

Methodendokument: KI-resistente Klausuren mit Case Studies

Entwickelt für schriftliche Prüfungen mit KI- und Internet-Zulassung

Prof. Dr.-Ing. Stefan Wagenpfeil, Software-Engineering
PFH Private Hochschule Göttingen

Sommersemester 2026 – Version 2.0

Contents

1	Zweck dieses Dokuments	3
2	Das Ausgangsproblem	3
2.1	Was KI gut kann	3
2.2	Was KI strukturell nicht kann	3
3	Die Methode: Drei-Ebenen-Architektur	4
3.1	Ebene 1: Die fiktive Case Study	4
3.1.1	Anforderungen	4
3.1.2	Empfohlene Struktur	5
3.2	Mehrdokument-Fallstudien (empfohlener Standard ab Version 2.0)	5
3.2.1	Empfohlene Dokumentstruktur	6
3.2.2	Warum das die KI-Resistenz erhöht	6
3.3	Ebene 2: Eingebaute Widersprüche	6
3.3.1	Die vier Typen	7
3.3.2	Techniken zur Platzierung	7
3.3.3	Erweiterte Widerspruchsmuster (aus der Mehrdokument-Praxis)	7
3.4	Ebene 3: Aufgabenformate	8
3.4.1	Format 1: Handschriftliches Artefakt (höchste KI-Resistenz)	8
3.4.2	Format 2: Fallentscheidung mit Begründung	9
3.4.3	Format 3: KI-Antwort kritisch bewerten	9
3.4.4	Format 4: Widerspruch identifizieren und benennen	9
3.4.5	Format 6: Fachliche Aussage eines Stakeholders kritisch bewerten	9
3.4.6	Format 7: Grafische Synthese aus Tabellendaten	9
3.4.7	Format 8: Offene strategische Synthesefrage	10
3.4.8	Format 5: Kurzantwort mit Zahlen-Abwägung	10
3.4.9	Extreme Wortlimits als Erschwernis	10
3.4.10	Zeitrichtwerte (45-Minuten-Klausur)	10
4	Schritt-für-Schritt-Anleitung	11
4.1	Schritt 1: Kontext klären	11
4.2	Schritt 2: Case Study konzipieren	11
4.3	Schritt 3: Selbsttest	12
4.4	Schritt 4: Aufgaben formulieren	12

4.5	Schritt 5: Musterlösung entwickeln	12
5	Prompt-Vorlage für neue KI-Sessions	12
5.1	Qualitätsprüfung vor Abgabe	13
6	Referenzbeispiele	13
6.1	BrakeTech Global – Hardware und Systemplattformen (SS 2026, Deutsch) .	13
6.2	PixVault – Software Engineering (SS 2026, Deutsch)	16
6.3	StreamForge – Software Engineering (SS 2026, Englisch)	16
6.4	Widerspruchs-Bibliothek	16
6.5	Muster für andere Fachgebiete	17
7	Beispiel-Klausurfragen als Formulierungsvorlagen	17
8	Häufige Fehler und Abhilfe	20

1 Zweck dieses Dokuments

Dieses Dokument beschreibt die Methodik zur Entwicklung KI-resistenter Klausuren auf Basis fiktiver technischer Case Studies. Es ist so formuliert, dass es direkt als Prompt-Grundlage für eine neue KI-Session genutzt werden kann: eine KI, die dieses Dokument als Kontext erhält, kann ohne weitere Erklärungen neue Klausuren auf demselben Qualitätsniveau erstellen.

Entstanden aus der iterativen Entwicklung der PixVault-Klausur (Software Engineering, SS 2026, Deutsch) und der StreamForge-Klausur (Software Engineering, SS 2026, Englisch). Version 2.0 erweitert die Methodik um Erkenntnisse aus der Entwicklung und Durchführung der BrakeTech-Global-Klausur (Hardware und Systemplattformen, SS 2026, Deutsch), insbesondere um die **Mehrdokument-Architektur** (Abschnitt 3.2), zusätzliche Widerspruchsmuster (Abschnitt 3.3.3) und zusätzliche Aufgabenformate (Abschnitt 3.4.5).

2 Das Ausgangsproblem

2.1 Was KI gut kann

Moderne KI-Tools (ChatGPT, Copilot, Claude) können folgende klassische Klausuraufgaben zuverlässig und korrekt lösen – auch wenn das Vorlesungsskript als Kontext mitgegeben wird:

- Konzepte erklären („*Was ist Komposition in UML?*“)
- Konzepte auf neue Szenarien anwenden („*Erkläre SOAP vs. REST anhand eines Beispiels*“)
- Texte analysieren und zusammenfassen
- Vorgegebene KI-Antworten kritisieren, wenn die Kritikpunkte aus allgemeinem Fachwissen ableitbar sind
- Multiple-Choice-Fragen beantworten, auch bei Verwechslungsfallen
- User Stories, Use Cases und einfache UML-Diagramme in Textform beschreiben

Kernbefund: Jede Aufgabe, die auf *erkläre*, *vergleiche*, *analysiere* oder *wende an* hinausläuft, wird von einer KI mit Zugang zum Vorlesungsskript korrekt und vollständig beantwortet. Das gilt auch für Aufgaben mit eingebauten Widersprüchen, wenn diese Widersprüche im Text explizit als solche markiert sind.

