



**WORKBOOK**

# Prompt-Labor – Generative KI in der Hochschullehre

## **Modul 3: Evaluationsphase mit Malte Persike**

---

Dieses Workbook begleitet die Live-Session am 24. November 2023 (10:00 bis 13:00 Uhr)

### **Autorinnen und Autoren**

PD Dr. Malte Persike

Lavinia Ionica

Stefan Göllner

***Liebe Teilnehmerinnen und Teilnehmer,***

*herzlich willkommen zum Modul 3 des Prompt-Labors!*

*Dieses Modul des Prompt-Labors soll Sie dazu befähigen, komplexe und passgenaue Prompts für Anwendungsszenarien in der Evaluationsphase von Lehrveranstaltungen zu entwickeln einzusetzen.*

*Auf den folgenden Seiten finden Sie Lernmaterialien und Übungen, die Sie durch das Modul 3 begleiten werden. Da Sie immer auch Wahlmöglichkeiten vorfinden werden, können Sie im Anschluss an die Live-Session sehr gern auch vertieft mit den hier angeführten Prompts experimentieren.*

*Viel Spaß und viel Erfolg!*

***Malte Persike, Lavinia Ionica und Stefan Göllner***

Das Workbook steht unter der Creative Commons Lizenz [CC BY SA 4.0](https://creativecommons.org/licenses/by-sa/4.0/).  
Hinweis: Teile dieses Dokuments wurden mit Chat-GPT erstellt

# Inhaltsverzeichnis

<b>Allgemeine Empfehlungen zur Arbeit und Verwendung generativer KI-Tools</b> .....	<b>4</b>
<b>Was ist das Prompt-Labor? – Das Prompt-Labor kurz erklärt</b> .....	<b>5</b>
<i>Was erwartet mich im Experimentierraum?</i> .....	5
<i>Welche Prinzipien leiten die Zusammenarbeit in Kleingruppen?</i> .....	5
<i>An wen wende ich mich für inhaltliche Fragen?</i> .....	6
<i>Wer bietet technischen Support im Prompt-Labor an?</i> .....	6
<i>An wen kann ich mich für Fragen zum Qualifizierungsangebot richten?</i> .....	6
<b>Modulziele</b> .....	<b>7</b>
<b>Aufgaben für die Live-Session im Prompt-Labor (Modul 3)</b> .....	<b>8</b>
<i>Arbeitsphase 1: Prompts zur Erstellung von Evaluationsfragen in Einzelarbeit schreiben (20 Min.)</i> .....	9
Worum geht es? .....	9
Ausgangspunkt: Das neue Lehrsetting .....	9
Schritt 1: Veränderungsdimensionen identifizieren .....	10
Schritt 2: Evaluationsfragen formulieren .....	11
<i>Arbeitsphase 2: Vorstellung der Ergebnisse und Austausch von Erfahrungen (40 Min.)</i> .....	13
Worum geht es? .....	13
Schritt 1: Erfahrungen teilen und diskutieren (max. 20 min) .....	13
Schritt 2: Prompts zur Auswertung von Evaluationsergebnissen formulieren (mind. 20 min) .....	14
<i>Arbeitsphase 3: Prompts zur Qualitätssicherung von Prüfungsaufgaben schreiben (40 Min.)</i> .....	18
Worum geht es? .....	18
Arbeitsauftrag 1: Prüfungsfrage auswählen (ca. 10 min) .....	18
Arbeitsauftrag 2: Prüfungsfrage qualitätssichern (ca. 25 min) .....	19
Arbeitsauftrag 3: Prüfungsleistung bewerten (ca. 5 min) .....	19
<b>Aufgabe nach der Live-Session: Dokumentation im offenen Prompt-Katalog</b> .....	<b>24</b>

# Allgemeine Empfehlungen zur Arbeit und Verwendung generativer KI-Tools

Hier haben wir einige Empfehlungen zusammengestellt, die wir Ihnen bei der Arbeit mit KI-Schreib-Tools ans Herz legen möchten. Weitere Informationen zu Chancen und Herausforderungen bei der Arbeit mit LLMs (Large Language Models) erhalten Sie im begleitenden [Moodle-Kurs auf dem KI-Campus](#).

## **Richtigkeit**

Viele Menschen nutzen textgenerative KI wie z.B. ChatGPT als Suchmaschine. Während eine Suchmaschine jedoch auf Grundlage von Schlagwörtern oder Fragen nach Quellen im Internet sucht, generiert textgenerative KI die Antworten aus sich heraus. Deshalb sollten diese Werkzeuge nicht als Suchmaschinen verwendet werden, da die Antworten gelegentlich fehlerhafte Fakten liefern oder "halluzinieren", indem sie gewünschte Suchergebnisse einfach erfinden.

## **Bias (Voreingenommenheit)**

Wer textgenerative KI nutzt, sollte sich darüber im Klaren sein, dass diese Systeme genauso wie Menschen über Voreingenommenheit verfügen. Damit ist eine oft unbewusste Neigung zur Diskriminierung von Menschen und Gruppen gemeint. Diese manifestiert sich in verzerrten oder falschen Ergebnissen, beispielsweise indem eine bestimmte Gruppe von Menschen falsch repräsentiert oder diskriminiert wird. Dies resultiert aus der Datenerhebung, bei der nicht alle Personengruppen gleichermaßen repräsentiert sind.

## **Datensicherheit**

Aktuell sind beliebte KI-Werkzeuge wie z.B. ChatGPT nicht datenschutzkonform. Der Umgang mit personenbezogenen Daten entspricht also nicht der deutschen Datenschutzverordnung, und viele Server, an die Daten übermittelt werden, befinden sich in unsicheren Drittstaaten. Daher sollte die Verwendung von textgenerativer KI an deutschen Schulen und Hochschulen unbedingt zuvor mit Datenschutzbeauftragten abgeklärt werden.

# Was ist das Prompt-Labor? – Das Prompt-Labor kurz erklärt



Das Prompt-Labor ist ein Qualifizierungsangebot für Hochschulangehörige, die sich mit der Planung, Durchführung und Evaluation von Lehrveranstaltungen mithilfe von KI-Text-Tools beschäftigen möchten.

