PAIHE

Practice of Artificial Intelligence in Higher Education
Jg. 1 (2026) · Artikel 1

GRUNDLAGENBEITRAG

Leistung ist nicht Lernen

Ein Minimalkonsens für die Gestaltung von Hochschullehre mit generativer KI

Sven Warnke[1], Harald Bolsinger[2], Tanja Mühling[1]

[1] Technische Hochschule Würzburg-Schweinfurt, Fakultät Angewandte Sozialwissenschaften

[2] Technische Hochschule Würzburg-Schweinfurt, THWS Business School

Zusammenfassung

Generative KI kann die Leistung steigern, ohne dass Kompetenz entsteht. Gute Arbeitsergebnisse sind deshalb für sich genommen keine hinreichenden Kompetenzbelege. Dieser Beitrag fragt, wie Hochschullehre gestaltet werden kann, damit die Nutzung generativer KI Kompetenz aufbaut, und worauf sich Lehrende verschiedener Fächer dabei gemeinsam stützen können. Ausgehend von der Leitidee intellektueller Souveränität bestimmen wir als Zielkompetenz einen urteilsfähigen Umgang mit generativer KI im Fach, der fachliche Kompetenz und KI-bezogene Urteilskompetenz verbindet. Constructive Alignment dient uns dabei als gemeinsamer Planungsrahmen, das Scholarship of Teaching and Learning mit Design-Based Research als Entwicklungsrahmen für Lehrveranstaltungen. Wo KI genutzt wird, verlangt die Prüfung nach Möglichkeit, das KI-bezogene Urteil zu zeigen. Den substanziellen Kern unseres Minimalkonsenses bilden drei Gestaltungsfunktionen: eigene Vorleistung vor der KI-Nutzung, KI als Feedback zu eigenen Zwischenständen sowie KI-Nutzung begründet wählen, dokumentieren und reflektieren. Die ersten beiden begründen wir mit belegten Lernmechanismen, die dritte normativ. Vier weitere Funktionen beschreiben wir als optionale Gestaltungsfunktionen, kurz Ausbauoptionen. Wie die Gestaltungsfunktionen beim Lernen mit generativer KI wirken, ist nach der von uns gesichteten Literatur bisher kaum untersucht. Die Gestaltungsfunktionen sind deshalb Gestaltungsvermutungen, die wir in weiteren Entwicklungszyklen untersuchen. Der Beitrag begründet und präzisiert einen bereits praktizierten Konsens. Methodisch ist der Beitrag eine theoriegeleitete, narrative Zusammenschau, keine systematische Übersichtsarbeit.

Schlagwörter: generative KI, Hochschuldidaktik, Constructive Alignment, Kompetenzorientierung, intellektuelle Souveränität, Scholarship of Teaching and Learning

Abstract

Performance is not learning: A minimal consensus for designing university teaching with generative AI

Generative AI can raise performance without competence developing. Good work products are therefore not, on their own, sufficient evidence of competence. This article asks how university teaching can be designed so that the use of generative AI builds competence, and what teachers from different disciplines can jointly rely on in doing so. Taking intellectual sovereignty as the guiding idea, we define the target competence as the judicious use of generative AI within a discipline, a use that combines disciplinary competence with AI-related judgement competence. Constructive alignment serves as the shared planning framework, and the scholarship of teaching and learning as the framework for further development of courses, with design-based research as its method. Where AI is used, assessment requires, where possible, that students show their AI-related judgement. The substantive core of our minimal consensus consists of three design functions: own prior work before using AI, AI as feedback on one's own interim work, and justifying, documenting and reflecting on AI use. We ground the first two in learning mechanisms supported by research and the third in normative reasoning. We describe four further functions as optional design functions, or extension options for short. In the literature we reviewed, how the design functions affect learning with generative AI has hardly been studied. The design functions

© 2026 Sven Warnke, Harald Bolsinger, Tanja Mühling · CC BY 4.0

are therefore design conjectures, which we examine in further development cycles. The article grounds and refines a consensus already in practice. Methodologically, the article is a theory-guided narrative synthesis, not a systematic review.

Keywords: generative AI, higher education teaching, constructive alignment, competence orientation, intellectual sovereignty, scholarship of teaching and learning

Veröffentlicht: 5. November 2026
Verfahren: Eingeladener Beitrag der Herausgeber:innen im Startverfahren der ersten Ausgabe; gemeinsam konzipiert und verantwortet, keine Prüfung durch nicht beteiligte Personen.
KI-Nutzung in der Prüfung: KI-erstellte Gutachten und Gegenlesungen (Claude, Anthropic) dienten als Sammlung möglicher Einwände. Alle Entscheidungen haben die Herausgeber:innen getroffen.

Zitierempfehlung: Warnke, S., Bolsinger, H. J., & Mühling, T. (2026). Leistung ist nicht Lernen: Ein Minimalkonsens für die Gestaltung von Hochschullehre mit generativer KI. *Practice of Artificial Intelligence in Higher Education, 1*(1), Artikel 1. https://doi.org/10.83397/PAIHE.2026.01

DOI: https://doi.org/10.83397/PAIHE.2026.01 · **Lizenz:** © 2026 Sven Warnke, Harald Bolsinger, Tanja Mühling. Dieser Beitrag steht unter der Lizenz CC BY 4.0 (https://creativecommons.org/licenses/by/4.0/deed.de).

1 Einleitung: Leistung ist nicht Lernen

Generative KI erzeugt aus gelernten Mustern neue Inhalte wie Texte, Bilder oder Code. Damit verändert sie, was Arbeitsergebnisse über das Lernen von Studierenden aussagen. Wie verbreitet ihre Nutzung ist, zeigt das Vorwort dieser Ausgabe. Dieser erste Beitrag setzt bei dem Problem an, das daraus für die Gestaltung von Hochschullehre folgt: Gute Arbeitsergebnisse zeigen nicht mehr verlässlich an, was Studierende selbst gelernt haben. Die Lernforschung unterscheidet seit Langem zwischen der Leistung, die während des Lernens beobachtbar ist, und dem Lernen als relativ dauerhafter, übertragbarer Veränderung von Wissen und Können. Beide können auseinanderfallen: Die Leistung kann steigen, ohne dass nennenswert gelernt wird, und umgekehrt kann gelernt werden, ohne dass die Leistung während des Lernens sichtbar steigt (Soderstrom & Bjork, 2015). Lernende halten ihre momentane Leistung, häufig fälschlich, für einen verlässlichen Hinweis auf ihr Lernen (Soderstrom & Bjork, 2015). Mit KI wird diese Unterscheidung dringlicher, denn KI kann Ergebnisse liefern, zu denen Lernende selbst wenig beitragen.

Erste Studien zum Lernen mit KI verdeutlichen dieses Problem. In einem Feldexperiment in der Schulmathematik mit fast tausend Schülerinnen und Schülern verbesserte der Zugang zu einem Sprachmodell die Ergebnisse beim Üben. In einer anschließenden Prüfung ohne KI schnitten jedoch diejenigen, die beim Üben eine übliche Chat-Variante ohne besondere Vorgaben zum Schutz des Lernens genutzt hatten, schlechter ab als eine Vergleichsgruppe, die ohne KI geübt hatte. Selbst bemerkt hatten sie diesen Nachteil nicht. Eine zweite Variante mit demselben Sprachmodell erhielt andere Vorgaben: Sie gab statt vollständiger Lösungen schrittweise Hinweise und stützte sich dabei auf Lösungen und typische Fehler, die Lehrkräfte bereitgestellt hatten. Diese Variante hob den Nachteil weitgehend auf, führte aber auch nicht zu besseren Prüfungsergebnissen als das Üben ohne KI (Bastani et al., 2025). In einer randomisierten Laborstudie mit Studierenden verbesserten diejenigen, die beim Überarbeiten eines Aufsatzes von einem Sprachmodell unterstützt wurden, ihre Texte stärker als Vergleichsgruppen, unterschieden sich aber nicht signifikant im Wissenszuwachs und im Transfer (Y. Fan et al., 2025). Beide Studien sind Einzelbefunde: Sie messen kurz nach dem Lernen, mit unterschiedlichen Aufgaben und Maßen, und nur eine stammt aus der Hochschule. Für den KI-Kontext liefern sie deshalb nur schwache Evidenz (zu den Evidenzstufen Abschnitt 4.1). Sie fügen sich aber in die etablierte Unterscheidung von Leistung und Lernen. Wenn die Leistung steigen kann, ohne dass gelernt wird, und Kompetenz erst durch Lernen entsteht (Abschnitt 2.2), folgt daraus das Leitproblem dieses

Beitrags: Generative KI kann die Leistung steigern, ohne dass Kompetenz entsteht. Gute Arbeitsergebnisse sind deshalb für sich genommen keine hinreichenden Kompetenzbelege.

Für die Lehre ist das folgenreich, denn Aufgaben und ihre Bewertung setzen voraus, dass ein Arbeitsergebnis die Tätigkeit anzeigt, durch die gelernt werden soll. Liefert generative KI das Ergebnis, ist offen, ob diese Tätigkeit stattgefunden hat. Ein gutes Ergebnis kann sowohl auf gelungenes Lernen hindeuten als auch auf die Auslagerung des Denkens: Eine überzeugende, mit KI erstellte Analyse eines Fachartikels zeigt nicht, ob die Studierenden dessen Argumente verstanden, die Analyse kontrolliert und eine eigene Position entwickelt haben. Weil sich am Ergebnis nicht mehr ablesen lässt, ob die Tätigkeit stattgefunden hat, lässt sich Lernen über Regeln zur KI-Nutzung allein nicht sichern. Es kommt deshalb weniger darauf an, ob KI genutzt wird, als darauf, wie Lehre die Lernaktivitäten gestaltet, durch die Kompetenz entsteht. Wir fragen daher, wie Lehre gestaltet werden kann, damit die Nutzung von KI Kompetenz aufbaut, statt nur die Leistung zu steigern, und worauf sich Lehrende verschiedener Fächer dabei gemeinsam stützen können.

Unsere Antwort ist ein Minimalkonsens, also das, was wir über unsere Fächer hinweg mindestens gemeinsam vertreten. Er besteht aus einem prozeduralen Kern, der festlegt, wie wir planen, und einem substanziellen Kern, der festlegt, was die Lernsituation leisten soll. Den prozeduralen Kern bildet Constructive Alignment: Lehre wird vom Kompetenzziel der Lehrveranstaltung her geplant, und Lernaktivitäten und Prüfung werden darauf abgestimmt. Weil das Kompetenzziel unter KI-Bedingungen das KI-bezogene Urteil einschließt (Abschnitt 2.3), heißt das: Wo KI genutzt wird, verlangt die Prüfung nach Möglichkeit, das KI-bezogene Urteil zu zeigen (Abschnitt 3.2). Den substanziellen Kern bilden drei Gestaltungsfunktionen, also Anforderungen an die Lernsituation, die Lernen auch bei KI-Nutzung ermöglichen sollen: „eigene Vorleistung vor der KI-Nutzung“, „KI als Feedback zu eigenen Zwischenständen“ und „KI-Nutzung begründet wählen, dokumentieren und reflektieren“. Wir nennen sie Kernfunktionen, weil jede Lehrveranstaltung sie nach unserem Minimalkonsens erfüllen soll. Ihre Reihenfolge folgt dem Ablauf einer Nutzung; die dritte Kernfunktion umspannt ihn: Die begründete Wahl steht am Anfang, die Rückschau am Ende und bestimmt die nächste Wahl. Wo die Bedingungen es erlauben, erweitern optionale Gestaltungsfunktionen, kurz Ausbauoptionen, diesen Kern.

Getragen wird der Konsens von uns drei Lehrenden, die diesen Beitrag verfassen und aus zwei Fakultäten kommen. Der Konsens beruht auf unserer eigenen Lehrerfahrung und auf Ergebnissen der Lern- und Kompetenzforschung, die wir gesichtet haben. Jede Gestaltungsfunktion knüpft dabei an einen Lernmechanismus an, also an einen untersuchten Zusammenhang zwischen einer Lernaktivität und dem Lernen. Wie weit die Forschung den Konsens trägt, legen wir deshalb für jede Funktion zweifach offen: wie gut ihr Mechanismus belegt ist und wie gut sie selbst beim Lernen mit KI untersucht ist (Kapitel 4). Für das Lernen mit KI sind die einzelnen Funktionen in der von uns gesichteten Literatur bisher durchgehend schwach belegt oder nur konzeptionell begründet (zu den Evidenzstufen Abschnitt 4.1).

Der Konsens bestand schon vor diesem Beitrag: Wir haben ihn in unserer Lehre praktiziert, er war die Voraussetzung unserer Zusammenarbeit und bestimmte, welche Lehrveranstaltungen in eine erste Befragung von Studierenden eingingen, über die Beitrag 3 (Mühling et al., 2026) berichtet. Theoretisch geprägt war diese Lehre zunächst eher implizit, durch Ansätze und Arbeiten, die uns vertraut waren, nicht durch eine systematische Aufarbeitung. Für diesen Beitrag haben wir die Ansätze ausdrücklich benannt, zusammengetragen und ihre tragenden Quellen kontrolliert. Während der Erarbeitung, bis

Oktober 2026, haben wir gezielt nach neueren Studien und Gegenbefunden zum Lernen mit KI gesucht: mit KI-gestützten Suchwerkzeugen (siehe Hinweis zur KI-Nutzung), mit punktuellen Datenbankrecherchen und ausgehend von Literatur, die wir bereits kannten. Dabei haben wir den Konsens präzisiert: Die Nutzungsarten von KI, die wir bisher angeboten haben, etwa „KI als Feedbackgeber“, verstehen wir jetzt als Gestaltungsfunktionen, die sichern sollen, dass die zugrunde liegenden Lernmechanismen auch bei KI-Nutzung wirken können. Für Studierende mit wenig Vorwissen haben wir zudem eine Bedingung ergänzt (Abschnitt 4.2). Die präzisierte Fassung ist Grundlage der nächsten Befragung.

Der Beitrag ist eine theoriegeleitete, narrative Zusammenschau, keine systematische Übersichtsarbeit oder Metaanalyse. Weil die Suche nicht systematisch war, können relevante Studien fehlen. Die Evidenz haben wir zudem selbst und ohne unabhängige Doppelbewertung eingestuft. Das Ergebnis des Beitrags sind keine Wirksamkeitsnachweise, sondern Gestaltungsvermutungen, also begründete, aber noch zu untersuchende Annahmen darüber, wie Lehre gestaltet werden kann, damit die Nutzung von KI Kompetenz aufbaut. Wie wir diese Vermutungen in Zyklen untersuchen und den Konsens dabei weiterentwickeln, beschreibt Abschnitt 3.3.

Dieser Beitrag bezieht sich auf keine bestimmte Lehrveranstaltung, denn der Konsens legt fest, was gewährleistet sein soll, nicht wie. Wie die Gestaltungsfunktionen in einzelnen Fächern Form annehmen, sollen verschiedene Praxisbeiträge zeigen. Den ersten enthält bereits diese Ausgabe: Beitrag 2 („The Practice of Reflective Teaching with AI: Future Skills Development with Offline LLMs“; Bolsinger et al., 2026). Beitrag 3 (Mühling et al., 2026) berichtet erste Ergebnisse der Befragung von Studierenden.