2.2 Was KI strukturell nicht kann

- Handschriftliche Artefakte produzieren (Diagramme, ausgefüllte Templates auf Papier)
- Versteckte Widersprüche erkennen, die nicht explizit markiert sind – insbesondere wenn die Widerspruchsinformation in einer Nebenspalte, Fußnote oder Warnbox versteckt ist
- Konkrete Zahlen aus einer Tabelle gegen andere Zahlen aus einer anderen Tabelle abwägen
- Die Standardempfehlung verweigern, wenn diese im konkreten Szenario falsch ist
- Den Scope eines Dokuments (z.B. eines Audits) aus einer Nebenspalte korrekt extrahieren

- Interne Aussagen einzelner Personen gegen offizielle Planungsdokumente abwägen
- Bei extrem engen Wortlimits (10–100 Wörter) zuverlässig genau die prüfungsrelevante Kerninformation auswählen, wenn die Antwort Wissen aus mehreren getrennten Dokumenten kombinieren muss
- Bewusst nicht bereitgestellte Informationen als solche erkennen und eigenständig recherchieren, statt sie aus dem Trainingswissen zu erfinden oder wegzulassen

Das entscheidende Prinzip: Nicht die Fragen müssen KI-sicher sein – sondern das Zusammenspiel aus Szenario, eingebauten Widersprüchen und Aufgabenformat. Eine KI, die das Szenario zusammenfasst, bekommt eine andere (schlechtere) Antwort als ein Student, der es wirklich liest.

3 Die Methode: Drei-Ebenen-Architektur

Jede nach dieser Methode entwickelte Klausur besteht aus drei Ebenen:

Ebene	Bestandteil	Funktion	KI-Resistenz
1	Fiktive Case Study (~20 Seiten)	Stresstest durch Informationsmenge	Mittel
2	Eingebaute Widersprüche	Nur durch echtes Lesen erkennbar	Hoch
3	Aufgabenformate	Papier-Artefakte + Zahlen + KI-Kritik	Sehr hoch

3.1 Ebene 1: Die fiktive Case Study

Die Case Study ist ein ca. 15–20-seitiges fiktives Dokument, das eine reale technische Situation beschreibt. Sie enthält bewusst **mehr Information als für die Beantwortung aller Aufgaben notwendig**. Studierende müssen KI einsetzen, um das Material zu erschließen, und dabei die richtigen Fragen stellen.

3.1.1 Anforderungen

- Fiktives Unternehmen mit konkretem Namen, Gründungsjahr, Mitarbeiterzahl – alles erfunden, aber intern konsistent
- Konkrete Versionsnummern, Bibliotheksnamen, Datumsangaben – niemals „*ein Framework*“ sondern „*React Native 0.73.2*“
- Reale technische Konzepte in fiktivem Kontext
- Mehrere Stakeholder mit namentlich genannten, voneinander abweichenden Einschätzungen
- Tabellen mit konkreten Zahlen, die in Aufgaben gegeneinander abgewägt werden müssen
- Mindestens 3 eingebaute Widersprüche (siehe Abschnitt 3.3)
- Abschnitte mit unterschiedlicher Relevanz – nicht alles ist für die Klausur relevant

Häufiger Fehler: Objektbeschreibungen, die UML-Begriffe direkt enthalten („*Objekt X existiert nur im Kontext von Y*“), werden von KI direkt auf die richtige Antwort gemappt.

Stattdessen: Verhalten und Prozesse beschreiben, aus denen der Student die Beziehung ableiten muss („Wenn *Y* gelöscht wird, werden im selben Prozessschritt alle *X* mitgelöscht“).

3.1.2 Empfohlene Struktur

Abschnitt	Inhalt	Länge	Widerspruchs-Potential
1. Hintergrund	Unternehmensgeschichte, Produkt	2–3 S.	Gering
2. Architektur	Stack, Versionen, Bibliotheken	3–4 S.	Mittel
3. Sicherheit / CVEs	Scan-Ergebnisse, Optionen	2–3 S.	Sehr hoch
4. Backup / Failover	Konzept, Messwerte, Lücken	2 S.	Hoch
5. Testabdeckung	Coverage-Zahlen, Lücken	2 S.	Hoch
6. Migration / Update	Pfade, Deadlines, Compliance	2–3 S.	Sehr hoch
7. Team / Kapazität	Personen, SP-Planung	2 S.	Hoch
8. Usability / UX	Bewertungen, Probleme	1–2 S.	Mittel
Anhang	Offene Fragen	1 S.	Gering

3.2 Mehrdokument-Fallstudien (empfohlener Standard ab Version 2.0)

Die ursprüngliche Methodik sieht ein einzelnes 15–20-seitiges Fallstudien-Dokument vor. Die Erfahrung mit der BrakeTech-Global-Klausur zeigt, dass eine Aufteilung auf **mehrere thematisch getrennte Dokumente** die KI-Resistenz zusätzlich erhöht und gleichzeitig realistischer wirkt, da reale Unternehmen ihre Informationen ebenfalls nicht in einem einzigen Dokument bündeln.

3.2.1 Empfohlene Dokumentstruktur

Dokument	Inhalt	Typische Widersprüche
Quartalsbericht	Kennzahlen, Standortübersicht, Strategie-Aussagen, Risikobericht	Aussage vs. Scope (Audit), Zahl vs. Zahl (Budget vs. Bedarf)
Standortprofile (je Standort ein Dokument)	Rollen/Personal, IT-Ausstattung, Netzwerk, Anwendungen	Standard vs. Kontext (Cloud vs. lokale Anforderung), Kapazitätsdaten
Interviews / Stakeholder-Statements	Frei formulierte Aussagen unterschiedlicher Länge und Textsorte	Person vs. Person, Person vs. Dokument, selbstrelativierende Aussagen
Beschaffungsangebot (mehrere Anbieter, unterschiedliches Layout)	Technische Daten, Preise, Konditionen	TCO-/Lizenz-Falle, Fallentscheidung mit Zahlen
Rechnungen / Abrechnungen	Reale Nutzungsdaten, die frühere Angebote oder Aussagen bepreisen	Auflösung früher platzierter Scope-Hinweise

3.2.2 Warum das die KI-Resistenz erhöht

- Eine KI-Sitzung, der nur ein einzelnes Dokument übergeben wird, kann Widersprüche, die erst beim Vergleich mit einem anderen Dokument sichtbar werden, prinzipiell nicht erkennen.
- Auch bei vollständigem Zugriff auf alle Dokumente müssen Studierende (und die KI) aktiv die richtigen Dokumente kombinieren – die Fragestellung nennt die Quellen bewusst nicht.
- Unterschiedliche Dokumenttypen erzwingen unterschiedliche Lesestrategien (Tabelle vs. Fließtext vs. Rechnung), was rein mechanisches Copy-Paste-Vorgehen erschwert.
- Unterschiedliche Layouts/Schriftarten bei vergleichbaren Dokumenten (z. B. mehrere Konkurrenzangebote) verhindern, dass eine KI Tabellenstrukturen blind zwischen Dokumenten überträgt.