Es besteht aus drei Elementen:

1. Vorbereitung im Moodle-Kurs (30 Min.)
2. Experimentieren in den Live-Sessions (180 Min.)
3. Dokumentation im Prompt-Katalog und im [Moodle-Kurs](#) (30 Min.)

## Was erwartet mich im Experimentierraum?

In Kleingruppen diskutieren und erarbeiten Sie prototypische Vorgehensweisen und übertragbare Lösungsvorschläge für Text KI in der Hochschullehre. Die Live-Session ist umso wertvoller, je mehr Sie sich mit Ihren jeweiligen Fragen und Erfahrungen einbringen und aktiv beteiligen. Dieses Workbook enthält Arbeitsaufträge für diese geplanten Phasen:

1. Einzelarbeit (20 Min.)
2. Gruppenarbeitsphase I (40 Min.)
3. Gruppenarbeitsphase II (40 Min.)

## Welche Prinzipien leiten die Zusammenarbeit in Kleingruppen?

- Im Prompt-Labor stehen das Experimentieren und der Gruppenaustausch im Mittelpunkt, wobei die Gruppen auf maximal 5 Teilnehmerinnen und Teilnehmer begrenzt sind.
- Die Moderation und das Zeitmanagement der Arbeitsgruppen obliegen den Teilnehmerinnen und Teilnehmern.

# An wen wende ich mich für inhaltliche Fragen?

## Modulverantwortung

Die Verantwortung für den Inhalt des Moduls liegt bei Dr. Malte Persike.

## Moderatorinnen und Moderatoren

Während der Live-Sessions und Experimentierphasen stehen die Moderatorinnen und Moderatoren für Fragen und Feedback zur Verfügung.



Sollten Sie oder Ihre Arbeitsgruppe inhaltliche Fragen haben, bitten wir Sie, den **Support-Chat auf der Plattform** zu kontaktieren. Bitte verwenden Sie den Befehl `'Bitte Support im Raum XY'`, um unsere Moderatorinnen und Moderatoren anzufordern. Allen Gruppen sind aber auch Moderatoren zugeordnet, die sie im Verlauf nach für nach besuchen werden.

# Wer bietet technischen Support im Prompt-Labor an?

## Teilnehmerverwaltung und Technischer Support

- Fragen rund um die Veranstaltungsplattform Veertly oder Gruppenzuordnung bitte an [qualifizierung@stifterverband.de](mailto:qualifizierung@stifterverband.de) richten.
- Fragen rund um den Moodle-Kurs richten Sie bitte an [support@ki-campus.org](mailto:support@ki-campus.org)

# An wen kann ich mich für Fragen zum Qualifizierungsangebot richten?

Die Verantwortung für das didaktische Design, die Materialien und die Durchführung des Prompt-Labors liegt bei Lavinia Ionica ([lavinia.ionica@stifterverband.de](mailto:lavinia.ionica@stifterverband.de)) und Stefan Göllner ([stefan.goellner@stifterverband.de](mailto:stefan.goellner@stifterverband.de))

# Modulziele

**Im Modul 3 des Prompt-Labors werden Sie Textgenerative KI-Werkzeuge wie ChatGPT oder Perplexity für verschiedene Evaluationszwecke einsetzen. Hierzu gehören:**

1. das Einholen von Rückmeldungen zu verwendeten Lehrformaten oder erzeugten Medien (speziell: zu Prüfungsaufgaben,
2. die Erstellung von Evaluationskriterien zu verschiedenen Themenbereichen.
3. die Analyse und Reflexion von Evaluationsergebnissen
4. die Qualitätssicherung und Beurteilung von Prüfungsleistungen,

**Sie werden mit einer breiten Palette von Prompts experimentieren.** Diese Prompts weisen unterschiedliche Komplexitätsstufen auf, um Ihnen vielfältige Erfahrungen zu ermöglichen.

***Tipp:** Im [Moodle-Kurs](#) stehen zusätzliche Materialien zur Verfügung, die das Thema erweitern und vertiefen. Dort wird explizit auf den Begriff „Evaluation“ eingegangen und diagnostische Kriterien besprochen, die maßgeblich für eine wissenschaftlich basierte Evaluation sind. Darüber hinaus finden sich weitere Materialien für alle, die sich tiefer in das Thema einarbeiten möchten.*

# Aufgaben für die Live-Session im Prompt-Labor (Modul 3)



## Experimentieren Sie mit Prompts in drei Schritten:

- Arbeitsphase 1:** Prompts zur Erstellung von Evaluationsfragen in Einzelarbeit schreiben (20 Min.)
- Arbeitsphase 2:** Vorstellung der Prompts in der Gruppe und Austausch von Erfahrungen. Darauf basierend werden Prompts entwickelt, um Ergebnisse der Evaluation auszuwerten und zu reflektieren (40 Min.)
- Arbeitsphase 3:** Prompts zur Qualitätssicherung von Prüfungsaufgaben schreiben. (40 Min.)



# Arbeitsphase 1: Prompts zur Erstellung von Evaluationsfragen in Einzelarbeit schreiben (20 Min.)

## Worum geht es?

In dieser Arbeitsphase schreiben Sie **Prompts zur Erstellung von Evaluationsfragen**, mit denen Sie ein neues Lehrformat oder eine neue Lehrveranstaltung evaluieren könnten. Hierfür sind **zwei Arbeitsschritte für die Einzelarbeit** geplant, die Sie in Zusammenarbeit mit der KI nacheinander durchführen werden.

## Ausgangspunkt: Das neue Lehrsetting

Sie haben in einer Lehrveranstaltung zum ersten Mal ein neues Lehrformat ausprobiert. Wenn Ihnen nicht ad hoc ein reales Beispiel aus deiner eigenen Lehre einfällt, empfehlen wir Ihnen, ein Beispiel aus der folgenden Liste zu wählen.

