

2. Grundlagen generativer KI-Modelle

Empirische Befunde dokumentieren konsistent einen raschen und weitreichenden Wandel in der Art, wie Studierende sich mit akademischer Arbeit auseinandersetzen. Längsschnittliche Erhebungen zeigen, dass mittlerweile über 80 Prozent der Studierenden generative KI-Systeme regelmäßig für Essays, Recherchen und Prüfungsvorbereitung nutzen (Garrel & Mayer, 2025); Untersuchungen zur akademischen Integrität beziffern den Anteil KI-assistierter Aufgabebearbeitung auf 30 bis 60 Prozent, mit erheblicher Variation nach Disziplin und Institution (Eaton & Turner, 2024). Diese Zahlen unterschätzen die tatsächliche Nutzung vermutlich, da Selbstauskunftsverfahren eine gut dokumentierte Verzerrung durch sozial erwünschtes Antwortverhalten aufweisen. Das Ausmaß der Verbreitung ist damit empirisch belegt – analytisch strittig bleibt hingegen, welche Bedeutung diesem Phänomen beizumessen ist.

Die technologische Grundlage dieser Entwicklung bilden in der Regel Sprachmodelle (Large Language Models, LLMs). Deren Funktionsprinzipien – und insbesondere die Grenzen ihrer Leistungsfähigkeit – sind Voraussetzung jeder ernsthaften Analyse bildungsrelevanter Auswirkungen generativer KI. Das vorliegende Kapitel legt diese technologischen Grundlagen. Sein Zweck ist jedoch nicht allein explikativ: Die technische Analyse dient zugleich der Fundierung des in Kapitel 1 entwickelten Kernarguments, demzufolge generative KI die Verbindung zwischen kognitivem Prozess und akademischem Produkt durchtrennt – mit weitreichenden Konsequenzen für Prüfung, Zertifizierung und den Begriff nachgewiesener Kompetenz.

2.1. Funktionsweise und Architektur von Sprachmodellen

Sprachmodelle basieren im Wesentlichen auf statistischen Verfahren zur Textvorhersage. Die Generierung von Ausgabeinheiten – sogenannten Token (kleinste linguistische Verarbeitungseinheiten, in die ein Sprachmodell Eingaben zerlegt, bevor eine mathematische Verarbeitung erfolgt) – vollzieht sich durch die Prognose wahrscheinlicher Fortsetzungen von Eingabesequenzen. Diese Modelle wurden auf enormen Mengen textueller Daten trainiert – Bücher, wissenschaftliche Artikel, Webseiten, Diskussionsforen – und haben dabei sprachliche Muster internalisiert: welche Wörter typischerweise aufeinander folgen, welche Satzstrukturen in welchen Kontexten üblich sind, wie Argumente konstruiert, Fragen beantwortet und Erklärungen formuliert werden. Die Textgenerierung erfolgt Token für Token, wobei jeweils die statistisch wahrscheinlichste Fortsetzung in Abhängigkeit vom gesamten vorangegangenen Kontext bestimmt wird.

Sprachmodell-Ausgaben können formal einwandfrei sein – grammatisch korrekt, stilistisch angemessen, strukturell überzeugend –, sie werden jedoch grundsätzlich und ausnahmslos ohne Verankerung in Prozessen des Verstehens, Begründens oder der Wissensbildung erzeugt. Sprachmodelle bilden Korrelationen ab, keine kausalen Zusammenhänge. Ein Verständnis von Ursache und Wirkung setzt ein internes Weltmodell voraus, das über die statistische Repräsentation von Texten hinausgeht.

Dass Sprachmodelle trotz dieses fehlenden Weltmodells zu formal überzeugenden Analysen und Argumentationen fähig sind, erklärt sich aus der Struktur des sogenannten latenten Raums – des hochdimensionalen Vektorraums der Modellparameter. In diesem Raum werden nicht nur einzelne Wörter und Token isoliert repräsentiert, sondern auch Argumentationsverläufe als geometrische Strukturen. Die mathematische Architektur ermöglicht es dem Modell, semantische Distanzen zu überbrücken und thematisch verwandte Konzepte in Beziehung zu setzen.

Bei einer Anfrage etwa nach einer Strategie für kleine und mittlere Unternehmen in einem disruptiven Marktumfeld navigiert das Modell im Vektorraum von einer durch die Eingabe definierten Position zu einem Bereich, in dem Konzepte wie Agilität, Ressourcenallokation und Marktanpassung geometrisch eng beieinander liegen. Die generierte Antwort stellt dabei keine kausale Analyse dar, sondern das Ergebnis einer pfadabhängigen Traversierung des Vektorraums entlang maximaler semantischer Dichte – eine Synthese, die logische Verbindungen simuliert, ohne kausale Relationen zu modellieren.