Praxishinweis: Die Materialien werden den Studierenden als einzeln herunterladbare Dateien (und zusätzlich als ZIP-Archiv) zur Verfügung gestellt, nicht als ein zusammengefügtes PDF. Das entspricht dem Vorgehen der BrakeTech-Global-Klausur und verstärkt den Effekt, dass die Kombination der Dokumente eine bewusste Leistung der Studierenden bleibt.

3.3 Ebene 2: Eingebaute Widersprüche

Ein eingebauter Widerspruch ist eine Information im Szenario, die einer anderen Information widerspricht – ohne dass dieser Widerspruch explizit benannt wird. Er ist nur erkennbar, wenn zwei verschiedene Stellen im Dokument gleichzeitig gelesen und verglichen werden. KI-Zusammenfassungen glätten diese Widersprüche systematisch.

3.3.1 Die vier Typen

Typ	Beschreibung	Beispiel	KI-Verhalten
Zahl vs. Zahl	Zwei Tabellen enthalten sich ausschließende Zahlen	SP: 240 verfügbar vs. 360 benötigt	KI nennt beide, erkennt Unmöglichkeit nicht
Aussage vs. Scope	Offizielles Ergebnis gilt nur für Teil des Systems	DSGVO-Audit bestanden – Scope nur eu-west-1	KI wertet „bestanden“ ohne Scope-Spalte
Person vs. Dokument	Namentliche Person widerspricht offiziellem Plan	Baumann: 6–9 Monate; Plan: 16 Wochen	KI ignoriert Warnbox-Aussagen
Standard vs. Kontext	KI-Standardempfehlung allgemein richtig, im Kontext falsch	CVE 9.1 → sofort updaten bricht Core	KI empfiehlt immer Standardantwort

3.3.2 Techniken zur Platzierung

- **Typ 1 (Zahl vs. Zahl):** Relevante Zahlen in verschiedene Tabellen in verschiedenen Abschnitten aufteilen. Nie in derselben Tabelle.
- **Typ 2 (Aussage vs. Scope):** Den Scope eines offiziellen Dokuments in einer Nebenspalte oder als kurzen Halbsatz verstecken.
- **Typ 3 (Person vs. Dokument):** Die abweichende Aussage in eine Warnbox verpacken – KI übersieht Warnboxen häufiger als Fließtext.
- **Typ 4 (Standard vs. Kontext):** Die kontextualisierende Information in einem technischen Detail einer Architektur-Tabelle verstecken.

Qualitätskriterium: Ein guter Widerspruch ist durch sorgfältiges Lesen findbar, wird von einer zusammenfassenden KI nicht genannt, wird in der Aufgabenstellung nicht verortet, und erfordert für volle Punkte eine Quellenangabe (Abschnitt X, Tabelle Y).

3.3.3 Erweiterte Widerspruchsmuster (aus der Mehrdokument-Praxis)

Zusätzlich zu den vier Grundtypen haben sich in Mehrdokument-Fallstudien vier weitere, praktisch sehr wirksame Muster gezeigt:

Muster	Beschreibung	Beispiel (BrakeTech Global)
Cross-Document-Widerspruch	Die beiden widersprüchlichen Aussagen stehen in <i>verschiedenen Dokumenten</i> , nicht nur verschiedenen Abschnitten desselben Dokuments	CAD-Speicherbedarf im Standortprofil vs. Bandbreitenangabe im Quartalsbericht
TCO-/Lizenz-Falle	Ein Angebot wirkt durch einen niedrigen Einmalpreis günstig; eine zwingend erforderliche laufende Zusatzkosten-Position steht nur als Fußnote oder in einem separaten Dokument	Cloud-PC-Hardwareangebot vs. separat abgerechnete Lizenzkosten in der Cloud-Rechnung
Selbstrelativierende Aussage	Eine Person äußert eine zugespitzte Behauptung und schränkt sie im selben Statement selbst wieder ein – nur wer den ganzen Text liest, erfasst die korrekte, differenzierte Aussage	Juristin vereinfacht CLOUD-Act-Wirkung zunächst zugespitzt, relativiert dies wenige Sätze später selbst
Fachlich zweifelhafte Stakeholder-Behauptung	Eine im Szenario auftretende Person äußert eine technisch klingende, aber bei genauerer Prüfung fragwürdige oder unbelegte Behauptung	Entwickler behauptet 80 % Datenkompression durch Huffman-Codierung auf bereits komprimierten CAD-Daten

Ergänzende Technik – bewusst nicht bereitgestelltes Material: Nicht jede für eine Aufgabe benötigte Information muss im Fallstudienmaterial enthalten sein. Wird ein aktuelles, real existierendes Produkt referenziert (z. B. zwei konkurrierende Hardware-Geräte), kann bewusst *kein* Datenblatt mitgeliefert werden. Die Studierenden müssen dann selbst im Internet recherchieren – was gleichzeitig prüft, ob sie überhaupt erkennen, *welche* technischen Kriterien für die gestellte Frage relevant sind (z. B. Kameraauflösung statt Display-Auflösung bei einer Frage zur Videoaufzeichnung).

3.4 Ebene 3: Aufgabenformate

3.4.1 Format 1: Handschriftliches Artefakt (höchste KI-Resistenz)

Studierende zeichnen Diagramme oder füllen Templates handschriftlich auf vorgedruckten Flächen aus. KI kann nicht schreiben. Das Artefakt muss szenariospezifisch sein – generische Diagramme ergeben halbe Punkte.

