

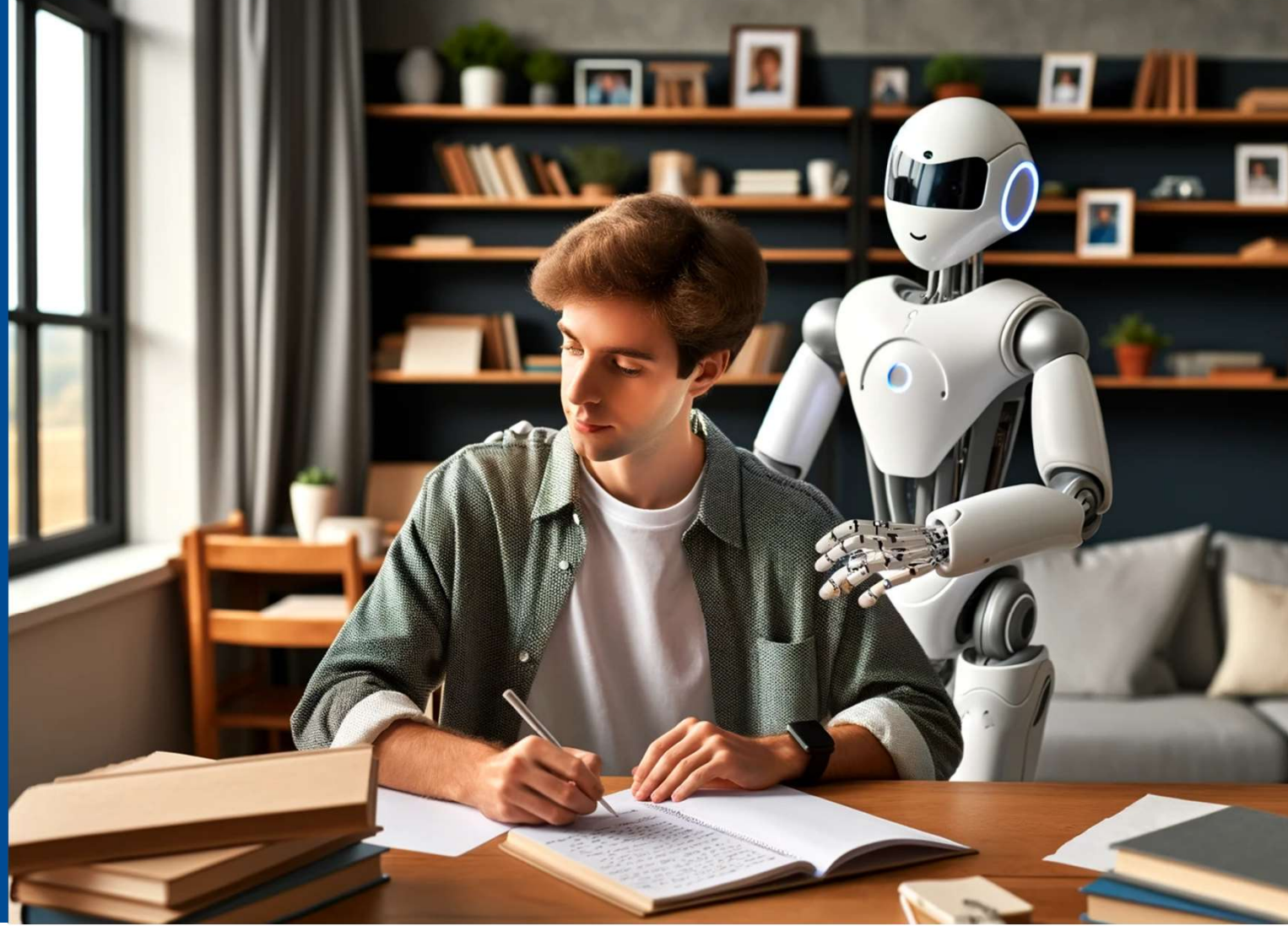
Prüfungsformen im KI-Zeitalter

Eine didaktische und operative
Bewertung