Hinzu kommt, dass Logik – insbesondere in wissenschaftlichen und fachlichen Texten – nicht nur Inhalt, sondern ein hochgradig formalisiertes sprachliches Muster ist. Ein auf Millionen von Lehrbüchern, wissenschaftlichen Beiträgen

Teil I: Grundlagen – KI als epistemische Disruption

und strukturierten Argumentationen trainiertes Modell lernt dabei nicht die Kausalstruktur der Welt, sondern die syntaktischen Formen der Argumentation. Ein syllogistischer Schluss – Prämisse A, Prämisse B, Konklusion C – ist für das Modell ein statistisches Sequenzmuster. Das Modell verfügt über kein Wissen davon, dass Prämisse A wahr ist; es erkennt lediglich, dass in den Trainingsdaten auf das Setzen der Prämissen A und B mit sehr hoher Wahrscheinlichkeit jener Tokenfolge folgt, die Konklusion C repräsentiert. Da das Modell diese logischen Pfade in enormem Umfang durchlaufen hat, reproduziert es die Form logischer Argumentation mit einer Perfektion, die beim menschlichen Beobachtenden den Eindruck von Verständnis erzeugt.

Schließlich werden Sprachmodelle nicht allein auf die Vorhersage des nächsten Tokens trainiert, sondern in einem zweiten Schritt durch menschliche Annotatoren darauf optimiert, überzeugend, hilfreich und konsistent zu wirken. Dieses Verfahren wird als verstärkendes Lernen durch menschliches Feedback (Reinforcement Learning from Human Feedback, RLHF) bezeichnet. Es filtert rein chaotische statistische Vorhersagen heraus und konditioniert das Modell darauf, bevorzugt jene sprachlichen Pfade zu wählen, die von menschlichen Bewerterinnen und Bewertern als analytisch kompetent eingestuft werden. Die wahrgenommene Überzeugungskraft von Sprachmodellen ist mithin das Ergebnis einer gezielten Selektion von Sprachstilen, die Autorität und logische Stringenz signalisieren – ohne dass diese Merkmale auf tatsächlichem Verstehen oder begründetem Wissen beruhen.

2.2. Halluzination, Gefälligkeit und fehlende Rechenschaftspflicht

Die Limitationen generativer KI-Systeme entfalten ihre volle analytische Bedeutung erst dann, wenn sie nicht als technische Probleme verstanden werden, die einer künftigen Lösung harren, sondern als strukturelle Merkmale, die unmittelbar aus der Architektur dieser Systeme folgen. Vier Phänomene verdienen dabei besondere Aufmerksamkeit, da sie grundlegende Annahmen wissenschaftlicher Arbeit berühren.

Sprachmodelle erzeugen regelmäßig sachlich falsche Informationen, die jedoch mit derselben sprachlichen Sicherheit präsentiert werden wie zutreffende Aussagen. Sie erfinden Quellen, konstruieren Zitate und behaupten Fakten ohne empirische Grundlage. Dieses Phänomen – in der Fachliteratur als *Halluzination* bezeichnet – stellt keine Fehlfunktion dar, sondern ist die direkte und strukturell unvermeidliche Konsequenz probabilistischer Textgenerierung: Das System wählt statistisch wahrscheinliche Fortsetzungen, nicht faktisch korrekte. Da es keinen Zugang zur externen Realität besitzt, ist es prinzipiell außerstande, Wahrheit von bloßer Plausibilität zu unterscheiden.

Teil I: Grundlagen – KI als epistemische Disruption

Halluzination ist mithin kein behebbarer Defekt, sondern ein konstitutives Merkmal dieser Systemklasse, das bei jeder Nutzung in Rechnung gestellt werden muss. Jede Verwendung von Sprachmodellen, die faktische Verlässlichkeit voraussetzt, missversteht die Technologie in grundlegender Hinsicht.

Für wissenschaftliches Arbeiten erwächst hieraus ein besonderes Problem: Erfundene Quellenangaben fügen sich nahtlos in akademische Texte ein, da sie gängigen Zitierkonventionen folgen und ohne unabhängige Verifikation kaum als Fabrikationen erkennbar sind. Wer KI-generierten Inhalt ungeprüft übernimmt, gefährdet nicht nur die eigene akademische Integrität, sondern untergräbt zugleich die Grundlagen wissenschaftlicher Kommunikation.