Geeignet: UML-Diagramme (Use Case, Sequenz, Klassendiagramm); Priorisierungsma-

trizen (2x2); Maßnahmenpläne mit SP-Rechnung.

3.4.2 Format 2: Fallentscheidung mit Begründung

Studierende wählen aus vorgegebenen Optionen (A, B, C, D) und begründen ihre Wahl mit konkreten Zahlen aus dem Szenario. Die richtige Antwort ist kontextabhängig.

- Entscheidungsoptionen so gestalten, dass die KI-Standardantwort eine der falschen Optionen ist
- Volle Punkte nur mit Szenariobezug (konkrete Zahlen, Abschnittsangaben)
- Differenzierung zwischen Plattformen / Teilsystemen erzwingen

3.4.3 Format 3: KI-Antwort kritisch bewerten

Eine vorab generierte KI-Antwort wird vorgelegt. Studierende müssen konkrete Fehler mit Quellenangabe benennen. Die KI-Antwort muss authentisch klingen – keine offensichtlichen Fehler. Bewertungsregel: *„Generische Aussagen werden nicht gewertet.“*

Generierungsanweisung: *„Beantworte [Frage] ohne Zugang zum Szenario, nur mit allgemeinem Fachwissen.“* Das Ergebnis ist typischerweise korrekt-generisch und enthält genau die Auslassungen, die echtes Szenarioverständnis erfordern.

3.4.4 Format 4: Widerspruch identifizieren und benennen

Studierende müssen ohne Hinweis zwei widersprüchliche Informationen finden, benennen und die Konsequenz erklären.

- Volle Punkte: beide Stellen mit Abschnittsangabe + Konsequenz
- Halbe Punkte: ein Teil korrekt, Konsequenz fehlt
- 0 Punkte: generische Problemanalyse ohne Fundort

3.4.5 Format 6: Fachliche Aussage eines Stakeholders kritisch bewerten

Analog zu Format 3, aber die zu bewertende Quelle ist keine KI-Antwort, sondern eine im Szenario auftretende Person mit Namen und Rolle. Die Studierenden müssen erkennen, an welcher Stelle eine plausibel klingende fachliche Behauptung technisch problematisch, unbelegt oder nur für einen Teilaspekt gültig ist – ohne dass das Szenario dies markiert.

Beispiel: *„Person X behauptet, mit Methode Y ließe sich Kennzahl Z um 80 % verbessern. Bezogen auf [Aspekt A] wäre das hervorragend – an welcher anderen Stelle könnte dies jedoch zu Problemen führen?“*

3.4.6 Format 7: Grafische Synthese aus Tabellendaten

Anders als beim handschriftlichen UML-/Prozess-Artefakt (Format 1) existiert hier keine vorgegebene Diagrammnotation. Studierende müssen selbst entscheiden, wie sie mehrere verstreute Tabellenwerte (z. B. Standort, Bandbreite, Nutzerzahl, Schwerpunkt) zu einer sinnvollen grafischen Übersicht verdichten. Das prüft nicht nur Lesekompetenz, sondern auch Informationsdesign – eine reine Textzusammenfassung durch KI hilft hier kaum weiter.

3.4.7 Format 8: Offene strategische Synthesefrage

Eine bewusst breit gestellte Abschlussfrage („Was würden Sie als neue Gesamtverantwortliche Person ändern?“), die kein einzelnes Faktum, sondern ein Gesamtverständnis der kompletten Fallstudie verlangt. Es gibt keine einzelne „richtige“ Antwort; bewertet wird, ob die genannten Punkte tatsächlich aus dem Material begründbar sind und nicht generisches Lehrbuchwissen ohne Szenariobezug darstellen.

3.4.8 Format 5: Kurzantwort mit Zahlen-Abwägung

Beispiel: „Kann Maßnahme X innerhalb des Budgets Y durchgeführt werden? Begründen Sie mit den Zahlen aus Abschnitt Z.“ Antworten ohne konkrete Szenariozahlen: maximal 50 % der Punkte.

3.4.9 Extreme Wortlimits als Erschwernis

Die BrakeTech-Global-Klausur zeigt, dass sich Wortlimits deutlich enger fassen lassen als ursprünglich angenommen – bis hinunter zu einem einzigen Satz. Je enger das Limit, desto weniger hilft es, eine lange KI-Antwort einfach zu kürzen; Studierende müssen bereits beim Lesen wissen, worauf es ankommt.

Limit	Punkte (Richtwert)	Geeignet für
1 Satz	5 P	Einzelne Berechnung mit eindeutigem Ergebnis
10 Wörter	5 P	Berechnungsergebnis + Einordnung (z. B. Auflösungsklasse)
20 Wörter	5 P	Ja/Nein-Bewertung mit knapper Begründung
50 Wörter	5–10 P	Fallentscheidung, kritische Bewertung, Stichwortliste (6 × 10 Wörter)
100 Wörter	10 P	Mehrdimensionale Bewertung unter Einbezug mehrerer Dokumente

Bewertungsregel: Wortlimits werden strikt durchgesetzt – Inhalte jenseits des Limits fließen nicht in die Bewertung ein. Das muss den Studierenden explizit in den Klausurhinweisen mitgeteilt werden.

3.4.10 Zeitrichtwerte (45-Minuten-Klausur)

Format	Zeit	Punkte	Max. pro Klausur
Handschriftliches Artefakt	8–12 Min.	10–14 P	1–2 Aufgaben
Fallentscheidung + Begründung	8–10 Min.	8–12 P	1–2 Aufgaben
KI-Antwort bewerten	6–9 Min.	6–10 P	1 Aufgabe
Widerspruch benennen	6–8 Min.	6–10 P	1–2 Aufgaben
Kurzantwort mit Zahlen	5–7 Min.	6–8 P	1–2 Aufgaben

Skalierung auf längere Klausuren: Die BrakeTech-Global-Klausur (90 Minuten, 90 Punkte,

4 Themenblöcke mit insgesamt 11 Teilaufgaben) bestätigt, dass sich die obigen Richtwerte grob proportional skalieren lassen. Bei längeren Klausuren empfiehlt es sich, mehrere Teilaufgaben zu einem Themenblock (z. B. „Konstruktions-Standort“) zu bündeln, die alle auf denselben Dokumenten aufbauen – das reduziert Kontextwechsel und erlaubt trotzdem 5–6 unterschiedliche Aufgabenformate innerhalb eines Blocks.