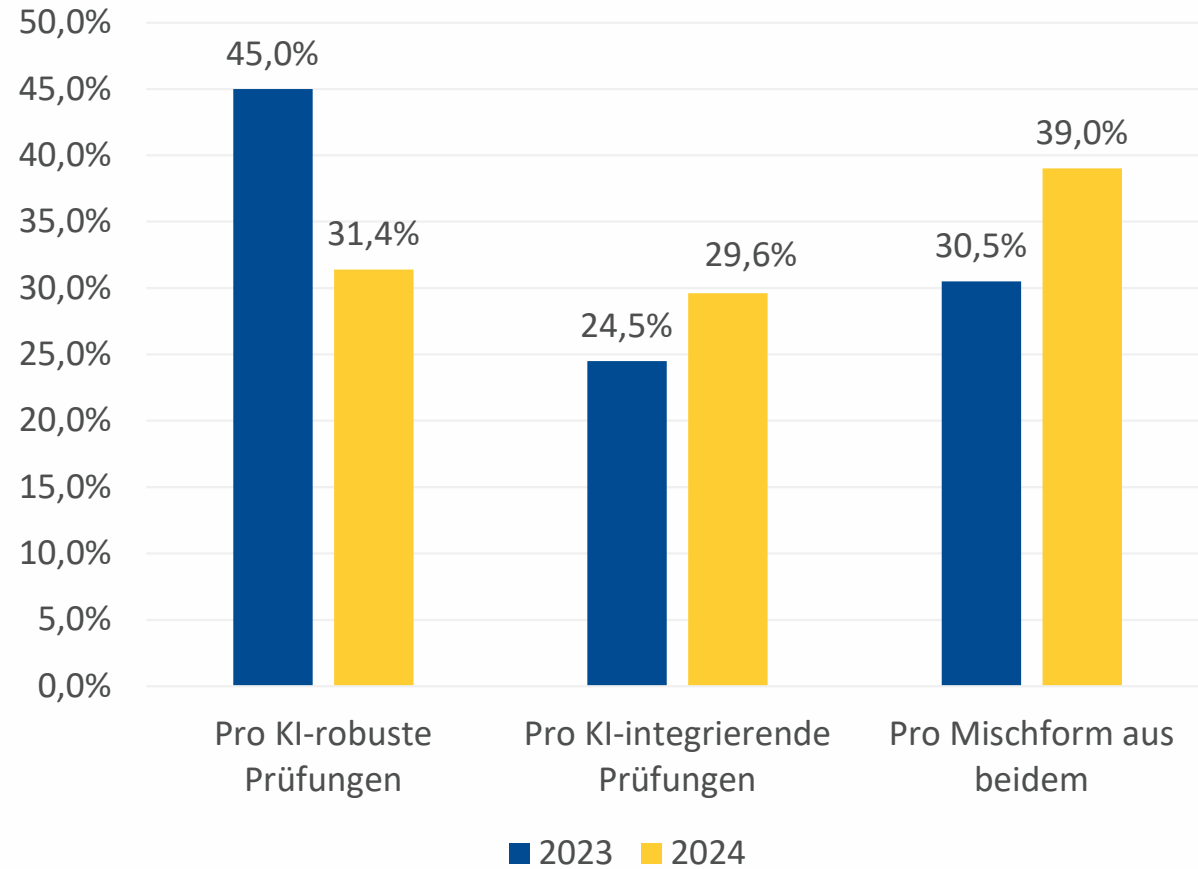
EURO-FH
UNIVERSITY OF APPLIED SCIENCES

Wie prüfen wir künftig?



