

Middendorf, William

**Zur Notwendigkeit der Neuausrichtung summativer Assessments in
Bildungseinrichtungen als Antwort auf Entwicklungen und Nutzungen von KI**
2026, 7 S.



Quellenangabe/ Reference:

Middendorf, William: Zur Notwendigkeit der Neuausrichtung summativer Assessments in
Bildungseinrichtungen als Antwort auf Entwicklungen und Nutzungen von KI. 2026, 7 S. - URN:
urn:nbn:de:0111-pedocs-360444 - DOI: 10.25656/01:36044

<https://nbn-resolving.org/urn:nbn:de:0111-pedocs-360444>

<https://doi.org/10.25656/01:36044>

Nutzungsbedingungen

Gewährt wird ein nicht exklusives, nicht übertragbares, persönliches und beschränktes Recht auf Nutzung dieses Dokuments. Dieses Dokument ist ausschließlich für den persönlichen, nicht-kommerziellen Gebrauch bestimmt. Die Nutzung stellt keine Übertragung des Eigentumsrechts an diesem Dokument dar und gilt vorbehaltlich der folgenden Einschränkungen: Auf sämtlichen Kopien dieses Dokuments müssen alle Urheberrechtshinweise und sonstigen Hinweise auf gesetzlichen Schutz beibehalten werden. Sie dürfen dieses Dokument nicht in irgendeiner Weise abändern, noch dürfen Sie dieses Dokument für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, aufführen, vertreiben oder anderweitig nutzen.

Mit der Verwendung dieses Dokuments erkennen Sie die Nutzungsbedingungen an.

Terms of use

We grant a non-exclusive, non-transferable, individual and limited right to using this document.

This document is solely intended for your personal, non-commercial use. Use of this document does not include any transfer of property rights and it is conditional to the following limitations: All of the copies of this documents must retain all copyright information and other information regarding legal protection. You are not allowed to alter this document in any way, to copy it for public or commercial purposes, to exhibit the document in public, to perform, distribute or otherwise use the document in public.

By using this particular document, you accept the above-stated conditions of use.

Kontakt / Contact:

peDOCS
DIPF | Leibniz-Institut für Bildungsforschung und Bildungsinformation
Informationszentrum (IZ) Bildung
E-Mail: pedocs@dipf.de
Internet: www.pedocs.de

Mitglied der


Leibniz-Gemeinschaft

William Middendorf: Zur Notwendigkeit der Neuausrichtung summativer Assessments in Bildungseinrichtungen als Antwort auf Entwicklungen und Nutzungen von KI

Mit der Beurteilung und Bewertung von (Prüfungs-)Leistungen (summative Assessments) im Bildungssystem verbinden sich nicht nur pädagogische Aufgaben (z.B. Motivation oder Diagnose), sondern auch und – zumindest aus (hoch-)schulrechtlicher Perspektive – vorrangig Berechtigungen, insofern die Bewertungen insbesondere in Form von Zeugnissen Qualifikationen zertifizieren sollen, die den Zugang zu weiterführenden Bildungseinrichtungen oder zum Beschäftigungssystem eröffnen.

Nun lässt sich seit Jahren bei den Leistungsbewertungen an Schulen und Hochschulen oftmals eine Tendenz zur Milde feststellen, so wenn etwa seit 2012 die in Schulleistungsvergleichen (z.B. PISA) gemessenen Kompetenzen sinken, während die Abiturdurchschnittsnoten sich in diesem Zeitraum eher verbessert haben. Besonders bemerkenswert stellt sich die Situation für eine Reihe von Studienfächern an etlichen Hochschulen dar, so wenn etwa über die Hälfte der Absolventen ihr Studium mit der Bestnote abschließen, die zweitbeste Note (oftmals als „erheblich über dem Durchschnitt“ definiert) dann tatsächlich aber nur eine unterdurchschnittliche Leistung angesichts der Gesamtverteilung der Noten darstellt. (Middendorf: 3f)

Eine solche Entwicklung birgt die Gefahr einer latenten Entwertung von Zertifikaten, insofern sie Skepsis an der Güte der den Zertifikaten zugrunde liegenden Leistungsbewertungen hervorruft.