4 Schritt-für-Schritt-Anleitung

4.1 Schritt 1: Kontext klären

Vor der Entwicklung werden folgende Informationen vom Dozenten eingeholt:

Information	Wichtig weil...	Beispiel
Fachgebiet / Themen	Bestimmt Konzepte im Szenario	SE: UML, APIs, Testing
Persönliche Meinungen	KI kennt diese nicht – stärkster Schutz	SOAP hat Compile-Zeit-Vorteile
Artefakte aus Vorlesung	Nur geübte Artefakte auf Papier	Use Case, Sequenz
Klausurdauer	Anzahl und Tiefe der Aufgaben	45 Min. → 5 Aufgaben
Anzahl Studierende	Beeinflusst Korrekturaufwand	15 → offene Fragen möglich
Prüfungsformat	Papier oder Rechner	Papier → handschriftliche Diagramme

4.2 Schritt 2: Case Study konzipieren

1. Branche und Produkt wählen: technisch genug für Fachfragen, nicht zu speziell
2. Unternehmensname, Gründungsjahr, Mitarbeiterzahl festlegen – alles fiktiv
3. Technologie-Stack mit konkreten Versionen und Bibliotheksnamen festlegen
4. **3–4 Widersprüche skizzieren – bevor die Case Study geschrieben wird**
5. Zahlen-Gerüst aufbauen: intern konsistent, außer bei den bewussten Widersprüchen
6. Mindestens 2 Stakeholder mit namentlich genannten, abweichenden Meinungen

Checkliste vor dem Weitergehen:

- ☐ Mindestens 3 eingebaute Widersprüche vorhanden
- ☐ Jeder Widerspruch ist in separaten Abschnitten versteckt (nicht in derselben Tabelle)
- ☐ Mindestens 2 Personen mit abweichenden Einschätzungen
- ☐ Alle Zahlen intern konsistent – außer den Widersprüchen
- ☐ Mindestens ein Abschnitt, der für keine Aufgabe relevant ist (Ablenkung)
- ☐ Alle Bibliotheksnamen / Versionsnummern sind plausibel
- ☐ Scope eines offiziellen Dokuments (Audit) ist eingeschränkt und versteckt

4.3 Schritt 3: Selbsttest

Bevor Klausurfragen formuliert werden: die geplanten Fragen mit einer KI testen (Szenario vollständig als Kontext übergeben).

- KI antwortet korrekt und vollständig? → Frage muss überarbeitet werden
- KI benennt Widersprüche ohne Hinweis? → Widerspruch ist zu offensichtlich
- KI empfiehlt die „falsche“ Standardantwort? → Gut – erzeugt Differenzierung
- KI antwortet generisch ohne Szenariozahlen? → Gut – Format funktioniert

Wichtig: Der Selbsttest ist ein obligatorischer Qualitätsschritt, kein optionaler.

4.4 Schritt 4: Aufgaben formulieren

Für jede Aufgabe wird festgelegt: Format, Thema (welcher Abschnitt), welcher Widerspruch abgefragt wird, Punktzahl und Zeitrictwert, Bewertungsregel (volle / halbe / null Punkte), typisches KI-Verhalten.

4.5 Schritt 5: Musterlösung entwickeln

- **Volle Punkte:** welche Kernaussagen müssen enthalten sein (mit Quellenangabe)
- **Halbe Punkte:** welche Teillösung genügt für 50 %
- **0 Punkte:** welche Antworttypen werden nicht gewertet
- **KI-Verhalten:** was würde eine KI typischerweise antworten – und warum reicht das nicht

5 Prompt-Vorlage für neue KI-Sessions

Das folgende Prompt zu Beginn einer neuen KI-Session verwenden:

Du hilfst mir dabei, eine KI-resistente Klausur fuer meine Hochschulvorlesung zu entwickeln. Ich habe eine spezifische Methodik entwickelt. Ich lade dir das Methodendokument hoch -- lies es vollstaendig, bevor du irgendetwas produzierst.

Deine Aufgaben:

1. Fiktive Fallstudie als Mehrdokument-Set entwickeln (nicht ein einzelnes Dokument, sondern mehrere getrennte Dateien wie Quartalsbericht, Standortprofile, Interviews, Angebote, Rechnungen -- siehe Abschnitt "Mehrdokument-Fallstudien")
2. Fuer Beschaffungs-/Vergleichsdokumente wo sinnvoll echte, recherchierte Produktdaten verwenden statt frei erfundener Spezifikationen; bei mehreren Konkurrenzangeboten unterschiedliche Layouts/Schriften/Absenderfirmen verwenden
3. Klausur mit Aufgaben in unterschiedlichen Formaten entwickeln (siehe Formatkatalog, inkl. Format 6-8), mit klaren, teils sehr engen Wortlimits je Teilaufgabe (1 Satz bis 100 Woerter, siehe Abschnitt "Extreme Wortlimits")
4. Vollstaendige Musterloesung mit Bewertungsrastern entwickeln, inkl. Spalte "typisches KI-/Studierenden-Verhalten"
5. Alles als druckfertige Word-Dokumente ausgeben; bewusst mindestens eine benoetigte Information NICHT liefern, wenn sie sich zu einem

aktuellen, recherchierbaren Realweltprodukt oder -Sachverhalt eignet

Bevor du beginnst -- frage mich nach:

- Fachgebiet und zentrale Vorlesungsthemen
- 2-3 persönliche Dozentenmeinungen, die die KI nicht kennt
- Welche Artefakte in der Vorlesung konkret geübt wurden
- Klausurdauer, Punktzahl, Anzahl Studierende, Prüfungsformat (Papier oder Rechner)
- Gewünschte Branche und Produkt (oder: du schlägst vor)

Qualitätskriterium: Vor jeder Aufgabe führst du intern einen Selbsttest durch. Beantworte die Frage mit allgemeinem KI-Wissen UND ohne Zugriff auf alle Dokumente gleichzeitig. Ist die Antwort vollständig korrekt, überarbeite die Frage -- sie ist noch nicht KI-resistent genug.

