser gewesen, hinsichtlich der materiell-rechtlichen Fragestellungen wird pauschal auf die Lösungsskizze verwiesen!

Klausurenkiste: Arbeits- bzw. Klausurtipps sind nicht sonderlich hilfreich!

DeepWrite: Die Korrektur ist m.E. viel zu lang, eine Korrektur sollte meiner Meinung nach auf die wesentlichen Punkte beschränkt werden. Für mich persönlich soll meine Korrektur nicht länger sein als meine Klausur. Zudem finde ich die Darstellung sehr unübersichtlich, im Vergleich zu den anderen **KI-Korrekturen**.

KI des Instituts für Rechtsdidaktik: Die Note geht in der Darstellung etwas unter! Ich persönlich würde es bevorzugen, wenn die Note groß und in fett am Ende der Bewertung steht! (Im Klausurenportal wurde bei mir hier z.B. auch die falsche Note eingetragen, vermutlich weil der Eintragende auch die Note nicht eindeutig herauslesen konnte)!

Ich verstehe nicht, wie man beispielsweise in Frage 2 dieser Klausur von jedem **KI-System** 15 von 16 Punkten bekommt, die Aufgabe also nahezu richtig löst, von der **menschlichen Korrektur** aber nur eine Anmerkung bekommt, dass die Systematik nicht nachvollziehbar ist. Daraus kann man leider überhaupt nichts lernen und es scheint einfach wie eine willkürliche Anmerkung.

Ich hatte schon lange keine so schlechte **menschliche Korrektur** mehr. Leider steht kein Name dabei. Die letzten Korrekturen, die ich hatte, fand ich besser. Mit der hier **menschlichen Korrektur** konnte ich nahe zu nichts anfangen.

Menschliche Korrektur: Insgesamt hängt die Qualität der Korrektur extrem vom jeweiligen Korrektor ab und ist damit glücksabhängig. Für den durchschnittlichen externen Korrektor gilt nach meiner Erfahrung: Die Notenvergabe schwankt stark und wirkt teilweise willkürlich. Eine nachvollziehbare Punkterechnung fehlt überwiegend. In einzelnen Klausuren ist die Anmerkungsichte sehr gering, und die Schlussvoten wirken stellenweise textbausteinartig und austauschbar; der Lerneffekt ist eher gering.

Klausurenkiste: Verkennt teilweise eine tatsächlich erfolgte Auseinandersetzung des Bearbeiters. Teilweise Lücken und Fehlplatzierungen der Anmerkungen. Der Wrapper kommt bei einem erneuten Durchgang teilweise zu einem anderen Ergebnis.

DeepWrite: Benotet systematisch zu streng. Es vergibt keinerlei Randbemerkungen mit Seiten-/Zeilenbezug; das Feedback bleibt thematisch, sodass schwer nachvollziehbar ist, wo genau im Text der Mangel liegt. Außerdem ist das Feedback/Votum viel zu lang - im Nachbereitungsalltag würde ich mir fast wieder eine Zusammenfassung durch ein LLM erstellen und diese dann den einzelnen Textbausteinen zuordnen lassen. ;) In Beispiel 2 hebt zudem eine für mich unerwartete kaufmännische Rundung (74,5 → 75 Rohpunkte) die Note über eine Notenpunktstufe.

Institut für Rechtsdidaktik: Das Layout der Randbemerkungen ist für meinen Geschmack unübersichtlich. Ich würde es bevorzugen, wenn die Anmerkungen – wie bisweilen üblich – an der Originalklausur verbleiben. Schön wäre außerdem eine Gewichtung der Anmerkungen, also bei Fehlern z.B. [gravierend]/[mittel]/[gering] und bei guten Ausführungen ebenso ein kurzer Tag mit farblicher Markierung. In Beispiel 3 wurde die Umrechnungstabelle falsch angewendet („62,5 Rp liegen im Bereich 63–66“ → 9 Punkte; korrekt wären 8 Punkte gewesen). Vereinzelt überschießende Monierungen – was ich für den Zweck der Klausuren, die eigene Selbstkorrektur, allerdings gar nicht so schlecht finde. Schön wäre des weiteren ein eigener Block "Arbeitsempfehlungen für die Nachbereitung".

bei allen **KI-Korrektoren:** sehr lange Sätze, nur um zu sagen, dass etwas i.O. ist, dass macht es unübersichtlich

KI des Instituts für Rechtsdidaktik

- schreibt mehrmals als Randbemerkung das Gesamtfazit zu einem Prüfungspunkt
- bewertet laut Lösung vertretbares (z.B. Ablehnung Anwendbarkeit § 346 IV) als falsch und schreibt dies bei jedem anschließenden Zwischenergebnis erneut

KI DeepWrite

- Korrektur nicht direkt neben Bearbeitung, erzählt gesamte Klausurbearbeitung nach, extrem unübersichtlich und viel zu ausführlich
- bewertet laut Lösung vertretbares (z.B. Ablehnung Anwendbarkeit § 346 IV) als falsch
- akzeptiert es nicht, am Anfang zu schreiben, dass nicht näher bezeichnete Zitate solche des BGB sind und markiert das immer wieder (aber auch nicht immer) als unvollständiges Zitat

Klausurenkiste

- keinerlei Anmerkungen zu in Lösung nicht vorgesehenen Punkten wie § 377 HGB
- gibt Rohpunkte für nicht geprüfte Sachen (obwohl es erkennt, dass sie nicht geprüft wurden...) (hat aber mein **menschlicher Korrektor** auch gemacht, letzterer ohne fehlende Prüfung zu erkennen)

menschlicher Korrektor

- es wurden im Rohpunkteschema oftmals für Aspekte 0 Punkte vergeben, die durchaus behandelt wurden → Note nicht nachvollziehbar (deutliche Abweichung zur KI Korrektur (3 Punkte (**Mensch**) vs. 8/8/6 Punkte (KI))
- gibt quasi keine Randbemerkungen und auch kein Abschlussvotum
- Arbeits- und Klausurtipps gibt es auch nicht

zur KI des Instituts für Rechtsdidaktik

- vllt. noch ein extra Absatz oder mehr Anmerkungen zu Stil/Schwerpunktsetzung (wie bei der Korrektur von **DeepWrite**)

zur DeepWrite

- Anmerkungen lieber - wie bei Instituts KI - in Tabelle

zur Klausurenkiste:

- bei Frage 3 kommen tlw Anmerkungen zu Frage 1

Die **KI des Instituts für Rechtsdidaktik** ist viel zu unübersichtlich, teilweise werden Bemerkungen wiederholt. Hier wäre eine strukturierte und kompaktere Bewertung zielführender. Die fast 17 Seiten der Korrektur könnten auf 10 Seiten zusammengefasst werden, wenn man die Wiederholungen heraus nehmen würde. Bei allen **KI-Korrekturen** fehlt mir die Vergleichbarkeit, weil ich zwei unterschiedliche Noten bekommen hätte.

Korrektur **Mensch**:

- es wurden bei dieser aber auch bei anderen Klausuren regelmäßig Prüfungspunkt die bearbeitet wurden übersehen und folglich negativ in die Gesamtbewertung eingeflossen
- es finden sich auf einer 14 seitigen Bearbeitung insgesamt 4 Anmerkungen ("zu knapp"; "zu knapp"; "an dieser Stelle falsch" und "zu viel")
- es gibt kein Votum, keine Verbesserungsvorschläge und es wird auch nicht auf falsche Normzitate oder unrichtige Subsumtionen hingewiesen
- es fehlt die Angabe der Note. Es wird lediglich das Rohpunkteschema angefügt

KI des Instituts für Rechtsdidaktik:

- das Format (Tabelle) eignet sich für einen guten Lesefluss und eine bessere Nachvollziehbarkeit leider nicht
- z.T. sind die Anmerkungen unnötig (eine Anmerkung, die das vom Bearbeiter Geschriebene wiedergibt um schlussendlich sinngemäß mit richtig unterstreicht ist nicht wirklich zielführend; hier würde es bereits ausreichen den Absatz mit "gut" oder "richtig" zu bezeichnen)
- hat der Bearbeiter Absätze für eine bessere Übersichtbarkeit und Lesbarkeit innerhalb eines Schreibblocks gesetzt, die keinen eigenen Gliederungspunkt darstellen, so wird für jeden Absatz die wortgleiche Anmerkung wiederholt => ist gerade im gewählten Tabellenformat schlecht

DeepWrite:

- die Bewertung arbeitet nicht direkt mit der vom Bearbeiter abgegebenen Lösung, sondern in einem separaten Dokument
- die Bewertung von Frage 2 und 3 wird zusammengefasst. Damit ist im Ergebnis teilweise nicht ganz klar wie viele Punkte pro Aufgabe verteilt wurden

Klausurenkiste:

- aus der Korrektur wird leider nicht ohne eigenes Nachrechnen ersichtlich, wie viele Punkte bei welcher Frage insgesamt vergeben wurden

DeepWrite: Sehr textlastig und stellenweise redundant; einzelne Abzüge wirken pingelig (z.B. 0,25/1 bei der Anspruchsgrundlage, wo der Handkorrektor volle Punkte gab). Frage 2 und 3 werden nur zusammen ausgewiesen.

KI des Instituts für Rechtsdidaktik: Bei der Benotung deutlich zu großzügig – 49 statt 33 Rohpunkte, also zwei Notenstufen über der Handkorrektur (5 statt 3 Punkte). Sie vergibt öfter volle Punkte, obwohl die eigene Anmerkung daneben noch einen Fehler benennt. Wer sich nur darauf verlässt, schätzt sich deutlich besser ein, als er ist.

Klausurenkiste: Ein paar heftige Ausreißer bei der Punktevergabe – vor allem wurde meine Zug-um-Zug-Lösung als einzige als gleichwertig zur Aufrechnung gewertet (volle Punkte), was Musterlösung, Handkorrektur und die anderen KIs anders sehen.

Insgesamt eine Notenstufe zu gut (4 statt 3 Punkte). Das eigene Punkteschema ohne Zwischensummen je Frage lässt sich schlecht mit dem offiziellen Rohpunkteschema vergleichen.

Menschliche Korrektur: Es wurde kein Abschlussvotum beigefügt, lediglich das Bewertungsschema ausgefüllt; die Randbemerkungen sind teils gut, teils nicht verständlich

Klausurenkiste: Randbemerkungen überwiegend gut, iRd Zitierweise wird "lit. a" als fehlerhaft angemerkt; bei § 441 BGB (welchen ich fehlerhaft direkt angewendet habe), wird die fehlende Analogie nicht bemängelt

DeepWrite: Die Bewertung wird nicht anhand der konkreten Abschnitte vorgenommen, dadurch leidet die Nachvollziehbarkeit und Übersichtlichkeit

KI des Instituts für Rechtsdidaktik: fehlende Analogie bei § 441 BGB nicht erkannt

- Fehlende Flexibilität und kein "Mitdenken": Die KI hängt sehr stark an der Lösungsskizze, wodurch Raum genommen wird, eigene Ideen außerhalb der Lösungsskizze zu entwickeln. Es ist der Eindruck entstanden, dass die KI-Note vor allem widerspiegelt, wie genau man die Lösungsskizze wiedergegeben hat, was ja eigentlich genau nicht Ziel der juristischen Ausbildung sein soll. Anders als (gute) **menschliche Korrekturkräfte** können Abweichungen von der Lösungsskizze nicht angemessen eingeordnet werden. Selbst Punkte, die eigentlich "Kleinigkeiten" sind oder gar nicht unbedingt erforderlich sind, werden als Fehler angestrichen, die teilweise auch zu Punktabzügen führen (so zB die **KI des Instituts für Rechtsdidaktik** bei vermeintlich falschen/fehlenden Normziten). Hinzu kommt, dass die Abhängigkeit von der Qualität der Lösungsskizze die Schwächen der Lösungsskizze reproduziert. Es geht die Möglichkeit verloren, dass "mitdenkende" Korrekturkräfte an einigen Stellen von der Lösungsskizze abweichen können, zB wenn es um Aufbaufragen geht.

- Teilweise erfasst die KI nicht korrekt, was man geschrieben hat. Beispielsweise wurde von der **KI des Instituts für Rechtsdidaktik** bemängelt, dass ich ungenau zitiert hätte (ohne "lit. a" oder "Fall 2") obwohl die Zitate genau waren (und die angesprochenen "lit. a" und "Fall 2" beinhalteten). Bei der **KI des Instituts für Rechtsdidaktik** sind auch ein paar inhaltlich falsche Anmerkungen vorhanden gewesen, beispielsweise, dass § 326 V BGB keine Halbsätze enthalten würde.

- Die KI kann keine Methodik: Die methodischen Anmerkungen waren teilweise nicht gut und widersprachen zum Teil explizit Hinweisen, die im Examenskurs und Einzelcoaching zur richtigen Methodik gegeben wurden. Beispielsweise wurde bemängelt, dass nicht immer Gutachtenstil verwendet wurde, sondern teilweise auch Feststellungsstil, was bei unproblematischen Punkten aber gerade nicht falsch ist. Auch wurden Formulierungen wie "Fraglich ist" ausdrücklich von der KI empfohlen, was eigentlich nicht in ein gutes, rechtsfolgenorientiertes Gutachten gehört. Auch hier zeigt sich, dass die KI zu sehr die Lösungsskizze reproduziert, die (wie bei Lösungsskizzen so

häufig) nicht methodisch einwandfrei verfasst ist, sondern eben als Lösungsskizze vor allem Inhaltliches transportieren soll.