In jüngster Zeit droht nun aber eine andere Entwicklung, welche individuelle Leistungsbewertungen und Zertifikate von Schulen und Hochschulen grundsätzlich in Frage stellen könnte: Diese Entwicklung ist den in letzten Jahren rapide gestiegenen allgemein zugänglichen Verwendungsmöglichkeiten Künstlicher Intelligenz (KI) geschuldet, die spätestens seit Mitte 2026 praktisch für jedermann den Einsatz von KI-Agenten erlaubt, mit denen sich auf einfache Weise diverse digitale Produkte generieren lassen, die den fachlichen Anforderungen von Bildungseinrichtungen an bestimmte individuelle Leistungsnachweise umfassend entsprechen.

Nachfolgend soll auf diese gewachsenen Möglichkeiten von KI, die hiermit verbundenen Herausforderungen und möglichen Konsequenzen für die Gestaltung formativer Assessments (Formen individueller Leistungserbringung) in Schulen und Hochschulen und deren Reaktionen eingegangen werden. Sodann soll ein Ausblick auf die zukünftige Relevanz von Leistungsbewertungen und Zertifikaten von Schulen und Hochschulen angesichts der (weiteren) Entwicklung von KI genommen werden.

Die rapide gewachsene und weiterhin wachsende Möglichkeit von KI, bestimmte individuelle an (Hoch-)Schulen erbrachte Prüfungsleistungen zu substituieren

Seit November 2022 sind KI-Sprachmodelle (Large Language Models) allgemein zugänglich und erfreuen sich einer zunehmenden und inzwischen breiten Nutzung.

Solche KI-Sprachmodelle (nachfolgend allgemeiner KI-Assistenten genannt) verfügen über eine Schnittstelle (Dialogfenster) für Eingaben in menschlicher Sprache (Prompts), können diese „verstehen“ und die in den Eingaben enthaltenen Anforderungen verarbeiten, um sodann Antworten in Form von Texten auszugeben. Soweit, sehr stark vereinfacht, die Anfänge.

In den letzten Jahren hat sich die Fähigkeit der KI-Assistenten in einem rasanten Tempo weiterentwickelt: Sie können heute unter Zugriff auf Trainingsdaten und Internet sehr komplexe Prompts bearbeiten und qualitativ hochwertige Ausgaben in textlicher, auditiver und visueller Form generieren.

So zeigten sich in einer 2022/23 durchgeführten Studie zur Leistungsfähigkeit von ChatGPT-3.5 im Hochschul-E-Assessment noch Licht und Schatten: Teilweise konnten in digitalen Klausuren mit Studierenden noch vergleichbare oder sogar etwas bessere Leistungen erzielt werden, zum Teil offenbarten sich jedoch auch Schwierigkeiten und Grenzen (Quelle: Moebert et al.: 108-110).

Eine 2024 durchgeführte Studie mit einer späteren und leistungsfähigeren Version, nämlich GPT-4 (die allerdings deutlich leistungsschwächer als aktuelle Versionen ist), ergab bereits, dass der KI-Assistent zur Erbringung von (überdurchschnittlichen) Leistungen bei Hochschulprüfungen in der Biomedizin in der Lage war (Stribling et al.: 2-4).

Rein schriftliche Leistungen in Textform auf Hochschulniveau lassen sich also spätestens seit 2024 mit Hilfe von allgemein zugänglichen KI-Assistenten generieren, sofern der KI-Assistent auf eine angemessene Datenbasis zurückgreifen kann und die Prompts hinreichend fundiert formuliert sind. Dementsprechend bedurfte es bislang lediglich einer gewissen Prompting-Kompetenz, um z.B. angemessene Mega-Prompts zu formulieren. Orientierung hierfür boten diverse Prompting Frameworks (Liu et al.: 6ff).