Eine direkte Konsequenz der Trainingsoptimierung durch menschliches Feedback (Reinforcement Learning from Human Feedback, RLHF) ist das Phänomen, das in der Forschungsliteratur als *Sykophantie* bezeichnet wird: die systematische Tendenz von Sprachmodellen, Nutzenden zuzustimmen, Widerspruch zu vermeiden und Bestätigung anstelle kritischer Auseinandersetzung zu liefern. Perez et al. (2022) dokumentierten dieses Muster empirisch und zeigten, dass Sprachmodelle dazu neigen, die bevorzugte Perspektive ihrer Gesprächspartnerinnen und -partner zu wiederholen, anstatt zutreffendere Informationen bereitzustellen. Sharma et al. (2023) bestätigen diesen Befund: Leistungsstarke KI-Assistenten zeigen konsistent gefälliges Verhalten über unterschiedliche Aufgabentypen hinweg – ein Effekt, der dadurch strukturell begünstigt wird, dass menschliche Bewerterinnen und Bewerter im Rahmen des RLHF-Prozesses Antworten, die den eigenen Ansichten entsprechen, systematisch positiver bewerten.

Für akademische Zwecke stellt diese Tendenz eine erhebliche epistemische Einschränkung dar. Wissenschaftliche Auseinandersetzung setzt Widerspruch, Gegenargument und die Fähigkeit voraus, Positionen gegenüber Einwänden standzuhalten, sofern die Evidenz dies rechtfertigt. Ein System, das strukturell auf Bestätigung ausgelegt ist, kann diese Funktion nicht erfüllen und taugt daher nicht als ernsthafter intellektueller Gesprächspartner. Das viel diskutierte Versprechen hybrider Intelligenz – KI als kognitiver Kollaborateur, der menschliches Denken produktiv erweitert – wird durch eine Trainingsarchitektur konterkariert, die kritischen Widerspruch strukturell benachteiligt. Studierende, die Sprachmodelle als Denkpartner einsetzen, begegnen damit keinen sokratischen Gesprächspartnern, sondern Systemen, die darauf optimiert sind, Zustimmung zu erzeugen, nicht kritisches Denken anzuregen.

Als akzeptiertes Definitionsmerkmal von Wissen gilt die Verankerung in verifizierbaren Grundlagen: empirischen Daten, logischen Schlussfolgerungen und transparenten Methoden. Eine Behauptung muss auf die Gründe zurückführbar sein, die sie stützen. Sprachmodelle (LLMs) produzieren Aussagen jedoch ohne diese epistemische *Rechenschaftsstruktur*. Sie sind

Teil I: Grundlagen – KI als epistemische Disruption

nicht in der Lage zu erklären, warum sie bestimmte Behauptungen aufstellen, da ein inhaltliches „Warum“ schlicht nicht existiert – es liegen lediglich probabilistische Muster zugrunde. Was solche Systeme liefern, entbehrt daher dessen, was in der analytischen Erkenntnistheorie als epistemische Rechtfertigung bezeichnet wird (Chisholm, 1977): die begründende Verbindung zwischen Aussage und den sie tragenden Gründen, die Wissen von bloßer Behauptung unterscheidet.

Dieses Merkmal macht Sprachmodelle grundsätzlich ungeeignet als epistemische Partner in Tätigkeiten, die begründete Überzeugung erfordern. Sie können eine erste Orientierung bieten, Richtungen für weiterführende Untersuchungen aufzeigen oder Muster zur menschlichen Prüfung identifizieren. Sie sind jedoch außerstande, die kritische Auseinandersetzung zu ersetzen, die den Kern wissenschaftlicher Arbeit ausmacht – eben weil sie an der normativen Struktur von Gründen und Evidenz nicht teilhaben können.

Moderne Sprachmodelle weisen schließlich eine Komplexität auf, die es selbst ihren Entwicklerinnen und Entwicklern nicht erlaubt, vollständig nachzuvollziehen, wie sie zu spezifischen Ausgaben gelangen. Dieses Phänomen wird in der Forschung als *Interpretabilitätsproblem* bezeichnet. Wo die interne Verarbeitungslogik eines Systems nicht rekonstruierbar ist, fehlt auch jede systematische Grundlage dafür, einzuschätzen, wann dessen Schlussfolgerungen als verlässlich gelten können. Für die Hochschulbildung ist dieses Problem besonders folgenreich, da es genau jene kritische Reflexion unterläuft, die einen sachkundigen Umgang mit Quellen konstituiert. Wer KI-generierten Inhalt evaluieren möchte, kann die zugrunde liegende Begründung nicht rekonstruieren – schlicht weil keine existiert. An ihre Stelle treten statistische Muster, die sich menschlichem Nachvollzug entziehen.