- Die Zusammenfassungen am Ende der Korrekturen aller KIs wiederholen eigentlich fast nur das, was bereits in den Randbemerkungen stand und sind daher kaum hilfreich. Generell werden Fehler oft an verschiedenen Stellen wiederholt angesprochen, was ebenfalls keinen Mehrwert liefert.

- Die **DeepWrite** Korrektur als Fließtext statt als Randbemerkungen ist sehr unhandlich zum Nacharbeiten der eigenen Fehler. Die Korrektur ist als Randbemerkung verfasst, wieso kann sie dann nicht einfach neben der Klausur stehen, so wie bei den anderen **KI-Korrekturen**? Der Text ist im Übrigen viel zu lang, so eine Korrektur liest man sich keinesfalls aufmerksam komplett durch. Auch hat Deepwrite scheinbar Probleme mit Zwischenüberschriften, die es teilweise für Obersätze hält und daher kritisiert.

- Deepwrite und die **KI des Instituts für Rechtsdidaktik** bewerten Aspekte als "Stärken", die nicht wirklich verallgemeinerungsfähig oder hilfreich sind. Es bringt nichts, wenn die KI einem sagt, eine eigene Stärke sei, dass man den § 720 BGB gesehen hat.

- **Menschliche Korrektur** enthält kaum Randbemerkungen und ist so nicht nachvollziehbar

- **Klausurenkiste** staut die Randbemerkungen zu fehlenden Inhalten am Ende der Klausur.

- **DeepWrite** ist wenig übersichtlich und teilweise repetitiv

- **KI des Instituts für Rechtsdidaktik** hat viele identische, sich wiederholende Randbemerkungen, welche nicht immer in engem Zusammenhang zum Text stehen.

Der **menschliche Korrektor** fällt besonders negativ auf, weil es so gut wie keine Randbemerkungen gibt. Die paar wenigen, die es gibt, sind nicht hilfreich. Auch das Ausfüllen des Rohpunkteschemas durch ihn ist in vielen Punkten nicht nachvollziehbar. Es sind oft einfach 0 Punkte eingetragen, obwohl ich die jeweiligen Punkte zumindest zum Teil in der Klausur angesprochen habe (man hat ein bisschen das Gefühl, dass die Klausur gar nicht richtig gelesen wurde und nur nach "Schlagwörtern" gesucht wurde). Dadurch bin ich bei der Klausur durch einen menschlichen Korrektor auch durchgefallen (3 Punkte), bei den **KI-Korrekturen** habe ich jedoch bestanden.

Die **menschliche Korrektur** besteht nur aus einer Art von Randbemerkung, nämlich, "zu kurz". Diese Bemerkung ist nicht sonderlich hilfreich. Außerdem gibt es am Ende kein Kommentar zur Bearbeitung, um etwaige Stärken und Schwächen aufzuzeigen oder Verbesserungsvorschläge

- **Mensch**: kein Votum und nur 2 kurze, völlig unleserliche Randbemerkungen

- **DeepWrite**: im Votum ziemlich demotivierend und hart formuliert (was ich von einer KI noch weniger gerne, als von einem **Menschen** höre - es reicht doch ein sachliches "falsch"); Randbemerkungen von der Formatierung her sehr unübersichtlich; insgesamt sehr viel Text, der sich inhaltlich wiederholt, es entsteht der Eindruck, dass auf Fehlern "herumgeritten" wird. Außerdem wird durchwegs im Votum kritisiert, ich könne nicht rechtsfolgenorientiert prüfen und es fehle die AGL im Obersatz, weil ich zunächst die Rechtsfähigkeit der GbR geprüft habe, und dannach mehrere AGL, so wie wir es bei Kuhn immer gemacht haben. Das wäre bei einem menschlichen Korrektor so nicht passiert. Der **Mensch** macht zwar auch Fehler oder sieht vor allem Fehler nicht. Aber mit völliger Überzeugung immer wieder lang und bereit auszuführen, der Aufbau sei total falsch und dann damit auch die ganze Methodik zu kritisieren, das wäre einem **Menschen** nicht passiert. Insoweit finde ich die Zulassung der **KI-Korrektur** eigentlich fahrlässig. Hätte ich weniger Erfahrung, hätte ich diese Kritik von **DeepWrite** ernst genommen und mich total verunsichern lassen.

Auch die Aufzählung dessen, was gut war in **DeepWrite** war nervig. zB "Du hast überwiegend ganze Sätze geschrieben" (ich habe selbstverständlich nur ganze Sätze geschrieben). Auch sonst wurden absolute Basics "KV erkannt" als positiv aufgezählt.

Ich muss sagen, gerade weil **DeepWrite** in der Besprechung so gelobt wurde, dass ich das überhaupt nicht teile.

Bei der **menschlichen Korrektur** war nicht einmal ein Abschlussvotum dabei, nur das ausgefüllte Rohpunkteschema

Mein Hauptproblem bei der **menschlichen Korrektur** ist eigentlich nur, dass sich der Korrektor nur selten mit meiner tatsächlichen Bearbeitung auseinandersetzt.

Mein Hauptproblem bei allen **KI-Korrekturen** ist die Masse. Eine Korrektur, die mehr sagt, als meine eigene Bearbeitung, erschlägt mich und kostet mich mehr Zeit, als mich mit einer kurzen, aber schlechten Korrektur auseinanderzusetzen. Es wird vieles zu oft wiederholt.

– Externe Korrektur: Meine Korrektur enthält kaum Randbemerkungen (kaum mehr als "schön"/ "etwas knapp" etc.) und ein Abschlussvotum fehlt

– **Klausurenkiste**: Die Fülle der Randbemerkungen könnte man m.E. reduzieren und nur noch Fehler/Verbesserungsvorschläge und besonders gute Passagen anmerken

– Deepwrite: Der Fließtext mit Feedback zum Inhalt ist nicht besonders übersichtlich; es wäre besser, wenn die Anmerkungen direkt an der Seite meiner Klausur stehen und auf einzelne Passagen Bezug nehmen

– Institut: Die Randbemerkungen wiederholen sich regelmäßig (d.h. es steht oft derselbe Inhalt an mehreren Passagen meiner Klausur), d.h. ist m.E. unnötig.

bei der KI Korrektur des Instituts für Rechtsdidaktik steht in der rechten Spalte ganz oft das gleiche, das einfach nur generell für den oberste Punkt (zB Stellvertretung, Rücktritt) gilt, aber keine Details zu den Unterpunkten.

Klausurenkiste versteht wohl genaue Normbezeichnung nicht so gut. Mir wurde § 434 III 1 Nr. 3 BGB als falsches Normzitat angestrichen. Begründet wurde dies damit dass § 434 III nr. 3 BGB einschlägig war.

Die **menschlichen Korrekturen** sind sehr vom individuellen Korrektor abhängig, aber nur in Einzelfällen gibt es wirklich hilfreiche Anmerkungen. Einzelne Haken am Rand helfen bei der Vorbereitung nicht viel mehr als das Durchgehen der Musterlösung.

Die **KI-Korrektur** ist teils etwas lang und damit unübersichtlich. Um die Anmerkungen wirklich jede Woche effektiv nachzuarbeiten, würde ich die Anmerkungen direkt am Rand als am besten einstufen.

Teilweise scheint die **KI-Korrektur** nicht besonders flexibel auf abweichende Lösungen reagieren zu können. Das könnte sich vor allem dann auswirken, wenn einzelne Punkte zwingend als falsch eingeordnet werden, auch wenn eine andere Ansicht evtl. vertretbar sein könnte. Um dieses Problem zu lösen, müsste ggf. wie im Testdurchlauf eine ausführliche Musterlösung für jede Klausur erstellt werden (wobei das Problem natürlich auch bei **menschlichen Korrekturen** auftreten kann, die sehr an der Musterlösung orientiert sind).

Ohne das Aufführen von Abweichungen wird es womöglich schwierig, einzelne Punkte wieder an anderer Stelle zu kompensieren, was ein menschlicher Korrektur meiner Ansicht nach flexibler gewährleisten könnte.

Mensch: Randbemerkungen sind quasi nicht vorhanden, auch ein Endvotum gar nicht. Es wurde nur das Klausurschema mit + oder - ausgefüllt. Dadurch ist der Mehrwert der Korrektur kaum/ gar nicht gegeben

DeepWrite: auch wenn die Ausführlichkeit toll ist und man dadurch vieles gut nachvollziehen kann, wirkt die Korrektur doch z.T. unübersichtlich und man braucht etwas, bis man sich orientieren kann. Da wäre eine "Gliederung" am Anfang toll

Sowohl die KI-Modelle als auch die **menschlichen Korrekturkräfte** gehen kaum auf Methodik ein. Wenn als Stärken und Schwächen die Darstellung und Lösung einzelner Probleme dieser Klausur aufgezeigt wird, ist das zwar als Nachbereitung sinnvoll. Aber was viel hilfreicher und wichtiger wäre, wäre das Aufzeigen von methodischen Stärken und Schwächen, weil diese Fähigkeiten auf jede Klausur übertragen werden können und vor allem auch unbekannte Probleme so gelöst werden können. Letztlich kommt es in der Praxis und im Examen darauf an, juristisch argumentieren zu können, diese

Argumente aufbauen zu können und mit ihnen überzeugen zu können. Vor allem muss man mit dem Gesetz arbeiten können und dazu die Methodik beherrschen. Diese Punkte sind schwer alleine zu lernen. Und dabei helfen aktuell weder die menschlichen noch die **KI-Korrekturen**.

Eine KI mag in der Lage dazu sein, Lösungswege, die von der Lösungsskizze abweichen, zu bewerten, allerdings nur sofern dieser Lösungsweg schon anderswo gewählt wurde. Wirklich neue Lösungsansätze kann eine KI qua ihrer technischen Natur nicht bewerten. Wir sollen im Studium bekanntlich lernen, das Recht "mit Verständnis erfassen und anwenden" zu können (§16 I 3 JAPO), dazu gehört auch die Entwicklung neuer Lösungswege und auf Korrekturseite das Honorieren eigenständiger Gedanken, auch wenn sie so noch nicht in der Literatur oder in der Rechtsprechung vertreten wurden. Genau das kann nur ein*e menschliche*r Korrektor*in leisten, nicht aber eine KI. Klausuren von KI korrigieren zu lassen, kann also nicht dazu führen, bessere Jurist*innen auszubilden, sondern primär Jurist*innen, die Meinungsstreits auswendig lernen und diese unkritisch wiedergeben.

Was außerdem bei einer **KI-Korrektur** logischerweise fehlt, ist die **Menschlichkeit**. Aufmunternde Worte einer KI haben nicht den gleichen Effekt wie aufmunternde Worte eines **Menschen**. Ebenso kann die KI nicht die Autorität in Anspruch nehmen, die ein juristisch ausgebildeter **Mensch** für sich in Anspruch nehmen kann. Das mindert den Lerneffekt und die Motivation im Klausurenkurs.

DeepWrite Korrektur:

- nennt Fehler, die keine sind, zB: ich schreibe "materielle Präklusionsfrist", er sagt mir das ist falsch und es muss heißen "prozessuale Frist, an deren Versäumung sich materiell-rechtliche

Folgen anschließen" / sagt mir § 275 I Alt. 2 BGB ist falsch zitiert, weil es bei § 275 I BGB keine Alternativen gibt

- hebt Aspekte, die nicht in der Lösung stehen als falsch und unnötig hervor (zB Hinweis darauf, dass A-GbR rechtsfähig ist und somit Anspruch geltend machen kann)

- abstrakt gehalten und nicht direkt mit eigener Klausur vergleichbar

- weist auf jeden kleinen Rechtschreibfehler hin

- sagt, dass es "besonders schwer wiegt", dass der Anspruch aus §§ 346 IV, 280 I, III, 283 BGB nicht gesehen und geprüft wurde und in der Lösung steht, dass dies nur überdurchschnittliche Bearbeiter prüfen werden

Rechtsdidaktik:

- ist unnötig lang gehalten dadurch, dass viele Aspekte einfach wiederholt werden (teilweise 4-5 Kästchen lang die gleiche Bemerkung)

- Schlussvotum etwas knapp gehalten und nicht viel erklärt, was falsch gemacht wurde

Klausurenkiste:

- in der Lösung steht, es ist nicht zu beanstanden, wenn auf § 346 II 1 Nr. 3 Hs. 2 BGB nicht eingegangen wird und es wird trotzdem beanstandet
 - könnte noch etwas besser erklären beim beanstanden und nicht nur sagen, dass es nicht gesehen wurde
-

Menschliche Korrektur (der eigenen Klausur) zeichnete sich durch wiederholte Korrekturfehler aus. Auch die (wenigen) Anmerkungen im Fließtext sind stellenweise falsch und widersprechen zentral der Musterlösung.

DeepWrite Korrektur wirkt aufgrund der Loslösung vom Fließtext unübersichtlich und teilweise schwer nachvollziehbar

Klausurenkiste überhäuft mit Bemerkungen

Menschliche Korrektur enthielt auf 12 Seiten lediglich 6 (wenig hilfreiche) Anmerkungen und kein wirkliches Abschlussvotum.