Inzwischen sind weitere erhebliche Entwicklungsschritte der KI zu verzeichnen: Seit etwa Mitte 2026 sind nun allgemein zugängliche Möglichkeiten von KI verfügbar, die die Nutzung von KI erleichtern und die Palette der von KI generierbaren Produkte auf einfach handhabbare Weise erweitern.

Die Erleichterung der Nutzung wird durch das sog. Loop Engineering erreicht, bei dem der Nutzer lediglich ein Ziel definiert und die KI dann die Aufgaben selbstständig erkennt, ausführt und ggf. korrigiert, bis das Ziel realisiert ist. Die Notwendigkeit, wie beim Prompt Engineering durch (fachlich tiefere und differenziertere) Nachfolgeeingaben die Antworten der KI im Sinne des angestrebten Ziels sukzessiv zu verbessern, entfällt also, da diese Aufgabe nun beim Loop Engineering der KI-Assistent übernimmt (vgl. Gahn: 4).

Die Erleichterung der Möglichkeiten, die Palette der von KI generierbaren Produktarten zu erweitern, wird durch den Einsatz von KI-Agenten erreicht.

Ein KI-Agent ist ein KI nutzendes Softwaresystem, das für vom Nutzer definierte Ziele entsprechende Aufgaben eigenständig plant und ausführt (Stahl et al.: 1f).

Hierzu benötigt und nutzt der KI-Agent ihm zugewiesenen *Kontext*, zu dem Tools (z.B. Programme wie Power Point und Ordner mit Dateien) gehören können, um Prozesse zu steuern oder unter Verwendung der ihm zur Nutzung zugewiesenen Programme und Dateien digitale Produkte zu erstellen.

Dementsprechend kann ein KI-Agent dann beispielsweise nicht nur textliche Aussagen (z.B. Bullet Points) für die Übernahme in Präsentationen formulieren, sondern – sofern ihm Zugriff auf etwa Power Point und angemessene digitale Literatur gestattet wurde – fachlich fundierte und kompetent designte Präsentationen selbstständig erstellen. Auf diese Weise lässt sich also die Erstellung professioneller Präsentationen komplett an die KI delegieren.

Dabei gilt für alle von KI-Assistenten und KI-Agenten generierten digitalen Produkte: Da der KI-Output nicht algorithmisch eindeutig und reproduzierbar festgelegt ist, lässt sich anhand des durch KI generierten Produkts die Autorenschaft des Produkts nicht belegen. Dies bedeutet insbesondere, dass für ein durch KI generiertes, aber als individuelle Leistung ausgegebenes Produkt die Täuschung nicht überzeugend nachgewiesen werden kann, ganz abgesehen davon, dass ein KI-generiertes Produkt auch menschlich nachbearbeitet werden kann, bevor es anderen Personen zugänglich gemacht wird.

Hieraus folgt, dass alle nicht unter Aufsicht zu erstellenden schriftlichen Leistungen (Referate, Präsentationen, Handouts, Poster, Videos usw.) in hochwertiger Qualität auch von KI generiert werden können, ohne dass die Autorenschaft der KI anhand des vorgelegten Produkts belegt werden könnte.

Angesichts dieser Möglichkeiten ist von hoher Relevanz, wie Schüler/innen und Studierende KI für ihre Arbeit in Schule und Hochschule nutzen und voraussichtlich zukünftig nutzen werden, um Entwicklungsbedarfe für die Gestaltung formativer Assessments abzuleiten.

Zur Nutzung von KI durch Lernende für Schule und Studium

Erfolgte die Nutzung von KI anfänglich eher zögerlich, so ist die Verwendung von KI durch Schüler/innen und Studierende heute weit verbreitet. Nach der JIM-Studie 2025, die das Mediennutzungsverhalten von 12- bis 19-Jährigen in Deutschland untersucht hat, nutzten 2025 rund drei Viertel der Jugendlichen KI-Assistenten für schulische Haus- und Lernaufgaben, wobei der Ki-Einsatz im Vergleich zum Vorjahr (2024) deutlich zugenommen hat (JIM: 62).