Dieser Beitrag beschreibt den Minimalkonsens als gemeinsamen Kern unserer Lehre und richtet sich an Lehrende aller Fächer. Begriffe der Lern- und Kompetenzforschung erläutern wir deshalb dort, wo wir sie einführen. Aufgebaut ist der Beitrag so, wie wir Lehre planen, nämlich vom Ziel her. Kapitel 2 bestimmt das Ziel: Die Leitidee der intellektuellen Souveränität gibt die Richtung vor, Kompetenz statt Leistung bildet den Maßstab, und das angestrebte Ergebnis des Lernens ist die Zielkompetenz, ein urteilsfähiger Umgang mit generativer KI im Fach. Kapitel 3 beschreibt den Rahmen, in dem wir Lehre planen und den Konsens weiterentwickeln: Constructive Alignment als gemeinsame Planungslogik und die Prüfung als Lernanreiz. Hinzu kommt das Scholarship of Teaching and Learning, in dem Lehrende ihre eigene Lehre untersuchen, mit Design-Based Research als Vorgehen, mit dem wir unsere Gestaltungsvermutungen in wiederholten Zyklen untersuchen. Kapitel 4 leitet die Gestaltungsfunktionen her, die eine Lehrveranstaltung unter KI-Bedingungen erfüllen soll, und fasst den Minimalkonsens zusammen. Dabei begründet es Kernfunktion 1 und 2 mit belegten Lernmechanismen, Kernfunktion 3 normativ, und weist für die Ausbauoptionen aus, welche Kriterien sie nicht erfüllen. Kapitel 5 zieht ein Fazit und kündigt einen Folgebeitrag an, der Spannungsfelder des Konsenses und Grenzen unseres Vorgehens ausführt. Ethik behandeln wir nicht in einem eigenen Kapitel, sondern denken sie in jeder Facette der Zielkompetenz und jeder Gestaltungsfunktion mit (Abschnitt 2.3 und Kapitel 4). Wer zuerst wissen möchte, was der Konsens für die eigene Lehrveranstaltung bedeutet, kann mit den Planungsfragen in Abschnitt 4.4 beginnen.

2 Intellektuelle Souveränität und Zielkompetenz

2.1 Intellektuelle Souveränität

Kapitel 1 hat gezeigt, dass gute Arbeitsergebnisse unter den Bedingungen generativer KI nicht mehr verlässlich anzeigen, was Studierende gelernt haben. Bevor wir bestimmen, woran sich Lehre stattdessen messen lässt, klären wir, worauf sie ausgerichtet ist. Der Wissenschaftsrat (2026) hat dafür intellektuelle Souveränität als übergreifende Leitidee für die Hochschulbildung in Zeiten generativer KI vorgeschlagen und damit an Prinzipien humaner Bildung angeknüpft. Intellektuelle Souveränität ist danach „die Fähigkeit zum freien und selbstbestimmten Handeln in der Wissensaneignung und Wissensproduktion sowie die Fähigkeit, autonom zu denken und Entscheidungen selbst argumentativ zu begründen“ (S. 8). Wir verstehen sie als normativen Bezugspunkt der Lehrgestaltung, nicht als Kompetenzkonstrukt. Zwar bestimmt die Definition den Begriff als Fähigkeit, der Wissenschaftsrat führt ihn aber als Leitidee ein, die den Hochschulen Orientierung geben soll. Zudem ist der Begriff für ein Kompetenzkonstrukt zu weit: Er umfasst Wissensaneignung und Wissensproduktion insgesamt, während Kompetenzen auf die Anforderungen eines bestimmten Bereichs bezogen werden.

Mit Blick auf KI verschiebt die Leitidee die Frage von der optimalen Nutzung hin zu Erhalt und Förderung unabhängiger Denk- und Urteilsfähigkeit (Wissenschaftsrat, 2026, S. 8). Weil die Leitidee weiter reicht als jede Kompetenz, folgt die Zielkompetenz, die wir in Abschnitt 2.3 bestimmen, nicht zwingend aus ihr, erhält von ihr aber Richtung und Begründung: Nach dem Wissenschaftsrat (2026) sind fachspezifisches Wissen und grundlegende KI-Kompetenzen „wesentliche Grundpfeiler intellektueller Souveränität“ (S. 27), und die Zielkompetenz greift beides auf. Umgekehrt konkretisiert die Zielkompetenz die Leitidee für das fachliche Arbeiten mit KI. Entscheidend für eine Lehrveranstaltung ist deshalb nicht, ob Studierende souverän sind, sondern ob sie diese Zielkompetenz aufbauen. Die Leitidee begründet zudem zwei Forderungen, denn selbstbestimmtes Handeln und autonomes Denken dürfen nicht vom Werkzeug abhängen: Fachliche Kompetenz muss, soweit sie für Kontrolle und eigenes Urteil nötig ist, auch ohne Werkzeug bestehen, und bewusster Verzicht gehört zur Einsatzentscheidung (Abschnitte 2.2 und 2.3). In dieselbe Richtung weist die Empfehlung des Wissenschaftsrats für KI-freie Räume, die er ebenfalls mit der Leitidee begründet (2026, S. 40 f.).

2.2 Leistung, Lernen und Kompetenzentwicklung

Zeigen Arbeitsergebnisse Kompetenz für sich genommen nicht mehr verlässlich an, kann sich die Gestaltung von Lehre nicht am Ergebnis allein ausrichten. Ihr Maßstab bleibt die Kompetenz selbst. Kompetenz entsteht erst durch Lernen, und die Lernforschung unterscheidet Lernen von der momentanen Leistung (Kapitel 1).

Kompetenz verstehen wir mit Weinert (2001/2014) in einem weiten Sinn, nämlich als „die bei Individuen verfügbaren oder durch sie erlernbaren kognitiven Fähigkeiten und Fertigkeiten, um bestimmte Probleme zu lösen, sowie die damit verbundenen motivationalen, volitionalen und sozialen Bereitschaften und Fähigkeiten um die Problemlösungen in variablen Situationen erfolgreich und verantwortungsvoll nutzen zu können“ (S. 27 f.). Kurz gesagt: Kompetenz ist, was eine Person weiß, kann und einzusetzen bereit ist, um Aufgaben eines Bereichs zu bewältigen. Wir übernehmen den weiten Begriff, weil die Zielkompetenz Bereitschaften einschließt: die Bereitschaft, vorhandenes Können auch einzusetzen, wenn

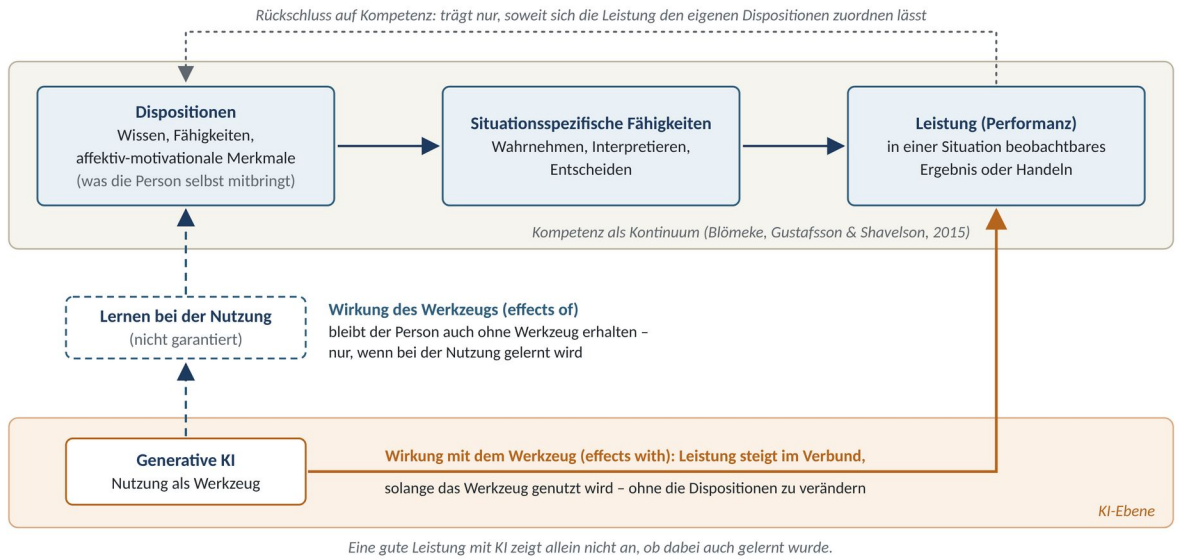
niemand es verlangt, etwa um eine überzeugend klingende KI-Ausgabe zu kontrollieren, und die Bereitschaft, für das eigene Ergebnis einzustehen. Die empirische Bildungsforschung grenzt den Begriff vielfach auf kognitive Leistungsdispositionen ein und begründet das forschungsstrategisch (Klieme & Hartig, 2007, S. 13). Die Eingrenzung auf das Kognitive übernehmen wir nicht, wohl aber die Kontextspezifität, die nach Klieme und Hartig (2007, S. 16) jede Kompetenzdefinition kennzeichnet: Die Zielkompetenz ist an das jeweils gelehrte Fach gebunden.

Wie Kompetenz und beobachtbares Handeln zusammenhängen, beschreiben Blömeke et al. (2015) als Kontinuum. Dispositionen, also Wissen, Fähigkeiten und affektiv-motivationale Merkmale wie Einstellungen und Motivation, wirken über situationsspezifische Fähigkeiten des Wahrnehmens, Interpretierens und Entscheidens auf die Leistung (Performanz), also das in einer Situation beobachtbare Ergebnis oder Handeln. Abbildung 1 zeigt dieses Kontinuum im oberen Teil. Für die Lehre folgt daraus: Leistung ist ein Indikator, von dem auf Kompetenz geschlossen wird (in Abbildung 1 der gepunktete Pfeil). Dieser Schluss trägt nur so weit, wie sich die Leistung den Dispositionen zuordnen lässt, um die es der Lehrveranstaltung geht: dem, was die Person selbst weiß, kann und einzusetzen bereit ist. Unter KI-Bedingungen wird diese Zuordnung unsicher, weil das Ergebnis auf eigenen Beiträgen und auf Beiträgen des Werkzeugs beruht.

Warum diese Zuordnung unsicher wird, erklärt eine Unterscheidung von Salomon et al. (1991): Ein Werkzeug kann auf zwei Arten wirken. Wirkungen mit dem Werkzeug (effects with) zeigen sich nur, solange es genutzt wird: Im Verbund mit KI entsteht ein besseres Ergebnis, die Leistung steigt. Wirkungen des Werkzeugs (effects of) bleiben der Person auch ohne das Werkzeug erhalten: Sie treten nur ein, wenn bei der Nutzung gelernt wird und sich dadurch die Dispositionen verändern. Abbildung 1 ergänzt das Kontinuum um eine KI-Ebene und zeigt beide Wege: Der durchgezogene untere Pfeil führt von der KI direkt zur Leistung, und dieser Teil der Leistung umgeht die Dispositionen. Der gestrichelte Pfeil führt über das Lernen bei der Nutzung zu den Dispositionen. Eine gute Leistung mit KI zeigt deshalb allein nicht an, ob dabei auch gelernt wurde. Über welche Lernmechanismen das geschieht und wie Lehre dafür sorgen kann, dass diese Mechanismen auch bei KI-Nutzung wirken, zeigt Kapitel 4.

Abbildung 1

Kompetenzkontinuum und KI-Ebene



Anmerkung. Die Abbildung zeigt eine Modellvorstellung, keine Wirkungsgrößen. Durchgezogen in der KI-Ebene: Wirkung im Verbund; durchgezogen im Kontinuum: Abfolge von den Dispositionen zur Leistung. Gestrichelt: Weg über das Lernen. Gepunktet: Rückschluss von der Leistung auf Kompetenz. Eigene Darstellung in Anlehnung an Blömeke et al. (2015) sowie Salomon et al. (1991). Entwurf mit Claude Opus 5.5 erstellt und von den Autor:innen überarbeitet.

Nach der Unterscheidung von Wirkungen mit dem Werkzeug und Wirkungen des Werkzeugs ist der Maßstab für Lernen, was der Person bleibt. Für die fachliche Kompetenz heißt das: was die Person ohne Werkzeug zeigen kann, denn erst daran lässt sich Lernen von einer Leistung im Verbund unterscheiden. Das ist eine Setzung, zu der es eine begründete Gegenposition gibt, die schon Salomon et al. (1991) erwägen: Nach der These des erweiterten Geistes gehört ein verlässlich gekoppeltes Werkzeug selbst zur Kognition (Clark & Chalmers, 1998). Die Zielkompetenz nimmt diese Gegenposition teilweise auf, denn ihre zweite Dimension, die KI-bezogene Urteilskompetenz (Abschnitt 2.3), ist eine Kompetenz für das Arbeiten im Verbund und zeigt sich im Umgang mit dem Werkzeug. Ohne Werkzeug bestehen muss die fachliche Kompetenz deshalb nur, soweit sie für die Kontrolle der KI-Beiträge und das eigene Urteil nötig ist.

2.3 Zielkompetenz

Mit der Leitidee als Richtung (Abschnitt 2.1) und Kompetenz als Maßstab (Abschnitt 2.2) bestimmen wir die Zielkompetenz wie folgt:

> Die Zielkompetenz umfasst die Fähigkeit und Bereitschaft, im fachlichen Arbeiten begründet zu entscheiden, ob, wann, wofür, womit und wie generative KI eingesetzt wird, ihre Beiträge auf der Grundlage eigenen Fachwissens zu kontrollieren und in ein eigenes, fachlich und ethisch verantwortetes Urteil einzubeziehen. Dazu gehört, KI so einzusetzen, dass die fachliche Kompetenz, die eine Aufgabe entwickeln soll, aufgebaut und nicht ersetzt wird.

Kurz gesagt geht es um einen urteilsfähigen Umgang mit generativer KI im Fach. Die Definition verbindet drei Tätigkeiten, nämlich das Entscheiden, das Kontrollieren und das

Einbeziehen in das eigene Urteil, mit einer Bedingung: Die fachliche Kompetenz, um die es in einer Aufgabe geht, wird aufgebaut und nicht ersetzt.

Die Zielkompetenz hat zwei Dimensionen. Die fachliche Kompetenz (D1) bestimmen wir hier nur ihrer Rolle nach, als das, was eine Aufgabe entwickeln soll; inhaltlich füllt sie jedes Fach selbst. Die KI-bezogene Urteilskompetenz (D2) wird fachspezifisch ausgeübt, ist in ihrer Struktur aber fachübergreifend: In jedem Fach wird entschieden, kontrolliert und einbezogen; hinzu kommt die Reflexion des eigenen Vorgehens, die sichern soll, dass die fachliche Kompetenz aufgebaut und nicht ersetzt wird. Dafür, dass beide Dimensionen zusammengehören, spricht zweierlei. Erstens ist kritisches Denken kein allgemeines Können, das sich beliebig übertragen lässt, sondern an Fachwissen und Übung im jeweiligen Fachgebiet gebunden (Willingham, 2007). Weil die fachliche Kontrolle von Ergebnissen kritisches Denken verlangt, ist sie ohne Fachwissen kaum möglich. Zweitens genügt fachliche Expertise allein für diese Kontrolle nicht: In einer kleinen Beobachtungsstudie beurteilten zehn professionelle Faktenprüfer:innen Online-Informationen treffsicherer als zehn promovierte Historiker:innen, zu deren Handwerk die Quellenkritik gehört, weil die Faktenprüfer:innen gezielt andere Quellen heranzogen, statt die vorliegende Seite genau zu lesen (Wineburg & McGrew, 2019). Fachliche Kompetenz ist damit für die Ergebniskontrolle in der Regel Voraussetzung, aber nicht hinreichend, denn es braucht zusätzlich Strategien der Kontrolle, wie sie die KI-bezogene Urteilskompetenz umfasst.