- **DeepWrite**: Bewertung ist sehr unübersichtlich —> besser wäre es, wenn diese Bemerkungen wie bei einer **menschlichen Korrektur** am geschriebenen Text im Korrekturrand stehen
 - **Klausurenkiste**: habe hier das Gefühl, dass die KI nicht so gut mit alternativen Lösungswegen umgehen kann wie die anderen KIs; Fehler, die **DeepWrite** und Institut gefunden haben, hat die **Klausurenkiste** nicht gefunden
 - Institut: viel zu viel und unübersichtlich
 - **Menschliche Korrektur**: fast keine Randbemerkungen außer Hacken; Schlussvotum (geschweige denn Verbesserungsvorschläge) fehlt komplett; steht keine Punktzahl unter der Klausur (ich muss selber in der Punktetabelle meine Punktzahl "suchen")
-

DeepWrite ist einfach zu lang, langfristig würde ich es nicht schaffen, immer alles nachzuarbeiten. Damit möchte ich aber keine Kritik an der Qualität der KI äußern, das ist einfach meine ehrliche Rückmeldung bzgl. einer effizienten Eingliederung dieser Korrektur in meinen realistischen Lernplan

Die **menschliche Korrektur** ist oft zu knapp und dadurch weniger hilfreich.

Die **KI-Korrekturen** sind bisher viel zu lange in ihren Ausführungen (insb. Rechtsdidaktik) und teilweise erzeugen sie Slop, der zwar auf den ersten Blick plausibel klingt, aber keiner professionellen Revision standhält.

Insgesamt fehlt mir bei der KI vor allem die Fähigkeit des Abwägens und die Möglichkeit, auch Punkte abseits der Lösungsskizze zu honorieren.

Mensch: ob die Korrekturen der **menschlichen Korrektoren** helfen, ich stark vom Korrektur abhängig. In dieser Klausur beispielsweise steht neben meiner 20-seitigen Lösung 7 Anmerkungen, davon 2x, dass eine Norm fehlt, 3x "knapp" oder "sehr knapp", 2x "bitte kürzer" - das "Votum" bestand aus einem abgearbeiteten Rohpunkteschema, ansonsten sind meine Korrektur nur Haken beigelegt, die zu meiner sehr mittelmäßigen Note ja gar nicht passen können. Bei Korrekturen solcher Art ist es schlichtweg unmöglich, die eigenen Fehler ausfindig zu machen und zu kategorisieren, sodass man sich auch gleich das Geld sparen könnte und die Korrektur den Studis selbst überlassen kann. So etwas frustriert einerseits weil mich das überhaupt nicht weiterbringt und sich die ganze Arbeit in diesem Studium gefühlt nicht auszahlt, andererseits verunsichert mich dies enorm weil ich nie weiß, ob die mir gegebene Note überhaupt so stimmen kann. Manche Korrektoren schreiben zudem schlichtweg Falsches an den Rand.

Klausurenkiste: Die **Klausurenkiste** fand ich unter den KI sehr starr. Ich habe den Eindruck gehabt, dass die KI sehr unflexibel korrigiert hat und zB Bearbeitungsansätze schnell negativ gewürdigt hat, wenn diese so nicht in der Lösungsskizze standen. Beispielsweise hatte ich den § 377 HGB angeprüft - auch mit hinreichender Begründung abgelehnt - dies wurde aber völlig falsch von der KI interpretiert und mit -1 Punkte versehen und der Gedankengang überhaupt nicht gewürdigt. Methodisch kommentiert die KI überhaupt nicht, was ich insoweit stark bemängelte, als dass einem im gesamten Studium zuvor eine richtige Methodik nie richtig beigebracht wird obwohl diese von hoher Relevanz für eine gute Bearbeitung ist.

DeepWrite: Diese Korrektur habe ich mir ehrlich gestanden überhaupt nicht zu Ende durchgelesen. Es ist ja ganz nett, dass man eine lange Korrektur bekommt aber einerseits finde ich die Korrekturen am Klausurenrand übersichtlicher, andererseits werde ich nie die Zeit aufbringen können und wollen, um mir für eine einzelne Klausur ein Feedback durchzulesen, dass ca. 3 mal so lang wie meine Bearbeitung ist. Die Tipps fand ich zT genauso unpräzise, nicht lösungsorientiert und nicht verständlich, so wie die KI meine Bearbeitung fand (zB "Formulierungen wie xy wirken nicht durchgehend klausurentypisch" oder "Ihre Lösung ist überwiegend in ganzen Sätzen formuliert"). Vieles wiederholt sich außerdem und es fehlt auch schlichtweg an positivem Feedback - das gilt aber ganz allgemein für die sämtlichen Korrekturen. Natürlich ist Sinn und Zweck einer Korrektur, dass man aus den eigenen Fehlern lernt, allerdings ist es jetzt rein psychologisch auch keine bahnbrechende Erkenntnis, dass zusätzlich positives Feedback zur Motivation anregt.

KI des Rechtsinstituts: hier zu bemängeln ist wieder die fehlende bzw. zu geringen methodische Auseinandersetzung mit meiner Bearbeitung.

Sonstiges: Ich würde mir insgesamt wünschen, wenn alternative Gedankengänge und Lösungsansätze, die so in der Lösung nicht zu finden, mehr Würdigung finden würden, wenn diese berechtigt und dem Grundsatz nach gut und richtig sind. Gleichzeitig würde ich mir aber auch Kommentierungen wünschen, die aufzeigen, wo ich vielleicht stark angefangen habe, wo ich aber dann falsch abgebogen ist. Grundsätzlich bin ich keine Freundin von KIs, sehe aber auch, dass darauf vA aus finanziellen Mitteln abgestellt wird. Sollte sich für eine KI entschieden werden, dann würde ich mir aber auch

wünschen, dass rein die des Rechtsinstituts verwendet wird und kontinuierlich an der Verbesserung gearbeitet wird.

DeepWrite: sehr unübersichtlich; Kommentare besser an den Rand oder in Tabelle wie von Institut

Institut: dadurch dass jeder Satz aufgeteilt wird, aber dennoch ein Kommentar pro Absatz erstellt wird, sehr viele Doppelungen; besser pro Absatz bzw. Prüfungspunkt kommentieren

Mensch: Layout der Kommentare (schwerer einsehbar)

Klausurenkiste: Randbemerkungen gehen meiner Wahrnehmung nach nicht über ein Abschreiben des Rohpunkteschemas hinaus; positive Randbemerkungen sind so ausführlich, dass die Länge abschreckt, bieten aber keinen Mehrwert.

DeepWrite: Wegen des Layouts der Korrektur ist sie schwerer nachvollziehbar; Feedback wie „Ihre Antwort ist überwiegend in ganzen Sätzen formuliert und dadurch grundsätzlich gut lesbar.“ halte ich für überflüssig (Beispiel für Kürzungspotential); die Länge der Korrektur wirkt zu Beginn abschreckend.

Institut: Das Layout mit dem zusammengestauchten Klausurtext ohne Absätze ist unübersichtlich;

Insgesamt waren mE zu wenige Randbemerkungen vorhanden (sechs auf 24 Seiten). Haken wurden nicht gesetzt, mir war stellenweise nicht ganz klar, was ich hätte verbessern können.

Mensch (allgemein zu dieser Art und Weise und Form von "Korrektur", NICHT auf den speziellen Korrektor bezogen!!): (Viel) zu wenig Randbemerkungen neben den entsprechenden Textpassagen und - wenn denn vorhanden - zu kurz oder schwer/kaum verständlich. Hinterher nur mit einem Korrekturbogen Punkte vergeben mit überhaupt nicht hilfreicher Begründung, sondern überhaupt nur zu nennen was wohl zu prüfen gewesen wäre und dann nicht einmal ein kurzes eigenes Votum zu schreiben, sondern nur "Die Bearbeitung ist mit ... Punkten zu bewerten" ist überhaupt kein hilfreiches und für viele Studierende ungenügendes und sehr frustrierendes Feedback was man sich auch sparen könnte. Denn trocken die Bearbeitung mit der Lösungsskizze vergleichen kann jeder Student auch selber. Worauf es vor allem ankommt ist ja gerade, dass man die Meinung eines erfahrenen Dritten bekommt, was man wo wie gut oder schlecht gemacht hat und warum und wie man es ggf. besser machen könnte und nicht nur "gut" oder "wieso" und nirgends eine richtige Begründung warum überhaupt.

KI: Zum Teil zitiert oder liegt die KI selber falsch, was dann sehr verwirrend ist. V.a.

DeepWrite finde ich auch wieder eher unübersichtlich bzgl. der "Randbemerkungen",

da es eben keine Randbemerkungen sind, sondern ein Text verfasst wird und man dann seine Bearbeitung daneben legen und erst suchen muss, welche Stelle der Bearbeitung gerade besprochen wird, anstatt direkt in/neben der Bearbeitung die Bemerkungen an den entsprechenden Stellen zu haben. Auch erkennt die KI z.T. alternative Möglichkeiten nicht oder erkennt manche (möglicherweise tatsächlich nicht 100% genaue) Formulierungen nicht und wertet sie als falsch oder nicht gesehen, obwohl sich aus dem Kontext ergibt, dass das Richtige gemeint ist und erkannt wurde.

Menschliche Korrektur:

- kaum Randbemerkungen bis auf wenig aussagekräftige "?" trotz nicht so guter Note
- große Diskrepanz zwischen dieser Note und allen KI-Noten
- keinerlei Feedback mit Tipps für die nächste Klausur und Verbesserungsvorschlägen
- eigentlich richtige Gedanken wurden als "fehlend" in der Bewertungstabelle eingetragen, obwohl sie nur mit anderen Worten aber sinngemäß entsprechend der Lösung im Gutachten bearbeitet wurden
- Fehler im Gutachten wurden nicht angestrichen (zB. Abstellen auf § 276 BGB in Aufgabe 2, Fristberechnung in Aufgabe 3; alle KI-Lösungen haben das erkannt und entsprechend verbessert)
- im Ganzen ist die Korrektur für mich völlig unbrauchbar und nicht zielführend

DeepWrite:

- etwas unhandlich, dass der Kommentar nicht direkt neben dem Gutachten steht, sondern als eigenes Dokument
-

Die **menschliche Korrektur** ist zu kurz. Allerdings sollte die **KI-Korrektur** Wiederholungen vermieden werden. Bei **DeepWrite** hat mir gestört, dass es keine Randbemerkungen gemacht hat.

d. Positives Feedback

DeepWrite: Exzellent nachvollziehbare Anmerkungen

KI des Instituts für Rechtsdidaktik: Dito

KI-Korrekturen sind sehr ausführlich und zeigen konkrete Fehler. Ferner ist ein Rohpunktesystem eine super Möglichkeit, die erhaltene Note nachzuvollziehen.

Sowohl positive als auch negative Aspekte des Gutachtens werden detailliert dargestellt und abgewogen, sodass das Endergebnis widerspruchslös nachvollziehbar ist.

Wenn die **menschlichen Korrektoren** Anmerkungen machen und Hinweise für die weitere Arbeit geben, sind diese "on point", treffen ziemlich genau das Problem. Gut finde ich die ausführliche Aufgliederung in der Korrektur des IRD, bei dem Stärken und Schwächen genannt werden, daneben aber auch Hinweise für bessere Ergebnisse gegeben werden.

Menschliche Korrektur: Grundsätzlich ist bei der **menschlichen Korrektur** der Bezug zum Einzelfall vorteilhafter als bei der **KI-Korrektur**. Der **menschliche Korrektor** kann unabhängig von der Lösungsskizze eine Einzelfallbewertung besser vornehmen und auf die Besonderheiten der einzelnen Bearbeitungen gezielter eingehen (Auch wenn ich mir hier bei meinem Korrektor nicht sicher bin xD)!

Klausurenkiste: Die Randbemerkungen sind deutlich besser als bei der **menschlichen Korrektur**!

DeepWrite: Die Arbeits- und Klausurtipps sind hier positiv hervorzuheben, sie sind sehr hilfreich und auf wenige wichtige Punkte beschränkt!

KI des Instituts für Rechtsdidaktik: Die Randbemerkungen sind hier am besten, das Schlussvotum ist knapp und klar nachvollziehbar! Insgesamt fand ich diese Korrektur am besten!

Die **KI-Korrekturen** sehen alle geschriebenen Punkte zuverlässig und lassen diese in die Bewertung einfließen und geben nützliche und motivierende Tipps.

Die **KI-Korrekturen** sind alle sehr ausführlich, das hilft total die eigenen Fehler zu erkennen und beim nächsten mal besser zu machen. Diese Korrekturen führen dazu, dass ich tatsächlich den Klausurenkurs wieder sinnvoll finde.

Klausurenkiste: Vom Layout der Randbemerkungen mit Abstand am übersichtlichsten und für meinen Geschmack gelungen.

DeepWrite: Das Stil-Feedback ist sehr hilfreich.

Institut für Rechtsdidaktik: Insgesamt von der inhaltlichen Qualität der Anmerkungen am überzeugendsten.

Klausurenkiste

- Angabe Rohpunkte direkt am jeweiligen Prüfungsort macht Bewertung übersichtlich und nachvollziehbar

-als einzige KI durchgehend in der Lage, laut Lösung vertretbare Auffassungen nicht als falsch zu markieren

DeepWrite: hilfreiche Anmerkungen zu in Lösung nicht zwingendem § 377 HGB

Institut für Rechtsdidaktik: schönes Gesamtfazit mit konkreten Vorschlägen

bei allen **KI-Korrekturen**

- unpräzise Normzitate werden markiert
 - KI ist höflicher und motivierender
-

KI des Instituts für Rechtsdidaktik

- sehr übersichtlich
- durch ausführliche Begründungen direkt im Rohpunkteschema sehr gute Nachvollziehbarkeit der Note (Anmerkungen und Begründungen doppeln sich zwar, allerdings so beste Nachvollziehbarkeit der Note)

KI DeepWrite

- extra Abschnitt bzw. viele Anmerkungen zu Stil/Schwerpunktsetzung
-

Inhaltliche Fehler werden in der **KI-Korrektur** aufgezeigt und erklärt. Das hilft viel mehr als die schlechte und nicht aufschlussreiche Korrektur durch einen **Menschen**.