Und für eine Studie im Auftrag des Branchenverbandes Bitkom gab über die Hälfte der befragten Schüler/innen an, KI für die Erbringung schulischer Leistungen (Präsentationen, Aufsätze usw.) einzusetzen, wobei sogar 9 % angaben, KI während Prüfungen, Tests oder Klassenarbeiten zu verwenden (Bitkom: 13).

Auch die Nutzung von KI durch Studierende für Studienzwecke ist längst Alltag. Nach einer Studie im Auftrag des CHE (Centrum für Hochschulentwicklung) im Wintersemester 2025/26 nutzten über 80 % der Studierenden KI regelmäßig für Studienzwecke wie Recherche, Vermittlung von Überblickswissen, Brainstorming, Zusammenfassung von Dokumenten oder Textgenerierung (Horstmann et al.: 1).

Über die individuelle anekdotische Evidenz hinaus darf davon ausgegangen werden, dass KI heute im Studium von einem Teil der Studierenden auch bei der Erstellung von Prüfungsleistungen eingesetzt wird, ohne dass dies immer hinreichend kenntlich gemacht wird. Das Ausmaß missbräuchlicher Nutzung von KI lässt sich aus naheliegenden Gründen empirisch kaum ermitteln. Es erscheint allerdings nicht lebensfremd, von einer nicht nur marginalen Quote missbräuchlicher KI-Nutzung im Sinne der Prüfungsordnungen auszugehen.

Umgang der Bildungseinrichtungen mit der Möglichkeit der Nutzung von KI bei der Erbringung von Prüfungsleistungen

Die vorstehend beschriebenen Potenziale der KI waren für Schulen und Hochschulen Anlass, Empfehlungen für die Nutzung von KI in Bildungseinrichtungen zu geben. Diese Empfehlungen hatten für Schulen zumeist den Charakter eines Verhaltenscodexes, der etwa die Beachtung datenschutz- und urheberrechtliche Aspekte, einen transparenten Umgang mit KI durch Schüler/innen und Schüler oder die Unterlassung missbräuchlicher KI-Verwendung bei der Erbringung schulischer Leistungen einforderte. Aber es gab auch Hinweise zur Gestaltung schulischer Leistungsüberprüfungen. So empfahl die Ständige Wissenschaftliche Kommission der Kultusministerkonferenz (SWK) 2024, den Fokus bei (Prüfungs-)Aufgaben auf projektartige und prozess-orientierte Formate sowie auf besondere Anforderungen an das Prompting oder die kritische Reflexion des KI-Outputs zu legen (SWK: 16f).

Und die Kultusministerkonferenz (KMK) empfahl ebenfalls 2024 ergänzend, Prüfungsformate, die juristisch nicht einwandfrei der in der Aufgabe geforderten eigenständigen Leistung einer Schülerin bzw. eines Schülers zugerechnet werden können, abzuschaffen oder grundlegend weiterzuentwickeln (KMK: 7).

Leitfäden aus dem Hochschulbereich wie etwa aus der Universität Köln fordern bei der KI-Nutzung die Beachtung rechtlicher Aspekte und der Grundsätze wissenschaftlichen Arbeitens und gaben zumeist Hinweise auf zulässige Nutzungen (z.B. Brainstorming oder vorbereitende Recherchen) und unzulässige Nutzungen (insbesondere die Übernahme KI-generierter Inhalte als Eigenleistung oder die Verwendung von KI als wissenschaftliche Quelle) (Uni Köln: 1f).

Intensiv mit der Thematik befasste sich das „Hochschulforum Digitalisierung“, welches empfiehlt, die Produktbewertung auf die Prozesssichtbarkeit zu verlagern, KI-resiliente Aufgaben (Transfer, Kontextbezug, kritische Abwägung) zu verwenden und mündliche Ergänzungen zur Absicherung fachlicher Durchdringung vorzusehen, wobei Präsenzprüfungen letztlich als einzige pragmatische Option betrachtet werden, sofern der KI-Einsatz ausgeschlossen werden soll (Becker et al.: 43).