Die KI-bezogene Urteilskompetenz gliedern wir in vier Facetten, also Teilfähigkeiten, die wir getrennt beschreiben, die aber zusammengehören. Sie folgen dem Ablauf einer KI-Nutzung und lehnen sich an die drei Phasen selbstregulierten Lernens an (Zimmerman, 2002): Die Einsatzentscheidung entspricht der Planung, die Ergebniskontrolle und die Urteilsintegration entsprechen der Ausführung, die Prozessreflexion entspricht der Selbstreflexion. Vor der Nutzung steht die Einsatzentscheidung (D2.1): ob, wann, wofür, womit und wie KI eingesetzt wird. Ziel dieser Entscheidung ist angemessenes Vertrauen, also ein Vertrauen, das der tatsächlichen Leistungsfähigkeit des Werkzeugs für diese Aufgabe entspricht (Lee & See, 2004). Verzicht gehört deshalb zu dieser Facette, ist für sich genommen aber noch kein Zeichen von Kompetenz, denn schon bei automatisierten Systemen gelten übermäßiges Sich-Verlassen (misuse) und unangemessene Nichtnutzung (disuse) beide als Fehlformen (Parasuraman & Riley, 1997). Bei der Ausgabe folgt die Ergebniskontrolle (D2.2), bei der Weiterverarbeitung die Urteilsintegration (D2.3), danach die Prozessreflexion und Verantwortung (D2.4), im Folgenden kurz Prozessreflexion. Weil die Prozessreflexion in die nächste Einsatzentscheidung einfließt, ist der Ablauf ein Zyklus. Zur Prozessreflexion gehört zudem der längerfristige Blick: Wer die eigene Arbeitsweise über längere Zeit betrachtet, kann eine schleichende Abhängigkeit von KI erkennen und die Arbeitsweise anpassen. Die Arbeiten zur Fachbindung kritischen Denkens und zur Expertise (Willingham, 2007; Wineburg & McGrew, 2019) und zu Vertrauen in Automation (Lee & See, 2004; Parasuraman & Riley, 1997) beziehen sich nicht auf generative KI, die von Anfrage zu Anfrage wechselnde und sprachlich überzeugende Ausgaben liefert. Dass diese Arbeiten auch für generative KI gelten, bleibt deshalb eine Annahme, die wir für begründet halten.

Wo wir im Folgenden von Facetten der Zielkompetenz sprechen, schließt das die fachliche Kompetenz (D1) ein. Tabelle 1 fasst diese Facetten zusammen und nennt für jede eine Leitfrage, die Tätigkeiten, um die es geht, die Theorieanker und die ethischen Gesichtspunkte. Für Ergebniskontrolle und Urteilsintegration knüpfen wir an Teilfähigkeiten kritischen Denkens an, die Facione (1990) als Analysieren und Bewerten sowie als

Schlussfolgern beschreibt, für die Prozessreflexion an Arbeiten zum Schreiben als Lernen (Nückles et al., 2020).

Tabelle 1

Facetten der Zielkompetenz, auf die der Minimalkonsens zielt

Facette	Leitfrage	Worum es geht	Theorieanker	Ethik
D1 Fachliche Kompetenz	Was soll ich in diesem Fach können?	wird aufgebaut, nicht ersetzt	Weinert (2001/2014) Blömeke et al. (2015) Salomon et al. (1991): effects of	Fachethik
D2.1 Einsatz-entscheidung	Ob, wann, wofür, womit und wie?	einschließlich Verzicht, Werkzeugwahl, Prompting als Urteil über das Wie	Planungsphase (Zimmerman, 2002) angemessenes Vertrauen (Lee & See, 2004) misuse und disuse (Parasuraman & Riley, 1997)	Daten und Rechte, Nachhaltigkeit, Herkunft und Trainingsdaten des Werkzeugs, Macht-verhältnisse
D2.2 Ergebnis-kontrolle	Stimmt das? Bringt es für die Aufgabe einen Mehrwert gegenüber meiner eigenen Lösung?	Richtigkeit, Relevanz und Verzerrung kontrollieren, an Quellen abgleichen, Passung zur Aufgabe und Mehrwert kritisch hinterfragen	Analysieren und Bewerten (Facione, 1990)	Fairness, Perspektivität
D2.3 Urteils-integration	Was bedeutet das für mein Ergebnis?	gewichten, annehmen oder verwerfen, eigenes Urteil bilden	Schlussfolgern (Facione, 1990)	Integrität, Urheberschaft
D2.4 Prozessreflexion und Verantwortung	Was habe ich getan und gelernt, und stehe ich dafür ein? Wie verändere ich mittel- und langfristig meinen Lern- und Arbeitsprozess?	dokumentieren, das eigene Vorgehen rückblickend beurteilen, offenlegen, eigene Arbeitsweise anpassen	Selbstreflexions-phase (Zimmerman, 2002) Schreiben als Lernen (Nückles et al., 2020)	Integrität, Verantwortung für Folgen

Anmerkung. In der Spalte „Facette“ steht D1 für die fachliche Kompetenz, D2.1 bis D2.4 stehen für die Facetten der KI-bezogenen Urteilskompetenz (D2). Leitfrage: Frage, an der sich Studierende orientieren können. Worum es geht: Tätigkeiten, die dazugehören. Theorieanker: Arbeiten, an die die Zeile anknüpft. Ethik: ethische Gesichtspunkte, die der Text nach der Tabelle erläutert. Eigene Darstellung.

Drei Klarstellungen sind für die Lehrgestaltung folgenreich. Erstens ist Kontrollieren nicht Einbeziehen: Wer jede Aussage der KI korrekt kontrolliert, aber die Gliederung und

Argumentation der KI ohne eigene Gewichtung vollständig übernimmt, hat das eigene Urteil ausgelagert. Deshalb sind Ergebniskontrolle und Urteilsintegration getrennte Facetten. Zweitens setzt die Frage nach dem Mehrwert einer KI-Ausgabe, ein Teil der Ergebniskontrolle, eine eigene Lösung voraus, mit der die Ausgabe verglichen werden kann. Hier setzt Kernfunktion 1 an, die eine eigene Vorleistung vor der KI-Nutzung verlangt, also keine Prüfungsvorleistung, sondern einen eigenen Lösungsversuch oder eine eigene Einschätzung (Abschnitt 4.2). Drittens meint das „Wie“ in der Einsatzentscheidung keine Bedienungs- oder Prompttechnik, sondern ein Urteil über die Art des Einsatzes, also darüber, welche Rolle die KI übernimmt, in welchem Umfang und für welche Teilaufgabe.

Die Definition verlangt ein fachlich und ethisch verantwortetes Urteil. Ethik bilden wir nicht als eigene Facette ab, sondern als Gesichtspunkt in jeder Facette, denn ethische Fragen stellen sich nicht in einem gesonderten Schritt, sondern in jedem Schritt einer KI-Nutzung. Wir unterscheiden fünf Inhaltsbereiche: Integrität und Urheberschaft, Daten- und Rechteschutz, Fairness und Perspektivität (also die Frage, aus welcher Sicht eine Ausgabe spricht), Nachhaltigkeit sowie gesellschaftliche Machtverhältnisse. Wo diese Bereiche ansetzen, zeigt die Ethikspalte in Tabelle 1. Die Spalte nennt zudem zwei weitere Gesichtspunkte: die Ethik des jeweiligen Fachs (Fachethik) und, bei der Prozessreflexion, die Verantwortung für die Folgen des eigenen Handelns. Zur Einsatzentscheidung gehört auch die Frage nach Herkunft und Trainingsdaten eines Werkzeugs. Dass sich aus Herkunft und Trainingsdaten eines bestimmten Werkzeugs bestimmte Perspektiven in dessen Ausgaben vorhersagen lassen, setzen wir dabei nicht als gesichert voraus.

Einordnung und Status. Die Zielkompetenz ist zugleich enger und weiter als AI Literacy. Long und Magerko (2020) verstehen darunter Kompetenzen, mit denen Menschen KI kritisch bewerten, mit ihr kommunizieren und zusammenarbeiten und sie als Werkzeug nutzen können. AI Literacy beschreibt damit, was alle im Umgang mit KI können sollten, und ist in der Regel fachübergreifend angelegt. Die Zielkompetenz richtet sich dagegen auf das Urteil im fachlichen Arbeiten und ist insofern enger. Weiter ist sie, weil sie die fachliche Kompetenz einschließt. Im Kontext der Diskussion um Zukunftskompetenzen greift die Zielkompetenz einzelne Zukunftskompetenzen des Future-Skills-Frameworks auf, etwa Kritisches Denken, Ethische Kompetenz und AI Literacy (Rampelt et al., 2025, S. 3), und teilt mit dem UNESCO-Kompetenzrahmen für Schülerinnen und Schüler die Betonung des kritischen Urteils (UNESCO, 2024). Nicht als Facetten abgebildet sind dagegen Kreativität sowie Kommunikations- und Kooperationskompetenz.

Die Zielkompetenz ist eine eigene theoretische Synthese und nicht validiert. Ihre Facetten sind analytisch unterschieden. Ob sie sich auch empirisch trennen lassen, ist offen. Auch die Messung ist ungelöst. Viele Instrumente zur AI Literacy beruhen vor allem auf Selbsteinschätzungen (Hornberger et al., 2023). Als Alternative zu Selbsteinschätzungen haben Hornberger et al. (2023) für Studierende einen validierten Multiple-Choice-Test zum Wissen über KI vorgelegt. Ein Wissenstest erfasst aber Facetten wie die Urteilsintegration höchstens mittelbar. Messansätze sind deshalb eine Aufgabe der Weiterentwicklung, ebenso Niveaustufen, die wir noch nicht ausgearbeitet haben.

3 Planungsrahmen und Entwicklungsrahmen

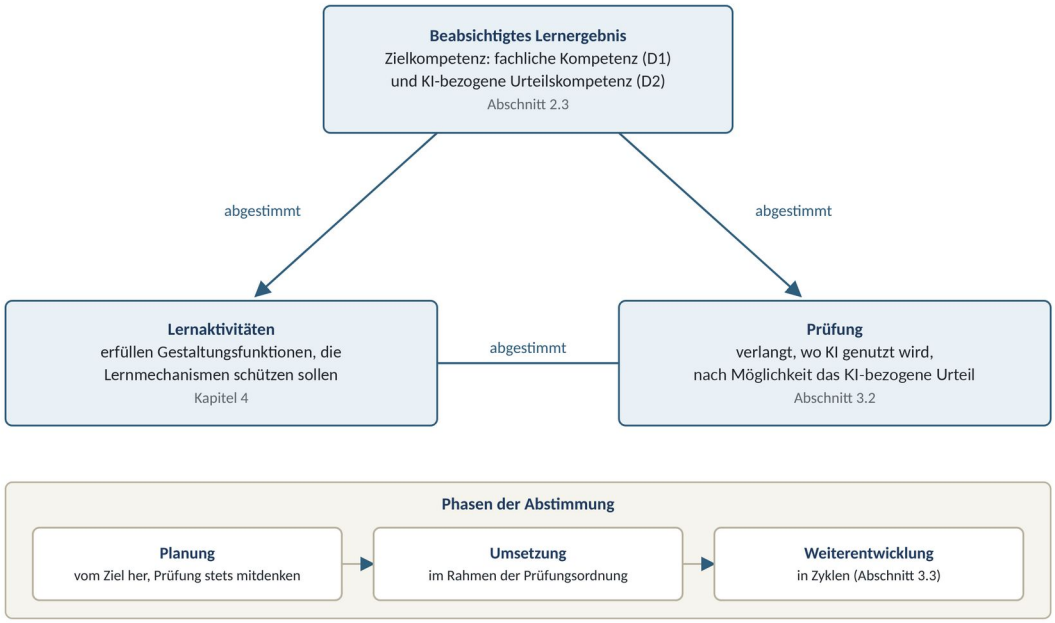
3.1 Constructive Alignment als Planungsrahmen

Das Kompetenzziel einer Lehrveranstaltung ist die für ihr Fach und ihre Aufgaben konkretisierte Zielkompetenz. Es sagt allein noch nicht, wie Lehre es erreichen kann. Einen Planungsrahmen dafür bietet Constructive Alignment, das eine konstruktivistische Auffassung des Lernens mit dem Grundsatz verbindet, beabsichtigte Lernergebnisse (bei uns: das Kompetenzziel), Lernaktivitäten und Prüfungsaufgaben aufeinander abzustimmen (Biggs, 1996b; Biggs et al., 2022). Konstruktivistisch heißt: Was gelernt wird, hängt vor allem davon ab, was Studierende selbst tun. Lehre muss deshalb die Lernaktivitäten hervorrufen, durch die das beabsichtigte Lernergebnis entsteht, und die Prüfung muss zeigen, ob es erreicht wurde. Unter den Bedingungen generativer KI gewinnt die Abstimmung an Gewicht: Liefert ein Werkzeug das Produkt, kann vor allem die Abstimmung von Lernaktivitäten und Prüfung dafür sorgen, dass die beabsichtigte Lernaktivität überhaupt stattfindet. Wir haben Constructive Alignment deshalb als gemeinsame Planungslogik und gemeinsames Grundverständnis und damit als prozeduralen Kern unseres Minimalkonsenses beschlossen.

Was die Abstimmung unter KI-Bedingungen bedeutet, zeigt Abbildung 2. Beabsichtigtes Lernergebnis ist das Kompetenzziel, also die für das Fach konkretisierte Zielkompetenz in beiden Dimensionen, der fachlichen Kompetenz und der KI-bezogenen Urteilskompetenz. Die Lernaktivitäten sollen die Gestaltungsfunktionen (Kapitel 4) erfüllen, deren Aufgabe es ist, die zugrunde liegenden Lernmechanismen unter KI-Bedingungen zu schützen. Die Prüfung verlangt, wo KI genutzt wird, nach Möglichkeit das KI-bezogene Urteil (Abschnitt 3.2). Ein gutes Produkt allein reicht für diese Abstimmung nicht mehr aus, weil es weder die Lernaktivität noch das Lernen anzeigt (Kapitel 1 und Abschnitt 2.2). Die Abstimmung geschieht in drei Phasen: Geplant wird vom Ziel her, umgesetzt im Rahmen der geltenden Prüfungsordnung, und was sich dabei zeigt, fließt in die Weiterentwicklung ein.