Mensch:

- ein Rohpunkteschema ist vorhanden

KI des Instituts für Rechtsdidaktik:

- die zusätzlich Tabelle mit den erreichten Rohpunkten pro Aufgabe hilft dabei, sich schnell einen Überblick über die gelungenen und die nicht gelungenen Teile zu verschaffen

DeepWrite:

- gute Nachvollziehbarkeit und hilfreiche Anmerkungen

Klausurenkiste (persönlicher Favorit):

- es wird direkt mit der Bearbeitung selbst an den relevanten Stellen gearbeitet
 - durch die farblichen Unterschied ist schnell erkennbar ob ein Hinweis, ein Fehler oder die an der Stelle korrekte Bearbeitung vorliegt
-

DeepWrite: Erklärt nicht nur, was falsch ist, sondern auch warum und wie man es besser macht – teilweise mit fertig ausformulierten Beispielsätzen (z.B. Obersatz zu § 174

BGB). Die Punktevergabe war am realistischsten und als einzige KI kam sie auf dieselbe Note wie die Handkorrektur (3 Punkte). Der Rat „nach jedem Anspruch fragen: entstanden – erloschen – durchsetzbar“ war für mich der hilfreichste Hinweis überhaupt.

KI des Instituts für Rechtsdidaktik: Zum Nachvollziehen der eigenen Fehler am angenehmsten, weil jede Anmerkung direkt an der jeweiligen Textstelle steht und jeder Punkt in der Rohpunktetabelle einzeln begründet wird. Inhaltlich deckt sich alles mit der Musterlösung; die Stärken/Schwächen/Empfehlungen am Ende sind kurz und treffend.

Klausurenkiste: Fühlt sich vom Format her am ehesten wie eine echte Examenskorrektur an – Randbemerkungen direkt am Text, Erwartungshorizont und Votum mit Ausblick. Bei vielen Einzelposten ziemlich treffsicher.

Alle drei: Haben meine eigentlichen Baustellen (Rückabwicklung/Aufrechnung, Wertsatz, § 174 BGB analog) zuverlässig und übereinstimmend erkannt.

KI des Instituts für Rechtsdidaktik: sehr übersichtlich und schön strukturiert

Menschliche Korrektur: die **menschliche Korrektur** erkennt insbesondere die fehlende Analogie bei § 441 BGB und kann bestimmte Lösungsansätze noch gezielter erkennen

In vielen Punkten (Genauigkeit/Ausführlichkeit der Anmerkungen und erkennbarer konkreter Bezug zu einzelnen Sätzen; mehr [positives] Feedback zu Punkten, die richtig sind) besser als viele der **menschlichen Korrekturkräfte**, jedoch sagt das mehr über die niedrige Qualität der **menschlichen Korrekturen** aus als über die Qualität der **KI-Korrekturen**.

Die Schnelligkeit der Korrekturen (in der Zukunft) ist ein positiver Aspekt, wobei eine Besprechung in der Folgewoche nicht unbedingt erforderlich ist, da es auch Vorteile hat, sich ein paar Wochen nach der Klausur nochmal in die dort behandelten Probleme eindenken zu müssen.

Die KI liest die Klausuren grundsätzlich genauer, das heißt es kommt tendenziell seltener vor, dass negative Aspekte bemängelt werden, die sich bei genauerem Lesen durch die Korrekturkraft erledigt hätten. Bei der KI besteht allerdings eher die Gefahr, dass sie das Gelesene nicht richtig versteht.

Die Darstellung der **Klausurenkiste** (Randbemerkungen mit den vergebenen Punkten direkt dahinter) war für die Nacharbeit am besten zu handhaben. Die **KI des Instituts für Rechtsdidaktik** war für methodische Anmerkungen noch am besten, aber auch nicht uneingeschränkt brauchbar.

- **KI des Instituts für Rechtsdidaktik** Schlussvotum ist außerordentlich aussagekräftig und hilfreich!

Vor allem die KI Korrektur von dem **Institut für Rechtsdidaktik** ist mir sehr positiv aufgefallen. Die Korrektur war sehr ausführlich und vor allem waren hilfreiche Anmerkungen dabei.

Alle **KI-Korrekturen** erklären genau, wo Stärken und wo Schwächen liegen und geben Hinweise und Verbesserungsvorschläge für künftige Klausuren.

- **Mensch**: Rohpunkte-Schema angehängt (ich hätte lieber das Rohpunkteschema selbst, als eine Neuformulierung dessen von einer KI)
 - If Rechtsdidaktik: formuliert positiv und freundlich, was motiviert. Hierzu: im Klausurenportal ist angegeben, meine IfR-Korrektur hätte 4 P. ergeben, es stehen jedoch 8 P. auf dem zugehörigen Pdf (in den anderen **KI-Korrekturen** hab ich 3 und 4 Punkte, laut Korrektor 8 P.)
-

Was ich grds. an der **KI-Korrektur** besser finde ist, dass es ein Votum gibt, dass sich tatsächlich mit der eigenen Klausur beschäftigt und nicht Copy-Paste vorgeschriebene Voten einfügt oder ggf. gar keines abgibt.

- Externe Korrektur: ./.
 - **Klausurenkiste**: M.E. ist das Layout dieser Korrektur (Anmerkungen direkt am Rand meiner Klausur und ein Schlussvotum am Ende) am besten.
 - Deepwrite: Dieses KI Modell gibt das beste Feedback von allen; insb. das Feedback zum Stil/ zu konkreten Formulierungsvorschlägen ist sehr gut
 - Institut: Das komprimierte Feedback am Ende der Klausur ist besonders übersichtlich und gut nachvollziehbar
-

Die **KI-Korrekturen** haben mir tatsächlich persönlich sehr geholfen. Fehler werden gut erklärt und ich kann teilweise bei Sachen, bei welchen ich sonst Schwierigkeiten habe, eine ganz andere Sichtweise und mehr Verständnis warum genau und inwiefern es nicht richtig ist. Allerdings vertraut man natürlich der **menschlichen Korrektur** mehr. Eine KI sieht auch so gut wie alles. **Menschen** übersehen wenigstens manchmal Rechtschreibfehler. Und der Korrektor hat nicht viel Zeit pro Klausur, deshalb habe ich auch da großes Verständnis.

Die **KI-Korrektur** kann die bei der **menschlichen Korrektur** oft fehlenden Anmerkungen leisten. An dieser Stelle fand ich persönlich das Konzept des **Institut für Rechtsdidaktik** am besten, das Anmerkungen direkt am Rand vornimmt (inkl. konstruktiver

Kritik, aber auch positiven Anmerkungen, die für die Motivation in der Examensvorbereitung ebenfalls wichtig sind). Die Klausurtipps, die alle **KI-Korrekturen** generieren, sind hilfreich und geben Orientierung bei der weiteren Vorbereitung. Tipps dieser Art finden sich bei der **menschlichen Korrektur** leider nur alle paar Wochen bei einzelnen Korrektoren.

Positiv an der **KI-Korrektur** ist auch, dass man ein "konstanteres" Feedback bekommen würde, also Woche für Woche von mehr oder weniger derselben Stelle korrigiert wird. Damit könnte man die eigene Leistung eher einordnen als bei den doch sehr unterschiedlich strengen Korrektoren, die teilweise entweder Punkte oder auch Fehler übersehen.

Institut für Rechtsdidaktik: es ist sehr hilfreich, dass wirklich zu jedem Absatz eine Rückmeldung geschrieben wird. Dadurch lässt sich die Korrektur toll konkret nachvollziehen

DeepWrite: Die Hinweise sind klar und klausurenorientiert formuliert, sodass klar ist, wie man sie in Zukunft umsetzen kann und was man besser machen kann

Da es bei dieser Klausur ein Rohpunkteschema gab, war die Bewertung bei allen Korrekturen nachvollziehbar.

DeepWrite:

- es wird verständlich darauf eingegangen, was nicht geprüft wurde und woran man es hätte erkennen können (sehr gut)

- weist auf falsche Zitation hin

Rechtsdidaktik:

- Bezug zur geschriebenen Klausur

Klausurenkiste:

- gute, verständliche Randbemerkungen

Inhaltliche Bewertung der Deep-Write Korrektur sowie der **KI des Instituts für Rechtsdidaktik** überzeugt. Auch die optische Darstellung der Anmerkungen sowie des Rohpunkteschematas am Ende überzeugt beim Institut, ebenso wie die abschließende Feststellung der Stärken / Schwächen / Empfehlungen.

Bestärkende Anmerkungen / Hervorhebung der Stärken fällt positiv auf - jedoch gerne nur an vereinzelten Stellen oder gebündelt am Ende, eine Überhäufung mit (inhaltsleeren) positiv-Feststellungen fällt eher negativ auf / erschwert die Nacharbeitung

- bei allen KI-Modellen: viel ausführlichere Korrektur, aus der man was für kommende Klausuren mitnimmt; die Möglichkeit, dass abstrakte Tipps zur Verbesserung gegeben werden, ist toll

- **DeepWrite**: Ausführungen zu „Feedback zum Inhalt“ und „Feedback zum Stil“ sehr ausführlich und hilfreich; Gesamtfeedback sehr hilfreich

Mir gefällt die Struktur der Korrektur meiner Textpassagen in Tabellenform besser, als die **DeepWrite**. Dadurch hat man die eigene Bearbeitung mehr vor Augen und kann noch besser die Korrektur verstehen.

Dogmatische Einordnungen der **menschlichen Korrekturen** sind überwiegend prägnant und korrekt.

Die Gefahr einer unbrauchbaren Korrektur ist bei einer KI-Lösung wohl niedriger als bei einer menschlichen.

Ich finde KI spannend, aber bei diesem Projekt habe ich erneut gemerkt, dass die KI Quantität zwar sehr gut kann, aber in einem juristischen Gutachten, welches nicht stringend und streng vorhersehbar verläuft, kommt es auch auf die einzelnen Auseinandersetzungen mit vertretbaren Argumenten und Lösungen an. Hieran scheitern die Programme offenkundig noch immer regelmäßig.

Allgemeines: Gut an den **KI-Korrekturen** fand ich im Allgemeinen, dass ich glaube ich so ziemlich zum ersten Mal kontinuierlich inhaltlich brauchbare Randbemerkungen erhalten habe. Mir wurden Fehler angestrichen, die ich wohl seit Jahren mache und dessen Vermeidung einfach auch sehr relevant ist, mir dies aber noch kein einziges Mal von einem **menschlichen Korrektur** angestrichen wurde.

KI des Rechtsinstituts: Die Anmerkungen konnten mir mit Abstand am besten helfen - zur Abwechslung wurde auch mal hervorgehoben, was man gut gemacht hat. Das Rohpunkteschema fand ich abschließend insoweit nett, als dass man sehen konnte, wie sich die Note auseinandersetzt und auch mit welcher Begründung.

Im Gegensatz zur **menschlichen Korrektur** wird mehr inhaltlich auf die Klausur eingegangen und Tipps gegeben

Mensch: Richtige Punkte werden ohne Schnickschnack mit „zutreffend.“ als solche markiert.

DeepWrite: Unterscheidung zwischen Feedback zum Inhalt und Feedback zum Stil.

Institut: Farbliche Hervorhebung besonders relevanter Anmerkungen

Im Gegensatz zu anderen Korrekturen war das Fazit ausführlich.

Mensch: Die einzig positiven Aspekte finde ich in den Korrekturanmerkungen von Prof. Harnos und Prof. Kuhn, da diese tatsächlich "RAND"bemerkungen sind und auch erklären wieso etwas falsch oder ungenau ist und man einfach ein ausführliches Feedback eines erfahrenen Dritten bekommt und nicht nur "das ist falsch", "das ist ungenau", "hier gibt es soundso viele Punkte" aber ohne jegliche Begründung, und dann überhaupt ein eigenes Votum, was ja gerade der Grund ist, warum man die Klausur geschrieben hat.

KI: Es ist v.a. bei **Klausurenkiste** und der InstitutsKI übersichtlich gestaltet, mit den Randbemerkungen dort, wo sie auch zur Bearbeitung gehören, damit man den Teil, worauf sie sich beziehen gleich daneben hat. Auch gibt es deutlich mehr Bemerkungen als beim menschlichen Korrektor (was aber vermutlich v.a. an dem programmierten KI Korrekturstil liegt), wobei natürlich trotzdem gilt, dass Quantität nicht gleich Qualität ist. Auch gibt es am Ende ein ausführliches Votum und Feedback. Zwar hat die InstitutsKI auch einen Bewertungsbogen, in diesem wird aber im Gegensatz zum menschlichen Korrektor nicht nur genannt, was zu prüfen gewesen wäre, sondern auch warum dann diese von der angegebenen Punktzahl erreicht wurde. Die Aufteilung in "Inhalt", "Stil" und solche Punkte finde ich recht interessant, da es bei mir z.B. oft nicht an beidem, sondern mal am Stil, mal am Inhalt hapert. Auch das Nennen von "positiven" und "negativen" Punkten bzw "Stärken" und "Schwächen" und so finde ich gut, da - sofern qualitativ verwertbar - hier gut aufgelistet ist, woran es v.a. gelegen haben könnte bzw. was man schon gut kann. Dadurch kann man besser auf seine Schwächen eingehen.