Befragung von Studierenden und Lehrenden

Sommer 2023/
Sommer 2024



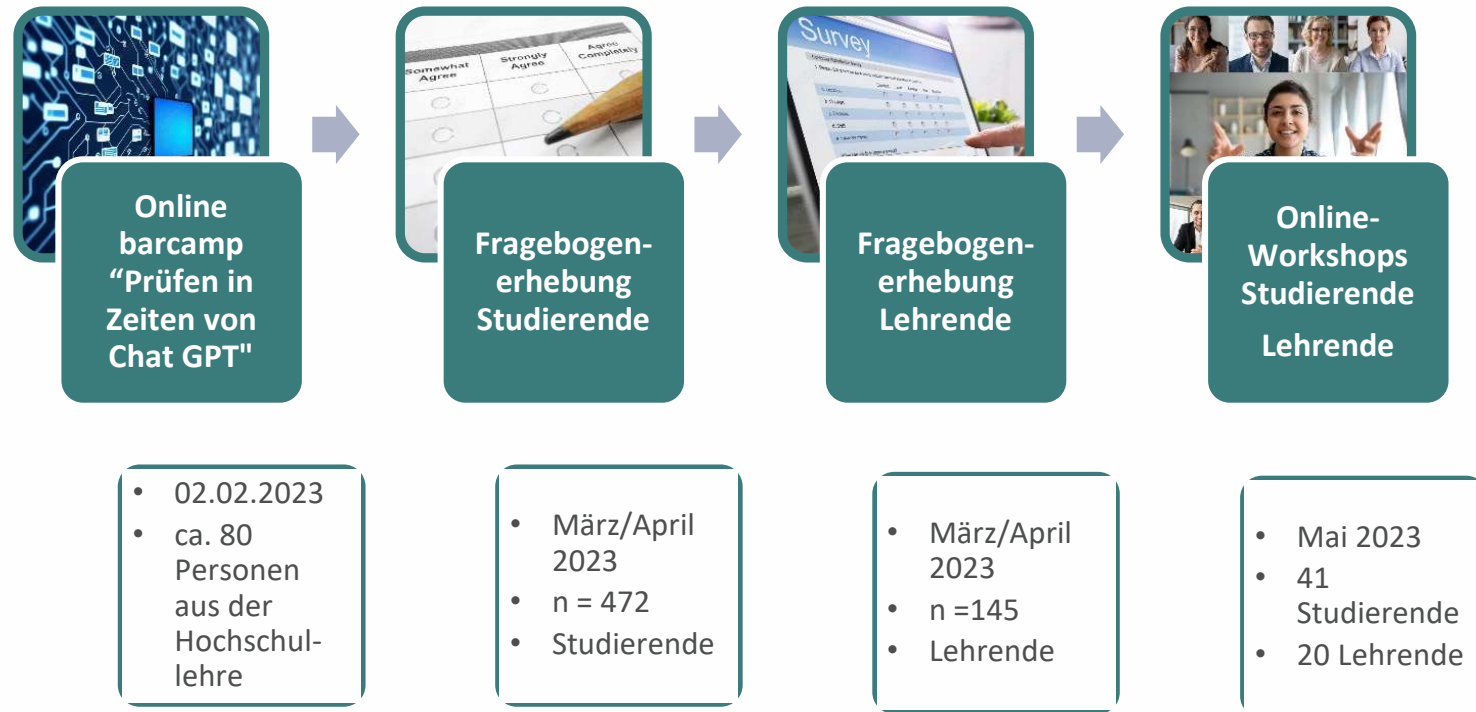
Zentrale Fragestellung

Welche Prüfungsformen sind im Zeitalter von textgenerierender KI möglich sowie didaktisch und praktisch sinnvoll?

Methodik

Zwei Gruppen:

- Studierende
- Lehrende



Mögliche Prüfungsformen



Überwachte Prüfungen

- Klausuren
- Mündliche Prüfungen
- Kombinationsprüfungen mündlich/schriftlich

Neue Prüfungsformen

- Videos
- Experimente
- Rollenspiele/Streitgespräche
- Formative Prüfungen (z.B. E-Portfolios)
- Peer Review
- Gruppenprüfungen
- Poster und Landkarten

Angepasste Fragestellung

- Transfer- und Anwendungsaufgaben
- Fokus auf Empirie und Quellenarbeit
- Fallstudien
- Fiktive Forschungsanträge
- Prozessdokumentation
- Reflexion

KI-Integration

- KI-Output überprüfen
- Prüfen von KI-Kompetenzen
- Prüfungen, bei denen die Nutzung von KI Teil der Aufgabenstellung ist

Didaktische Bewertung: Lerntaxonomie

Erinnern	Verstehen	Anwenden	Analysieren	Bewerten	Erschaffen
• erkennen • abrufen	• interpretieren • erläutern • einordnen • zusammenfassen • schlussfolgern • vergleichen • erklären	• ausführen • implementieren	• differenzieren • organisieren • zuordnen	• überprüfen • kritisieren	• generieren • planen • produzieren

(Anderson & Krathwohl, 2001)

Was funktioniert in der Praxis?