Zusammenfassend lässt sich festhalten, dass sowohl für den Schul- als auch den Hochschulbereich ein Schwerpunkt auf Verhaltensregeln gesetzt wird; ergänzend wird für „KI-resiliente“ Prüfungsformate plädiert, die also weitgehend unzugänglich für eine KI-Nutzung sind.

Zur Forderung und Umsetzbarkeit KI-resilienter Prüfungsformate

Alle Prüfungen, die unter Aufsicht absolviert werden und für die daher die Nutzung unzulässiger Hilfsmittel wie KI ausgeschlossen werden kann, können als klar KI-resilient charakterisiert werden. Dies gilt auch für digital gestützte Prüfungen an digitalen Endgeräten, sofern hierbei der Zugriff auf nicht zugelassene digitale Quellen wie z.B. das Internet wirksam gesperrt werden kann.

Bei mündlichen Prüfungen und Klausuren unter Aufsicht kann damit die Autorenschaft der vorgelegten Leistung gesichert werden.

Allerdings ist die Organisation und Durchführung von mündlichen Prüfungen zeitlich aufwendig und insbesondere im Hochschulbereich (wegen des Zwei-Prüfer-Prinzips) personalintensiv.

Klausuren bleiben auch im KI-Zeitalter ein geeignetes Prüfungsformat, wobei digital gestützte Prüfungen mit gebundenen Aufgaben einschließlich digitaler Auswertungen und zukünftig eventuell KI-gestützte (Vor-)Bewertungen von Leistungen bei Prüfungen mit auch offenen Aufgaben sogar Entlastungspotenziale für Lehrende versprechen.

Ambivalent ist hier das Prüfungsformat „Präsentation“ zu beurteilen. Sofern die Präsentation aus Folien mit textlichen und visuellen Aussagen und hierauf bezogenen ergänzenden mündlichen Erläuterungen der vortragenden Person besteht, ist im KI-Zeitalter nur für diese Erläuterungen eine individuelle Leistung eindeutig zuordbar.

Präsentationen herkömmlicher Art lassen daher nur eingeschränkt Rückschlüsse auf die individuelle Leistung des Prüflings zu. Dies gilt besonders dann, wenn Prüflinge sich eng an die Folienaussagen halten und diese weitgehend nur repetieren.

Allerdings lässt sich hier das Prüfungsformat „Präsentation“ weitgehend KI-resilient gestalten. Dafür ist zum einen darauf hinzuwirken, dass die Folien jeweils möglichst nur eine Kernaussage mit einer visuellen Unterstützung aufweisen. Der mündliche Vortrag kann sich dann nicht in einer Repetition der Folienaussagen erschöpfen; vielmehr muss die jeweilige Kernaussage mündlich erläutert und argumentativ vertreten werden. Auf diese Weise steigt der Stellenwert des mündlichen Vortrags. Darüber hinaus kann dem Prüfling vorgegeben werden, den Vortrag mit einer Folie zu beenden, auf der sich 3-5 gehaltvoll (kontrovers) diskutierbare Thesen befinden, zu denen dann eine Diskussion stattzufinden hat, die der Prüfling angemessen zu moderieren hat. Auch diese Moderationstätigkeit kann Teil einer (KI-unabhängigen) Prüfungsleistung sein. Schließlich kann ein weiterer Teil der Prüfungsleistung darin bestehen, die authentische Diskussion schriftlich zu analysieren und zu reflektieren. Auch bei dieser Aufgabe kann KI kaum behilflich sein.

Die Präsentation behält also nicht nur ihre Berechtigung als ein Baustein im Unterricht oder in einer Seminarsitzung, sie kann auch als KI-resiliente Prüfungsleistung gestaltet werden.