5.1 Qualitätsprüfung vor Abgabe

Kriterium	Prüffrage	Ziel
Widersprüche	Mind. 3 eingebaute, nicht markierte Widersprüche?	Szenario-Lektüre erzwingen
KI-Standardantwort	Mind. 1 Aufgabe, bei der KI-Standard falsch ist?	Kontext-Verständnis prüfen
Zahlen-Abwägung	Mind. 1 Aufgabe mit Zahlenvergleich aus verschiedenen Tabellen?	Echtes Lesen prüfen
Papier-Artefakt	Mind. 1 handschriftliches Artefakt vorhanden?	KI-sicherer Kern
Musterlösung	Jede Aufgabe mit „KI-Verhalten“-Spalte im Raster?	Faire Korrektur
Selbsttest	Jede Aufgabe mit KI ohne Szenario getestet?	Qualitätssicherung
Szenariobezug	Jede Aufgabe verlangt explizit Abschnittsangaben?	Copy-Paste verhindern
Mehrdokument-Bezug	Mind. 1 Aufgabe erfordert Daten aus zwei getrennten Dokumenten?	Cross-Referenzieren erzwingen
Wortlimit	Jede Teilaufgabe mit explizitem, engem Wortlimit versehen?	Kernaussage statt Lauftext erzwingen

6 Referenzbeispiele

6.1 BrakeTech Global – Hardware und Systemplattformen (SS 2026, Deutsch)

Umfangreichstes Beispiel und erste vollständige Umsetzung der Mehrdokument-Architektur (Abschnitt 3.2). Fallstudie: fiktiver Automotive-Zulieferer mit acht Standorten auf fünf Kontinenten, verteilt auf fünf separate Dokumente (Quartalsbericht, Standortprofile, In-

terviews, Angebote, Cloud-Rechnung). Klausur: 90 Minuten, 90 Punkte, 4 Themenblöcke.

Aufg.	Format	Thema	Widerspruch / Anforderung	KI-Falle
1	Grafische Synthese (Format 7)	Standortübersicht	Bandbreite, Nutzerzahl, Schwerpunkt aus mehreren Tabellen kombinieren	KI liefert Text statt Grafik, übersieht Tabellenzeilen
2a	Kurzantwort, 1 Satz	CAD- Migration Cloud	Datenvolumen (Standortprofil) vs. Bandbreite (Quartalsbericht)	Zahlen aus verschiedenen Dokumenten nicht kombiniert
2b	Stichwortliste (6×10 W.)	Cloud- Migration Alltag	Pro/Contra aus Interviews + Standortprofil ableiten	Generische Pro/Contra ohne Szenariobezug
2c	Format 6 (Person vs. Person)	Ohlendorf vs. Reiter	Zwei Interviewpartner, gegensätzliche Positionen, je eigene Datenbasis	KI referiert nur eine Seite oder harmonisiert unzulässig
2d	Fallentscheidung (Format 2)	Angebotsvergleich CAD	Drei Angebote, unterschiedliches Layout, versteckte Eignungsunter- schiede	KI vergleicht nur Preis, nicht Tauglichkeit
2e	Format 6	Huffman- Kompression	Fachlich zweifelhafte Stakeholder- Behauptung	KI übernimmt 80 %-Zahl unkritisch
3a	Berechnung, 10 W.	VR-Streaming- Auflösung	Bandbreite / (Streams × FPS) – cross-document	KI findet Bandbreitenwert nicht oder rechnet ohne Einordnung
3b	Offene Recherche (Format 6- Vorstufe)	Vision Pro vs. Quest 3	Datenblatt bewusst nicht mitgeliefert	KI erfindet Spezifikationen statt zu recherchieren
3c	Kurzantwort, 20 W.	Streaming- Bandbreite	Kamera- vs. Displaywerte, Netzabhängigkeit	KI verwechselt Display- mit Kameraauflösung
4a	Format 6	CLOUD Act	Selbstrelativierende Juristen-Aussage; Folgefrage prüft Tiefenverständnis	KI übernimmt Zuspitzung oder verneint Folgefrage pauschal
4b	Fallentscheidung, 100 W.	Cloud-PC- Beschaffung	TCO-/Lizenz- Falle: Hardwarepreis vs. Cloud-Rechnung	KI vergleicht nur Anschaf- fungspreis
4c	Format 8	Strategische Synthese	Gesamtbild aus allen fünf Dokumenten	KI liefert generische IT-Trends ohne

6.2 PixVault – Software Engineering (SS 2026, Deutsch)

Aufg.	Format	Thema	Widerspruch	KI-Falle
A1	Fallentscheidung	CVE-11847	Update bricht Core	KI wählt Option A (sofort updaten)
A2	KI-Kritik	Testabdeckung	73 % Coverage, 0 % für krit. Libs.	KI bewertet 73 % positiv
A3	Widerspruch	OS + DSGVO	Audit bestanden – Scope nur Backend	KI liest „bestanden“ ohne Scope
A4	Kurzantwort	Kapazität	240 SP verfügbar vs. 360+ SP nötig	KI nennt keine Zahlen
A5	Artefakt (Papier)	Tech. Schulden	V2 (200 SP) > Budget (240 SP)	KI kann nicht zeichnen

6.3 StreamForge – Software Engineering (SS 2026, Englisch)

Task	Format	Topic	Contradiction	AI Trap
T1	Contradiction	GDPR Scope	Audit passed – eu-west-1 only	AI reads „Pass” without scope column
T2	AI Critique	CI/CD Maturity	CFR 34 % = Low Performer	AI stops at Frequency + Lead Time
T3	Artifact (Paper)	Data Residency	DSR deletion broken in ap-southeast-1	AI cannot draw by hand
T4	Decision	OSS Licence	Scan excluded transitive dependencies	AI accepts „no conflicts”
T5	Artifact (Paper)	Tech Debt	TD-04 needs DevOps at 98 %	AI ignores capacity constraint