Abbildung 2

Constructive Alignment unter KI-Bedingungen



Anmerkung. Alle drei Verbindungen stehen für die Abstimmung der Elemente; die Pfeilspitzen zeigen zusätzlich die Planungsrichtung vom Lernergebnis aus. Die drei Phasen gelten für alle drei Elemente. Die Weiterentwicklung betrifft die Prüfungsordnung (Abschnitt 3.2) und den Konsens (Abschnitt 3.3). Eigene Darstellung in Anlehnung an Biggs (1996b) und Biggs et al. (2022). Entwurf mit Claude Opus 5.5 erstellt und von den Autor:innen überarbeitet.

Planung. Geplant wird vom Kompetenzziel her. Die Prüfung wird bei jeder Planung mitgedacht, weil Studierende ihr Lernen auf die erwartete Prüfung ausrichten (Abschnitt 3.2). Ein Beispiel: Soll eine Lehrveranstaltung Studierende befähigen, Fachliteratur mit KI-Unterstützung zu erschließen, fragt die Planung zuerst, was am Ende gezeigt werden soll, etwa die begründete Kontrolle einer KI-Zusammenfassung am Originaltext. Die zweite Frage gilt der Lernaktivität, die dazu befähigt, etwa: Die Studierenden schreiben vor der KI-Nutzung eine eigene Zusammenfassung, gleichen sie mit der KI-Ausgabe ab und kontrollieren die KI-Ausgabe am Originaltext. Zuletzt wird festgelegt, wie die Prüfung die Kontrolle sichtbar macht und wie die Studierenden vorher erfahren, was die Prüfung verlangt. Die beiden anderen Phasen zeigt Abschnitt 3.2 an der Prüfung, weil sie dort am stärksten an die Prüfungsordnung gebunden sind: die *Umsetzung* im Rahmen der geltenden Prüfungsordnung und die *Weiterentwicklung* der Prüfungsordnung. Wie wir den Konsens selbst in Entwicklungszyklen weiterentwickeln, beschreibt Abschnitt 3.3.

Als Gesamtprinzip ist Constructive Alignment nach der von uns gesichteten Literatur allerdings vor allem theoretisch begründet. Wir behandeln es deshalb nicht als belegten Lernmechanismus, sondern als gemeinsame Planungslogik, auf die sich eine Anforderung an die Lehre berufen kann, etwa die KI-spezifische Prüfungsanforderung (Abschnitt 3.2). Das Kriterium der Begründungsgüte (Abschnitt 4.1) lässt eine solche prozedurale Begründung auch für Gestaltungsfunktionen zu.

Die Kritik an Constructive Alignment richtet sich auf zwei Stellen: den Gebrauch und die Grundlagen. Den Gebrauch kritisieren Loughlin et al. (2021): Lernergebnisse und Abstimmung würden zur bloßen Formalie, wenn Constructive Alignment von oben verordnet und zur Qualitätssicherung genutzt werde. Wir folgen, wie Loughlin et al. es vorschlagen, der

konstruktivistischen Lesart, die vom Lernen der Studierenden ausgeht, und nutzen Constructive Alignment, das wir selbst beschlossen haben, als Planungshilfe für die eigene Lehre. Die Grundlagen kritisieren Newby und Cornelissen (2025) und treffen damit auch die Lesart, der wir folgen: Schon die zugrunde liegende Theorie enthalte problematische Annahmen, etwa über die Fähigkeiten Studierender. Wo Constructive Alignment verbindlich vorgeschrieben werde, schwäche es die Eigenständigkeit der Studierenden und begrenze andere Lehransätze. Der Konsens legt fest, dass Studierende ein Urteil zeigen, nicht welches, und kommt damit dem Einwand zur Eigenständigkeit ein Stück entgegen. Ausgeräumt ist die grundsätzliche Kritik aber nicht (Kapitel 5).

3.2 Prüfung als Lernanreiz

Studierende richten ihr Lernen auf die erwartete Prüfung aus (Biggs, 1996a). Diese Rückwirkung der Prüfung auf das Lernen ist bisher schwach belegt, vor allem durch Befragungen: In einer Befragung gaben Studierende an, für eine Hausarbeit eher vertiefend, für eine Klausur mit Auswahlantworten eher oberflächlich zu lernen (Scouller, 1998). Lernförderlich ist diese Rückwirkung vor allem dann, wenn die Prüfung den angestrebten Prozess verlangt und diese Anforderung vorher bekannt ist. Bewertet eine unbeaufsichtigte Prüfung dagegen nur das Produkt, lässt sich die Prüfungsleistung mit KI erbringen, ohne dass es auffällt: In einem verblindeten Test in fünf Bachelormodulen eines Psychologiestudiengangs blieben 94 % vollständig KI-geschriebener Arbeiten aus häuslichen Online-Prüfungen unentdeckt (Scarfe et al., 2024). Auch authentische Aufgaben schützen allein nicht davor: In einer kleinen explorativen Studie erkannten Lehrende KI-Anteile bei authentischeren Aufgaben nicht besser; die Autor:innen schließen daraus, dass Authentizität allein KI-Nutzung auch nicht vorbeugt (Kofinas et al., 2025).

Deshalb stellen wir die KI-spezifische Prüfungsanforderung auf: Wo KI genutzt wird, verlangt die Prüfung nach Möglichkeit, das KI-bezogene Urteil zu zeigen, also den Einsatz zu begründen, eine Ausgabe zu kontrollieren und das Ergebnis zu vertreten. Geprüft wird damit das Urteil, nicht die Dokumentation allein. Damit berührt die Anforderung alle vier Facetten der KI-bezogenen Urteilskompetenz (D2.1 bis D2.4). Die Anforderung folgt aus der Abstimmung (Abschnitt 3.1): Das beabsichtigte Lernergebnis schließt das KI-bezogene Urteil ein, und die Prüfung muss zeigen, ob dieses Lernergebnis erreicht wurde. Die Anforderung ist deshalb Teil des prozeduralen Kerns und keine vierte Kernfunktion: Die Kernfunktionen gestalten die Lernsituation, die Prüfungsanforderung gestaltet den Leistungsnachweis. Ob die Anforderung das KI-bezogene Urteil auch aufbaut, ist unseres Wissens nicht untersucht. Die Evidenz im KI-Kontext ist deshalb konzeptionell.

Die Anforderung fügt sich in die Empfehlungen des Wissenschaftsrats (2026, S. 44–46) ein. Er hält weder ein pauschales Verbot noch eine generelle Freigabe generativer KI in Prüfungen für zielführend und empfiehlt, Prüfungsszenarien mit und ohne KI zu kombinieren. Wie unsere Anforderung setzt er dabei auf Transparenz: Wird generative KI bei häuslichen schriftlichen Prüfungen zugelassen, verlangt er eine klare Darstellung der erlaubten Hilfsmittel und transparente Information über die Bewertungsmaßstäbe, und für formative Prüfungen empfiehlt er, dass Lehrende Erwartungen und Regeln zur KI-Nutzung transparent darlegen. Er gibt zudem zu bedenken, dass nicht jede Aufgabe in eine benotete Prüfungsleistung münden muss.

Umsetzung. Gehandelt wird im Rahmen der geltenden Prüfungsordnung. Sie legt oft die Prüfungsform fest, aber nicht alles, was die Rückwirkung der Prüfung bestimmt:

Aufgabenstellung, Bewertungskriterien und vorab bekannt gegebene Regeln zur KI-Nutzung können, je nach Prüfungsordnung und Zuständigkeit, Spielraum für die Abstimmung auf das Kompetenzziel bieten. So kann auch eine Klausur verlangen, eine vorgelegte KI-Ausgabe fachlich zu kontrollieren. Weil die Verantwortung für Prüfungsformen zudem häufig nicht bei einzelnen Lehrenden liegt, ist die Abstimmung auch Aufgabe des Studiengangs.

Weiterentwicklung der Prüfungsordnung. Zeigt sich, dass eine andere Prüfungsform die Zielkompetenz besser sichtbar machen würde, sammeln wir dazu Erkenntnisse, um die Prüfungsordnung auf dieser Grundlage weiterzuentwickeln. Diese Erkenntnisse richten sich auch an die Verantwortlichen im Studiengang. Die Grenze der Prüfungsordnung ist damit kein Grund, auf die Abstimmung zu verzichten, sondern ein Anlass für die Entwicklungszyklen (Abschnitt 3.3).

Wie in jeder Facette und jeder Gestaltungsfunktion denken wir auch bei der Prüfung die Ethik mit. Hier geht es um Chancengleichheit: Ungleicher Zugang zu Werkzeugen darf sich nicht in Noten niederschlagen. Wie wir untersuchen, ob unsere Gestaltung trägt, beschreibt Abschnitt 3.3.

3.3 Scholarship of Teaching and Learning als Entwicklungsrahmen

Wir verorten diesen Beitrag im Scholarship of Teaching and Learning (SoTL): Lehrende untersuchen das Lernen in ihrer eigenen Lehre und machen die Ergebnisse öffentlich (siehe Vorwort). Der Wissenschaftsrat (2026, S. 51 f.) empfiehlt den Hochschulen, SoTL zu fördern. Im Sinne von SoTL begründet und präzisiert dieser Beitrag einen Konsens, den wir in unserer eigenen Lehre bereits praktiziert haben. Was wir dabei präzisiert haben, beschreibt Kapitel 1.

Innerhalb dieses Rahmens nutzen wir Design-Based Research als Vorgehen: Der Konsens wird in wiederholten Zyklen aus Erprobung, Auswertung und Überarbeitung weiterentwickelt (McKenney & Reeves, 2019). Dieser Beitrag steht zwischen dem ersten und dem zweiten Zyklus: Er stützt sich auf die Praxis des ersten und legt die präzisierte Fassung für den zweiten vor.

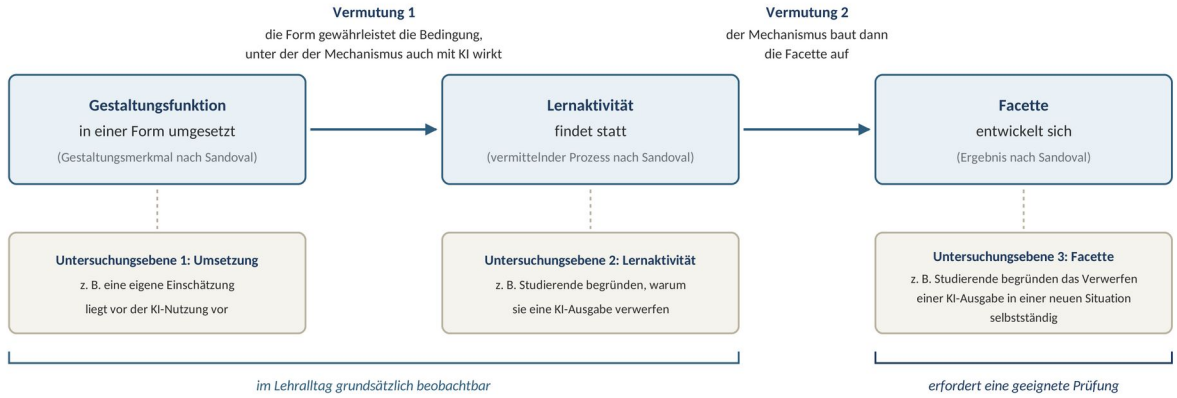
Untersucht werden in den Zyklen Gestaltungsvermutungen (Kapitel 1). In Anlehnung an Sandoval (2014, S. 21–24) enthält jede Gestaltungsfunktion zwei Vermutungen: dass eine Form, also die konkrete Umsetzung einer Gestaltungsfunktion im Fach (Abschnitt 4.1), die Bedingung gewährleistet, unter der der Mechanismus auch mit KI wirkt, und dass der Mechanismus dann die Facette aufbaut. Bei Sandoval heißt die erste Vermutung Gestaltungsvermutung, die zweite theoretische Vermutung. Wir fassen beide unter dem ersten Begriff zusammen. Alle Gestaltungsfunktionen in Kapitel 4 sind in diesem Sinn Gestaltungsvermutungen. Auch die KI-spezifische Prüfungsanforderung (Abschnitt 3.2) ist eine Gestaltungsvermutung, denn ob sie das KI-bezogene Urteil aufbaut, ist unseres Wissens nicht untersucht.

Abbildung 3 zeigt, wie die beiden Vermutungen drei Glieder verbinden: die in einer Form umgesetzte Gestaltungsfunktion, die Lernaktivität, in der der Mechanismus wirkt, und die Facette. Sandoval nennt diese Glieder embodiment (bei uns: Gestaltungsmerkmale), vermittelnde Prozesse und Ergebnisse. Daraus ergeben sich für die Lehrpraxis drei Untersuchungsebenen, die in der Abbildung unter den Gliedern stehen: ob eine Funktion umgesetzt ist (etwa: eine eigene Einschätzung liegt vor der KI-Nutzung vor), ob die angestrebte Lernaktivität stattfindet (etwa: Studierende begründen, warum sie eine KI-Ausgabe verwerfen) und ob sich die Facette entwickelt (etwa: Studierende begründen das

Verwerfen einer KI-Ausgabe in einer neuen Situation selbstständig). Die ersten beiden lassen sich im Lehralltag grundsätzlich beobachten. Die dritte erfordert eine geeignete Prüfung, denn ob sich die Facette entwickelt, zeigt sich erst am selbstständigen Urteil in einer neuen Situation; dass die ersten beiden zutreffen, genügt dafür noch nicht.

Abbildung 3

Gestaltungsvermutungen und Untersuchungsebenen



Anmerkung. Eigene Darstellung in Anlehnung an Sandoval (2014), Beispiele aus dem Text. Entwurf mit Claude Opus 5.5 erstellt und von den Autor:innen überarbeitet.

Im ersten Zyklus folgten Lehrveranstaltungen der früheren Fassung des Konsenses. Aus diesem Zyklus berichten zwei Beiträge dieser Ausgabe, und beide beschreiben erste Durchläufe: Beitrag 2 (Bolsinger et al., 2026) zeigt eine dieser Lehrveranstaltungen, Beitrag 3 (Mühling et al., 2026) berichtet, wie die erste Befragung von Studierenden, ein Pretest des Fragebogens am Semesterende, umgesetzt wurde und was sie ergab. Der zweite Zyklus untersucht die präzisierte Fassung, erstmals mit einer Befragung zu Beginn und am Ende des Semesters, die sich über einen freiwilligen Code verknüpfen lässt. Die Befragung erfasst vor allem die ersten beiden Untersuchungsebenen, und zwar so, wie Studierende sie berichten. Weil eine Vergleichsgruppe fehlt, erlauben die Daten Beschreibungen und einen Abgleich mit den Vermutungen, aber keine Wirkungsaussagen. Als Selbstauskünfte messen die Daten zudem keine Kompetenz, und Studierende können darin unterschätzen, wie stark sie sich auf KI-Ausgaben stützen, wie ein Experiment mit Studierenden bei Entscheidungsaufgaben zeigt (Ren, 2026).

Zum Entwicklungsrahmen gehört auch, wie über den Konsens entschieden wird. Die inhaltlichen Leitlinien haben wir gemeinsam beschlossen und den Beitrag gemeinsam konzipiert. Der Konsens macht diese Leitlinien zur gemeinsamen Arbeitsgrundlage, belegt aber nicht ihre Wirksamkeit. Weil wir ihn selbst tragen und praktizieren, sind unsere Einschätzungen zudem nicht unabhängig. Auch der Konsens selbst ist revidierbar: Was eine Leitlinie berührt, entscheiden wir gemeinsam, und stellen Ergebnisse der Entwicklungszyklen eine Leitlinie infrage, entscheiden wir, ob und wie wir sie ändern.