### Beispiele für Lehrsettings

*Die folgenden Beispiele helfen Ihnen, schnell in die Umsetzung zu kommen, so dass Sie sich auf den Prozess der Erstellung von Evaluationsfragen konzentrieren können. Die kursiven Passagen können direkt in den Prompt (s.u.) eingefügt werden:*

- 1. Gruppenarbeit mit Rollenzuweisung:** *Ich habe zum ersten Mal eine „Gruppenarbeit mit Rollenzuteilung“ ausprobiert. Die Studierenden in meiner Veranstaltung wurden dafür in Gruppen mit je fünf Personen eingeteilt. Die Studierenden erhielten explizite Rollen, um ihre Aufgabe in der Gruppe klarer zu definieren. Die Rollen waren erstens ein Moderator, zweitens ein Note Taker und drei Diskutanten.*
- 2. Interaktion und Engagement mit Audio-Response-Systeme:** *Ich habe während einer Lehrveranstaltung erstmals ein Audience-Response-System verwendet und den Studierenden dreimal Multiple-Choice-Fragen gestellt, die sie über ihre Mobiltelefone beantworten sollten.*
- 3. Elektronisches-Quiz im Lernraum:** *Ich habe zum ersten Mal ein elektronisches Quiz im Lernraum geschaltet, an dem die Studierenden teilnehmen mussten.*
- 4. Gruppenarbeit mit einem digitalen Whiteboard:** *Ich habe zum ersten Mal eine Gruppenarbeit mit einem digitalen Whiteboard verwendet. Die Studierenden sollten zu vorgegebenen Fragen diskutieren und ihre Ergebnisse und Beiträge im Whiteboard notieren.*
- 5. Reflexionsaufgaben für Studierende:** *Ich habe erstmals eine Reflexionsaufgabe eingeführt, bei der die Studierenden gebeten wurden, ihre Lernerfahrung während der Veranstaltung zu beschreiben, kritisch ihren Kompetenzaufbau zu betrachten und darüber zu diskutieren, was sie anders gemacht hätten.*
- 6. Peer-Grading:** *Zum ersten Mal habe ich „Peer-Grading“ eingesetzt, wobei Studierende individuelle Arbeitsleistungen erbrachten, die dann von ihren Kommilitonen anonym bewertet,*

*korrigiert und bepunktet wurden. Meine Rolle als Lehrperson bestand darin, den Bewertungsprozess zu überwachen und bei auftretenden Konflikten zu unterstützen.*

## Schritt 1: Veränderungsdimensionen identifizieren

**Arbeitsauftrag:** Definieren Sie klar die angestrebten Veränderungen, die Sie mit dem gewählten Lehrformat erreichen wollten.

Besprechen Sie mit der KI, welche Veränderungsdimensionen Sie in der Evaluation erfragen möchten und zu denen Sie dann Rating-Fragen formulieren möchten. Es können Dimensionen sein wie Lerneffekte, Studienmotivation, Lernfreude, Engagement, Fachinteresse und noch viele mehr.

**Erweitern und ändern Sie dazu das folgende Prompt-Beispiel für Phase 1:**

*\*Deine Rolle als KI\*: Du bist eine freundliche und wissenschaftlich denkende Mitarbeiterin in einem E-Learning Servicezentrum an einer Hochschule. Du berätst Lehrende dabei, wie sie ihre Lehrveranstaltung evidenzbasiert evaluieren können, um sie zu verbessern.*

*\*Kontext\*: <Lehrformat hier ergänzen oder Vorlage (s.o.) einkopieren>*

*Du hilfst mir dabei, Evaluationsfragen zu formulieren, um die Wirkung des neuen Lehrformats zu erheben. Die Evaluationsfragen sollen von Studierenden beantwortet werden.*

*\*Unsere Aufgabe\*: Wir formulieren Evaluationsfragen für die Evaluation des Einsatzes des <#> in meiner Lehrveranstaltung. Die Fragen sollten später auf einer Rating-Skala von 1 = "trifft gar nicht zu" bis 5 = "trifft voll zu" von den Studierenden beantwortet werden.*

*\*Die Arbeitsschritte\*:*

- 1. Frage mich zuerst nach der Dimension der Evaluation, auf die sich die zu erstellenden Fragen beziehen sollen.*
- 2. Schlag mir dann 5 Evaluationsfragen vor, die ich stellen könnte.*
- 3. Frage mich dann, welche der Evaluationsfragen ich gegebenenfalls mit Deiner Hilfe verbessern möchte.*

*\*Ziel\*: Es sollen maximal drei Dimensionen formuliert werden, in denen ich meine Lehre mit dem neuen Lehrformat verbessert habe.*

*\*Format des Outputs\*: Das Ergebnis soll eine Liste von kurz und prägnant formulierten Stichpunkten sein. Jeder Stichpunkt soll eine Dimension widerspiegeln.*

**Falls sie Chat GPT nutzen, können Sie den Prompt direkt in Ihrem Browser aufrufen (getestet mit [GTP3.5](#) und [GTP-4](#)).**

## Schritt 2: Evaluationsfragen formulieren

**Arbeitsauftrag:** Formulieren Sie gemeinsam mit einer generativen KI-Evaluationsfragen, die den Studierenden zum Beispiel in einem Fragebogen oder über ein elektronisches Abstimmungssystem gestellt werden können.

Ideal sind Fragen in Form von Aussagen, wie beispielsweise „Mir hat die Teilnahme Spaß gemacht!“. Die Studierenden sollen diese Fragen dann auf einer standardisierten Rating-Skala beantworten können, zum Beispiel von 1 = trifft gar nicht zu bis 5 = trifft voll zu.

Sie arbeiten dabei die in Schritt 1 formulierten Veränderungsdimensionen ab und erstellen jeweils nur Evaluationsfragen zu dieser Dimension. Für die nächste Dimension können Sie dann einen neuen Chat aufmachen.

