

Kapitel 10

Krise der Prüfungsvalidität

Wenn Noten nichts mehr beweisen



Zentrales Argument: Generative KI erodiert die Prüfungsvalidität als systemisches Problem, nicht als Einzelfall akademischen Fehlverhaltens: Schriftliche Prüfungsformate können ihre eigene erkenntnistheoretische Grundannahme – Produktqualität als Indikator kognitiver Prozesse – unter KI-Bedingungen nicht mehr erfüllen. Detektionsstrategien adressieren Symptome statt das Validitätsproblem selbst; KI-resistente und KI-integrative Formate sind die konzeptuell tragfähigen Antworten.

10.1 Strukturelle Vulnerabilität reproduktiver Prüfungsformate

Generative KI zerstört die epistemische Grundlage schriftlicher Prüfungsformate: die inferentielle Verbindung zwischen Produktqualität und kognitivem Prozess. Ein Essay, eine Fallstudie oder ein Bericht, der sämtliche Beurteilungskriterien erfüllt, lässt sich ohne jeden entsprechenden Lernprozess produzieren. Dies ist kein Betrugs-, sondern ein **Validitätsproblem** im technischen Sinne: Das Instrument misst nicht mehr, was es zu messen beansprucht. Alle herkömmlichen schriftlichen Hausarbeiten, Seminararbeiten und Klausuren ohne Aufsicht sind gleichermaßen betroffen. Detektionsstrategien – KI-Erkennungssoftware, TurnItIn, Stilanalysen – kämpfen gegen Symptome, ohne die Validitätsgrundlage wiederherzustellen. Da LLM-generierte Texte zunehmend von menschlicher Autorschaft ununterscheidbar werden und Falsch-Positiv-Raten unschuldige Studierende belasten, ist die Detektionslogik keine skalierbare Lösung.

10.2 Defensive Strategien und ihre Grenzen

Beaufsichtigte Prüfungen, mündliche Prüfungen ohne Vorbereitung und Handyverbote reagieren auf das Problem mit einer **Ausschlusslogik**: KI wird strukturell ferngehalten. Diese Strategien haben reale Grenzen. Beaufsichtigte Klausuren messen eine spezifische Variante von Kompetenz – Wissensabruf unter Zeitdruck –, die häufig nicht jene ist, die Lernziele eigentlich beschreiben. Technische Überwachungsmaßnahmen (Browser-Locking, Gesichtserkennung) sind invasiv, skalieren schlecht und adressieren die Validitätsfrage nur partiell. Mündliche Prüfungen bleiben das robusteste defensive Format, da sie direkte epistemische Auseinandersetzung erzwingen, die nicht delegiert werden kann; ihre Skalierbarkeit bei großen Kohorten und die Interrater-Reliabilität bleiben jedoch offene Herausforderungen, die explizite Protokollierung und Schulung der Prüfenden erfordern.

10.3 KI-resistente Prüfungsformate

KI-resistente Formate bauen nicht auf Ausschluss, sondern auf **strukturelle Unsubstituierbarkeit**: Sie verlangen kognitive Leistungen, die KI nicht zu erbringen vermag. Drei Mechanismen sind zentral: Erstens **Prozessdokumentation** – Lerntagebücher, Arbeitsportfolios, Revisionshistorien und Begründungsprotokolle erzeugen kognitive Kohärenzanforderungen, die oberflächliche KI-Nutzung sichtbar machen: Argumentationspolyzyklen, plötzliche Einsichten ohne stützende Recherche und Kurswechsel ohne Zwischenschritte sind Indizien für fehlende Eigenleistung. Zweitens **authentische Aufgaben** mit situiertem Urteilsvermögen: Probleme aus realen Kontexten, deren Lösung Stakeholderwissen, lokale Bedingungen und normative Abwägungen erfordert, entziehen sich musterbasierten KI-Lösungen. Drittens **kollaborative und präsentative Formate** – Gruppenpräsentationen mit unmittelbarer Befragung, Poster-Sessions und strukturierte Peer-Reviews – erfordern interaktive Verantwortung, die keine KI übernehmen kann.

10.4 KI-integrative Prüfung

Der fortschrittlichste Ansatz verschiebt die Prüfungslogik grundlegend: Statt KI-Nutzung zu unterbinden, wird die Fähigkeit **kompetenter Mensch-KI-Kollaboration** selbst zum Bewertungsgegenstand. Studierende müssen schriftlich ausweisen, welche KI-Tools sie eingesetzt haben, zu welchem Zweck, und wie sie die KI-Ausgaben kritisch bewertet und ggf. verworfen haben. Bewertet werden Prompting-Kompetenz, kritische Distanz gegenüber KI-Ergebnissen und die Qualität der eigenen Urteilsleistung jenseits der KI-Unterstützung. Dies misst nicht mehr ausschließlich disziplinäres Wissen, sondern **epistemische Handlungsfähigkeit** im KI-Zeitalter – eine Kompetenz, über die Studierende erst angeleitet werden müssen und die nicht durch bloße KI-Nutzung entsteht. KI-integrative Prüfung setzt voraus, dass diese Kompetenz explizit gelehrt, geübt und rückgemeldet wird – nicht erst im Prüfungskontext erstmals eingefordert wird.

Kernaussage: Prüfungsvalidität ist keine administrative, sondern eine epistemische Frage: Das Prüfungssystem entscheidet darüber, welche Kompetenzen tatsächlich entwickelt werden. Hochschulen, die auf KI-Detektion setzen statt auf Formatinnovation, adressieren ein Validitätsproblem mit Integritätsinstrumenten – und scheitern strukturell. Wer valide prüfen will, muss neu definieren, was geprüft werden soll.

► Warum das vollständige Kapitel lesen?

Kapitel 10 enthält eine vollständige Typologisierung aller gängigen defensiven und konstruktiven Prüfungsstrategien mit Wirksamkeitsbewertung auf Basis verfügbarer Evidenz, entwickelt konkrete Gestaltungsprinzipien für KI-resistente und KI-integrative Formate und zeigt, warum Detektionsstrategien konzeptuell unzureichend sind. Für Prüfungsverantwortliche bietet das Kapitel eine direkt handlungsfähige Grundlage.