4 Gestaltungsfunktionen und Lernmechanismen

4.1 Vom Kompetenzziel zur Gestaltungsfunktion

Die Herleitung der Gestaltungsfunktionen geht, wie die Planung (Abschnitt 3.1), vom Kompetenzziel aus. Wie Niedermeier (2026, S. 31) verstehen wir die Integration generativer

KI als „weniger eine technologische als eine didaktische Gestaltungsaufgabe“, ähnlich Brandstetter und Weiand (2026, S. 4) sowie Warnke (2026, S. 19). Diese Aufgabe übersetzen wir in Gestaltungsfunktionen und leiten jede in vier Schritten her. Ausgangspunkt ist (1) eine Facette der Zielkompetenz. Dazu kommt (2) der Lernmechanismus, der diese Facette aufbaut, mit seiner Wirkbedingung, unter der er schon ohne KI wirkt. Es folgen (3) die Frage, ob KI diese Bedingung bestehen lässt oder aufhebt, und (4) die Gestaltungsfunktion, die diese Bedingung auch bei KI-Nutzung gewährleistet. Die Herleitung rekonstruiert einen Konsens, den wir bereits praktiziert haben: Sie klärt, ob sich die gewählten Funktionen so begründen lassen, nicht, wie sie entstanden sind.

Eine Gestaltungsfunktion ist damit eine didaktische Anforderung an die Lernsituation: Sie benennt, was gewährleistet sein soll, damit die Wirkbedingung auch bei KI-Nutzung gegeben ist. Wie das geschieht, bestimmt die Form. Diese Trennung von Funktion und Form übernehmen wir sinngemäß von Hawe et al. (2004), die für randomisierte Studien komplexer Interventionen in der Gesundheitsförderung vorschlagen, die Funktion zu vereinheitlichen und die Form dem Kontext anzupassen. Für die Lehre heißt das: Die Kernfunktionen sind allgemein formuliert, ihre Formen fachspezifisch; Beispiele gibt Beitrag 2 (Bolsinger et al., 2026). Indem die Gestaltungsfunktion die Wirkbedingung gewährleistet, schützt sie den Lernmechanismus: KI kann an die Stelle von Instruktion oder Feedback treten, bringt deren Wirkbedingungen aber nicht von selbst mit.

Ein Beispiel, fortgeführt aus Abschnitt 3.1, zeigt die vier Schritte. (1) Facette: Studierende sollen Fachliteratur verstehen (fachliche Kompetenz), um eine KI-Zusammenfassung am Originaltext kontrollieren zu können. (2) Lernmechanismus: Wer selbst eine Lösung versucht, bevor sie erklärt wird, versteht die Erklärung besser. Seine Wirkbedingung: Ein eigener Versuch liegt vor, und die Erklärung knüpft daran an. (3) Einfluss der KI: Liefert die KI die Zusammenfassung sofort, fehlt der Versuch. (4) Gestaltungsfunktion: Kernfunktion 1 gewährleistet den Versuch, und die KI-Nutzung bezieht sich darauf. Eine mögliche Form: Die Studierenden schreiben vor der KI-Nutzung eine kurze eigene Zusammenfassung, gleichen sie anschließend mit der KI-Ausgabe ab und kontrollieren die KI-Ausgabe am Originaltext.

Ob Lernmechanismen unter KI-Bedingungen weiter wirken, ist selten untersucht. Die Evidenz stufen wir deshalb getrennt für den Lernmechanismus und für den KI-Kontext ein, jeweils danach, wie gut die Befunde Aussagen über Wirkungen tragen: stark (mehrere Metaanalysen oder systematische Reviews mit konsistenten, auf die Hochschule übertragbaren Befunden), mittel (eine Metaanalyse oder mehrere [quasi-]experimentelle Studien mit überwiegend konsistenten Befunden), schwach (Einzelstudien, Korrelations- oder Querschnittsdaten) und konzeptionell (begründet, aber empirisch nicht untersucht). Bei uneinheitlicher Lage gilt die vorsichtigere Stufe. Die Stufen sind eine vereinfachte, von uns gesetzte Heuristik zur Orientierung in diesem Beitrag, keine formale Evidenzbewertung. Berücksichtigt haben wir Zahl, Richtung und Übertragbarkeit der Befunde, nicht systematisch die Qualität der Einzelstudien.

Über Kern und Ausbau entscheiden die vier Kriterien in Tabelle 2, nicht die Stufe der stützenden Evidenz im KI-Kontext, denn der Konsens formuliert Gestaltungsvermutungen, die erst in den Entwicklungszyklen (Abschnitt 3.3) untersucht werden. Tabelle 2 nennt die Kriterien mit ihren Kürzeln und der Frage, die jedes stellt: Kompetenzbezug (KK1), Begründungsgüte (KK2), Formunabhängigkeit (KK3) und keine gewichtige Gegenevidenz (KK4). Zum Kern gehört eine Funktion, die alle vier Kriterien erfüllt, zum Ausbau eine, die den Kompetenzbezug erfüllt, aber Begründungsgüte oder Formunabhängigkeit nicht. Ohne

Kompetenzbezug oder bei grundsätzlicher Gegenevidenz gehört eine Funktion zu keinem von beiden. Die beiden ersten Kriterien fragen nach Verschiedenem: der Kompetenzbezug danach, ob die Funktion mindestens eine Facette der KI-bezogenen Urteilskompetenz schützt, ohne die fachliche Kompetenz zu gefährden, die Begründungsgüte danach, worauf sich eine Funktion stützt. Die Begründungsgüte lässt neben dem Mechanismus normative und prozedurale Begründungen zu, setzt sie aber nicht mit der Stützung durch einen Mechanismus gleich und verlangt, sie als solche zu kennzeichnen. Prozedural heißt: begründet aus Constructive Alignment (Abschnitt 3.1), normativ: begründet aus der Zielkompetenz oder einer Norm wie der Offenlegung, beides nicht aus einer Evidenzstufe.

Tabelle 2
Kriterien für Kern und Ausbau

Kriterium	Frage
KK1 Kompetenzbezug	Schützt die Funktion mindestens eine Facette von D2, ohne D1 zu gefährden?
KK2 Begründungsgüte	Ist der Mechanismus, auf den sich die Funktion als Ganzes unmittelbar stützt, mindestens mittel belegt, oder ist die Funktion ausdrücklich prozedural oder normativ begründet und als solche gekennzeichnet? Stützt sie sich nur auf einzelne Komponenten oder überträgt sie einen Mechanismus per Analogie, zählt die Stufe der Gesamtstruktur bzw. der Übertragung.
KK3 Formunabhängigkeit	Lässt sich die Funktion in jedem Fach und jedem regulären Lehrformat umsetzen?
KK4 Keine gewichtige Gegenevidenz	Gibt es Befunde, die die Funktion grundsätzlich infrage stellen? Grundsätzlich ist Gegenevidenz, die mindestens die Stufe der stützenden Befunde erreicht und sich nicht auf benennbare Bedingungen (etwa Vorwissen, Aufgabenart, Zielgruppe) eingrenzen lässt. Einschränkende Befunde führen nicht zum Ausschluss, sondern zu einer Geltungsaussage, die in Abschnitt 4.2 gebündelt ausgewiesen wird.

Zwei Prüffragen aus Tabelle 2 legen wir für die Anwendung genauer fest. Ob eine Funktion denselben Mechanismus nutzt oder ihn per Analogie überträgt (KK2), entscheiden wir danach, ob die Lernaktivität dieselbe bleibt und die Funktion die Wirkbedingung des Mechanismus selbst herstellt. Fehlt eine der beiden Bedingungen, zählt nicht die Stufe des Mechanismus, sondern die Stufe, auf der die Übertragung selbst belegt ist. Ob eine Funktion formunabhängig ist (KK3), beurteilen wir aus unseren Fächern (zur Reichweite dieses Urteils Abschnitt 4.2).

Nach diesen Kriterien behandelt Abschnitt 4.2 die drei Kernfunktionen, Abschnitt 4.3 die vier Ausbauoptionen, und Abschnitt 4.4 fasst den Minimalkonsens zusammen.

4.2 Gestaltungsfunktionen des Minimalkonsenses

Drei Gestaltungsfunktionen erfüllen alle vier Kriterien und gehören damit zum Kern. Kernfunktion 1 und 2 erfüllen die Begründungsgüte über ihre belegten Mechanismen, die vor allem fachliche Kompetenz aufbauen, Kernfunktion 3 dagegen nicht über ihren Mechanismus, sondern über ihre ausdrücklich normative Begründung. Diese Kernfunktionen stellen wir jeweils in derselben Folge vor: Lernmechanismus und Wirkbedingung, Gefährdung durch KI, Funktion, Evidenz im KI-Kontext und ethischer Aspekt. Am Ende des Abschnitts prüfen wir gebündelt die einschränkenden Befunde zu Kernfunktion 1 und 2

(*Geltung und Grenzen*) sowie *Formunabhängigkeit* und *Kompetenzbezug* aller drei Kernfunktionen.

Kernfunktion 1: Eigene Vorleistung vor der KI-Nutzung. Der erste Mechanismus ist Problemlösen vor der Instruktion: Wer ein Problem zu lösen versucht, bevor die Lösung erklärt wird, lernt im Mittel mehr als bei umgekehrter Reihenfolge. Eine Metaanalyse über 53 Studien fand einen moderaten Vorteil bei konzeptuellem Verständnis und Transfer, nicht aber bei Verfahrenswissen (Sinha & Kapur, 2021). Wirksam wird der Versuch vor allem dann, wenn die Instruktion an ihn anknüpft (Loibl et al., 2017). Der Mechanismus baut fachliche Kompetenz auf und ist mittel belegt.

Liefert generative KI die Lösung sofort, entsteht kein eigener Versuch, an den die Instruktion anknüpfen könnte. Die Leistung kann dann steigen, ohne dass gelernt wird. Die schützende Funktion ist Kernfunktion 1: Ein eigener Lösungsversuch oder eine eigene Einschätzung wird festgehalten, und die KI-Nutzung bezieht sich darauf. Die Lernaktivität bleibt dabei dieselbe: Studierende versuchen zuerst selbst eine Lösung, und die KI-Ausgabe tritt an die Stelle der Instruktion. Anders als eine Instruktion ist die KI-Ausgabe aber nicht verlässlich richtig, und sie knüpft nicht von selbst an den Versuch an. Für das Anknüpfen sorgt die Funktion: Die KI-Nutzung bezieht sich auf den festgehaltenen Versuch, etwa indem die Studierenden die Ausgabe mit ihrem Versuch vergleichen. Anders als bei Loibl et al. (2017) stellen damit die Studierenden das Anknüpfen her, nicht die Instruktion. Die Lernaktivität bleibt aber dieselbe, und die Funktion nutzt damit denselben Mechanismus, keine Analogie (Abschnitt 4.1). Die fehlende Verlässlichkeit bleibt dagegen eine Einschränkung: Sie begrenzt, wie weit die Befunde zum Mechanismus für die KI-Nutzung gelten. Begrifflich lässt sich zudem begründen, dass die Vorleistung die Urteilsintegration stützt: Sie gibt einen eigenen Bezugspunkt, an dem sich eine KI-Ausgabe gewichten lässt. Mittelbar stützt die Vorleistung auch die Ergebniskontrolle, denn ob eine KI-Ausgabe einen Mehrwert bringt, lässt sich erst im Vergleich mit einer eigenen Lösung erkennen (Abschnitt 2.3).

Zu generativer KI sind uns keine direkten Befunde zur Reihenfolge bekannt. Ein indirekter Befund stammt aus der Internetsuche: In einem Experiment mit einer Programmieraufgabe verbesserte ein eigener Lösungsversuch vor der Suche das spätere Abschneiden bei Transferfragen (Giebl et al., 2021). Die Evidenz im KI-Kontext ist deshalb schwach und indirekt. Zur Bedingung bei wenig Vorwissen siehe *Geltung und Grenzen*.

Ethisch geht es um Integrität und Urheberschaft: Die festgehaltene Vorleistung macht den eigenen Anteil sichtbar und damit offenlegbar.

Kernfunktion 2: KI als Feedback zu eigenen Zwischenständen. Der zweite Mechanismus ist Feedback zum eigenen Arbeitsstand. Feedback gehört zu den am besten untersuchten Einflussgrößen auf Lernen: Eine Metaanalyse über 435 Studien fand einen mittleren Effekt, der mit dem Informationsgehalt des Feedbacks steigt (Wisniewski et al., 2020). Der Effekt streut aber stark, wie schon bei Kluger und DeNisi (1996): Dort verschlechterte Feedback in über einem Drittel der Fälle die Leistung. Wirksam wird Feedback erst, wenn Lernende es aufnehmen, einordnen und in einer eigenen Überarbeitung nutzen (Winstone et al., 2017). Bei KI-Feedback gehört dazu, die Hinweise kritisch zu hinterfragen. Der Mechanismus baut fachliche Kompetenz auf. Trotz der vielen Studien ist er mittel belegt, weil der Effekt stark streut und Hochschule sowie verzögerte Lernmaße nicht gesondert ausgewertet sind.

KI kann den Mechanismus verstärken, weil sie jederzeit Feedback gibt. Sie kann ihn aber auch aushöhlen, wenn sie selbst überarbeitet oder ihre Hinweise unkontrolliert übernommen

werden. Dazu passt ein Befund: Beim KI-gestützten Überarbeiten zeigten Studierende nach Einschätzung der Studienautor:innen weniger Bewertungsprozesse als mit Feedback einer Fachperson oder einer Checkliste (Y. Fan et al., 2025). Die Funktion ist Kernfunktion 2: Die KI antwortet auf einen eigenen Stand, und die Studierenden entscheiden, ob die Hinweise der KI zutreffen, ob sie diese übernehmen und wie sie überarbeiten. Damit stellt die Funktion die Wirkbedingung selbst her. Die Lernaktivität bleibt dieselbe, nur die Quelle des Feedbacks ändert sich. Die Funktion nutzt also denselben Mechanismus, keine Analogie (Abschnitt 4.1). Anders als Feedback von Lehrenden sind die Hinweise der KI aber nicht verlässlich richtig. Das begrenzt, wie weit die Befunde zum Mechanismus für die KI-Nutzung gelten. Begrifflich lässt sich außerdem begründen, dass die Kontrolle des Feedbacks Ergebniskontrolle und Urteilsintegration übt.

Ob Feedback durch generative KI über die unmittelbare Leistung hinaus das Lernen fördert, ist selten untersucht. In der Laborstudie aus Kapitel 1 verbesserten Studierende mit KI-Unterstützung beim Überarbeiten ihre Aufsätze stärker als die Vergleichsgruppen, unterschieden sich aber nicht signifikant im Wissenszuwachs und im Transfer (Y. Fan et al., 2025). Die Evidenz im KI-Kontext ist deshalb schwach. Zum Vergleich mit Feedback von Lehrenden siehe *Geltung und Grenzen*.