Klausurenkiste und KI des Instituts für Rechtsdidaktik:

- sehr hilfreiche Anmerkungen
- sehr übersichtlich direkt neben dem Gutachtentext

Klausurenkiste:

- Tipps und Feedback für die nächste Probeklausur unter "Ausblick"

KI des Instituts für Rechtsdidaktik:

- nützliche Aufteilung in Stärken, Schwächen und Empfehlungen

DeepWrite:

- neben vielen Kritikpunkten sind konkrete Zitate aus dem Gutachten aufgeführt, so dass sich der Fehler direkt anhand eines Beispiels nachvollziehen lässt
- ausführliches und hilfreiches End-Feedback und besondere Hinweise bezüglich "wiederkehrenden methodischen Problemen"

Mir gefällt sehr bei der **KI-Korrektur**, dass sie so ausführlich ist. Mir gefällt, dass die **KI-Korrektur** inhaltlich weniger Fehler macht als die **menschliche Korrektur**. Ich finde ihre Notenbildung fairer.

e. Sonstige Anmerkungen

Gerade an Stellen wo die Bearbeitung von der Musterlösung abweicht aber dennoch vertretbar und gerade auch juristisch gut geschrieben ist scheitert teilweise die KI aber auch oft ein **menschlicher Korrektor**. Ich würde sagen die **KI-Korrekturen** wirken dennoch neutraler und sind viel ausführlicher und genauer während die **menschlichen Korrektoren** oft nur „???“ oder „gut“ schreiben. Insgesamt waren die **KI-Korrekturen** überraschend gut. Wenig nachvollziehbar war allerdings die KI welche Punkte abgezogen hat für gewisse Fehler??

Für den ersten Korrektur Durchgang und die Note würde ich ausschließlich einen **Menschen** nehmen (ohne Bias) und danach erst KI für eine ausführliche Anmerkung heranziehen. KI ist super was den Mehrwert der Korrektur angeht. Ich denke nur, dass ein **Mensch** mit Ermessen ggfs. an der ein oder anderen Stelle großzügiger gewesen wäre. So lässt sich nicht nachvollziehen, wie eine Korrektur davon spricht, dass einige Stellen mit zu den besten Bearbeitungen des Kurses zählen, die Note aber eindeutig aussagen soll, dass man unterdurchschnittlich schlecht wäre. V.A. ergibt sich nicht, wie Punkte die man laut Lösungsskizze nur von besseren Bearbeitungen verlangen kann, für das Bestehen relevant wären.

Im Ergebnis wäre ich für eine hybride Lösung. Eine reine KI Korrektur würde ich nach dem momentanen technischen Stand ablehnen.

Ich habe in dieser Klausur leider eine schlechte Erfahrung mit der **menschlichen Korrektur** gemacht. Ich muss jedoch anmerken, dass eine solche, meiner Meinung nach schlechte **menschliche Korrektur**, erst ein zweites Mal vorgekommen ist (ich habe mittlerweile knapp 50 Klausuren geschrieben). Daher würde ich vorschlagen weiterhin mit **menschlichen Korrektoren** zu arbeiten und, wie hier, eine zusätzliche **KI-Korrektur** anzubieten.

Des Weiteren wäre es wünschenswert von den Professoren ein Feedback zu erhalten, wie aussagekräftig die **KI-Korrekturen** sind, inkl. ihrer Notenvergabe.

Schade ist wirklich, dass man teilweise nur 4 Randbemerkungen in einer ganzen Klausur bekommt. Diese sind meistens aber auch nicht ausführlich, sondern beinhalten nur Angaben wie "Systematik nicht nachvollziehbar". Warum ist die Systematik aber nicht nachvollziehbar? Als Student generiert man hieraus keinen Mehrwert. Genauso könnte ich die Musterlösung neben meine Lösung legen und selbst vergleichen, was

ich richtig und was ich falsch gemacht habe. Auch ist die Notengebung überhaupt nicht nachvollziehbar, wie dieses Beispiel nun bestens zeigt. Wie kann es denn sein, dass 3 **KI-Systeme** die Leistung zwischen 5 und 7 Punkten bewerten, der **menschliche Korrektor** mich aber durchfallen lässt? Es wäre schön, wenn die Note begründet werden würde, gerade wenn es doch darum geht, ob man die Klausur besteht oder nicht.

Am überzeugendsten wäre m.E. fast ein Zusammenschnitt der drei Wrapper: ein Vorgehen, das die inhaltliche Qualität der Anmerkungen des Instituts, das Layout der KlausurenKiste und die ausführliche Stilanalyse des **DeepWrite**-Projekts vereint.

DeepWrite ist völlig unübersichtlich, das bitte nicht verwenden.

Die KI **Klausurenkiste** und die des Instituts für Rechtsdidaktik sind hilfreich, wenn auch sehr lang und mit merkbaren Fehlern. Ich würde aber behaupten, dass die Hälfte der **menschlichen Korrektoren** besser ist. Die KIs sind gerade mit abweichenden Ansichten überfordert.

Wenn ich mich unbedingt entscheiden muss, würde ich die **KI-Korrekturen** vermutlich einem **Menschen** vorziehen, weil man hier wenigstens nachvollziehen kann, an welchen Stellen die KI Fehler macht (z.B. Bewertung vertretbares Ergebnis als falsch, Orientierung an Rohpunkten). Bei den schlechteren Korrektoren gibt es kaum Anmerkungen, was noch weniger hilfreich ist als die falschen Anmerkungen der KI.

Ich würde mich sehr freuen, wenn auch Personen, die im März 27 schreiben wollen, schon die Möglichkeit bekämen, noch **KI-Korrekturen** zu erhalten!

Hinweis1: die von der KI vergebenen Noten werden im Portal selbst teilweise nicht korrekt angezeigt

Hinweis 2: **KI des Instituts für Rechtsdidaktik** schreibt an mehreren Sachen wie "Als einzige Bearbeitung erkennt sie, dass ...". Das erscheint mir nicht ganz plausibel...

- es müsste technisch sichergestellt werden, dass alternative Lösungsansätze auch Beachtung finden, auch wenn diese gerade nicht in der Form von der Lösungsskizze gesehen worden sind

- falls möglich sollten pro Durchlauf stichprobenartige Kontrollen stattfinden um so die Richtigkeit und gewünscht Qualität zu garantieren

Die **menschliche Korrektur** fällt in diesem Beispiel allerdings auch negativer auf als z.B. iRd Probeexamens. Leider fehlt es hier an einem Schlussvotum und an konkreten

Tipps (s.o.), sodass diese Korrektur hier deutlich schlechter abschneidet als andere **menschliche Korrekturen**.

Insgesamt ist eine **KI-Korrektur** allerdings wohl deutlich ausführlicher und insofern auch hilfreicher.

Vorschlag für den Einsatz der **KI-Korrektur** im Klausurenkurs: Die **KI-Korrekturen** sind relativ gleichwertig zu einigen (!) Korrekturen der **menschlichen Korrekturkräfte** im Klausurenkurs, weshalb ich es legitim fände, wenn teilweise (aber nicht komplett!) auf **KI-Korrekturen** im Examenskurs umgestellt wird. Als Kompromiss würde ich daher vorschlagen, ein Modell einzuführen, bei dem nur jede 2. Klausur von einem **Menschen** korrigiert wird. So könnte man auch die guten Korrekturkräfte (die gibt es im Examenskurs auch und durch sie bekommt man wirklich hilfreiches Feedback, das das Feedback der KI qualitativ um Längen übertrifft!) weiterhin im Examenskurs einsetzen und wäre gleichzeitig nicht mehr (so sehr) auf die schlechteren Korrekturkräfte angewiesen. Es ist in jedem Fall wichtig, in der Examensvorbereitung von **Menschen** korrigiert zu werden, wenn man im Examen selbst auch von **Menschen** korrigiert wird! Daher sollte auf keinen Fall auf eine ausschließliche **KI-Korrektur** umgestellt werden. Bei einer ausschließlichen Nutzung von KI sähe ich die Gefahr, dass die **KI-Korrekturen** (auch qua ihrer Länge) von neueren Studierenden im Klausurenkurs nicht richtig eingeordnet werden können und der KI zu schnell "blind" vertraut wird. Bei den Korrekturen durch die **menschlichen Korrekturen** ist die teilweise niedrige Qualität bereits bekannt und man kommt auch ohne das Vorwissen einfacher darauf, dass die Korrektur eines **Menschen**, die aus ein paar Haken und 5 Randbemerkungen besteht, nicht die aussagekräftigste Bewertung ist.

Die **menschliche Korrektur** hat meiner Meinung nach unsauber gearbeitet. Es fehlen Begründungen und Aspekte die im Punkteschema als fehlend gekennzeichnet werden sind in manchen Fällen in Wirklichkeit tatsächlich vorhanden. Es gibt auch keinerlei Feedback oder Begründung für die Bewertung.

Die KI hat nachvollziehbar begründet und Feedback gegeben.

Deshalb finde ich eine KI Korrektur sehr sinnvoll. Nur mit KI zu korrigieren ist aber definitiv problematisch. Am Ende sollte ein **Mensch** zumindest einmal die Arbeit durchlesen, um zu überprüfen, ob alles stimmig ist und zusammenpasst, denn auch die KI kann Fehler machen.

Die KI wird sehr gehyped. In meinen Augen liegt der Vorteil vor allem darin, dass durch KI-Verwendung der Prüfungsmaßstab rationalisiert werden muss und es detaillierte Rohpunkteschemata gibt. Es würde mehr Vertrauen schaffen, dann einfach die Rohpunkteschemata direkt herauszugeben.

Klausurtaktische und methodische Tipps kommen durch die KI kaum. Ich finde Kritik von der KI ferner schwerer zu akzeptieren.

Der wirkliche Mehrwert der KI ist im Vergleich zu einem Rohpunkteschema in meinen Augen v. a., dass sie mir die ca. 15 Min. Zeit abnimmt, die Connection zwischen den Punkten in meinem Fall und dem Rohpunkteschema zu finden. Das ist sehr hilfreich.

Ich finde es trotzdem weiterhin wichtig, dass unsere Bearbeitungen weiterhin von **Menschen** korrigiert werden. Am besten unter der Prämisse, dass sie evtl. nicht zu sehr inhaltlich auf die Fehler eingehen, die sehe ich ja, wenn ich mit der Lösung vergleiche oder in der Besprechung sitze. Ich fände es viel hilfreicher, wenn tatsächlich mehr auf den Eindruck eingegangen wird, positiv wie negativ: Schwerpunktsetzung, Argumentation, Gutachtenstil, Sprache, Zitation, Gesamteindruck. Wenn ich ein Votum lese, in dem diese Dinge (wenn auch nur eines) angesprochen werden, ob negativ oder positiv, kann ich damit viel mehr anfangen. Ich denke, dass wenn sich die **menschlichen Korrekturen** an so etwas halten würde, würde sich der Mehrwert der Korrekturen auch ohne KI erheblich steigern.

Bei den **KI-Korrekturen** ist tatsächlich auch nicht das inhaltliche, das auf das ich mehr achte, sondern das Eingehen auf meine Arbeit unabhängig von der Richtigkeit zur Lösung. Ich habe Zweifel, dass eine KI sich so von Sprache, Argumentation und dem Gesamteindruck lenken und beeindrucken lassen kann, wie ein **Mensch**. Zudem zweifle ich auch sehr an den inhaltlichen Bemerkungen der KI. Über die letzten Jahre konnte ich auch eigene Erfahrungen mit KI in der juristischen Praxis sammeln und ich habe gemerkt, dass ich dadurch kein Vertrauen in die inhaltliche Richtigkeit einer juristischen Aussage einer KI habe. Natürlich kann das (wohl?) durch Prompts oÄ gesteuert werden, jedoch meinte Prof. Harnos gestern, dass eine KI wohl auch bei der Korrektur ins Internet gegangen ist und selbst etwas „recherchiert“ hat. Das finde ich super beängstigend, weil ich weiß, dass sich KIs im juristischen Bereich gerne mal Normen oder sogar Gerichtsentscheidungen aus der Nase ziehen.

Daher merke ich, dass ich kein Vertrauen in die KI habe und somit in das Votum habe. Dann ist es mir lieber, dass ein **Mensch** meine Bearbeitung durchliest und dann evtl. aussteigt, weil ich von der Lösung abweiche, aber wenigstens weiß ich, dass inhaltlich nichts falsches bei mir am Rand steht. Und dann lese ich lieber selbst in der Lösung nach, an welcher Stelle ich falsch abgebogen bin, anstatt mir diesen Schritt auch noch von der KI abnehmen zu lassen.

Ich denke auch, dass der Eindruck, den die KI vermittelt gerade einfach nur trügt, dadurch dass man selbst ein Rohpunkte-Schema hat. Das zu haben, finde ich, hilft mir selbst noch einmal super viel weiter, die Klausur und die Schwerpunktsetzung nachzuvollziehen.

Des Weiteren sehe ich die Gefahr, dass man sich von den „goodies“ der KI einlullen lässt. Ich finde es zur Nachbereitung super wichtig, mich inhaltlich mit meiner Bearbeitung und der Lösung auseinanderzusetzen, zu vergleichen und selbst noch einmal darüber nachzudenken, wieso ich damals so abgebogen bin und warum die Lösung das

anders macht. Ich habe Bedenken, dass wenn die KI in Zukunft genau diesen Schritt übernimmt, dass man sich dadurch weniger mit seiner eigenen Bearbeitung auseinandersetzt, weil das die KI ja bereits übernommen hat und sich nur noch mit den Anmerkungen oder Vorschlägen der KI beschäftigt. Natürlich ist eine gute Anmerkung auch toll, aber ich denke ein wichtiger Lernpunkt ist auch, sich dogmatisch und denklogisch noch einmal mit Lösung, Sachverhalt, eigener Bearbeitung auseinanderzusetzen. Das sehe ich durch eine umfassende KI Korrektur gefährdet.