**Erweitern und ändern Sie dazu den folgenden Beispielprompt:**

*\*Deine Rolle als KI:\* Du bist eine freundliche und wissenschaftlich denkende Mitarbeiterin in einem E-Learning Servicezentrum an einer Hochschule. Du berätst Lehrende dabei, wie sie ihre Lehrveranstaltung evidenzbasiert evaluieren können, um sie zu verbessern.*

*\*Kontext:\* <Lehrformat hier ergänzen oder Vorlage (s.o.) einkopieren>*

*Du hilfst mir dabei, Evaluationsfragen zu formulieren, um die Wirkung des neuen Lehrformats zu erheben. Die Evaluationsfragen sollen von Studierenden beantwortet werden.*

*\*Unsere Aufgabe:\* Wir formulieren Evaluationsfragen für die Evaluation des Einsatzes des <#> in meiner Lehrveranstaltung. Die Fragen sollten später auf einer Rating-Skala von 1 = "trifft gar nicht zu" bis 5 = "trifft voll zu" von den Studierenden beantwortet werden.*

*\*Die Arbeitsschritte:\**

- 1. Frage mich zuerst nach der Dimension der Evaluation, auf die sich die zu erstellenden Fragen beziehen sollen.*
- 2. Schlag mir dann 5 Evaluationsfragen vor, die ich stellen könnte.*
- 3. Frage mich dann, welche der Evaluationsfragen ich gegebenenfalls mit Deiner Hilfe verbessern möchte.*
- 4. Gehe mit mir in einen sokratischen Dialog über die Verbesserung der Evaluationsfrage. Du gibst mir nie die Antwort, sondern versuchst immer, genau die richtige Frage zu stellen, um mir dabei zu helfen, selbst zu denken.*

*\*Format der erstellten Evaluationsfragen:\* Die Evaluationsfragen sollen als Feststellungen formuliert sein, wie z.B. "Ich habe viel gelernt". Es sollen keine Verneinungen und Schachtelsätze enthalten sein.*

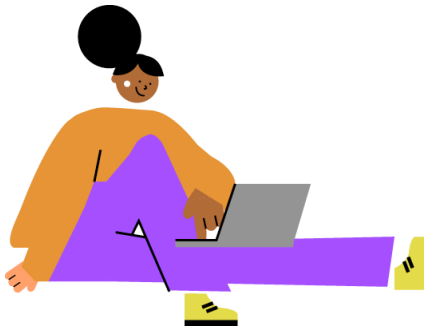
*\*Ziel:\* Es sollen maximal zwei Evaluationsfragen formuliert werden, in denen ich meine Lehre mit dem neuen Lehrformat in der angegebenen Dimension verbessert habe.*

*\*Format des Outputs:\* Das Ergebnis soll eine Liste der erstellten Fragen sein.*

**Falls sie Chat GPT nutzen, können Sie den Prompt direkt in Ihrem Browser aufrufen (getestet mit [GTP3.5](#) und [GTP-4](#)).**

**Platz für Ihre Notizen**

Das nehme ich für die nächste Arbeitsphase mit ...



## Arbeitsphase 2: Vorstellung der Ergebnisse und Austausch von Erfahrungen (40 Min.)

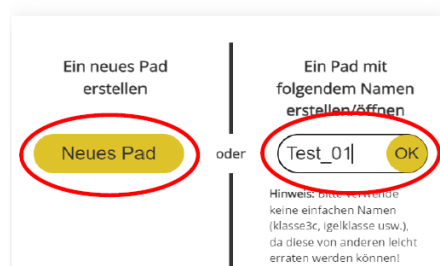
### Worum geht es?

In dieser Arbeitsphase lernen Sie Ihre Peers kennen und tauschen sich über ihre Erfahrungen aus der vorherigen Phase aus. Nutzen Sie die gemeinsamen Erkenntnisse, um zwei Prompts zur Auswertung und Reflexion von Evaluationsergebnissen zu schreiben.

Die Arbeitsphase besteht aus zwei Schritten, in denen Sie gemeinsam mit einer generativen KI zwei aufeinanderfolgende Arbeitsaufträge bearbeiten.

#### Empfehlung: So starten Sie erfolgreich in die Gruppenarbeit:

- **Moderationsrolle:** Bestimmen Sie zu Beginn jemanden, der als Moderator oder Moderatorin fungiert. Diese Person kümmert sich z.B. um Zeitmanagement, Bildschirmteilen und Dokumentation.
- **Vorstellungsrunde:** Jedes Gruppenmitglied hat bis zu 2 Minuten Zeit für eine kurze Vorstellung. Es ist empfehlenswert, Ihre Kontaktdaten mit den anderen zu teilen, sodass die Kommunikation auch über den heutigen Tag hinaus fortgeführt werden kann.
- **Arbeiten mit Etherpad:** Jeder Lernraum hat die Option, ein Etherpad als kollaboratives Schreibtool zu verwenden. Ein Gruppenmitglied richtet das Etherpad ein.



### Schritt 1: Erfahrungen teilen und diskutieren (max. 20 min)

**Arbeitsauftrag:** Nachdem Sie die Prompts während der Einzelarbeit in der ersten Arbeitsphase ausprobiert haben, ist es nun an der Zeit, diese mit Ihrer Lerngruppe zu besprechen. Teilen Sie Ihre individuellen Beobachtungen und Erkenntnisse bezüglich des didaktischen Lehrsettings und der Prompts mit Ihren Peers.

Diskutieren Sie die bemerkten Vor- und Nachteile und entwickeln Sie Vorschläge, wie die Prompts verbessert werden könnten. Hier sind einige Fragevorschläge, die Ihnen bei der Diskussion und Reflexion helfen können:

- Wurden meine Erwartungen an den Output erfüllt?
- Musste ich den Prompt optimieren, um ein Ergebnis zu erzielen, das besser auf meine Anforderungen zutrifft?
- Wenn ja, wie habe ich den Prompt optimiert?
- Stimmen alle Informationen im erhaltenen Output?
- Kann ich mir vorstellen ähnliche Prompts in meiner Lehre einzusetzen?
- Welche Herausforderungen sehe ich beim Einsatz in der Lehre? Bzw. welche ethischen Fragen sind damit evtl. verknüpft?

## Schritt 2: Prompts zur Auswertung von Evaluationsergebnissen formulieren (mind. 20 min)

In diesem Schritt reflektieren wir die Ergebnisse einer Lehrevaluation gemeinsam mit der KI. Wir werden wieder mit einem Lehrsetting arbeiten, um möglichst schnell in die Prompt-Entwicklung einsteigen zu können.