Mögliche Problemstellungen

- Didaktischer und administrativer Aufwand für Einführung
- Anpassung der Prüfungsordnungen
- Neudefinition Bewertungskriterien
- DSGVO-konforme Umsetzung
- Technische und didaktische Infrastruktur
- Hoher Schulungsbedarf
- Laufender Anpassungsbedarf

Klausuren

Pro

- Einfach umsetzbar, da etabliert
- „KI-sicher“

Contra

- Didaktisch wenig sinnvoll, da reine Wissensabfrage auf unteren Taxonomiestufen („Bulimielernen“)
- Lehrt kein eigenständiges wissenschaftliches Arbeiten

Erinnern

Verstehen

Anwenden

Analysieren

Bewerten

Erschaffen

Mündliche Prüfungen und Kombinations- prüfungen

Pro

- Großes didaktisches Spektrum möglich
- „KI-sicher“

Contra

- Gefahr, dass Prüfung reine Wissensreproduktion bleibt
- Lehrt kein wissenschaftliches Arbeiten
- Hoher administrativer Aufwand
- Dauerhaft hohe Aufwände seitens Lehre
 - Ständig neue Fragen nötig
 - Bei Kombinationsprüfung Einarbeitung in individuelle Prüfungsinhalte

Erinnern

Verstehen

Anwenden

Analysieren

Bewerten

Erschaffen

Angepasste Fragestellung/ Bewertung

- Transfer- und Anwendungsaufgaben
- Fokus auf Empirie und Quellenarbeit
- Fallstudien
- Fiktive Forschungsanträge
- Prozessdokumentation
- Reflexion

Pro

- Umsetzbar innerhalb von Haus- und Projektarbeiten, d.h. vorhandener Prüfungsformate
- Transfer und Anwendung sind didaktisch sinnvoll.
- Didaktisch wertvolle Aufgabenstellungen sind möglich.

Contra

- Hoher Aufwand bei Einführung.
- Mit hoher Wahrscheinlichkeit ständiger Anpassungsbedarf an die sich weiterentwickelnde KI
- Kampf „gegen die KI“, der vermutlich verlorenght
- Die Nutzung von KI wird nicht ausgeschlossen, Studierende können der KI trotzdem das Schreiben überlassen.

Erinnern

Verstehen

Anwenden

Analysieren

Bewerten

Erschaffen

Neue Prüfungsformate

- Videos
- Experimente
- Rollenspiele/
Streitgespräche
- Formative Prüfungen
(z.B. E-Portfolios)
- Peer Review
- Gruppenprüfungen
- Poster und Landkarten

Pro

- Ermöglichen vielfältige didaktische Szenarien

Contra

- Hoher Aufwand bei der Einführung (administrativ und seitens der Lehre)
- Notwendigkeit, Bewertungskriterien neu zu definieren
- Teilweise hoher Aufwand bei der Durchführung und Bewertung
- Nicht per se „KI-sicher“
- Einige Formate bevorzugen technikaffine Studierende

Erinnern

Verstehen

Anwenden

Analysieren

Bewerten

Erschaffen

KI-Integration

- KI-Output überprüfen
- Prüfen von KI-Kompetenzen

Pro

- Didaktisch wertvolle Prüfungsform
- KI-Kompetenzen werden vermittelt → zukunftsweisend
- Arbeitet mit der KI anstatt gegen sie

Contra

- Hoher Aufwand bei Einführung (administrativ und technisch)
- Schulungsbedarfe
- Notwendigkeit, Bewertungskriterien neu zu definieren
- Ggf. Notwendigkeit zur Weiterentwicklung bei Entwicklung der KI