Im Großen und Ganzen würde ich daher lieber weiterhin von **Menschen** korrigiert werden, wie gesagt, vielleicht mit der neuen Arbeitsanweisung nicht nur auf den Inhalt sondern mehr auch auf den Gesamteindruck zu achten. Am Ende sehe ich den Mehrwert im Klausurenkurs darin, dass ich sehe, ob ich mit meiner juristischen Arbeitsweise einen **Menschen** überzeugen kann oder woran ich noch arbeiten kann. Im Examen muss ich ja auch einen **Menschen** überzeugen und keine KI mit bestimmten Formulierungen triggern.

Ich sehe außerdem nicht wirklich, dass der Einsatz von KI eine Arbeitserleichterung sein kann. So wie ich das verstanden habe, muss dafür die Lösung sehr stark angepasst und viel abstrakter an mehrere Lösungsansätze denken. Ich weiß nicht wie die Motivation in der Fakultät ist, aber ich denke nicht, dass viele Profs die Lust haben, sich so zeitintensiv mit dem Klausurenkurs zu beschäftigen. Und dann kann es ja in Zukunft doch so bleiben wie bisher, nur dass man einmal- zweimal im Semester zusätzlich zur **menschlichen Korrektur** auch eine abgespeckte, sich nicht allzuoft wiederholende **KI-Korrektur** erhält.

Ich frage mich nämlich auch gerade, ob es sich überhaupt lohnt als KI so viele Anmerkungen zu schreiben und wer gerade in der Examensvorbereitung die Zeit und Lust hat sich mit so viel zu beschäftigen.

Zusammenfassend fände ich es super, wenn weiterhin **Menschen** meine Klausuren korrigieren und meine Arbeit nicht nur inhaltlich, sondern auch unter anderen Gesichtspunkten bewerten und das auch in einem Abschlussvotum kommunizieren, ich ein Rohpunkte-Schema bekomme, es eine gute Lösung gibt und die Besprechung von den Profs in Präsenz gehalten wird. Eine **KI-Korrektur** kann ggf. zusätzlich bei ein- zwei Klausuren angeboten werden, sofern die Anmerkungen sich kurz und nicht wiederholend auf meine Arbeitsweise und den richtigen Inhalt ausschließlich aus der Lösungsskizze beziehen.

Im Grunde halte ich eine Korrektur durch gut "trainierte" KI-Modelle für sinnvoll und gewinnbringend (gerade wenn die **menschlichen Korrekturen** im Vergleich schlechte Korrekturen abliefern). M.E. gibt es allerdings ein großes Problem mit dem Umstieg auf **KI-Korrekturen**: Man nimmt dadurch z.B. Leuten, die gerade ihr Examen geschrieben haben und gerne Klausuren korrigieren möchten, diese Chance weg. Das Korrigieren von Klausuren hätte ich bspw. nach meinem Examen auch gerne genutzt, um weiter "im Stoff drin zu bleiben" und auch noch mehr dazu zu lernen.

Natürlich sehe ich den Punkt, dass gerade neue Korrektoren, die zuvor noch (fast) nie korrigiert haben, keine exzellente Korrektur abliefern können und die **KI-Korrekturen** dann ggf. inhaltlich besser sind. Beim Korrigieren von Klausuren hat man aber auch die Chance, sich selbst zu verbessern und dazu zu lernen. Diese Möglichkeit besteht dann nicht mehr, wenn dieser Job gänzlich von der KI "übernommen wird".

Daher ein konkreter Vorschlag, wie man damit in Zukunft umgehen könnte: Vielleicht wäre es denkbar, nicht gänzlich auf KI umzusteigen, sondern lediglich überwiegend. Was ich damit meine: Es könnte z.B. nur noch eine Klausur im Monat von **Menschen** korrigiert werden und der Rest durch KI. Dann würde man trotzdem etwas Geld sparen und man bekommt als Student*in in der Examensvorbereitung trotzdem noch ein paar **menschliche Korrekturen**. Dadurch entsteht auch ein geringerer Bedarf an **menschlichen Korrektoren**, was auch dazu führt, dass man nicht mehr "jeden" nehmen muss, sondern nur noch diejenigen, die gute Arbeit leisten.

Bitte in Zukunft die KI von **DeepWrite** für den Klausurenkurs zum Einsatz bringen. Nur diese KI hat auch kleinere Ungenauigkeiten erkannt. Die Korrektur durch den echten **Menschen** war - wie so oft im Uni-Klausurenkurs - völlig unbrauchbar.

Wäre sehr hilfreich, wenn eine zusätzliche KI Korrektur zugänglich ist. Aber eine menschliche wäre auch wünschenswert. Wobei ich eine KI Korrektur bevorzugen würde, solange alle paar Klausuren auch eine menschliche erfolgt.

Nie wieder will ich eine Klausur schreiben für die ich mich bei der Vhb anmelden muss. Lasst mich bitte Jura machen keinen bock es auf noch mehr langweile Passwörter und unnötige internetseiten die nur zeit fressen und niemand braucht.

P.S. das war allgemein gemeint nicht bezogen auf die konkrete Klausur.
sowas nie wieder bitte.

Ich teile die von Kommilitoninnen und Kommilitonen geäußerten Bedenken hinsichtlich Abhängigkeit von KI und die Gefahr, sich selbst als Jurist „abzuschaffen“. Die pragmatische Einwände dagegen verstehe ich aber genauso. Im Ergebnis sehe ich, wie überlegen die KI in Korrekturen sein kann. Bei meinen erhaltenen Korrekturen war die **menschliche Korrektur** gar nicht zu gebrauchen, die KI hingegen überraschend gut. Einen gesunden Mittelweg von (wenn möglich guten) menschlichen und **KI-Korrekturen** fände ich gut.

Ich habe den Eindruck, dass die Entscheidung, **menschliche Korrekturkräfte** durch KI zu ersetzen, bereits feststeht.

Nur weil die **menschlichen Korrekturkräfte** schlecht korrigieren, muss die Lösung nicht sein, eine KI einzusetzen. Das größte Problem, das ich bei den **KI-Korrekturen** sehe, ist, dass sich Studierende sehr auf die KI verlassen und sich dann insbesondere methodische Fehler etablieren.

Die beste Lösung wäre, bessere **menschliche Korrekturkräfte** einzustellen. Wenn das Geld dazu nicht reicht, wäre es vielleicht eine Lösung, dass man beispielsweise zwei Klausuren pro Monat von KI korrigieren lässt und zweimal pro Monat werden gute (!) **menschliche Korrekturkräfte** eingesetzt.

Der Anspruch sollte sein, eine bestmögliche Lösung für die Studierenden zu schaffen. Dazu wäre es hilfreich, wenn die KI nicht mit schlechten bis durchschnittlichen Korrekturkräften verglichen wird, sondern mit guten Korrekturkräften. Erst dann kann die **KI-Korrektur** in meinen Augen tatsächlich einen Fortschritt bringen.

Ein Mix aus allen drei wäre vermutlich perfekt. Mir persönlich hat für die Verständlichkeit die Korrektur von **DeepWrite** am meisten geholfen, weil auch gleich erklärt wird, warum es falsch ist bzw. was richtig wäre und wenn man etwas nicht gesehen hat, woran man es hätte erkennen können (zumindest teilweise). Auch wenn dadurch die Korrektur etwas länger ist, macht es vermutlich keinen großen Unterschied, wenn man sich stattdessen erspart den jeweiligen Aspekt in der Lösung herauszusuchen. Ansonsten finde ich an der **Klausurenkiste** gut, dass es stark einer **menschlichen Korrektur** ähnelt vom Aufbau, aber eben mehr und verständlichere Randbemerkungen macht. Natürlich weist alles seine Schwächen auf. Jedoch war alles nachvollziehbarer und ausführlicher als die **menschliche Korrektur**, welche überhaupt keine Tipps o.ä. gegeben hat.

Verbessert werden könnte meiner Meinung nach aber die Fehlererkennung (s.o. als Beispiele was beanstandet wurde).

ggf. wäre es bei Einführung der **KI-Korrektur** möglich, eine konstant bestehende Feedbackplattform für wiederkehrende "Bewertungsfehler" und Verbesserungsvorschläge einzurichten

Sehr interessanter Test, welcher m.E. nur allzu deutlich zeigt, dass man die finanziellen Mittel für die Korrektur besser verwenden könnte. Dank des Rohpunkteschemas auch sehr vergleichbare Notenstufen. Interessant wäre auch ein Durchgang ohne Rohpunkteschema. Vielen Dank.

- Designtechnisch eher an der **Klausurenkiste** orientieren, aber die inhaltliche Qualität der beiden anderen KI-Modelle übernehmen

- Allgemeine Verbesserungstipps am Ende der Klausur helfen, um sich wirklich verbessern zu können

Danke für dieses Projekt, allein anhand dieser Klausur konnte ich so viel mitnehmen. Vorher haben mir die Samstagsklausuren nicht wirklich viel, bis auf eine Einschätzung, ob ich grds. bestehen würde, gebracht.

Solange scharfe Klausuren noch von **menschlichen Korrektoren** korrigiert werden und auch gute Korrektoren dazu eingesetzt werden, wäre ich mit der Korrektur durch KI im Examensklausurenkurs einverstanden.

Sorgen bzgl. eines Einsatzes von KI in der öffentlich-rechtlichen Bearbeitung: im Nachgang mache ich mir außerdem große Sorgen, ob eine **KI-Korrektur** auch nur annähernd an die Qualität der zivilrechtlichen Korrektur rankommt: im öffentlichen Recht kann sich der Aufbau einer Bearbeitung sehr stark nach individuellem Belieben unterscheiden. Mit der jetzigen fehlenden Flexibilität der KI in Sachen Aufbau und alternative Lösungsansätze sehe ich den Einsatz von KI im öffentlichen Recht nicht wirklich als sinnvoll

Die KI von Prof. Harnos – hier nicht aufgeführt – war sehr schön hinsichtlich Länge der Kommentare, Übersichtlichkeit, Tonfall und Brauchbarkeit des Votums/Tipps.

Ich schätze Ihren Einsatz, Herr Prof. Beurskens und Herr Prof. Hannos, für gute Studienbedingungen sehr, auch wenn ich nicht immer derselben Meinung bin, was das im Einzelnen bedeutet.

Ein wirklich schönes Projekt.

Die Korrekturen durch KI fand ich sehr aufschlussreich und konnte viel für mich mitnehmen.

Dass die KI noch nicht perfekt ist, war zu erwarten. Ich bin jedoch erstaunt darüber, wie präzise sie doch bereits ist.

Zunächst: Die Bewertung oben bei "**Mensch**" ist nur deshalb mittelmäßig gut bewertet, da die Nachkorrektur und Anmerkungen von Prof. Harnos und Prof. Kuhn meiner Meinung nach genau dem anderen Ende (der 5 Sterne Bewertung) entspricht und deshalb die gesamte Bewertung etwas anhebt. Ansonsten hätte ich überall 1 Stern vergeben. Dazu sofort mehr.

Ob nur noch **Mensch** oder KI korrigieren soll, kann ich nach diesen Korrekturen nicht einfach generell mit **Mensch** oder KI entscheiden, was aber v.a. an dieser Art der

menschlichen "Korrektur" liegt, sofern man es "Korrektur" nennen kann. Die KI müsste inhaltlich und "auslegungstechnisch" deutlich zulegen, v.a. was ungenaue Formulierungen oder eigene Ungenauigkeiten oder Widersprüche angeht.

Ich bevorzuge generell den **Menschen** als Korrektor, wenn es eine brauchbare Korrektur gibt wie die von Prof. Harnos oder Prof. Kuhn, was bei den Klausuren die ich geschrieben habe, überwiegend der Fall war. Wenn es jedoch Korrekturen wie diese gibt, wo nahezu keine Randbemerkungen vorhanden sind, die dann wenig hilfreich und kaum verständlich (was bitte ist "vert.??") sind und anschließend stumpf Punkte in einem Bewertungsbogen vergeben werden ohne jegliche Begründungen (was das Ganze eher willkürlich aussehen lässt und - mit Verlaub - in Jura ist ja sonst auch regelmäßig die Begründung gerade das Entscheidende und das Ergebnis zweitrangig, solange es noch vertretbar ist..) und außer "Die Bearbeitung ist mit ... Punkten zu bewerten" überhaupt kein eigenes Feedback, ganz zu schweigen von einem Votum gegeben wird, worauf es dem Bearbeiter ja gerade ankommt, was DER Grund für das Schreiben der Klausur war, dann ziehe ich doch die Korrektur der KI vor, auch wenn sie momentan in einigen Belangen noch zu wünschen übrig lässt. Denn die KI gibt einem immerhin ein Feedback und Votum, wohingegen eine **menschliche Korrektur** wie die Vorliegende, mit prinzipiell überhaupt keiner eigenen Meinung/Sicht des Korrektors und nur Angabe der Lösungsskizze, meiner Meinung nach - so hart und möglicherweise kränkend es klingen mag - als "Feedback" für den Bearbeiter unbrauchbar und mehr oder weniger für die Tonne ist.