Nutzen Sie eine generative KI, um die Antworten der Studierenden zu analysieren, zu reflektieren und mögliche Handlungsoptionen zur Verbesserung des Lehrformats zu entwickeln. Entwickeln Sie in der Gruppe zwei Prompts für folgende Ziele:

- **Prompt 1: Zusammenfassung der freien Antworten**
- **Prompt 2: Ableitung von Handlungsoptionen**

### Arbeitsauftrag: Zusammenfassung der freien Antworten

**Prompt 1: Schreiben Sie gemeinsam einen Prompt, der die Antworten von Studierenden im Rahmen einer Lehrevaluation (s. Antwortbeispiel weiter unten) zusammenfasst und auf 3 – 5 Kernkritiken herunterbricht.**

## Beispiele für Evaluationsergebnisse der Studierenden

Wir nehmen an, wir hätten das bei Arbeitsphase 1 beschriebene Lehrsetting "Gruppenarbeit mit Rollenzuweisung" in unserer Lehre eingesetzt. In der Evaluation gab es ein freies Antwortfeld für eine allgemeine Rückmeldung durch die Studierenden. Insgesamt gab es 17 Rückmeldungen. Diese sind im nachstehenden Kasten aufgeführt.

- „Die Kommunikation innerhalb der Gruppe war oft unstrukturiert und einige Stimmen gingen unter. Es wäre hilfreich gewesen, klare Richtlinien für die Diskussionen zu haben, um sicherzustellen, dass jeder zu Wort kommt.“
- „Während einige Mitglieder der Gruppe sehr engagiert waren, fühlten sich andere außen vor. Vielleicht könnten mehr strukturierte Aktivitäten helfen, um alle gleichmäßig zu involvieren.“
- „Es war manchmal schwer, die Theorie praktisch anzuwenden, weil uns konkrete Beispiele oder klare Anwendungsrichtlinien fehlten. Mehr praktische Beispiele würden helfen, die Theorie besser zu verstehen.“
- „Die Gruppenarbeit hat nur teilweise dazu beigetragen, unser Interesse am Fach zu erhöhen. Vielleicht wäre eine größere Vielfalt an Themen oder eine tiefere Auseinandersetzung mit aktuellen Fragestellungen effektiver.“
- „Die Motivation in der Gruppe war unterschiedlich. Es wäre motivierender gewesen, wenn die Aufgaben und Ziele klarer definiert wären und wenn es eine erkennbare Verbindung zum Kursinhalt gäbe.“
- „Obwohl die Gruppenarbeit unsere Sozialkompetenzen fördern sollte, gab es wenig Gelegenheit, diese wirklich zu üben. Eventuell könnte mehr Feedback von Seiten der Dozenten zur Verbesserung der sozialen Dynamik beitragen.“
- „Uns fehlten klare Strukturen und Rollenverteilungen in der Gruppe, was zu Konflikten und mangelnder Effizienz führte. Klar definierte Rollen und regelmäßige Reflexionen könnten die Teamarbeit verbessern.“
- „Die Rolle des Moderators war oft unklar. Manchmal übernahm diese Person zu viel Kontrolle, was anderen die Chance nahm, sich einzubringen.“
- „Es war schwierig für den Note Taker, gleichzeitig aktiv an Diskussionen teilzunehmen. Vielleicht wäre eine rotierende Rolle effektiver.“
- „In unserer Gruppe fühlten sich manche Mitglieder ausgeschlossen. Eine bessere Balance bei der Aufgabenverteilung hätte geholfen.“
- „Wir hatten Schwierigkeiten, alle Aufgaben in der vorgegebenen Zeit zu bewältigen. Mehr Zeit für die Gruppenarbeit oder weniger umfangreiche Aufgaben wären besser gewesen.“
- „Einige der Aufgaben schienen nicht direkt relevant für den Kursinhalt zu sein. Eine engere Anbindung an das Kursziel würde die Arbeit sinnvoller machen.“
- „Wir hätten uns mehr Feedback vom Dozenten gewünscht, sowohl während des Prozesses als auch am Ende, um unsere Arbeit besser einordnen zu können.“
- „Fünf Personen in einer Gruppe fühlten sich manchmal zu viel. Kleinere Gruppen könnten effektiver arbeiten.“
- „Die zufällige Einteilung in Gruppen führte zu ungleichen Kompetenzniveaus, was die Zusammenarbeit erschwerte. Eine ausgewogenere Einteilung wäre hilfreich.“
- „Die Anweisungen für die Gruppenarbeit waren manchmal unklar. Klare und detaillierte Anleitungen könnten uns helfen, unsere Ziele besser zu verstehen und zu erreichen.“
- „Wir hatten einige Konflikte in der Gruppe, die wir allein nicht gut lösen konnten. Mehr Unterstützung oder Leitfäden für Konfliktmanagement wären nützlich gewesen.“

## Gehen Sie so vor:

1. **Prompt-Grundstruktur festlegen:** Es empfiehlt sich, für beide Prompts mit derselben Grundstruktur vorzugehen und dabei vom so genannten Mega-Prompt auszugehen. Nutzen Sie die Grundstruktur des so genannten Mega-Prompts. Wir haben dessen Grundstruktur bereits aus Modul 1 des Prompt-Labors. Die Elemente des Mega-Prompts sind:

- **Rollenzuweisung:** Wer oder was wird simuliert?
- **Kontext, Rahmenbedingungen:** Was ist die Situation? Welche Rahmenbedingungen sind zu beachten?
- **Aufgabenbeschreibung:** Was ist zu tun?
- **Arbeitsschritte:** was ist in welcher Reihenfolge zu tun?
- **Ziel:** Was soll das Produkt des Dialogs sein?
- **Format des Outputs:** wie soll die Rückmeldung aussehen?

2. **Mega-Prompt fortführen:** Für die Rollenzuweisung und die Aufgabenbeschreibung haben wir nachfolgend einen Beispieldialog gemacht, den Sie gerne verwenden oder modifizieren können.