Erinnern

Verstehen

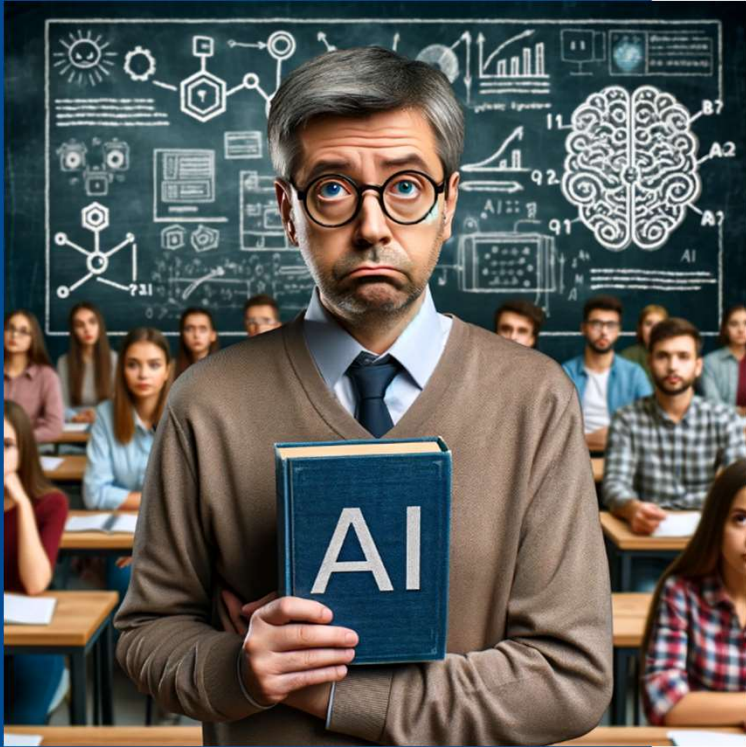
Anwenden

Analysieren

Bewerten

Erschaffen

Takeaways



- Es gibt nicht die eine (einfache) Lösung.
 - Die Mischung macht es?
 - Prüfungsformate werden in den nächsten Jahren einem ständigen Anpassungsprozess unterlegen sein.
 - Es ist vermutlich zielführender, mit der KI zu arbeiten als gegen sie.
- Zentral ist, Prüfungen so zu konzipieren, dass KI-Nutzung nicht bedeutet, dass keine Eigenleistung erbracht werden muss.

KI-integrative Prüfungen: Was muss beachtet werden?



DSGVO-konformer Zugang

- Wenn KI-Tools durch die Studierenden genutzt werden sollen, muss ein DSGVO-konformer Zugang zur Verfügung gestellt werden.
- Erhebung und Verarbeitung nur der unbedingt notwendigen personenbezogenen Daten.
- Schutz der Identität der Studierenden durch Anonymisierungsmaßnahmen.
- Transparente Datenverarbeitung (Datenschutzerklärung)

Prüfungsgestaltung

- Entwicklung von Aufgaben, die trotz KI-Nutzung eine Eigenleistung erfordern.
- Klare Richtlinien, welche Tools in welchem Umfang genutzt werden dürfen.

Prüfungsordnung

- Überprüfung und ggf. Aktualisierung

Bewertungskriterien

- Entwicklung von tragfähigen Beurteilungskriterien
- Klare Kommunikation der Kriterien und deren Gewichtung an die Studierenden.
- Schulung der Prüfenden, um eine konsistente Anwendung der Kriterien zu gewährleisten.

Arten von KI- integrierenden Prüfungen

1. Dialog/Diskussion/Fachdiskurs/Prüfungsgespräch mit Bot

- interaktives Gespräch zwischen dem Studierenden und einem KI-gesteuerten Chatbot.
- Ziel ist es, Fachwissen zu demonstrieren, kritisches Denken anzuwenden und die Fähigkeit zur Interaktion mit KI-Systemen zu zeigen.

2. Entwicklung mit KI mit anschließender kritischer Reflexion/Bearbeitung

- KI-Tools werden eingesetzt, um ein bestimmtes Produkt oder eine Lösung zu entwickeln.
- Anschließend reflektieren die Studierenden kritisch über den Entwicklungsprozess und das Ergebnis, identifizieren Verbesserungsmöglichkeiten und nehmen gegebenenfalls Anpassungen vor.
- Die Prüfung bewertet sowohl die Kompetenz im Umgang mit KI-Technologien als auch die Fähigkeit zur kritischen Analyse und Weiterentwicklung der erstellten Arbeit.

3. Analyse und Kritik von KI-Erzeugnissen

- von KI generierte Inhalte sind zu analysieren und zu bewerten (z.B. Texte, Bilder, Datenanalysen).
- Die Aufgabe besteht darin, die Qualität, Genauigkeit und Relevanz der KI-Erzeugnisse zu prüfen, mögliche Fehler oder Verzerrungen zu identifizieren und deren Implikationen zu diskutieren.
- Ziel ist es, das Verständnis für die Funktionsweise von KI sowie die Fähigkeit zur kritischen Bewertung von KI-generierten Inhalten zu fördern.

1. Dialog/Diskussion/ Fachdiskurs/Prüfungs- gespräch mit Bot

Descartes vertrat die Ansicht, dass die Fähigkeit zur Sprachverwendung ein "unreduzierbares" Merkmal des Geistes sei, dass es das Kennzeichen menschlicher Mentalität sei und dass es "nicht vorstellbar" sei, dass eine Maschine dazu in der Lage sein könne. Beweist ChatGPT, dass er falsch lag?