Es bleibt die Frage, ob und ggf. unter welchen Voraussetzungen schriftliche Referate, Hausarbeiten und Abschlussarbeiten (z.B. Bachelor- oder Masterarbeiten) noch als individuelle Prüfungsleistungen akzeptiert werden können. Dass die mit solchen Arbeiten intendierte intensive Auseinandersetzung mit einer fachlich anspruchsvollen Problemstellung eine bedeutsame Leistung ist, zu der insbesondere im Rahmen eines Studiums möglichst ein Befähigungsnachweis (Befähigung zu wissenschaftlicher Arbeit) erbracht werden sollte, liegt auf der Hand. Nur stellt sich im KI-Zeitalter die Frage, wie weitgehend gesichert werden kann, dass der Prüfling auch tatsächlich Autor der abgelieferten Arbeit ist.

Die diesbezüglichen Empfehlungen, dass schriftliche Arbeiten (ohne Aufsicht) sich auf praktische authentische Prüfungssettings (z.B. aktuelle lokale Praxisfälle oder persönliche Erfahrungen) beziehen sollen, begegnen allerdings erheblichen didaktischen Herausforderungen. Denn authentische Prüfungssettings dürften nicht über Medien vermittelbar sein, da diese der KI zugänglich wären und von ihr analysiert und interpretiert werden könnten. Authentische Prüfungssettings müssten also für den Prüfung nicht medial vermittelte, sondern unmittelbar erfahrene Realität sein, die er deshalb nur ohne KI-Unterstützung erfassen und analysieren könnte. Solche Prüfungssettings sind möglich (so z.B. eine von einer konkreten Unterrichtserfahrung als Hospitant ausgehende Prüfung), sie wären aber nicht für jedes Seminar- oder Prüfungsthema in einem beliebigen Studienfach verfügbar und erst recht nicht in der Anzahl, wie sie eine Vielzahl von Prüflingen erfordern würde. Auch könnte nicht ausgeschlossen werden, dass ein Prüfling seine direkte Realitätserfahrung (authentischer Fall) zwar selbst angemessen erfasst (kriterienorientiert beobachtet und zutreffend beschreibt), das Ergebnis dann aber anschließend der KI zugänglich macht und durch diese sodann analysieren und interpretieren lässt.

Der mögliche Einwand, dass die Aufgaben der Analyse und Interpretation so gestaltet werden müssten, dass die KI hier keine wirksame Unterstützung leisten könne, überzeugt nicht. Denn in der Analyse und Interpretation soll der Prüfling ja gerade fachliche Kompetenzen nachweisen. Und diese Kompetenzen werden ja nicht akut, individuell und subjektiv definiert. Sie stellen vielmehr ein anhand fachlicher Indikatoren intersubjektiv feststellbares Verhalten dar, das üblicherweise in einschlägiger Fachliteratur beschrieben ist und damit der KI zugänglich ist. Kurzum: Sofern schriftliche Arbeiten ohne Aufsicht sich auf authentische Prüfungssettings beziehen, steht die Autorenschaft des Prüflings für einen Teil der Prüfungsleistung in Frage.

Dies gilt auch für Prüfungssettings, die den Fokus auf Prozessbewertungen (statt Produktbewertungen), Reflexionen und Argumentationen legen. Inzwischen sind KI-Assistenten in der Lage, (auf Wahrscheinlichkeitsberechnungen basierend) nicht nur deklaratives Wissen zu generieren, sondern auch konditionales Wissen. Sie können damit auch Prozesse des Verstehens, des Argumentierens, des Reflektierens und des strategischen Denkens generieren.

Zur Zukunft von KI und die Relevanz von Bildungszertifikaten – ein Ausblick mit Unbehagen