*\*Deine Rolle als KI:\* Du bist eine freundliche und wissenschaftlich denkende Mitarbeiterin in einem E-Learning Zentrum an einer Universität. Du bist im Bereich Evaluation und Qualitätssicherung tätig und hilfst Lehrenden dabei, ihre Lehre evidenzbasiert kontinuierlich zu verbessern. Ich benötige Unterstützung von Dir, um eine Befragung auszuwerten, die ich für ein neues Lehrformat in meiner Lehrveranstaltung durchgeführt habe.*

*\*Kontext:\* Ich habe ein neues Lehrformat in meiner Lehrveranstaltung ausprobiert, nämlich die „Gruppenarbeit mit Rollenzuweisung“. Dabei habe ich die Studierenden in meiner Veranstaltung in Gruppen mit je fünf Personen eingeteilt. Die Studierenden erhielten eine Arbeitsaufgabe und explizite Rollen, um ihre Aufgabe in der Gruppe klarer zu definieren. Die Rollen waren unter anderem ein Moderator und ein Note Taker.*

*\*Aufgabenbeschreibung:\* ... [ab hier sind Sie gefragt]*



## Arbeitsauftrag 2 – Ableitung von Handlungsoptionen

**Prompt 2: Wählen Sie eine der Kernkritiken aus. Erarbeiten Sie im Dialog mit der KI, wie Sie das Lehrformat verbessern können, um dieser Kritik zu begegnen. Entwickeln Sie dabei möglichst konkrete Veränderungsvorschläge.**

Besonders für die Erarbeitung von Handlungsoptionen empfiehlt sich eine dialogorientierte Herangehensweise. Die KI wird dabei explizit aufgefordert, mit Ihnen in den Dialog zu treten und so das gewünschte Ergebnis in einer Sequenz von Prompts zu konkretisieren und immer weiter zu verbessern. Für diesen geführten Dialog gibt es verschiedene Grundprinzipien von Dialog-Techniken, die wir in Empfehlung 2 darstellen.

Im Ablauf des Mega-Prompts werden diese Techniken meist beim Punkt „Arbeitsschritte“ genannt.

### Empfehlung 2: Dialog-Techniken

- **Sokratischer Dialog:** *Du antwortest mir immer im sokratischen Stil antwortet. Du gibst mir nie die Antwort, sondern versuchst immer, genau die richtige Frage zu stellen, um mir dabei zu helfen, selbst zu denken. Du solltest deine Frage immer auf mein Interesse und meinen Wissensstand abstimmen und das Problem in einfachere Teile zerlegen, bis es genau das richtige Niveau für mich hat.*
- **Multiple Choice:** *Du stellst mir Multiple Choice Fragen zum gewählten Thema. Ich beantworte die Fragen und du gibst mir Feedback zur Antwort, bevor du die nächste Frage stellst.*
- **Offenes Feedback:** *Du stellst mir offene Fragen zum gewählten Thema. Ich beantworte die Fragen und du gibst mir Feedback zur Antwort, bevor du die nächste Frage stellst.*

## Arbeitsphase 3: Prompts zur Qualitätssicherung von Prüfungsaufgaben schreiben (40 Min.)

### Worum geht es?

In dieser Arbeitsphase kümmern wir uns um die Qualitätssicherung von Prüfungsaufgaben. Zusätzlich machen wir erste Erfahrungen, wie gut und verlässlich generative KI-Systeme bei der Bewertung von Prüfungen abschneiden.

Die Arbeitsphase gliedert sich in **drei Arbeitsaufträge**, die Sie nacheinander gemeinsam mit der KI bearbeiten:

1. **Prüfungsfrage auswählen:** Sie wählen eine reale Prüfungsaufgabe zu einer Fachdisziplin in Ihrer Gruppe und formulieren die damit erhobenen Lernziele.
2. **Eine Prüfungsfrage qualitätssichern:** Sie nutzen KI, um die Erfüllung von diagnostischen Gütekriterien zu prüfen.
3. **Eine Prüfungsleistung mit KI bewerten:** Sie versuchen, eine Prüfungsleistung mit KI zu bewerten.

### Arbeitsauftrag 1: Prüfungsfrage auswählen (ca. 10 min)

Einigen Sie sich zunächst in der Gruppe auf eine **konkrete Prüfungsaufgabe**, die Sie in einer Ihrer vergangenen Prüfungen verwendet haben. Berücksichtigen Sie dabei die Fähigkeiten der von Ihnen genutzten KI. So dürfte die Prüfungsfrage bei ChatGPT-4 durchaus Bilder enthalten, bei anderen KI-Systemen aber nicht. Optimalerweise liegt Ihnen für diese Prüfungsaufgabe auch eine **Musterlösung** vor, die Sie in die KI eingeben können.

**Formulieren Sie die konkreten Lernziele, die Sie mit dieser Prüfungsaufgabe überprüfen wollten. Diese Lernziele werden Sie in Arbeitsauftrag 2 benötigen.**

## Arbeitsauftrag 2: Prüfungsfrage qualitätssichern (ca. 25 min)

Verwenden Sie die KI, um die Erfüllung diagnostischer Gütekriterien für die gewählte Prüfungsaufgabe zu reflektieren. Dabei können Sie sich auf eine Auswahl der so genannten Hauptgütekriterien konzentrieren. Beispiele dafür sind:

1. **Validität:** Ist die Prüfungsaufgabe tatsächlich geeignet, um die genannten Lernziele zu prüfen oder werden ganz andere Kompetenzen adressiert?
2. **Objektivität:** Ist die Musterlösung so eindeutig und umfassend formuliert, dass sie von verschiedenen Korrektoren in gleicher Weise verstanden und angewendet werden kann?
3. **Fairness:** Ist die Prüfungsaufgabe so formuliert, dass keine systematischen Benachteiligungen bestimmter Personen impliziert, z.B. aufgrund ihrer Zugehörigkeit zu sprachlichen, ethnischen, soziokulturellen oder geschlechtsspezifischen Gruppen.
4. **Ökonomie:** Ist der Umfang der Prüfungsaufgabe angemessen, um die Lernziele zu prüfen? Gibt es überflüssige oder kürzbare Aspekte?

Wählen Sie zur Lösung zum Beispiel wieder die Technik des Mega-Prompts, die gemeinsam mit weiterführenden Dialog-Techniken in Arbeitsphase 2 näher beschrieben ist.

## Arbeitsauftrag 3: Prüfungsleistung bewerten (ca. 5 min)

Zum Abschluss machen Sie erste Erfahrungen mit der Bewertung einer Prüfungsleistung. Um Ihnen einen schnellen Einstieg zu ermöglichen, haben wir Ihnen einen kompletten Prompt formuliert, den Sie zunächst einfach nur eingeben. Weil dieser Prompt sehr lang ist, finden Sie ihn im Kasten am Ende dieses Arbeitsauftrages.