Die Studierenden führen ein Gespräch mit einem LLM über dieses Thema (sie könnten es sogar bitten, die "Stimme" von Descartes zu imitieren), und prüfen, welche Aspekte und Argumentationen in dem Gespräch für die Sicht von Descartes und welche dagegen sprechen. Ergeben sich aus dem Dialog andere unreduzierbare Merkmale menschlicher Mentalität?

Die Studierenden führen einen Turing-Test durch (d.h. spielen das "Imitationsspiel,,) mit ChatGPT. Besteht es oder fällt es durch? Wie und warum?

Welche Arten von Gesprächen würden es wahrscheinlicher machen, dass es besteht oder durchfällt? Was zeigt das über den Turing-Test? Was zeigt das über ChatGPT? Die Studierenden diskutieren dies schriftlich.

1. Dialog/Diskussion/ Fachdiskurs/Prüfungs- gespräch mit Bot

Eine Studentin interagiert mit einem KI-Bot, der als **fiktiver CEO eines konkurrierenden Unternehmens** auftritt. In einer simulierten Verhandlungssituation müssen beide Parteien Entscheidungen über Preise und Produktionsmengen treffen. Der Bot agiert strategisch gemäß den Prinzipien des Nash-Gleichgewichts, fordert die Studentin heraus und zwingt sie, ihre Entscheidungen zu begründen und die möglichen Reaktionen des Gegenübers zu antizipieren.

Im Prüfungsgespräch kommuniziert ein Student mit einem KI-Bot, der die Persönlichkeit von **Adam Smith** verkörpert. Der Bot fordert den Studenten auf, die Unterschiede und Gemeinsamkeiten zwischen der unsichtbaren Hand des Marktes und dem Nash-Gleichgewicht zu diskutieren. Dabei stellt der Bot provokative Fragen, um das Verständnis des Studenten für die Entwicklung der Wirtschaftstheorie zu testen und ihn zur kritischen Analyse anzuregen.

1. Dialog/Diskussion/ Fachdiskurs/Prüfungs- gespräch mit Bot

Ein Psychologiestudent interagiert mit einem Bot, der die **Rolle eines Kindes** annimmt, das im Rahmen eines Sorgerechtsstreits begutachtet wird. Der Bot legt typische kindliche Verhaltensweisen an den Tag, mit denen der Student geschickt umgehen muss, um die Aussagen nicht zu beeinflussen, gleichzeitig aber den tatsächlichen Verhältnissen auf den Grund zu gehen.

Im Prüfungsgespräch tritt der KI-Bot als **Patientin mit möglicher psychischer Störung** auf. Die Studentin muss ein diagnostisches Gespräch führen, um Hinweise auf eine psychische Erkrankung zu erkennen. Dabei wendet sie Kenntnisse über Störungsbilder und Diagnostikkriterien an, um eine fundierte Einschätzung zu geben.

Eine Studentin führt ein Interview mit einem KI-Bot, der die Rolle eines **Opferzeugen in einem fiktiven Kriminalfall** einnimmt. Ihre Aufgabe ist es, unter Anwendung rechtspsychologischer Befragungstechniken die Aussage des Zeugen auf Glaubwürdigkeit und Konsistenz zu prüfen. Der Bot simuliert dabei mögliche Erinnerungsverzerrungen, emotionale Reaktionen und versucht, die Studentin auf die Probe zu stellen.

2. Entwicklung mit KI mit anschließender kritischer Reflexion/Bearbeitung

Eine Jurastudentin nutzt eine KI-Anwendung, um **Allgemeine Geschäftsbedingungen** zu erstellen. Anschließend prüft sie das Produkt auf rechtliche Vollständigkeit, identifiziert fehlende oder unklare Bestimmungen und achtet auf die Einhaltung gesetzlicher Vorgaben. Sie überarbeitet die AGB, fügt spezifische Klauseln hinzu und stellt sicher, dass sie auf den Sachverhalt zugeschnitten sind. In ihrer Reflexion bewertet sie die Qualität des KI-generierten Vertrags, diskutiert Grenzen und Potenziale der KI im juristischen Kontext und überlegt, inwieweit solche Tools die juristische Praxis beeinflussen können.