6.4 Widerspruchs-Bibliothek

Bewährte Muster, die in beliebigen Case Studies eingesetzt werden können:

Muster	Beschreibung	Versteckort
Update-Paradoxon	Sicherheits-Update schließt CVE, bricht Abhängigkeit ohne Nachfolger	Architektur-Tabelle + Sicherheits-Tabelle
Audit-Scope-Lücke	Audit bestanden – Scope-Hinweis zeigt: krit. Teilsystem nicht geprüft	Ergebnis prominent, Scope als Nebenspalte
Experten-Plan-Diskrepanz	Person widerspricht offiziellem Plan mit anderer Zeitschätzung	Plan als Tabelle, Abweichung in Warnbox
Coverage-Illusion	Gesamtzahl klingt gut, Detailtabelle zeigt 0 % für kritische Pfade	Gesamtzahl prominent, Details nächste Seite
Kapazitäts-Arithmetik	Summe der Aufwände übersteigt verfügbares Budget	Aufwände in Schulden-Tabelle, Budget separat
Datenschutz-Abbruch	Migration korrekt, aber Abbruch erzeugt DSGVO-Problem	Migrationsplan detailliert, Abbruch in Dangerbox
Skalierungs-Grenzwert	Prozess versagt ab Volumen, das X % der Nutzer überschreiten	Performance-Tabelle + Nutzerdaten getrennt

6.5 Muster für andere Fachgebiete

Fachgebiet	Case-Study-Idee	Typische Widersprüche
Projektmanagement	IT-Großprojekt öffentliche Hand	Projektplan vs. Fortschritt; Scope-Creep ohne Change Request
IT-Recht / DSGVO	EU-Expansion eines Startups	Einwilligung nicht konform; Aufbewahrungsfristen vs. Geschäftszweck
WI allgemein	ERP-Migration Mittelstand	Migrationsplan technisch korrekt, betrieblich unrealistisch
Cloud Computing	SaaS-Startup skaliert 1k → 1M	Cost-Estimate skaliert nicht mit Traffic
IT-Security	Fintech nach Sicherheitsaudit	Patch schließt CVE, bricht Compliance-Zertifizierung

7 Beispiel-Klausurfragen als Formulierungsvorlagen

Die folgenden Vorlagen sind aus der BrakeTech-Global-Klausur abgeleitet, aber bewusst generisch formuliert, sodass sie sich auf andere Fachgebiete und Case Studies übertragen lassen. Zu jedem Format 1–8 steht zunächst die generische Vorlage (Platzhalter in eckigen Klammern), danach die konkrete Originalformulierung als Beleg.

Hinweis: Nicht jeder eingebaute Widerspruch muss in eine Frage münden. In der BrakeTech-

Global-Klausur wurde beispielsweise der Audit-Scope-Widerspruch (Standort 2, ISO-27001-Geltungsbereich) nicht direkt abgefragt – er blieb als Tiefeninformation im Material, die aufmerksame Studierende bei anderen Fragen (z. B. zur IT-Sicherheit) zusätzlich einordnen können. Ablenkungsmaterial darf ungenutzt bleiben.

Format 1 – Handschriftliches Artefakt

Vorlage: „Zeichnen Sie [Diagrammtyp] für [Szenario-Ausschnitt]. Nutzen Sie ausschließlich die im Material verfügbaren Informationen aus [Dokument/Abschnitt].“

Beispiel (PixVault): „Skizzieren Sie einen Maßnahmenplan zum Abbau der technischen Schulden inklusive SP-Rechnung. Nutzen Sie die Zahlen aus Abschnitt 5.“

Format 2 – Fallentscheidung mit Begründung

Vorlage: „Sie stehen vor der Entscheidung zwischen [Option A], [Option B] und [Option C]. Für welche Option entscheiden Sie sich? Begründen Sie kurz mit konkreten Angaben aus [Dokument]. (max. [X] Wörter, [Y] Punkte)“

Original (BrakeTech, Aufg. 2d): „Analysieren Sie die vorliegenden Angebote der Firmen „Tech Supply“, „NordIT“ und „Alpin Workstation“ in Bezug auf Tauglichkeit im Thema CAD. Welches Angebot würden Sie beauftragen? Begründen Sie das kurz. (max. 50 Wörter, 5 Punkte)“

Format 3 – KI-Antwort kritisch bewerten

Vorlage: „Nachfolgend finden Sie eine KI-generierte Antwort auf die Frage „[Frage]“. Nennen Sie [Anzahl] konkrete Fehler oder Auslassungen mit Quellenangabe (Dokument/Abschnitt). (max. [X] Wörter)“

Beispiel (StreamForge, T2): „Die folgende KI-Antwort bewertet die CI/CD-Reife des Teams. Nennen Sie zwei Kennzahlen, die die KI-Antwort ignoriert, mit Quellenangabe. (max. 60 Wörter)“

Hinweis: In der BrakeTech-Global-Klausur wurde dieses Format durch Format 6 ersetzt – statt einer KI-Antwort wird eine Aussage einer Szenario-Person bewertet (realistischer und schwerer im Vorfeld durch KI-generisches Faktencheck-Wissen zu erkennen).