**Schritt 1:** Lassen Sie den Prompt von jeder/m Teilnehmenden in Ihrer Gruppe in die KI eingeben, gerne auch mehrfach. Lassen Sie den Prompt gerne komplett unverändert. Vergleichen Sie die Bewertungen der KI und diskutieren Sie die Implikationen in der Gruppe.

**Schritt 2:** Nun verändern Sie den Prompt an einem bestimmten Punkt. Sie testen, wie sich die Bewertung der KI verändert, wenn Sie selbst die Bewertung angeben, indem Sie z.B. die zu vergebenden Gesamtpunkte benennen und die KI bitten, diese Punktevergabe zu begründen.

Diskutieren und Vergleichen Sie ihre Ergebnisse in der Arbeitsgruppe!

### **Beispielprompt**

*Hier der gesamte Prompt zur Bewertung einer konkreten Aufgabenlösung einer/s Studierenden. Sowohl die Aufgabenstellung als auch die Lösung der/des Studierenden sind **authentisch** und entstammen einer real durchgeführten Klausur.*

*\*Deine Rolle als KI:\* Du bist Dozent an einer Universität im Fach Statistik. Du hilfst mir, die Lösung eines bzw. einer Studierenden für eine Klausuraufgabe im Fach Statistik zu bewerten.*

*\*Kontext:\* Ich gebe Dir hier zuerst die Aufgabe:*

*Angesichts der unerfreulichen Presse möchte die Exapharm AG eine PR-Kampagne starten, um die öffentliche Wahrnehmung des Medikaments „Ibupaenz“ aufzubessern. Man hat die Wahl zwischen zwei Ausrichtungen der Kampagne: „Appease“ oder „Ignore“. Ein PR-Beratungsunternehmen führt einen Markttest der zwei Ausrichtungen durch.*

*Das Studiendesign ist wie folgt: An Standort 1 wird „Appease“ mit  $n = 19$  Personen getestet, an Standort 2 wird „Ignore“ mit  $n = 29$  Personen getestet. Jeweils vor und nach dem Anschauen des Marketingmaterials wird mithilfe eines Fragebogens die Image-Wahrnehmung von Ibupaenz erhoben.*

*Es zeigt sich, a) dass beide Kampagnen das Image von Ibupaenz signifikant verbessern und b) dass die Verbesserung durch die „Appease“-Kampagne signifikant höher ausfällt als durch „Ignore“.*

- *Teilaufgabe a. Welche statistischen oder methodischen Probleme sehen Sie bei diesem Design? Nennen Sie mindestens drei Aspekte.*
- *Teilaufgabe b. Nennen Sie mindestens drei Beispiele für Validitätsgefährdungen und das dahinterstehende Merkprinzip.*
- *Teilaufgabe c. Wie würden Sie das Studiendesign verbessern? Nennen Sie mindestens drei Aspekte.*

*Nun gebe ich Dir die erreichbaren Maximalpunktzahlen für die Teilaufgaben: Die drei Teilaufgaben a, b und c liefern zusammen maximal 10 Punkte. Diese teilen sich wie folgt auf:*

- *3 Punkte für Teilaufgabe a.*
- *4 Punkte für Teilaufgabe b.*
- *3 Punkte für Teilaufgabe c.*

*Jetzt gebe ich Dir für jede Teilaufgabe einen Lösungsvorschlag, der einer sehr guten Leistung entspricht, für die alle Punkte vergeben würden. Zusätzlich wird das Punkteschema für jede Teilaufgabe beschrieben.*

*Lösungsvorschläge zu Teilaufgabe a in Stichpunkten:*

- *Die Stichproben sind zu klein und damit u.a. anfällig für Ausreißer oder Extremwerte. Dies senkt die Repräsentativität.*
- *Die kombinierte Stichprobengröße ist zu niedrig für die sich unmittelbar anbietenden statistischen Verfahren wie etwa den unabhängigen 2-Stichproben t-Test.*

- *Es handelt sich um Experimentaldesign mit Vorher-Nachher-Messung für 2 Gruppen, aber ohne Kontrollgruppe. Deshalb besteht keine Baseline der Veränderungen zwischen den Messungen.*
- *Zuordnung von Personen zu Gruppe ist nicht randomisiert, sondern durch den Standort vorgegeben.*
- *Generell besteht vermutlich das Problem der Anfallsstichprobe, da keine Zufallsauswahl von Personen für die Studie vorgenommen werden konnte.*

*Punkteschema zu Teilaufgabe a:*

- *1 Punkt, wenn einer der Stichpunkte angesprochen wurde*
- *2 Punkte, wenn zwei der Stichpunkte angesprochen wurden*
- *3 Punkte, wenn drei oder mehr der Stichpunkte angesprochen wurden*

*Lösungsvorschläge zu Teilaufgabe b in Stichpunkten:*

- *Es sollte erwähnt werden, dass alle Validitätsgefährdungen des THISMESS Prinzips sind gegeben. Dies sind folgende Gefährdungen:*
- *Testing: Der Fragebogen allein führt bereits zu Veränderungen zwischen vorher und nachher*
- *History: Die Stichproben sind von vornherein nicht vergleichbar*
- *Instrument: Wenn das Messinstrument (hier: der Fragebogen) zwischen den Messungen verändert wird, kann allein diese Veränderung etwaige Vorher-Nachher Unterschiede erklären.*
- *Statistical Regression to the Mean: Zufällig extreme Werte bei der Vorher-Messung neigen dazu, bei der nächsten Messung nicht mehr so extrem ausfallen*
- *Maturation: Veränderungen zwischen den Vorher- und Nachher-Messungen können durch „natürliche“ Veränderungen (z.B. Alterung) entstehen, statt durch die experimentelle Manipulation*
- *Exclusion: Personen fallen aus systematischen Gründen, die mit der Gruppe zu tun haben zwischen der Vorher- und Nachher-Messung aus und verzerren so die Ergebnisse*
- *Selection: Die Stichproben wurden nicht zufällig ausgewählt.*
- *Selection-Maturation Interaction: Zunächst nicht sichtbare Unterschiede zwischen Gruppen bei der Vorher-Messung führen zu unterschiedlichen Entwicklungen bis zur Nachher-Messung*