Ethisch geht es um Fairness und Perspektivität: Lehrende können fragen lassen, wessen Maßstäbe die KI anlegt und welche Schreib- und Denkweisen sie als Abweichung behandelt.

Kernfunktion 3: KI-Nutzung begründet wählen, dokumentieren und reflektieren. Der dritte Mechanismus ist Selbstregulation. Selbstreguliertes Lernen verläuft als Zyklus aus Planung, Ausführung und Selbstreflexion: Die Rückschau auf das eigene Vorgehen bestimmt die nächste Planung (Zimmerman, 2002). Reflexionsinterventionen verbesserten in einer Metaanalyse den Lernerfolg mit mittlerem Effekt (Guo, 2022). Wirksam ist Reflexion, wenn sie angeleitet wird und in die nächste Planung zurückfließt (Nückles et al., 2020). Von selbst wirkt sie nicht: Ein strukturiertes Lerntagebuch ohne begleitende Information zur Selbstregulation verbesserte die Selbstregulation Studierender nicht und ging mit sinkender Motivation einher (Fabriz et al., 2014). Dieser Befund begrenzt die Form der Reflexion, nicht die Funktion. Der Mechanismus ist mittel belegt. Übertragen auf die eigene KI-Nutzung zielt er unmittelbar auf Einsatzentscheidung und Prozessreflexion, mittelbar auf fachliche Kompetenz.

KI kann Regulation weniger anregen als andere Unterstützung: Beim KI-gestützten Überarbeiten waren nach Einschätzung der Studienautor:innen neben den Bewertungs- auch die Orientierungsprozesse seltener als mit Feedback einer Fachperson oder einer Checkliste (Y. Fan et al., 2025). Die Funktion ist Kernfunktion 3: Die KI-Nutzung wird begründet gewählt, einschließlich des Verzichts. Dokumentiert, also festgehalten, wird die Nutzung als Grundlage von Rückschau und Offenlegung, nicht als bloßes Protokoll. Für die Rückschau genügt, was Studierende selbst festhalten. Die Nutzung wird angeleitet reflektiert, und diese Rückschau bestimmt die nächste begründete Wahl. Die drei Teile sind Phasen desselben Zyklus. Die Funktion stellt damit die Wirkbedingung selbst her, aber die Lernaktivität ändert sich: Reflektiert wird die eigene KI-Nutzung, nicht das eigene Lernen insgesamt. Ob angeleitete Reflexion über die eigene KI-Nutzung das Lernen so fördert wie Reflexion über das eigene Lernen, ist unseres Wissens nicht untersucht. Wir stufen die Übertragung deshalb als konzeptionell ein (Abschnitt 4.1). Mit dieser Stufe erfüllt die Funktion die Begründungsgüte nicht über den Mechanismus, sondern über ihre ausdrücklich

normative Begründung: Die Reflexion der eigenen KI-Nutzung ist Teil der Zielkompetenz, die Offenlegung folgt aus der Integrität wissenschaftlichen Arbeitens.

Zur Reflexion im Umgang mit KI liegt ein experimenteller Befund vor: In einem randomisierten Experiment mit 342 Studierenden senkte eine kurze Reflexionsaufforderung die Übernahme falscher KI-Ratschläge, während richtige ähnlich häufig übernommen wurden. Untersucht wurden allerdings begrenzte Entscheidungsaufgaben, nicht Lernen (Ren, 2026). Ob das Dokumentieren Lerngewinne bringt, ist unseres Wissens nicht untersucht. Insgesamt ist die Evidenz im KI-Kontext schwach; sie betrifft Entscheidungen, nicht Lernen, und lässt die Einstufung der Übertragung als konzeptionell unberührt.

Ethisch geht es um Integrität und Verantwortung für Folgen: Offenlegen heißt sichtbar machen, was eigenes Urteil ist, und dafür einstehen. Offengelegt werden dabei Art und Umfang der Nutzung, nicht notwendig vollständige Verläufe. Festgehaltene Eingaben können personenbezogene Daten enthalten, und auch der Umgang mit diesen Daten gehört zur Reflexion.

Geltung und Grenzen. Der Konsens gilt für alle Studierenden. Für Studierende mit wenig Vorwissen gilt aber bei Kernfunktion 1 eine Bedingung. Nach dem Expertise-Reversal-Effekt nützt Anleitung vor allem Lernenden mit wenig Vorwissen und verliert mit wachsendem Vorwissen ihren Nutzen oder wirkt sogar nachteilig (Kalyuga et al., 2003). Ohne ausreichende Wissensgrundlage kann ein eigener Versuch deshalb weniger wirksam sein als Instruktion zuerst. Befunde aus der Hochschule sind dazu uneinheitlich: Ein Experiment deutet an, dass ein eigener Versuch vor allem Studierenden mit etwas Vorwissen nützt (Giebl et al., 2021). Eine andere Studie mit zwei randomisierten Experimenten in Lehrveranstaltungen fand beim nahen Transfer das umgekehrte Muster: Dort nützte der eigene Versuch Studierenden mit weniger Vorwissen, Instruktion zuerst denen mit mehr Vorwissen (He et al., 2025). Solange diese Frage offen ist, gilt vorsorglich: Die eigene Vorleistung wird so angeleitet und begrenzt, dass sie an vorhandenes Wissen anknüpfen kann; fehlt diese Grundlage, geht eine angeleitete Einführung voraus.

Offen ist zudem, wie weit der Mechanismus von Kernfunktion 1 in anderen Fächern trägt, denn die Studien der Metaanalyse von Sinha und Kapur (2021) stammen überwiegend aus Mathematik, Naturwissenschaften und Medizin. Für fächerübergreifende Fähigkeiten zeigte sich sogar eine Tendenz zugunsten von Instruktion zuerst.

Zu Kernfunktion 2 gehört eine Abgrenzung: Dass KI-Feedback das Feedback von Lehrenden ersetzen kann, ist nicht belegt. In einem randomisierten Feldexperiment, in dem die Studierenden die Quelle des Feedbacks nicht kannten, verbesserte KI-Feedback die Argumentation beim Überarbeiten schriftlicher Arbeiten weniger als Feedback von Lehrenden, obwohl Studierende es als fairer wahrnahmen (Weidlich et al., 2025). Eine Replikation ist uns nicht bekannt. Die Befunde zu Vorwissen, fächerübergreifenden Fähigkeiten und KI-Feedback begrenzen Kernfunktion 1 und 2, stellen sie aber nicht grundsätzlich infrage: Die Befunde zum Vorwissen und zu fächerübergreifenden Fähigkeiten lassen sich auf benennbare Bedingungen eingrenzen. Der Befund zum KI-Feedback stammt aus einer Einzelstudie und betrifft nur die Annahme, KI-Feedback könne das Feedback von Lehrenden ersetzen; diese Annahme macht Kernfunktion 2 nicht. Zu Kernfunktion 3 haben wir keine einschränkenden Befunde gefunden; der Befund von Fabriz et al. (2014) betrifft die Form der Reflexion, nicht die Funktion.

Formunabhängigkeit. Nach dem Urteil, das die beteiligten Lehrenden aus ihren Fächern heraus treffen, sind die Kernfunktionen formunabhängig. Die Reichweite dieses Urteils ist begrenzt: Ob sich die Kernfunktionen auch in nicht textzentrierten Fächern umsetzen lassen, ist zu untersuchen (Kapitel 5). Formunabhängig heißt zudem nicht aufwandsneutral: Jede Kernfunktion lässt sich zwar mit geringen Mitteln erfüllen, etwa durch Notizen, die Studierende selbst führen, der Aufwand hängt aber von Format und Gruppengröße ab.

Kompetenzbezug. Jede Kernfunktion schützt mindestens eine Facette der KI-bezogenen Urteilskompetenz, wie bei den einzelnen Funktionen dargelegt: bei Kernfunktion 1 und 2 begrifflich begründet, bei Kernfunktion 3 über ihren Mechanismus, dessen Übertragung wir als konzeptionell einstufen. Keine der drei Kernfunktionen gefährdet zudem die fachliche Kompetenz: Kernfunktion 1 und 2 bauen sie über ihren Mechanismus auf, Kernfunktion 3 stützt sie mittelbar.

4.3 Optionale Gestaltungsfunktionen

Vier weitere Funktionen nennen wir optionale Gestaltungsfunktionen oder kurz Ausbauoptionen: Sie erfüllen den Kompetenzbezug, aber Begründungsgüte oder Formunabhängigkeit nicht; bei den vier hier vorgestellten Optionen sind es beide. Grundsätzliche Gegenevidenz haben wir für keine von ihnen gefunden. Ausbau bedeutet keine Abwertung, sondern eine begründete Erweiterung unter günstigen Bedingungen. Jede Ausbauoption baut auf einer Kernfunktion auf und erweitert sie, ohne die fachliche Kompetenz zu gefährden. Wir stellen die Optionen jeweils in derselben Folge vor: Funktion und erweiterte Kernfunktion, Kompetenzbezug, Begründungsgüte, Evidenz im KI-Kontext, Formunabhängigkeit und ethischer Aspekt.

Soziale Verankerung. In der Gruppe diskutieren Studierende erst die eigenen Einschätzungen, die sie nach Kernfunktion 1 festgehalten haben, dann die KI-Ergebnisse. Die Option erweitert damit Kernfunktion 1. Begrifflich lässt sich begründen, dass sie Ergebniskontrolle und Urteilsintegration übt: Abweichende Beiträge decken Fehler auf, und das eigene Urteil muss begründet oder revidiert werden. Der Lernmechanismus der Option ist die individuelle Vorbereitung auf die Zusammenarbeit. Ein systematisches Review verglich die individuelle Vorbereitung mit sofortiger Zusammenarbeit: Vorbereitung förderte meist das gemeinsame Schlussfolgern, beeinträchtigte aber meist das gemeinsame Erinnern (Mende et al., 2021). Nach dem Review können Vorbereitungsaufgaben, bei denen Studierende selbst etwas erzeugen, die Vorteile verstärken; direkt verglichen wurde das aber selten. Die Option nutzt diesen Mechanismus selbst, keine Analogie: Die Lernaktivität bleibt dieselbe (erst eigene Vorbereitung, dann Austausch), und die festgehaltenen Einschätzungen stellen die Wirkbedingung her. Weil die Befunde uneinheitlich sind und gerade die selbst erzeugte Vorbereitung, auf die die Option setzt, selten direkt verglichen wurde, ist der Mechanismus nach der vorsichtigeren Stufe (Abschnitt 4.1) schwach belegt (Begründungsgüte nicht erfüllt). Der Nachteil beim Erinnern ist keine grundsätzliche Gegenevidenz (Tabelle 2), weil die Option nicht auf das Erinnern zielt. Für das gemeinsame Einprägen von Stoff eignet sie sich deshalb weniger. Eine Metaanalyse fand für KI in kollaborativen Lehrformen große Effekte, allerdings auf Grundlage von nur elf Effektstärken, und unterschied nicht, ob mit oder ohne KI gemessen wurde (C. Fan et al., 2026). Die Evidenz im KI-Kontext ist deshalb schwach. Zudem braucht die Option ein Gruppenformat (Formunabhängigkeit nicht erfüllt). Ethisch geht es um Fairness und Perspektivität: Wessen Sicht wird verdrängt?

Iteration in fachlogischen Stufen. Ein komplexes Ergebnis entsteht in fachlogischen Stufen mit je eigener Feedback- und Überarbeitungsschleife, sodass die KI kein Gesamtprodukt auf einmal liefert. Die Option erweitert damit Kernfunktion 2, denn jede Stufe nutzt deren Feedbackmechanismus. Begrifflich lässt sich begründen, dass die Option neben der fachlichen Kompetenz die Urteilsintegration übt, weil die Ergebnisse der Stufen zu einem stimmigen Gesamtergebnis zusammenzuführen sind. Neu ist nur die Folge der Stufen. Belegt sind deshalb die Komponenten, also die einzelnen Schleifen (Abschnitt 4.2), nicht aber ihre Verbindung zur Gesamtstruktur. Diese Verbindung ist konzeptionell begründet (Begründungsgüte nicht erfüllt), und auch die Evidenz im KI-Kontext ist konzeptionell. Zudem braucht die Option eine geplante Folge mehrerer Aufgaben, die nicht jedes Format vorsieht (Formunabhängigkeit nicht erfüllt). Ethisch geht es um Integrität: Was in Stufen entsteht, macht den eigenen Anteil nachvollziehbar.

Vergleich mehrerer Werkzeuge. Studierende vergleichen die Ausgaben mehrerer Werkzeuge und gleichen sie an Quellen ab. Das vertieft die begründete Wahl, erweitert also Kernfunktion 3. Begrifflich lässt sich begründen, dass der Vergleich Einsatzentscheidung und Ergebniskontrolle übt: Unterschiede zwischen den Ausgaben machen sichtbar, dass das Ergebnis vom gewählten Werkzeug abhängt. Solche Unterschiede lassen sich vor allem an Quellen klären. Der Lernmechanismus der Option ist das Lernen durch Fallvergleich, das in einer Metaanalyse mittel belegt ist; am stärksten wirkte er dort, wo Lernende nach Gemeinsamkeiten suchten (Alfieri et al., 2013). Die Lernaktivität ändert sich aber: Verglichen werden Ausgaben von Werkzeugen statt fachlicher Fälle. Der Vergleich überträgt den Mechanismus deshalb nur per Analogie, und für die Begründungsgüte zählt die Stufe der Übertragung (Abschnitt 4.1). Diese Stufe ist konzeptionell (Begründungsgüte nicht erfüllt), ebenso die Evidenz im KI-Kontext. Zudem setzt der Vergleich Zugang zu mehreren Werkzeugen voraus (Formunabhängigkeit nicht erfüllt). Ethisch geht es um Daten und Rechte sowie Nachhaltigkeit: Jedes weitere Werkzeug erhält Eingaben und verbraucht Ressourcen.