Ein Informatikstudent verwendet eine KI-Codierungsplattform, um ein Programm zu entwickeln, das einen **Sortieralgorithmus** implementiert. Der Student analysiert den Code, identifiziert mögliche Optimierungen, behebt Fehler. Zudem überprüft er die Laufzeitkomplexität und testet das Programm mit verschiedenen Datensätzen. In seiner Reflexion diskutiert er die Leistungsfähigkeit der KI beim Programmieren, mögliche Risiken durch fehlerhaften Code und die Bedeutung des menschlichen Faktors bei der Softwareentwicklung.

2. Entwicklung mit KI mit anschließender kritischer Reflexion/Bearbeitung

Eine BWL-Studentin lässt von einer KI einen **Businessplan für ein Start-up** im Bereich erneuerbare Energien erstellen. Die KI liefert einen Entwurf mit Marktanalyse, Marketingstrategien und Finanzprognosen. Die Studentin überprüft den Businessplan auf Realismus, Genauigkeit der Marktanalysen und Plausibilität der Finanzdaten. Sie passt die Strategien an, fügt eigene Recherchen hinzu und korrigiert Annahmen, die nicht zutreffen. In ihrer Reflexion bewertet sie die Nützlichkeit der KI bei der Erstellung von Businessplänen und diskutiert die Notwendigkeit menschlicher Expertise.

Eine Pädagogikstudentin lässt von einer KI einen **Unterrichtsentwurf für eine Schulstunde** zum Thema Klimawandel erstellen. Die KI liefert einen groben Plan mit Zielen, Methoden und Materialien. Die Studentin prüft den Entwurf auf didaktische Qualität, Altersangemessenheit und pädagogische Wertigkeit. Sie ergänzt differenzierte Aufgabenstellungen, interaktive Elemente und Bewertungsmethoden. In ihrer Reflexion bewertet sie die Unterstützung durch die KI bei der Unterrichtsplanung und reflektiert über die Grenzen der KI in der Bildung.

3. Analyse und Kritik von KI-Erzeugnissen

Ein Psychologiestudent untersucht eine von einer KI erstellte **Analyse eines Fallbeispiels** aus der Klinischen Psychologie. Er prüft die diagnostischen Schlussfolgerungen, die angewandten Theorien und die vorgeschlagenen Interventionen. Er identifiziert mögliche Fehlinterpretationen von Symptomen oder unbegründete Annahmen. In seiner Kritik diskutiert er die Eignung von KI für diagnostische Prozesse und therapeutische Empfehlungen.

Eine Studentin bewertet einen von einer KI generierten **Fragebogen zur Messung eines Persönlichkeitskonstrukts**. Sie analysiert die inhaltliche Validität der Items, die theoretische Fundierung und die ethischen Implikationen. Sie identifiziert kulturelle Biases oder unangemessene Fragestellungen. In ihrer Kritik reflektiert sie über die Verantwortung bei der Entwicklung psychologischer Tests und die Rolle von KI dabei.

3. Analyse und Kritik von KI-Erzeugnissen

Eine Studentin analysiert ein von einer KI verfasstes **Rechtsgutachten** zu einem Haftungsfall. Sie bewertet die juristische Argumentation, die Zitierung relevanter Gesetzestexte und Urteile sowie die Schlussfolgerungen. Sie identifiziert fehlerhafte Interpretationen oder logische Fehlschlüsse. In ihrer Kritik reflektiert sie über die Grenzen der KI im juristischen Denken und die Notwendigkeit menschlicher Expertise.

Eine Studentin analysiert einen von einer KI erstellten **Finanzprognosebericht** für ein mittelständisches Unternehmen. Der Bericht enthält Umsatz-, Kosten- und Gewinnprognosen für die nächsten fünf Jahre. Sie prüft die Annahmen, auf denen die Prognosen basieren, wie Markttrends, Kostenentwicklungen und Investitionspläne. Sie identifiziert mögliche Unstimmigkeiten, wie überoptimistische Wachstumsraten oder nicht berücksichtigte Risiken. In ihrer Kritik bewertet sie die Zuverlässigkeit von KI in der Finanzplanung und diskutiert die Notwendigkeit menschlicher Expertise bei der Interpretation finanzieller Daten.

**Vielen Dank für Ihre
Aufmerksamkeit**



Prof. Dr. Vera Lenz-Kesekamp
Präsidentin der Euro-FH Hamburg
vera.kristina.lenz-kesekamp@euro-fh.de



Julia Jochim (MBA)
Teamleiterin Digitale Lehre & KI
Euro-FH Hamburg
julia.jochim@euro-fh.de