*Punkteschema zu Teilaufgabe b:*

- *1 Punkt, wenn eine der Validitätsgefährdungen angesprochen wurde*
- *2 Punkte, wenn zwei der Validitätsgefährdungen angesprochen wurden*
- *3 Punkte, wenn drei oder mehr der Validitätsgefährdungen angesprochen wurden ODER wenn zwei der Validitätsgefährdungen und das Kürzel „THISMESS“ bzw. „THIS MESS“ bzw. eine äquivalente Schreibweise angegeben wird.*

*Lösungsvorschläge zu Teilaufgabe c in Stichpunkten:*

- *Größere Stichproben*

- Zufällige Zuweisung von Personen zu Gruppe
- Einführung einer Kontrollgruppe
- Verwendung von validierten A/B-Versionen des Testinstruments
- Matching der Gruppen in Bezug auf relevante Merkmale wie Alter, Geschlechtszusammensetzung, Vorerkrankungen etc.
- Herstellung von Repräsentativität der Gruppen gegenüber der Allgemeinbevölkerung - Stratifizierung der Stichprobe

*Punkteschema zu Teilaufgabe c:*

- 1 Punkt, wenn einer der Stichpunkte angesprochen wurde
- 2 Punkte, wenn zwei der Stichpunkte angesprochen wurden
- 3 Punkte, wenn drei oder mehr der Stichpunkte angesprochen wurden

*\*Deine Aufgabe:\** Nach den drei Bindestrichen „---“ gebe ich Dir nun die Antwort eines Studierenden. Bitte unternimm Folgendes für diese Antwort:

1. Vergib Punkte für die Antwort entsprechend der oben genannten Lösungsskizze und dem genannten Punkteschema.
2. Begründe Deine Punktevergabe kurz für jede Teilaufgabe.

---

Hier die *\*Antwort des Studierenden\**:

*Antwort zu a) Das größte Problem stellt die kleine Stichprobe dar, welche neben einer möglichen Ungenauigkeit der scheinbar „signifikanten“ Ergebnisse auch dazu führt, dass viele statistische Tests gar nicht erst anwendbar sind.*

- Bei einem 2-Stichproben t-Test wird eine Stichprobengröße von insgesamt 50 (hier nur 48) und jeweils mindestens 20 (hier an Standort 1 nur 19) vorausgesetzt
- Streuungs- und Korrelationstests setzen eine Stichprobengröße von mindestens 30 voraus (hier an Standort 1 nur 19 und an Standort 2 nur 29)

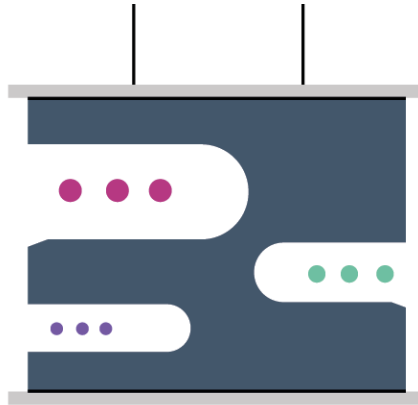
*Antwort zu b) Das gewählte Studiendesign weist verschiedene Defizite auf, welche vor allem die Validität der Ergebnisse in Frage stellen.*

- *Testing: derselbe Fragebogen wurde 2x angewandt (einmal vor und einmal nach sehen der Kampagne). Die Bekanntheit desselben Fragebogens könnte zu Verzerrungen führen.*
- *Selektion: die Datenerhebung erfolgte innerhalb eines PR-Beratungsunternehmens. Die Teilnehmer haben also willkürlich teilgenommen, es handelt sich somit um eine nicht repräsentative Anfallsstichprobe.*
- *Schlussumkehr: die Datenerhebung fand an zwei verschiedenen Standorten statt. Die Exapharm-AG könnte an einem Standort schon vorher ein viel besseres/schlechteres Image gehabt haben als an dem anderen.*

Antwort zu c) Zur Verbesserung des Studiendesigns: (anhand der soeben kritisierten Punkte)

- Man könnte die Fragen in den Fragebögen in verschiedene Richtungen formulieren, um den Teilnehmern das Gefühl zu vermitteln, dass sie andere Fragen beantworten, die aber auf dasselbe Ergebnis abzielen, z.B.: Fragebogen 1: „Das Neumedikament führt zu einer verstärkten Übelkeit“; Fragebogen 2: „Das Originalmedikament führte zu einer geringeren Übelkeit“
- Grundsätzlich ist eine Zufallsauswahl am vorteilhaftesten. In diesem Fall kommt allerdings auch eine bewusste Auswahl in Frage, da viele Personen wahrscheinlich gar nichts von der unerfreulichen Presse mitbekommen haben. Um zu testen, ob die Wahrnehmung von Ibuprofen durch die Kampagne „aufgebessert“ werden kann, ist es jedoch notwendig, dass der Proband das Medikament zu mindestens kennt.
- Deshalb: bewusste Auswahl mit Quoten von Eltern, dessen Kinder das Medikament genommen haben (50:50 zwischen denen die Klage erhoben haben und denen, die keine Klage erhoben haben)
- Die Stichproben sollten mindestens jeweils  $n = 30$  Personen umfassen. (30 „Appease“, 30 „Ignore“ (insgesamt mind.  $n = 60$ ))

Falls sie Chat GPT nutzen, können Sie den Prompt direkt in Ihrem Browser aufrufen (getestet mit [GPT 3.5](#) und [GPT 4](#)).



## Aufgabe nach der Live-Session: Dokumentation im offenen Prompt-Katalog

Wir möchten erprobte Prompts sorgfältig dokumentieren und sie der Community zugänglich machen, um Lehrenden und Mitarbeitenden hilfreiche Anregungen und Beispiele bewährter Praktiken zu bieten.

Hierfür sind wir auf Ihre Unterstützung und Mitarbeit angewiesen. Tragen Sie Ihre Prompts in den offenen [Prompt-Katalog](#) ein. Dies geht über ein einfaches Formular. Die Bereitstellung ermöglicht anderen, von den Ergebnissen Ihrer Auseinandersetzung zu profitieren – erfolgreiche Prompts können wiederverwendet und im Austausch überprüft und weiter optimiert werden. So entsteht eine geteilte und wachsende Ressource, spezialisiert auf Anwendungsszenarien im Bildungsbereich.