Fading. Gemeint ist, KI-Unterstützung schrittweise auszublenden: Innerhalb einer Lehrveranstaltung wird sie geplant und am Lernfortschritt ausgerichtet zurückgenommen, bis hin zu Aufgaben ohne KI. Warum das sinnvoll ist, begründet Abschnitt 2.2: Maßstab für die fachliche Kompetenz ist, was ohne Werkzeug bleibt, und Unterstützung, die mit wachsender Kompetenz zurückgenommen wird, bewahrt diesen Maßstab. Fading gilt deshalb dort, wo fachliche Kompetenz ohne Werkzeug bestehen muss. Die Option erweitert Kernfunktion 3, weil die Lehrenden die Unterstützung zurücknehmen und die Studierenden so Schritt für Schritt selbst entscheiden müssen, ob und wofür sie KI noch einsetzen. Begrifflich lässt sich begründen, dass Fading vor allem die Prozessreflexion stützt, weil Studierende ihre Arbeitsweise an die wachsende eigene Kompetenz anpassen, und mittelbar die Einsatzentscheidung, bis hin zum begründeten Verzicht. Der Mechanismus der Option ist das schrittweise Ausblenden von Unterstützung: Das schrittweise Ausblenden vorgerechneter Lösungsschritte in Beispielaufgaben förderte den nahen Transfer, zusammen mit Aufforderungen zur Selbsterklärung auch den fernen Transfer (Atkinson et al., 2003). Eine Metaanalyse zu computergestützten Lernhilfen in MINT-Fächern fand für ausgeblendete Unterstützung keinen signifikanten Unterschied zu gleichbleibender Unterstützung (Belland et al., 2017). Auch die Art der Anpassung (nach Leistung, nach festem Plan oder nach Wahl der Lernenden) machte keinen Unterschied. Die Option nutzt diesen Mechanismus selbst, keine Analogie: KI-Unterstützung ist eine Lernhilfe, und sie wird wie andere Lernhilfen zurückgenommen, während die Aufgabe dieselbe bleibt. Die

Befundlage zum Ausblenden ist damit uneinheitlich, und nach der vorsichtigeren Stufe (Abschnitt 4.1) ist der Mechanismus schwach belegt (Begründungsgüte nicht erfüllt). Eine normative Begründung wie bei Kernfunktion 3 beanspruchen wir für Fading nicht: Der Maßstab aus Abschnitt 2.2 begründet, wo Fading gilt, nicht, dass es in jeder Lehrveranstaltung gewährleistet sein muss. Der Befund der Metaanalyse ist keine grundsätzliche Gegenevidenz (Tabelle 2): Er zeigt keinen Nachteil des Ausblendens, Fading kam in den einbezogenen Studien selten vor (rund ein Sechstel der Ergebnisse), und ob ohne Unterstützung gemessen wurde, also an dem Maßstab, auf den Fading zielt, berichtet die Metaanalyse nicht. Den Nachteil des Ausblendens nach festem Plan, den eine frühere Pilot-Metaanalyse derselben Gruppe gefunden hatte, bestätigte die Metaanalyse von Belland et al. nicht. Im KI-Kontext haben wir nur einen Befund zur Gefährdung gefunden, die Schulstudie aus Kapitel 1: Wer beim Üben eine übliche Chat-Variante genutzt hatte, schnitt in einer anschließenden Prüfung ohne KI schlechter ab als eine Vergleichsgruppe, die ohne KI geübt hatte (Bastani et al., 2025). Zur Funktion ist die Evidenz konzeptionell. Fading braucht eine geplante Folge mehrerer Aufgaben, die nicht jedes Format vorsieht (Formunabhängigkeit nicht erfüllt). Ethisch geht es um Machtverhältnisse: Wer ohne Werkzeug arbeiten kann, ist weniger von Anbietern abhängig.

4.4 Der Minimalkonsens im Überblick

Der Minimalkonsens besteht aus einem prozeduralen und einem substanziellen Kern. Der prozedurale Kern ist Constructive Alignment als gemeinsame Planungslogik, im Rahmen der geltenden Prüfungsordnung und mit dem Ziel, die Prüfungsordnung weiterzuentwickeln (Abschnitte 3.1 und 3.2). Für die Prüfung konkretisieren wir Constructive Alignment unter KI-Bedingungen in der KI-spezifischen Prüfungsanforderung: Wo KI genutzt wird, verlangt die Prüfung nach Möglichkeit, das KI-bezogene Urteil zu zeigen (Abschnitt 3.2). Den substanziellen Kern bilden die drei Kernfunktionen (Abschnitt 4.2). Ausbauoptionen erweitern den substanziellen Kern, wo die Bedingungen es erlauben (Abschnitt 4.3). Die Gestaltungsfunktionen und die Prüfungsanforderung sind Gestaltungsvermutungen (Abschnitt 3.3). Gesetzt sind dagegen Constructive Alignment als Planungslogik (Abschnitt 3.1), die Leitidee als normativer Bezugspunkt (Abschnitt 2.1), die an ihr ausgerichtete Zielkompetenz (Abschnitt 2.3) und der Maßstab, dass fachliche Kompetenz ohne Werkzeug bestehen muss, soweit sie für Kontrolle und Urteil nötig ist (Abschnitt 2.2). Der Konsens ist eine Orientierung für die Lehrgestaltung, kein Modell akademischer Bildung insgesamt. Tabelle 3 fasst die Einordnung der Bestandteile von Kern und Ausbau zusammen: Jede Zeile nennt, welche Facetten eine Funktion oder Anforderung schützt, wie gut ihr Mechanismus belegt ist, wie gut sie selbst beim Lernen mit KI untersucht ist und ob sie zum Kern oder zum Ausbau gehört.

Tabelle 3

Der Minimalkonsens im Überblick

Ebene	Gestaltungsfunktion bzw. Anforderung (Abschnitt)	Facetten	Mechanismus	KI-Kontext	Zuordnung
Prozeduraler Kern: Constructive Alignment (3.1)	KI-spezifische Prüfungsanforderung (3.2)	D2.1–D2.4	schwach (Rückwirkung)	konzeptionell	prozeduraler Kern, prozedural begründet (Constructive Alignment), keine vierte Kernfunktion
Substanzieller Kern	1 Eigene Vorleistung vor der KI-Nutzung (4.2)	D1, D2.3 (D2.2)	mittel	schwach (indirekt)	Kern, Bedingung bei wenig Vorwissen, Fächerreichweite offen
Substanzieller Kern	2 KI als Feedback zu eigenen Zwischenständen (4.2)	D1, D2.2, D2.3	mittel	schwach	Kern, Ersatz für Feedback von Lehrenden nicht belegt
Substanzieller Kern	3 KI-Nutzung begründet wählen, dokumentieren und reflektieren (4.2)	D2.1, D2.4 (D1)	mittel, Übertragung konzeptionell	schwach (Dokumentation nicht untersucht)	Kern, normativ begründet (Zielkompetenz, Offenlegung)
Ausbau (optional)	Soziale Verankerung (4.3)	D2.2, D2.3	schwach	schwach	Ausbau, nicht erfüllt: Begründungsgüte (Mechanismus schwach), Formunabhängigkeit (Gruppenformat)
Ausbau (optional)	Iteration in fachlogischen Stufen (4.3)	D1, D2.3	mittel (Komponenten), Gesamtstruktur konzeptionell	konzeptionell	Ausbau, nicht erfüllt: Begründungsgüte (Gesamtstruktur konzeptionell), Formunabhängigkeit (Aufgabenfolge)
Ausbau (optional)	Vergleich mehrerer Werkzeuge (4.3)	D2.1, D2.2	mittel (Fallvergleich), Übertragung konzeptionell	konzeptionell	Ausbau, nicht erfüllt: Begründungsgüte (Übertragung konzeptionell), Formunabhängigkeit (Zugang)
Ausbau (optional)	Fading (4.3)	D1, D2.4 (D2.1)	schwach (uneinheitlich)	konzeptionell	Ausbau, nicht erfüllt: Begründungsgüte (Mechanismus schwach), Formunabhängigkeit (Aufgabenfolge)

Anmerkung. Kürzel der Facetten wie in Tabelle 1: D1 fachliche Kompetenz, D2.1 Einsatzentscheidung, D2.2 Ergebniskontrolle, D2.3 Urteilsintegration, D2.4 Prozessreflexion und Verantwortung. Ebene: Teil des Minimalkonsenses. Facetten: was die Funktion oder Anforderung unter KI-Bedingungen schützt oder über ihren Mechanismus aufbaut, in Klammern mittelbar Gestütztes. Mechanismus und KI-Kontext: Evidenzstufe nach Abschnitt 4.1 für den Lernmechanismus und für den Kontext generativer KI, in Klammern erläutert. Eine zweite Stufe nach einem Komma gilt für Übertragung oder Gesamtstruktur und entscheidet, ob die Begründungsgüte über den Mechanismus erfüllt ist. Zuordnung: im Kern Besonderheiten wie die Bedingung bei wenig Vorwissen oder die normative Begründung, im Ausbau die nicht erfüllten Kriterien aus Tabelle 2, in Klammern der Grund. Der Kompetenzbezug ist bei allen Gestaltungsfunktionen erfüllt, gewichtige Gegenevidenz haben wir zu keiner gefunden. Belege in den genannten Abschnitten. Eigene Einschätzung.

Liest man die Spalten nebeneinander, bestätigt sich, was Abschnitt 4.1 festlegt: Die Zuordnung der Gestaltungsfunktionen zu Kern und Ausbau folgt aus den Kriterien in Tabelle 2, nicht aus der Evidenz im KI-Kontext, die in allen Zeilen schwach oder konzeptionell ist.

Planungsfragen. Für die eigene Lehrveranstaltung lassen sich die Kernfunktionen, die Vorwissensbedingung und die Prüfungsanforderung als Fragen lesen:

– Bezieht sich die KI-Nutzung auf eine festgehaltene eigene Vorleistung?
– Ist die Vorleistung bei wenig Vorwissen so angeleitet und begrenzt, dass sie an vorhandenes Wissen anknüpfen kann, und geht eine angeleitete Einführung voraus, wo diese Grundlage fehlt?
– Gibt KI Feedback zu einem eigenen Zwischenstand, und bleiben Kontrolle, Entscheidung über die Übernahme und Überarbeitung bei den Studierenden?
– Wird die KI-Nutzung vorab begründet gewählt, einschließlich des Verzichts, festgehalten und angeleitet reflektiert, und bestimmt die Rückschau die nächste begründete Wahl?
– Verlangt die Prüfung bei KI-Nutzung im Rahmen der geltenden Prüfungsordnung nach Möglichkeit, das KI-bezogene Urteil zu zeigen, und ist diese Anforderung den Studierenden vorab bekannt?

Die Fragen legen fest, was gewährleistet sein soll, nicht wie. Für das Beispiel aus Abschnitt 3.1, fortgeführt in Abschnitt 4.1, hieße das: Zuerst halten die Studierenden eine eigene Zusammenfassung des Fachtextes fest und begründen, wofür sie die KI nutzen (Einsatzentscheidung). Dann nutzen sie die KI als Feedback zu dieser Zusammenfassung und gleichen die Hinweise der KI am Originaltext ab (Ergebniskontrolle), bevor sie selbst überarbeiten (Urteilsintegration). Zuletzt halten sie fest und begründen, welche Hinweise sie übernommen oder verworfen haben und wie sie beim nächsten Text vorgehen würden (Prozessreflexion). Bei wenig Vorwissen geht eine angeleitete Einführung in den Fachtext voraus, und die Prüfung verlangt wie in Abschnitt 3.1 die begründete Kontrolle einer KI-Zusammenfassung am Originaltext. Das Beispiel veranschaulicht eine mögliche Form, belegt aber keine Wirksamkeit.

5 Fazit und Ausblick

Generative KI kann die Leistung steigern, ohne dass Kompetenz entsteht. Von diesem Leitproblem ausgehend haben wir eine Zielkompetenz bestimmt und einen gemeinsamen Planungs- und Entwicklungsrahmen beschrieben. Zu diesem Rahmen gehört die KI-spezifische Prüfungsanforderung: Wo KI genutzt wird, verlangt die Prüfung nach Möglichkeit, das KI-bezogene Urteil zu zeigen. Von der Zielkompetenz aus haben wir drei Kernfunktionen hergeleitet, die jede Lehrveranstaltung nach unserem Minimalkonsens erfüllen soll. Kernfunktion 1 und 2 erfüllen die Begründungsgüte über ihre belegten Lernmechanismen, Kernfunktion 3 über ihre ausdrücklich normative Begründung (Abschnitt 4.2). Vier Ausbauoptionen erweitern diesen Kern, wo die Bedingungen es erlauben (Abschnitt 4.3). Der Konsens legt fest, was gewährleistet sein soll, nicht wie, und die Formen bleiben fachspezifisch, wie Beitrag 2 (Bolsinger et al., 2026) an einer Lehrveranstaltung veranschaulicht. Aus dem Konsens folgt auch keine allgemeine Regel, KI möglichst häufig oder möglichst selten einzusetzen (Kapitel 1 und Abschnitt 2.3). Entscheidend ist, welche Kompetenzen eine Aufgabe aufbauen soll und ob die dafür nötigen Lernaktivitäten stattfinden.

Die Grenzen dieses Vorgehens sind deutlich. Die Evidenz zu den Lernmechanismen ist bei den Kernfunktionen mittel, bei den Ausbauoptionen teils schwach, im Kontext generativer KI aber durchgehend schwach oder konzeptionell, und die Forschung zu den Mechanismen unter KI-Bedingungen ist in Bewegung. Die Einstufungen sind offengelegte eigene Einschätzungen, keine formale Evidenzbewertung. Viele Befunde stammen aus anderen Bildungsstufen, Fächern oder Aufgaben. Dass sie auch für die Hochschullehre in unseren Fächern und bei unseren Aufgaben gelten, nehmen wir begründet an, untersucht ist es nicht. Die Bezüge der Gestaltungsfunktionen zur KI-bezogenen Urteilskompetenz sind begrifflich begründet, bei Kernfunktion 3 über eine nur konzeptionell eingestufte Übertragung; empirisch geprüft ist keiner. Die Zielkompetenz ist nicht validiert; Niveaustufen fehlen, und Messansätze, die ihre Facetten jenseits von Selbsteinschätzungen unmittelbar erfassen, stehen aus. Weil die Evidenz im KI-Kontext schwächer ist als die zu den Mechanismen selbst, untersuchen wir zuerst, ob die Formen, in denen Lehrende die Gestaltungsfunktionen umsetzen, die Wirkbedingungen ihrer Lernmechanismen unter KI-Bedingungen gewährleisten. Am wenigsten geklärt ist das bei den Ausbauoptionen. Offen ist auch, ob sich die Kernfunktionen in nicht textzentrierten Fächern umsetzen lassen.

Drei Linien setzen wir fort. Erstens untersucht die nächste Befragung anhand der Angaben der Studierenden, ob Lehrveranstaltungen nach der präzisierten Fassung des Konsenses die Wirkbedingungen der Kernfunktionen gewährleisten und ob die angestrebten Lernaktivitäten stattfinden. Befragt wird dafür erstmals zu Beginn und am Ende des Semesters. Weil eine Vergleichsgruppe fehlt, erlauben auch diese Daten keine Wirkungsaussagen (Abschnitt 3.3). Zweitens sammeln wir Erkenntnisse dazu, welche Prüfungsformen das KI-bezogene Urteil sichtbar machen, und richten diese Erkenntnisse auch an die Verantwortlichen im Studiengang. Drittens führen wir in einem Folgebeitrag in einer der nächsten Ausgaben von PAIHE fünf Spannungsfelder aus, die der Konsens nicht auflöst und die dieser Beitrag deshalb bewusst offenlässt: die Reichweite einer dauerhaften Kopplung von Person und Werkzeug, die kulturelle Prägung von Modellen und von fachlichen Bewertungsmaßstäben, die Abhängigkeit vom Werkzeug und von Anbietern, den Ressourcenverbrauch sowie das Verhältnis von Steuerung durch Lehrende und eigenständigem Urteil der Studierenden. Dort behandeln wir auch die Grenzen unseres Vorgehens und die grundsätzliche Kritik an den Annahmen von Constructive Alignment, etwa über die Fähigkeiten Studierender. Diese Kritik trifft auch die Lesart, der wir folgen (Abschnitt 3.1). Lehrende anderer Fächer laden wir ein, eigene Formen und Erfahrungen beizutragen.

Hinweis zur KI-Nutzung

Bei der Erarbeitung dieses Beitrags haben wir generative KI eingesetzt.