Wie oben bereits erwähnt, waren solche Korrekturen in den Klausuren, die ich schon geschrieben habe, zum Glück in der doch relativ kleinen Minderheit und man bekommt doch in ca. 8-9 von 10 Fällen Korrekturen, die quantitativ und v.a. QUALITATIV (mal mehr, mal weniger [aber immerhin haben diese dann zumindest überhaupt ein Votum und es kommt die Sicht/Meinung des Korrektors hervor und Begründungen sind vorhanden]) brauchbar sind.

Falls meine Formulierungen z.T. etwas sehr harsch und unfreundlich/unsachlich aufgenommen werden und möglicherweise den menschlichen Korrektor kränken, ich möchte um Gottes Willen niemanden persönlich angreifen, geschweige denn kränken, beleidigen, herabwürdigen oder ähnliches. Mir geht es ganz generell um diese Art und Weise und Form (!!!) von "Korrektur", egal von wem.

Aber wenn ich eine Klausur schreiben will ich gerade das Feedback des Korrektors haben, von jemandem, der mehr Ahnung und Erfahrung hat als ich, damit ich weiß was ich gut/schlecht gemacht habe, wo ich überhaupt stehe mit dem derzeitigen Stand. Und wenn ich dann die Zeit und Konzentration investiere und dann aber eine Korrektur bekomme, die nahezu keine Randbemerkungen (und wenn, dann wenig brauchbare), ÜBERHAUPT KEINE Begründungen um die Punktevergabe nachzuvollziehen und keinen einzigen Satz eigene Meinung des Korrektors enthält, sondern im Prinzip nur die Lösungsskizze beigelegt und Punkte (für mich ohne Begründung willkürlich) eingetragen werden, und am Ende nur steht ich habe ... Punkte, dann fühle ich mich frustriert und v.a. leider etwas veräppelt. Denn meine Bearbeitung mit der Lösungsskizze vergleichen kann ich auch selber und die Punktzahl am Ende interessiert mich auch deutlich weniger, da ich ja eigentlich mal wissen will, was ich denn gut/schlecht bzw. richtig/falsch

gemacht habe, damit ich sehe wo ich mich eben verbessern kann, wo ich momentan stehe, was ja gerade der Sinn des Ganzen ist.

Nochmals, es ist keinesfalls meine Absicht in irgendeiner Form jemanden persönlich anzugreifen, zu kränken, zu beleidigen, herabzuwürdigen oder ähnliches, aber ich hoffe Sie verstehen meine Sicht und warum ich hier rein auf diese ART UND WEISE und FORM der **menschlichen Korrektur** so deutlich und harsch formuliere.

Da solche Korrekturen bei mir eben bisher nur die Ausnahme waren, habe ich oben "Künftig soll nur durch **Menschen** korrigiert werden" gewählt.

Ich finde die **KI-Korrekturen** waren bisher die besten Korrekturen, die ich bei Klausurenkursklausuren bekommen habe.

2. Antworten der 31 Personen die nicht teilgenommen haben

a. Künftig soll die Korrektur erfolgen durch...

14: Nur Mensch	3: Offen	14: Nur KI
-----------------------	-----------------	-------------------

b. Vergleich der Korrekturen

	Mensch	Klausuren-Kiste	DeepWrite	Institut für Rechtsdidaktik
Qualität und Quantität der Randbemerkungen	2.11 Sterne	3.56 Sterne	2.93 Sterne	3.75 Sterne
Übersichtlichkeit und Verständlichkeit des Abschlussvotums	2.39 Sterne	3.63 Sterne	3.63 Sterne	3.88 Sterne
Nachvollziehbarkeit der Note	2.67 Sterne	3.07 Sterne	3.2 Sterne	3.47 Sterne
Nützlichkeit von Arbeits-/Klausurtipps	2.29 Sterne	3.2 Sterne	3.38 Sterne	3.63 Sterne
Gesamtbeurteilung	2.53 Sterne	3.19 Sterne	3.12 Sterne	3.53 Sterne

c. Negatives Feedback

Mensch: auf jeder Seite 15-20 Hacken, nur meine 2 größten Schnitzer wurden mit Fehlzeichen und Stichwort angezeigt (sonst nur Hacken überall)

Kein Endvotum

Mensch:

- Teilweise Seitenlang keine Randbemerkung
- unklar welchen Lerneffekt verschiedenartige Unterstreichungen haben sollen, die nicht weiter erklärt werden

- Tabelle am Ende macht zwar transparent, wo Punkte fehlen, allerdings teilweise keine konstruktive Erklärung, wie man diese hätte erreichen können

Klausurenkiste:

- ist sehr von der Musterlösung abhängig, es ist zweifelhaft ob kreative Lösungswege verstanden und angemessen bewertet werden können

DeepWrite:

- keine direkten Anmerkungen an der Klausur

Rechtsdidaktik:

- einzelnes Feedback bei Randbemerkungen fast zu ausführlich

- benennt Zitierungenaugigkeiten, gibt aber nicht an welche

Schneller Nebenverdienst der Korrektoren, sodass eher Quantität vor Qualität im Vordergrund steht.

Die Randbemerkungen sind pauschal und bringen wenig Mehrwert.

Die Noten hängen vom guten Willen der Korrektoren ab, sodass man sich vor den Examenkorrektoren fürchten muss, je nachdem wie der Tag gelaufen ist.

Dabei merkt man riesige Unterschiede in der Notengebung, die die Karriere und Zukunft junger Juristen wortwörtlich zerstören können.

Ganz besonders **DeepWrite** aber auch die **Klausurenkiste** überfordern komplett mit ihrem Feedback. Dadurch, dass zu jedem einzelnen Abschnitt etwas gesagt wird (und bei **DeepWrite** sogar doppelt und dreifach), gehen wichtige Fehler komplett unter. Ich muss nicht 20 mal in umständlichen Formulierungen lesen, wieso meine Lösung so schon in Ordnung ist, sondern ich brauche diesen Umfang nur, wenn ich etwas gerade nicht gut gelöst habe.

Durch die schiere Menge an Anmerkungen vergeht mir jede Lust, mich in die Lösung einzuarbeiten und meine Fehler zu analysieren. Man wird dazu verleitet, nur die Note zu sehen und sich zu denken: "Passt schon".

Die **KI-Korrekturen** sind durch die Bank viel zu lang. (In Relation zum Inhalt, den sie vermitteln.)

Die Korrekturen des Instituts und von **DeepWrite** ferner äußerst unübersichtlich und schon aus diesem Grunde kaum zu gebrauchen.

Und ich möchte mich über folgendes Kriterium beschweren: "Qualität und Quantität der Randbemerkungen". Seit wann ist Quantität von Bemerkungen ein Kriterium für die Güte einer Korrektur? Sinnlose Prosa kann kein Selbstzweck sein. Und die Randbemerkungen zu einer Klausur sollten jedenfalls nicht länger sein als die Klausur selbst.

Und leider ist die Prosa der Klen nicht nur sinnlos, sondern auch noch erstaunlich schlecht.

Folgender Satz zum Beispiel: "Die Bearbeitung zeigt ein insgesamt solides Problembewusstsein und eine erkennbare Struktur." Ich persönlich ziehe es ja vor, wenn die Struktur, die eine Bearbeitung zeigt, nicht erkennbar ist.

Und noch ein Beispiel: "Eine Fristsetzung nach § 323 Abs. 1 BGB als grundsätzliches Erfordernis des Rücktritts wird nicht geprüft. Sie steigen unmittelbar in § 326 Abs. 5 BGB ein. Gerade der Ausgangspunkt, dass normalerweise zunächst eine angemessene Frist zur Nacherfüllung zu setzen ist, hätte kurz benannt werden müssen." [Kontext: Der Bearbeiter hatte es unterlassen festzustellen, dass § 326 V BGB keiner Fristsetzung bedarf, den Rohpunkt für "Keine Fristsetzung" daher verpasst.]

Äußerst Wortreich führt die KI aus: "Fristsetzung fehlt". Nicht nur braucht die KI für diese Aussage statt Zweien, stolze 37 Wörter (1 850 %). Die Bemerkung ist nichtmal richtig. Die Fristsetzung zu prüfen wäre hier schlicht falsch gewesen, da sie entbehrlich ist und in der Sache sinnlos wäre. Allenfalls hätte man dies feststellen können. Ferner ist es auch richtig, mit § 326 V BGB einzusteigen, da dieser erst auf § 323 BGB verweist. Alles andere wäre unlogisch.

Aber zu schlechter Letzt ist die viel zu lange, inhaltlich unzutreffende Ausführung auch noch in einer schrecklichen Prosa verfasst. Lesen Sie aufmerksam: "Gerade der Ausgangspunkt, dass normalerweise zunächst eine angemessene Frist zur Nacherfüllung zu setzen ist, hätte kurz benannt werden müssen". Was ist das bitte für ein verquerer Satzbau? Und auch das Vokabular ist (nett gesagt) ungewöhnlich. Das Wort "gerade" impliziert übriges ein besonderes Hervorragen: "Gerade der Ausgangspunkt" impliziert also das bestehen Mehrerer paralleler Ausgangspunkte. Indes zeichnet einen AusgangspUNKT aus, dass er singulär ist.

Ein letztes Beispiel, weil es recht lustig ist: "Unklar bleibt aber die dogmatische Präzision: [...]"

Diese Beispiele sind übrigens keine Rosinenpickerei. Ich darf Ihnen versichern: ich habe mir bei der Suche nach Beispielen keine Mühe gegeben. Sie sind das erst Beste, das mir unter die Augen kam. Und mir geht es eigentlich auch garnicht um die einzelnen Fehler, die ich oben so pedantisch breit getreten habe. Mein Punkt ist folgender: KI denkt nicht. Es handelt es um einen mit Statistik angefütterten Textgenerator. Text ist nicht Sinn.

Zwar ist KI inzwischen sehr gut darin, Texte zu generieren, die Sinn stark ähneln und ihn zuweilen zufälligerweise sogar enthalten. Aber dabei handelt es sich um genau das: Zufall.

Auf Zufall kann ich nicht vertrauen. Und mag die Korrektur auch zufälligerweise richtig sein, woher weiß ich das? Finden Sie es wirklich besser, wenn wir Student:innen uns durch Seitenweise sinnlos zusammengewürfelte Gemeinplätze und leere Worthülsen quälen müssen, in der Hoffnung etwas verwertbares zu finden? Wo ist da der Vorteil zu einer Selbstkorrektur (oder einem Korrekturzirkel)? Sie mögen dem vielleicht entgegenhalten, dass man ebenso wenig wissen könne, ob eine echte Korrektur richtig sei, aber das stimmt nicht. Zum einen sehe ich durch die Gestaltung des Voti (Votums) und der Randbemerkungen, ob sich der Korrektor hinreichend Zeit genommen hat, den Text zu durchdringen. Zum anderen hat sich der Korrektor, auch wenn es falsch war oder

misslich ausgedrückt wurde, bei jeder Ausführung etwas gedacht. Als ebenfalls denkende Wesen (und angehende Jurist:innen) sind meine Kommiliton:innen und ich in der Lage, dies nachzuvollziehen; zu ermitteln was gesagt werden sollte und warum. Bei einem automatisch generierten Text gibt es keinen Sinn zu ermitteln, keine dahinterliegende Idee. Und die Länge des Textes zeigt mir nicht, dass die KI sich besonders viel überlegt hätte (denn das kann sie nicht), sondern nur, wie viele Bäume für meine Korrektur brennen werden.

Ich bitte Sie, setzen sie sich einmal hin, und lesen Sie die KI-Ausführungen aufmerksam durch, Satz für Satz.

Ich unterstelle, dass Sie um die Minderwertigkeit der automatischen Korrekturen wissen. Argumente, wie, die automatischen Korrekturen seien gerechter, sind offensichtlich vorgeschoben. Inhaltlich kann das Argument übrigens schon darum nicht überzeugen, weil KI-Ergebnisse nicht wiederholbar sind. Es handelt sich nichtmal um Willkür, da es an einem Willen ja leider fehlt. Ob man Zufall gerechter findet als Willkür ist eine spannende Frage, über die es sich nachzudenken lohnt. Aber nicht hier.

In der Sache ist das Argument, die KI sei gerechter als der **Mensch**, selbst wenn es stimmte, aber schlicht verfehlt. Das Examen wird von **Menschen** korrigiert. Wenn eine automatische Korrektur gerechter wäre als eine echte, dann wäre sie für den Zwecke der Examensvorbereitung schlechter.

vvv

KI erkennt bloße Tippfehler nicht, abweichende Lösungsansätze werden teilweise negativ bewertet

-
- **Mensch**: fehlende Erläuterungen
 - Rechtsdidaktik: etwas knappes Gesamtfeedback

d. Positives Feedback

Klausurkiste habe ich die Korrektur von Professor Harnos Prompt vorgenommen

Ich fand diesen Prompt übersichtlicher in der Durcharbeit, da nur die relevanten Stellen angezeigt wurden.

Bei dem Institut gingen die Bemerkungen oft unter.

Ich hatte das Gefühl, weniger "Schnitzer" zu haben, da die KI jeden Satz mehr oder weniger lobt und dadurch der Blick auf die Probleme etwas schwieriger ist.

Insgesamt bin ich hellauf begeistert von den **KI-Korrekturen**.

Ich konnte meine Fehler sehr (!) gut nacharbeiten und vor allem auch erstmal bemerken.

Leider finde ich in der **menschlichen Korrektur** gelingt dies meist nicht so, da die einzelnen Fehler oft nicht angestrichen werden und wenn doch, fehlt es oft an einem Rat-schlag oder einem Lösungshinweis.