Format 4 – Widerspruch identifizieren und benennen

Vorlage: „Im Material finden sich zwei Angaben zu [Thema], die sich widersprechen. Benennen Sie beide Fundstellen (Dokument, Abschnitt/Tabelle) und erläutern Sie die Konsequenz. (max. [X] Wörter, [Y] Punkte)“

Beispiel (StreamForge, T1): „Das Audit-Ergebnis wird im Bericht als „bestanden“ geführt. Finden Sie die Stelle, die den tatsächlichen Geltungsbereich einschränkt, und erklären Sie die Konsequenz für die übrigen Regionen. (max. 60 Wörter)“

Format 5 – Kurzantwort mit Zahlen-Abwägung

Vorlage (Berechnung): „Wie lange würde [Vorgang] unter den im Material genannten Bedingungen dauern? (1 Satz, [Y] Punkte)“

Original (BrakeTech, Aufg. 2a): „Wie lange würde eine Migration der CAD-Daten am Standort Sindelfingen in die Cloud dauern? (1 Satz, 5 Punkte)“

Vorlage (Stichwortliste): „Nennen Sie die [N] wichtigsten Positiv- und die [N] wichtigsten Negativ-Aspekte von [Maßnahme]. ([2N] Stichworte mit je max. [X] Wörtern Erklärung, [Y] Punkte)“

Original (BrakeTech, Aufg. 2b): „Welche Auswirkungen hätte das auf die tägliche Arbeit in der Konstruktion? Nennen Sie die drei wichtigsten Positiv-Aspekte und die drei wichtigsten Negativ-Aspekte. (6 Stichworte mit jeweils max. 10 Wörtern Erklärung, 10 Punkte)“

Format 6 – Fachliche Aussage eines Stakeholders kritisch bewerten

Vorlage: „[Name], [Rolle], äußert: „[Zitat]“. Bewerten Sie diese Aussage fachlich. Ist sie so haltbar? (max. [X] Wörter, [Y] Punkte)“

Original (BrakeTech, Aufg. 2e): „Jonas Trierweiler behauptet, mit der Huffman-Codierung eine Kompression von 80 % erreichen zu können. Bezogen auf die Daten- und Übertragungsmenge wäre das zweifellos hervorragend – an welcher anderen Stelle könnte dies jedoch zu Problemen führen? (max. 50 Wörter)“

Variante (Zwei Personen gegenüberstellen): „Was halten Sie von den Vorschlägen von [Person A] und [Person B]? Wie beurteilen Sie die Situation? (max. [X] Wörter, [Y] Punkte)“ – Original: Aufg. 2c (Ohlendorf vs. Reiter).

Format 7 – Grafische Synthese aus Tabellendaten

Vorlage: „Erstellen Sie eine grafische Übersicht von [Dimension 1], [Dimension 2] und [Dimension 3] für alle in [Dokument] genannten [Einheiten]. (1/2 Seite, [Y] Punkte)“

Original (BrakeTech, Aufg. 1): „Erstellen Sie eine grafische Übersicht der Standorte inklusive verfügbarer Bandbreiten, Nutzerzahlen und Schwerpunkt des Standorts. (1/2 Seite, 10 Punkte)“

Format 8 – Offene strategische Synthesefrage

Vorlage: „Sie übernehmen die Rolle [Position]. Welche strategischen Optimierungspotenziale sehen Sie? Was würden Sie als erstes ändern? (max. [X] Wörter, [Y] Punkte)“

Original (BrakeTech, Aufg. 4c): „Welche strategischen Optimierungspotenziale an der derzeitigen IT-Hardware/-Infrastruktur sehen Sie? Was würden Sie ändern, wenn Sie der neue IT-Gesamtverantwortliche werden? (max. 50 Wörter, 10 Punkte)“

Zusätzliche Vorlage: Cross-Document-Berechnung mit Einordnung

Vorlage: „Unter den in [Dokument A] und [Dokument B] genannten Bedingungen: Berechnen Sie [technische Kennzahl] und ordnen Sie das Ergebnis ein (max. [X] Wörter, [Y] Punkte)“

Original (BrakeTech, Aufg. 3a): „Der Service-Leiter möchte für den Mobilen Service VR-Brillen beschaffen [...] In Spitzenzeiten können 70 Streams gleichzeitig anfallen (50 FPS). Berechnen Sie, in welcher Auflösung das während des normalen Betriebs maximal möglich wäre, und ordnen Sie diese ein. (max. 10 Wörter, 5 Punkte)“

Diese Vorlage kombiniert Format 5 (Berechnung) mit einer Einordnungsleistung und erfordert Werte aus mindestens zwei getrennten Dokumenten (Netzwerk-/Bandbreitendaten und Anzahl gleichzeitiger Nutzungsfälle).

8 Häufige Fehler und Abhilfe

Fehler	Symptom	Lösung
Widerspruch zu offensichtlich	KI benennt Widerspruch ohne Aufgabe	Information in Nebenspalte; nie in derselben Tabelle wie der Gegenpart
Objektbeschreibungen zu direkt	KI ordnet UML-Beziehung korrekt zu	Verhalten beschreiben: „wird mitgelöscht“ statt „existiert nur im Kontext“
Aufgabe fragt Konzept statt Anwendung	KI gibt lehrbuchkorrekte Antwort	Frage auf konkrete Zahlen / Personen / Stellen im Szenario beziehen
KI-Antwort für Format 3 zu falsch	Kein Prüfungswert	KI-Antwort muss wie kompetenter Analyst klingen – nur Szenariobezug fehlt
Zu viele Aufgaben	Studierende schaffen nicht alle	45 Min. → max. 5 Aufgaben; Zeitrichtwert pro Abschnitt angeben
Artefakt zu generisch	Auch ohne Case Study korrekt	Aufgabe muss explizit szenariospezifische Elemente verlangen
Musterlösung ohne KI-Verhalten	Korrektor kann schlechte KI nicht von schlechtem Verständnis unterscheiden	Spalte „Typisches KI-Verhalten“ in jeder Aufgabe der Musterlösung
Alles in einem Dokument	KI mit vollem Kontext beantwortet auch Cross-Referenz-Fragen korrekt	Fallstudie auf mehrere getrennte Dokumente aufteilen (Abschnitt 3.2)
Wortlimit zu großzügig	Lange, generische KI-Antworten passen trotzdem hinein	Limit auf das Minimum senken, das die Kernaussage gerade noch zulässt (Abschnitt 3.4.9)