Literaturrecherche. Claude Opus 5.5 (Anthropic), Consensus (Webdienst, Stand August 2026), ChatGPT (OpenAI, GPT-5), Suche bis Oktober 2026. Jede verwendete Quelle haben wir mindestens an der Titelei geprüft, die zitierten Aussagen, soweit zugänglich, an Abstract oder Volltext.

Ideen und Gliederung. Claude Opus 5.5, September–Oktober 2026. Gliederung und Argumentationsgang haben wir festgelegt.

Formulieren und Überarbeiten. Claude Opus 5.5, im November auch Claude Fable 5.1, September–November 2026: Formulierungsentwürfe und Formulierungsvorschläge aus Gegenlesungen. Alle Formulierungen haben wir geprüft, wo nötig überarbeitet und freigegeben.

Übersetzen. Claude Opus 5.5, Oktober–November 2026: englischer Titel, englisches Abstract und Keywords. Diese haben wir geprüft.

Abbildungen und Tabellen. Claude Opus 5.5, September–November 2026: Entwürfe und Überarbeitungen der Abbildungen 1–3 und der Tabellen. Inhalt und Darstellung haben wir festgelegt und geprüft.

Prüfen von Belegen. Claude Opus 5.5, im November auch Claude Fable 5.1, September–November 2026: Abgleich von Belegen, KI-erstellte Gutachten zu Zwischenfassungen und KI-Gegenlesungen überarbeiteter Abschnitte, genutzt als Sammlung möglicher Einwände. Über jeden Einwand haben wir selbst entschieden.

Förderung und Interessenkonflikte

Für diesen Beitrag wurden keine Fördermittel in Anspruch genommen. Die Autor:innen sind Herausgeber:innen von PAIHE. Darin liegt ein Interessenkonflikt, denn im Startverfahren der ersten Ausgabe haben keine unbeteiligten Personen den Beitrag geprüft. Weitere Interessenkonflikte bestehen nicht.

Beiträge der Autor:innen (CRediT): Sven Warnke: Konzeption, Methodik, Visualisierung, Schreiben – Erstentwurf, Schreiben – Überarbeitung. Harald Bolsinger: Konzeption. Tanja Mühling: Konzeption.

Literatur

Alfieri, L., Nokes-Malach, T. J., & Schunn, C. D. (2013). Learning through case comparisons: A meta-analytic review. *Educational Psychologist, 48*(2), 87–113. https://doi.org/10.1080/00461520.2013.775712

Atkinson, R. K., Renkl, A., & Merrill, M. M. (2003). Transitioning from studying examples to solving problems: Effects of self-explanation prompts and fading worked-out steps. *Journal of Educational Psychology, 95*(4), 774–783. https://doi.org/10.1037/0022-0663.95.4.774

Bastani, H., Bastani, O., Sungu, A., Ge, H., Kabakcı, Ö., & Mariman, R. (2025). Generative AI without guardrails can harm learning: Evidence from high school mathematics. *Proceedings of the National Academy of Sciences, 122*(26), Artikel e2422633122. https://doi.org/10.1073/pnas.2422633122

Belland, B. R., Walker, A. E., Kim, N. J., & Lefler, M. (2017). Synthesizing results from empirical research on computer-based scaffolding in STEM education: A meta-analysis. *Review of Educational Research, 87*(2), 309–344. https://doi.org/10.3102/0034654316670999

Biggs, J. (1996a). Assessing learning quality: Reconciling institutional, staff and educational demands. *Assessment & Evaluation in Higher Education, 21*(1), 5–16. https://doi.org/10.1080/0260293960210101

Biggs, J. (1996b). Enhancing teaching through constructive alignment. *Higher Education, 32*(3), 347–364. https://doi.org/10.1007/BF00138871

Biggs, J., Tang, C., & Kennedy, G. (2022). *Teaching for quality learning at university* (5. Aufl.). Open University Press.

Blömeke, S., Gustafsson, J.-E., & Shavelson, R. J. (2015). Beyond dichotomies: Competence viewed as a continuum. *Zeitschrift für Psychologie, 223*(1), 3–13. https://doi.org/10.1027/2151-2604/a000194

Bolsinger, H. J., Warnke, S., & Mühling, T. (2026). The practice of reflective teaching with AI: Future skills development with offline LLMs. *Practice of Artificial Intelligence in Higher Education, 1*(1), Artikel 2. https://doi.org/10.83397/PAIHE.2026.02

Brandstetter, B., & Weiand, A. (2026). Künstliche Intelligenz in der Lehre an Hochschulen – einige Gedanken. *Didaktiknachrichten (DiNa)*, (6), 4–6. https://bayziel.de/wp-content/uploads/DiNa_26-06.pdf

Clark, A., & Chalmers, D. (1998). The extended mind. *Analysis, 58*(1), 7–19. https://doi.org/10.1093/analys/58.1.7

Fabriz, S., Dignath-van Ewijk, C., Poarch, G., & Büttner, G. (2014). Fostering self-monitoring of university students by means of a standardized learning journal: A longitudinal study with process analyses. *European Journal of Psychology of Education, 29*(2), 239–255. https://doi.org/10.1007/s10212-013-0196-z

Facione, P. A. (1990). *Critical thinking: A statement of expert consensus for purposes of educational assessment and instruction. Research findings and recommendations* (ED315423). ERIC. https://eric.ed.gov/?id=ED315423

Fan, C., Ke, L., Chen, Z., & Lv, P. (2026). Exploring the effect of GenAI on learning outcomes in higher education: A three-level meta-analysis. *Frontiers in Psychology, 17*, Artikel 1758670. https://doi.org/10.3389/fpsyg.2026.1758670

Fan, Y., Tang, L., Le, H., Shen, K., Tan, S., Zhao, Y., Shen, Y., Li, X., & Gašević, D. (2025). Beware of metacognitive laziness: Effects of generative artificial intelligence on learning motivation, processes, and performance. *British Journal of Educational Technology, 56*(2), 489–530. https://doi.org/10.1111/bjet.13544

Giebl, S., Mena, S., Storm, B. C., Bjork, E. L., & Bjork, R. A. (2021). Answer first or Google first? Using the Internet in ways that enhance, not impair, one's subsequent retention of needed information. *Psychology Learning & Teaching, 20*(1), 58–75. https://doi.org/10.1177/1475725720961593

Guo, L. (2022). How should reflection be supported in higher education? – A meta-analysis of reflection interventions. *Reflective Practice, 23*(1), 118–146. https://doi.org/10.1080/14623943.2021.1995856

Hawe, P., Shiell, A., & Riley, T. (2004). Complex interventions: How "out of control" can a randomised controlled trial be? *BMJ, 328*(7455), 1561–1563. https://doi.org/10.1136/bmj.328.7455.1561

He, C.-W., Fiorella, L., & Lemons, P. P. (2025). Does instruction-first or problem-solving-first depend on learners' prior knowledge? *Educational Psychology Review, 37*(1), Artikel 18. https://doi.org/10.1007/s10648-025-09993-3

Hornberger, M., Bewersdorff, A., & Nerdel, C. (2023). What do university students know about artificial intelligence? Development and validation of an AI literacy test. *Computers and Education: Artificial Intelligence, 5*, Artikel 100165. https://doi.org/10.1016/j.caeai.2023.100165

Kalyuga, S., Ayres, P., Chandler, P., & Sweller, J. (2003). The expertise reversal effect. *Educational Psychologist, 38*(1), 23–31. https://doi.org/10.1207/S15326985EP3801_4

Klieme, E., & Hartig, J. (2007). Kompetenzkonzepte in den Sozialwissenschaften und im erziehungswissenschaftlichen Diskurs. *Zeitschrift für Erziehungswissenschaft, 10*(Sonderheft 8), 11–29. https://doi.org/10.1007/978-3-531-90865-6_2

Kluger, A. N., & DeNisi, A. (1996). The effects of feedback interventions on performance: A historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological Bulletin, 119*(2), 254–284. https://doi.org/10.1037/0033-2909.119.2.254

Kofinas, A. K., Tsay, C. H.-H., & Pike, D. (2025). The impact of generative AI on academic integrity of authentic assessments within a higher education context. *British Journal of Educational Technology, 56*(6), 2522–2549. https://doi.org/10.1111/bjet.13585

Lee, J. D., & See, K. A. (2004). Trust in automation: Designing for appropriate reliance. *Human Factors, 46*(1), 50–80. https://doi.org/10.1518/hfes.46.1.50_30392

Loibl, K., Roll, I., & Rummel, N. (2017). Towards a theory of when and how problem solving followed by instruction supports learning. *Educational Psychology Review, 29*(4), 693–715. https://doi.org/10.1007/s10648-016-9379-x

Long, D., & Magerko, B. (2020). What is AI literacy? Competencies and design considerations. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (S. 1–16). ACM. https://doi.org/10.1145/3313831.3376727

Loughlin, C., Lygo-Baker, S., & Lindberg-Sand, Å. (2021). Reclaiming constructive alignment. *European Journal of Higher Education, 11*(2), 119–136. https://doi.org/10.1080/21568235.2020.1816197

McKenney, S., & Reeves, T. C. (2019). *Conducting educational design research* (2. Aufl.). Routledge. https://doi.org/10.4324/9781315105642

Mende, S., Proske, A., & Narciss, S. (2021). Individual preparation for collaborative learning: Systematic review and synthesis. *Educational Psychologist, 56*(1), 29–53. https://doi.org/10.1080/00461520.2020.1828086

Mühling, T., Warnke, S., & Bolsinger, H. J. (2026). Reflektierte KI-Nutzung in Lehrveranstaltungen: Befunde aus dem Pretest am Ende des Sommersemesters 2026. *Practice of Artificial Intelligence in Higher Education, 1*(1), Artikel 3. https://doi.org/10.83397/PAIHE.2026.03

Newby, L., & Cornelissen, L. (2025). A neutral toolkit? For a fundamental critique of constructive alignment. *Teaching in Higher Education, 30*(7), 1687–1702. https://doi.org/10.1080/13562517.2025.2465987

Niedermeier, S. (2026). Künstliche Intelligenz in der Hochschullehre zwischen Kompetenzentwicklung und didaktischer Gestaltung. *Didaktiknachrichten (DiNa)*, (6), 27–33. https://bayziel.de/wp-content/uploads/DiNa_26-06.pdf

Nückles, M., Roelle, J., Glogger-Frey, I., Waldeyer, J., & Renkl, A. (2020). The self-regulation-view in writing-to-learn: Using journal writing to optimize cognitive load in self-regulated learning. *Educational Psychology Review, 32*(4), 1089–1126. https://doi.org/10.1007/s10648-020-09541-1

Parasuraman, R., & Riley, V. (1997). Humans and automation: Use, misuse, disuse, abuse. *Human Factors, 39*(2), 230–253. https://doi.org/10.1518/001872097778543886

Rampelt, F., Matthes, W., Hannken-Illjes, K., Sandmeir, A., Gehrs, V., Horstmann, N., Kunz, A. M., Eigbrecht, L., Johannsen, T., Blum, S., Frank, S., Sutter, C., & Koch, H. (2025). *Future Skills 2030: Ein aktualisiertes Framework für Zukunftskompetenzen*. Stifterverband für die Deutsche Wissenschaft. https://doi.org/10.5281/zenodo.17910748

Ren, S. (2026). College students' metacognitive awareness of generative-AI reliance: An experimental study of decision confidence and attribution bias. *Frontiers in Psychology, 17*, Artikel 1926110. https://doi.org/10.3389/fpsyg.2026.1926110

Salomon, G., Perkins, D. N., & Globerson, T. (1991). Partners in cognition: Extending human intelligence with intelligent technologies. *Educational Researcher, 20*(3), 2–9. https://doi.org/10.3102/0013189X020003002

Sandoval, W. (2014). Conjecture mapping: An approach to systematic educational design research. *Journal of the Learning Sciences, 23*(1), 18–36. https://doi.org/10.1080/10508406.2013.778204

Scarfe, P., Watcham, K., Clarke, A., & Roesch, E. (2024). A real-world test of artificial intelligence infiltration of a university examinations system: A "Turing Test" case study. *PLOS ONE, 19*(6), Artikel e0305354. https://doi.org/10.1371/journal.pone.0305354

Scouller, K. (1998). The influence of assessment method on students' learning approaches: Multiple choice question examination versus assignment essay. *Higher Education, 35*(4), 453–472. https://doi.org/10.1023/A:1003196224280

Sinha, T., & Kapur, M. (2021). When problem solving followed by instruction works: Evidence for productive failure. *Review of Educational Research, 91*(5), 761–798. https://doi.org/10.3102/00346543211019105

Soderstrom, N. C., & Bjork, R. A. (2015). Learning versus performance: An integrative review. *Perspectives on Psychological Science, 10*(2), 176–199. https://doi.org/10.1177/1745691615569000

UNESCO. (2024). *AI competency framework for students*. https://doi.org/10.54675/JKJB9835

Warnke, S. (2026). KI-Nutzung und wissenschaftliches Arbeiten: Wie ein strukturiertes Lehrdesign Studierende zu einer reflektierten KI-Nutzung befähigen soll. *Didaktiknachrichten (DiNa)*, (6), 17–20. https://bayziel.de/wp-content/uploads/DiNa_26-06.pdf

Weidlich, J., Gotsch, F., Schudel, K., Marusic-Würscher, C., Mazzarella, J., Bolten, H., Bütler, D., Luger, S., Wohlfender, B., & Maag Merki, K. (2025). Teacher, peer, or AI? Comparing effects of feedback sources in higher education. *Computers and Education Open, 9*, Artikel 100300. https://doi.org/10.1016/j.caeo.2025.100300

Weinert, F. E. (2014). Vergleichende Leistungsmessung in Schulen – eine umstrittene Selbstverständlichkeit. In F. E. Weinert (Hrsg.), *Leistungsmessungen in Schulen* (3., aktual. Aufl., S. 17–31). Beltz. (Originalarbeit veröffentlicht 2001)

Willingham, D. T. (2007). Critical thinking: Why is it so hard to teach? *American Educator, 31*(2), 8–19.

Wineburg, S., & McGrew, S. (2019). Lateral reading and the nature of expertise: Reading less and learning more when evaluating digital information. *Teachers College Record, 121*(11), 1–40. https://doi.org/10.1177/016146811912101102

Winstone, N. E., Nash, R. A., Parker, M., & Rowntree, J. (2017). Supporting learners' agentic engagement with feedback: A systematic review and a taxonomy of recipience processes. *Educational Psychologist, 52*(1), 17–37. https://doi.org/10.1080/00461520.2016.1207538

Wisniewski, B., Zierer, K., & Hattie, J. (2020). The power of feedback revisited: A meta-analysis of educational feedback research. *Frontiers in Psychology, 10*, Artikel 3087. https://doi.org/10.3389/fpsyg.2019.03087

Wissenschaftsrat. (2026). *Intellektuelle Souveränität: Empfehlungen für die Hochschulbildung in Zeiten von generativer KI* (Drs. 3319-26). https://doi.org/10.57674/1evx-t906

Zimmerman, B. J. (2002). Becoming a self-regulated learner: An overview. *Theory Into Practice, 41*(2), 64–70. https://doi.org/10.1207/s15430421tip4102_2