Ich könnte mir wirklich vorstellen, dass durch diese Art der Korrektur ein wertvolles Tool geschaffen werden kann, das den Studierenden sehr helfen würde.

Mensch:

- ist vergleichbarer mit aktueller Examenspraxis
- Tabelle zeigt Punkteverteilung transparent auf

Klausurenkiste:

- ausführliches Feedback
- übersichtlich

DeepWrite:

- ausführliches erklärendes Feedback

Rechtsdidaktik:

- hilfreiche Endübersicht über Stärken und Schwächen, Empfehlungen

KI generell:

- endlose Chancen, wie Feedback angepasst werden kann, bzw. Feedback-"Ge-spräch" über Chat wäre denkbar -> Prüfling könnte Rückfragen zur Korrektur stellen
-

Stimulierung noch mehr zu arbeiten und sich nicht von den zwar gerechteren "besse-ren" Noten blenden zu lassen.

Die drei **KI-Systeme** sind vom Ergebnis her oft näher an der Lösung der Profs als die externen Korrektor*innen (gerade bei Beispiel 3). Allerdings sollte man seine Leistungen insbesondere in der Examensvorbereitung nicht nach den Noten bewerten, son-der konkret auf inhaltliche Fehler und Fortschritte achten.

Das gelingt mMn am besten bei **menschlichen Korrekturen**, da diese zu durch-schnittlich gut gelösten Abschnitten einfach nichts sagen. So gelingt der Blick für das Wesentliche selbst bei schlechten **menschlichen Korrekturen** noch am besten. Wie-der zu Beispiel 3: Die menschliche externe Korrektur hier ist wirklich nicht gut, aber gibt mir mehr als die der **KI-Systeme**, da ich mich nicht erst durch "Slop" durchkämp-fen muss. Am ehesten kommt für mich die **KI des Instituts für Rechtsdidaktik** da heran, aber auch nicht wirklich.

Die automatischen Korrekturen waren (zumindest noch und rein pekuniär) vermutlich recht billig in der Herstellung.

Künftig soll nur noch das **Institut für Rechtsdidaktik** selber korrigieren, sie vergeben die besten Punkte und machen sich am meisten Arbeit damit. Leider kann ein Korrek-tor für Niedriglöhne nicht diese Arbeit verrichten. Daher am besten alles selber am Institut mit institutseigenem Personal korrigieren.

Es wird gezeigt, an welchen Stellen explizit wie viel Punkte verloren gegangen sind (könnte man bei **menschlichen Korrekturen** einführen, um diese aufzuwerten)

- **Klausurenkiste** und Rechtsdidaktik: Anmerkungen am Lösungstext
- **DeepWrite**: Umfang des Feedbacks

e. Sonstige Anmerkungen

im Probeexamen würde ich mir wünschen, dass ggfs. weiterhin durch **Mensch** und KI korrigiert wird. Auch wenn die KI weniger willkürlich korrigiert, wird das echte Examen immer noch von **Menschen** korrigiert, die manchmal willkürlicher korrigieren. Damit man sich nicht auf die Ergebnisse der KI verlässt, sondern ggf. eine zweite (subjektivere) Meinung hat und auch mit vermeintlichen negativen Erfahrungen aus dieser Korrektur umgehen kann, fände ich dies wichtig. Auch, falls die KI einmal Fehler macht, wäre zumindest für das Probeexamen eine **menschliche Korrektur** bzw. Überprüfung meines Erachtens sinnvoll.

Auch frage ich mich, ob ich durch das Hochladen meiner Klausuren nicht selbst zu einer Überflüssigkeit von mir beitrage, indem ich der KI weitere Datenpunkte liefere, um Klausuren besser und schneller zu lösen - was wird dies für Auswirkungen auf die Zukunft des juristischen Studiums haben?

Solange die echten Examensklausuren von **Menschen** korrigiert werden ist es essentiell, dass auch die Korrektur im Rep durch **Menschen** erfolgt. Die KI wird immer streng bei der Lösungsskizze bleiben, **Menschen** haben jedoch ihre Eigenheiten. Dadurch lernt man, was anderes Juristen als Maßstab für Genauigkeit o.Ä. ansetzen. Das mag zwar zu unfairen Korrekturen führen, aber immerhin hat man die Chance sich darauf einzustellen und wird nicht im Examen überrascht.

Was die Ausführlichkeit der Bewertung angeht, so sollten "Best Practices" ausgearbeitet werden, wie z.B. ein ausformuliertes Fazit, Randbemerkungen bei Punktabzug, und vertraglich gegen die Korrekturassistenten durchgesetzt werden.

Bestenfalls finden sich Gelder, um ein hybrides Modell zu ermöglichen. Dies sollte im Interesse der Uni Passau hinsichtlich ihres Rufes für die Juristenausbildung sein. Auch könnte es im Interesse von bereits berufstätigen Absolventen sein ein solches Pilotprogramm zu fördern.

Ein Umdenken bei der Korrektur der echten Examensklausuren wird erst erfolgen, wenn das Modell getestet ist, dies kann jedoch nicht zu Lasten der Examensvorbereitung gehen. Vielleicht kann die **KI-Korrektur** für Probe-/Klausuren im Grund- und Hauptstudium angewendet werden

Die **KI-Korrektur** ist in jedem Bereich um Welten besser als Einzelkorrekturen. Angesichts dessen, dass die KI ständig trainiert wird, hat sie auch noch stetiges Verbesserungspotenzial.

Solange jedoch die Examensklausuren von **Menschen** korrigiert werden, muss man sich im Kopf behalten, dass die KI-Noten nicht die (im Durchschnitt ungerechteren) realen Examensnoten widerspiegeln, da der Unterschied zwischen **Mensch** und KI immerhin beträchtlich ist.

Ihr maximales Potenzial kann die KI nur erreichen, wenn sie auch die Examensklausuren korrigieren wird. Denn ich weiß, wenn ich bei der Vorbereitung im Schnitt 7- 8 Punkte habe, dass die auch die potenzielle Examensnote abbilden. Noch muss ich darauf hoffen, dass der **Mensch** seine Arbeit nicht sabotiert. Interessant wird dabei auch die blinde Doppelkorrektur.

Ich bin für eine **KI-Korrektur**. Sie kann den zunehmend schlechter werdenden Korrekturen Abhilfe schaffen und eine gewisse Mindestqualität sicherstellen. Trotzdem bin ich dafür, in geregelten Abständen (Z.B. 1 im Monat/ 2-3 im Semester) weiterhin menschliches Feedback zu erhalten, schließlich wird auch das Staatsexamen von **Menschen** korrigiert.

Ich habe mir aus Zeitgründen nur Beispiel 3 anschauen können.

Die externe menschliche Klausur ist hier ganz besonders kurz und (mMn) schlecht geraten, ein Votum fehlt komplett. Eine Korrektur in so geringem Umfang ist mir bis jetzt selten untergekommen. Deshalb fand ich besonders spannend, dass ich mit dieser Korrektur immer noch mehr anfangen konnte als mit allen 3 **KI-Korrekturen**. Den **KI-Systemen** fehlt mMn komplett der Fokus aufs Wesentliche. Natürlich ist in einer 17-seitigen (!) Korrektur (**DeepWrite**) jeder Fehler einmal angesprochen, aber man muss ihn da drin auch erst einmal finden.

Ich verstehe, dass die Uni Mittel sparen möchte, indem sie auf KI setzt, und ich finde das auch gerechtfertigt. Aber solange KI es nicht wie **Menschen** schafft, den Fokus auf das Wesentliche zu legen, ist mir selbst die schlechteste **menschliche Korrektur** hundertmal lieber als die einer KI.

Positiv finde ich die Auswertung der KI inklusive der Verbesserungsvorschläge an den Bearbeiter, was bei einigen **menschlichen Korrektoren** leider oft zu kurz kommt.

Ferner sehe ich ein großes Plus bei der Nachbereitung der Klausur, da die KI sich genauer auf einzelne bearbeitete Passagen bezieht und so nicht mühsam die Lösung der Klausur im Detail abgeglichen werden muss. Gerade deshalb war oft die **menschliche Korrektur** besonders bei wenigen Randbemerkungen wenig hilfreich.

Dass das "freiwillige Pilotprojekt" der **KI-Korrektur** nicht ohne Hintergedanken durchgeführt wurde, war klar. Ich finde es aber gerade darum in einem ärgerlichen Maße unaufrichtig, dass Sie in der diesbezüglichen PowerPoint schrieben: "Freiwillige Teilnahme für interessierte Studierende → kein KI-Zwang!" Denn wenn die hiesige Abstimmung im Sinne des Institutes ausgeht, dann wird die **KI-Korrektur** für die Teilnehmer:innen des Klausurenkurses zum Zwang, ganz gleich ob sie daran "interessiert"

sind. Ob das dem Willen einer Mehrheit entspricht (hoffentlich nicht) ändert hieran nichts.

Mir ist übrigens in der Formulierung Ihrer (Jura Beurskens Admin) E-Mail vom 27. Juli, mit dem Ruf zur Abstimmung, folgende Formulierung nicht entgangen: "Ob wir diesen Schritt wagen, hängt auch von Ihnen ab: Stimmen Sie [...] ab!" Auch von uns heißt vermutlich letztlich nicht von uns, sofern das Ergebnis nicht passt. Auch das ist ein unwürdiger Zug.

Die Lösung wird für mich die Teilnahme am Klausurenkurs eines privaten Repertori-ums sein. Für die Universität Passau habe ich mich ursprünglich entschieden, weil ich von der Qualität des "Unireps" gehört hatte, und hoffte, private Repetitoren vermeiden zu können. Für mich wäre die de facto Abschaffung des Klausurenkurses also höchst frustrierend. Für zukünftige Schulabgänger wäre das natürlich kein Problem: Denn von der Qualität des Passauer Unireps würden sie wohl nicht mehr hören.

Künftig soll nur noch das **Institut für Rechtsdidaktik** selber korrigieren, sie vergeben die besten Punkte und machen sich am meisten Arbeit damit. Leider kann ein Korrektor für Niedriglöhne nicht diese Arbeit verrichten. Daher am besten alles selber am Institut mit institutseigenem Personal korrigieren.

Auffällig ist, dass die **KI-Korrekturen** bedeutend besser ausfallen als die menschlichen. Im echten Examen wird ein **Mensch** korrigieren. Daher sind die **KI-Korrekturen** für mich nicht aussagekräftig und vermitteln ein falsches Bild des Leistungsstandes. Lieber habe ich ab und zu eine weniger gute Korrektur von einem **Menschen**, als mich die ganze Zeit auf zu Gute Notenpunkte zu verlassen und im Examen dann das "böse Erwachen zu haben".

Ich bin an die Universität Passau zurück gewechselt, aufgrund des exzellenten Rufes des Uni Rep. Sollten die Korrekturen tatsächlich rein von KI demnächst korrigiert werden, ist für mich der einzig vernünftige Schritt, einen kommerziellen Klausurenkurs zu suchen, was ich persönlich sehr schade fände, da ich das Niveau der Sachverhalte sehr gut finde.

Grundsätzlich habe ich keine Einwände, dass KI lediglich unterstützend! als zweite Korrektur mit angeboten wird, so dass man selbst besser einschätzen kann, ob der Korrektor vielleicht einen weniger guten Tag hatte oder ob die KI die Auffassung teilt. Dass finanzielle Mittel begrenzt sind, ist ein durchaus valider Punkt, jedoch kann dieser für mich persönlich, nicht auf Kosten einer guten Examensvorbereitung erfolgen. Natürlich lernt die KI mit der Zeit dazu, jedoch schreibe ich mein Examen nicht in 2 Jahren, wenn die KI ausgelernt ist und benötige jetzt realistische Klausuren und nicht erst irgendwann. Ich fände es schön, wenn KI als Unterstützung lediglich verwendet würde und nicht als einziges Mittel. Daher werde ich meine Zustimmung für **KI-Korrekturen** meiner Klausuren auch erst einmal nicht erteilen, wenn aufgrund dieser Daten die **menschlichen Korrektur** gestrichen wird.

Vielleicht wäre es ja eine Möglichkeit (sofern dies rechtlich und tatsächlich möglich ist), sich mit anderen Universitäten in die Ausarbeitung einer KI als Zweitkorrektur zu teilen, um so die Kosten zu begrenzen und mehr Daten sammeln zu können.

Anzumerken ist auch, dass ich im Probeexamen neben einer wenig nützlichen Korrektur auch unter anderem eine herausragend gute Korrektur hatte, bei welcher sich der Korrektor intensiv Zeit genommen hat und die aufgrund der überragenden Anmerkungen mein Klausurenschreiben tatsächlich um einiges verbessert hat.

f. Warum haben Sie nicht teilgenommen?

Ich war am Tag der Klausur leider verhindert und konnte nicht mitschreiben, würde aber in Zukunft gerne an weiteren möglichen KI-Tests teilnehmen.

zeitlich

Ich wollte an der **KI-Korrektur** teilnehmen, habe hierfür auch meine Zustimmung erteilt, jedoch wurde sie nur durch den **Mensch** korrigiert.

Ich werde mich erst demnächst auf das Examen vorbereiten und im September 2026 starten.

Ich war im Urlaub, als die Klausur gestellt wurde, und habe sie deshalb gar nicht mitgeschrieben.

Ich bin leider nicht dazugekommen.

Für die Umwelt. Denn für diese wertlose ... soll Bordeaux nicht brennen!

vvv

Ich habe die Klausur nicht mitgeschrieben

Noch nicht am Klausurenkurs teilgenommen.

Keine Zeit