Die bisherigen Prüfungsleistungen in Schule und Hochschule haben alle ihre jeweilige Bildungsberechtigung, insofern sie geeignet sind, Lernprozesse und den Erwerb von Kompetenzen zu befördern und letztere nachzuweisen. Handlungsnotwendigkeit ist also nicht wegen der Bildungsintention angezeigt, sondern wegen der sehr einfachen Möglichkeit der Vortäuschung individueller Leistung bei bestimmten Prüfungsformaten. Hierbei sind die Bereiche Schule und Hochschule getrennt zu betrachten. Denn im Bereich von Schule gibt es zwar bislang auch nicht unter Aufsicht erstellte schriftliche Leistungen (z.B. schriftliche Referate), doch fallen diese bei Zeugnisnoten üblicherweise wenig ins Gewicht, da i.d.R. schriftliche Arbeiten unter Aufsicht (Klassenarbeiten, Klausuren) und sonstige Leistungen im Unterricht den dominierenden Einfluss auf die Leistungsbewertung haben (von Projekt- und Praxisleistungen in bestimmten Jahrgangs-

stufen/Bildungsgängen abgesehen). Schulen können sich also ohne große Umstellungen bei Leistungsüberprüfungen auf KI-resiliente Prüfungsformate beschränken.

Anders verhält es sich an Hochschulen, zumindest für eine Reihe von Studiengängen. Hier sind oftmals Studien- und Prüfungsleistungen wie Präsentationen, Poster, Essays, schriftliche Hausarbeiten sowie insbesondere Bachelor- und Masterarbeiten traditionelle Formate einer gewachsenen Prüfungskultur. Dies gilt besonders für Hochschulseminare, bei denen bislang – wie ein kursorischer Blick in diverse auch aktualisierte Modulbeschreibungen besonders für geistes- und sozialwissenschaftliche Studiengänge zeigt – schriftliche Hausarbeiten und Präsentationen ein gängiges Prüfungsformat darstellen.

Bislang ist noch kein allgemeiner Trend erkennbar, von „KI-anfälligen“ Prüfungsformaten Abstand zu nehmen. Die Erfahrung zeigt allerdings, dass auch Appelle und Empfehlungen von Hochschulen zum verantwortlichen und transparenten Umgang mit KI die Vortäuschung einer Prüfungsleistung durch missbräuchliche Verwendung von KI nicht immer verhindern. Und da die Performance von KI-Assistenten kontinuierlich zunimmt, werden auch die von KI erbrachten Prüfungsleistungen immer besser. Dies könnte unter Umständen die ärgerliche Folge haben, dass die ehrlichen Studierenden bei der Leistungsbewertung auch noch das Nachsehen haben könnten, wenn nämlich ihre individuelle Leistung mit der von KI generierten Leistung (z.B. bei der Vorlage von Bachelorarbeiten) verglichen werden. Es liegt auf der Hand, dass solche Aussichten die Verwendung von KI bei der Erbringung von Prüfungsleistungen nicht reduzieren, sondern eher befördern würden.

Zudem könnte die Beibehaltung „KI-anfälliger“ Prüfungsformate perspektivisch Folgen für die Relevanz sowie Akzeptanz entsprechender Hochschulzeugnisse und auch die Reputation der jeweiligen Hochschule bzw. ihrer entsprechenden Studienfächer haben. Denn wenn die Zeugnisse aufgrund zugrunde liegender Prüfungsformate Zweifel an der Autorenschaft der zertifizierten Leistungen und damit korrespondierender Kompetenzen der Absolventen begründen, reduziert sich die Aussagekraft von Hochschulzeugnissen, was nicht nur Auswirkungen auf die Reputation des absolvierten Studiengangs, sondern auch auf die Berufschancen der Absolventen hätte. Stark überspitzt formuliert: Wenn die zertifizierte Kompetenz gar nicht die des Absolventen, sondern die der KI ist, könnte sich der betriebliche Einsatz der KI statt des Absolventen anbieten.

Es gibt also reichlich Anlass und Grund, KI-anfällige zeitnah durch KI-resiliente Prüfungsformate zu ersetzen, auch wenn damit in der Vergangenheit bewährte Formen der Leistungserbringung und -bewertung (leider) betroffen sind.

Literatur

Becker, S./Leifeld, J./ Lüthi, R. et al. (2026): Leitlinien-Check 2026: Ein Update zu generativer KI an Hochschulen. Berlin: Hochschulforum Digitalisierung; verfügbar unter:

https://hochschulforumdigitalisierung.de/wp-content/uploads/2026/05/Blickpunkt_KI-Leitlinien2026.pdf

Bitkom: Wintergest, R. im Auftrag von Bitkom e.V.: Digitale Schule – so wollen Schülerinnen und Schüler lernen. Berlin 2025; verfügbar unter: <https://www.bitkom.org/sites/main/files/2025-05/Bitkom-Praesentation-Digitale-Schule-2025.pdf>

Gahn, F.: Prompt Engineering 2026: Vom perfekten Prompt zum Loop Engineering; verfügbar unter: <https://florian-gahn.de/blog/prompt-engineering-loop-engineering> (Zugriff am 09.08.2026)

Horstmann, N./Hüsch, M./Breiter, A.: Künstliche Intelligenz im Studium – die Sicht von Studierenden im Wintersemester 2025/26; verfügbar unter: <https://hochschuldaten.che.de/kuenstliche-intelligenz-im-studium-die-sicht-von-studierenden-im-wintersemester-2025-26/>

JIM: Medienpädagogischer Forschungsverbund Südwest (Hg.): JIM 2025. Jugend, Information, Medien. Basisuntersuchung zum Medienumgang 12- bis 19-Jähriger in Deutschland. Stuttgart 2025; verfügbar unter: https://mpfs.de/app/uploads/2025/11/JIM_2025_PDF_barrierearm.pdf

KMK: Kultusministerkonferenz (Hg.): Handlungsempfehlung für die Bildungsverwaltung zum Umgang mit Künstlicher Intelligenz in schulischen Bildungsprozessen. Themenspezifische Handlungsempfehlung. Beschluss Bildungsministerkonferenz vom 10.10.2024. Berlin 2024; verfügbar unter: https://www.kmk.org/fileadmin/Dateien/veroeffentlichungen_beschluesse/2024/2024_10_10-Handlungsempfehlung-KI.pdf

Liu, X./Wang, J./Sun, J.: Prompting Frameworks for Large Language Models: A Survey; verfügbar unter: <https://arxiv.org/pdf/2311.12785>

Middendorf, W.: Bildungssystem auf abschüssiger Bahn – heutige Herausforderungen und Handlungsnotwendigkeiten für die verschiedenen stake-holder. 2025; verfügbar unter: https://www.pedocs.de/volltexte/2025/32380/pdf/Middendorf_2025_Bildungssystem_auf_abschuessiger.pdf

Moebert, T./Samoilova, E./Wiepke, A.: ChatGPT-3.5 im Hochschul-E-Assessment: Prüfungsleistung, akademische Integrität und zukünftige Bewertungsansätze. In: Informatik Spektrum (2024) 47:105–112; verfügbar unter: <https://doi.org/10.1007/s00287-024-01575-x>

Stahl, T./Schwab, J./Strohm, F.: KI Agent – Einführung und Umsetzung; verfügbar unter: <https://www.ipa.fraunhofer.de/de/loesungen/digitalisierung-und-ki/strategie-und-smart-services/ki-agent.html> (Zugriff am 09.08.2026)

Stribling, D./Xia, Y./Amer, M.K. et al.: The model student: GPT-4 performance on graduate biomedical science exams. Sci Rep 14, 5670 (2024). <https://doi.org/10.1038/s41598-024-55568-7>

SWK: Ständige Wissenschaftliche Kommission der Kultusministerkonferenz (SWK) (Hg.): Large Language Models und ihre Potenziale im Bildungssystem. Impulspapier der Ständigen Wissenschaftlichen Kommission der Kultusministerkonferenz. Bonn 2024; verfügbar unter: https://www.pedocs.de/volltexte/2024/28303/pdf/SWK_2024_Large_Language_Models.pdf

Uni Köln: Universität zu Köln – Humanwissenschaftliche Fakultät (Hg.): Leitfaden für Studierende: Verantwortungsvoller Umgang mit KI. Stand 5.5.2026; verfügbar unter: https://www.hf.uni-koeln.de/data/bisuiia/File/KI-Leitfaden%20fur%20Studierende_05-05-2026.pdf