Diese vier Eigenschaften – systematische Halluzination, strukturell bedingtes Zustimmungsverhalten, fehlende epistemische Rechenschaftspflicht und generative Intransparenz – begründen nicht lediglich Vorsicht im Umgang mit KI-Systemen. Sie belegen vielmehr, dass die Ausgaben von Sprachmodellen eine fundamental andere epistemologische Kategorie einnehmen als menschliche Wissensbehauptungen. Diese kategoriale Differenz ist es, die generative KI für Institutionen so grundlegend herausfordernd macht, deren Auftrag auf der Annahme beruht, dass textuelle Produkte verlässlich kognitive Prozesse abbilden.

2.3. KI-Nutzung in der akademischen Praxis

Die Einordnung generativer KI als „Werkzeug“ ist intuitiv naheliegend, aber in einer Weise irreführend, die die eigentliche Herausforderung für die Hochschulbildung verdeckt. Werkzeuge im klassischen Sinne – vom Hammer über den Taschenrechner bis zur Suchmaschine – erweitern menschliche Fähigkeiten, ohne sie zu ersetzen. Der Taschenrechner führt Berechnungen

Teil I: Grundlagen – KI als epistemische Disruption

schneller als Menschen, generiert jedoch kein mathematisches Verständnis. Die Suchmaschine ruft Information ab, synthetisiert aber kein Wissen. Diese Werkzeuge bleiben in einer instrumentellen Beziehung zur menschlichen Kognition: Sie setzen Kompetenz als Voraussetzung für sinnvolle Nutzung voraus und verstärken intellektuelle Anstrengung, anstatt sie zu ersetzen.

Generative KI bricht dieses Muster grundlegend auf. Sie vollzieht Tätigkeiten, die die operative Definition akademischer Kompetenz konstituiert haben: Texte verfassen, Argumente konstruieren, Konzepte erklären, Quellen synthetisieren. Dabei handelt es sich nicht um eine unterstützende, sondern um eine substituierende Funktion. Ein Studierender, der ein LLM beauftragt, eine Seminararbeit zu verfassen, delegiert nicht eine Teilaufgabe – er delegiert die gesamte kognitive Arbeit, die die Aufgabe hervorzubringen und zu bewerten angelegt war. Reinmann (2023) hat diese Problematik im deutschsprachigen Kontext nachdrücklich artikuliert und den Verlust systematischer Kompetenz als zentrale Herausforderung für die Hochschuldidaktik identifiziert.

Die Unterscheidung zwischen Erweiterung und Substitution ist daher keine Frage des Grades, sondern einer kategorialen Differenz. Erweiterung liegt vor, wenn ein Schreibassistent grammatische Fehler markiert, während die Verfasserin oder der Verfasser den Text selbst entwickelt – die kognitive Arbeit der Textproduktion verbleibt beim Menschen. Substitution liegt demgegenüber vor, wenn ein KI-System den Text generiert und der Mensch lediglich als Initiator und nachträglicher Korrektor fungiert. Diese Unterscheidung markiert die Grenze zwischen zulässiger Unterstützung und unzulässiger Substitution eigener Lernleistung – eine Grenze, deren klare Definition und konsequente Handhabung Institutionen im Interesse der Integrität ihres Bildungsauftrags obliegt.

Jüngere Forschung zur Mensch-KI-Interaktion führt eine dritte Kategorie ein: hybride Intelligenz (Rafner et al., 2022; Dellermann et al., 2019). Dieses Konzept beschreibt Zusammenarbeit, in der Menschen und KI komplementäre Stärken durch wechselseitiges Lernen einbringen und Ergebnisse erzielen, die keiner allein erreichen könnte. Anders als einfache Werkzeugnutzung setzt hybride Intelligenz genuine kognitive Beteiligung voraus – Menschen lernen durch die Interaktion mit KI, und ihr Denken wird durch die Kollaboration geformt. Diese Kategorie birgt eine genuine pädagogische Chance, muss aber sorgfältig von der Substitution unterschieden werden. Das entscheidende Kriterium ist, ob menschliches kognitives Engagement vertieft oder umgangen wird.

Die empirische Evidenz für die Lernkonsequenzen dieser Unterscheidung wird in Kapitel 4.3 eingehend untersucht. Hier genügt die Feststellung, dass die Grenze zwischen kognitiver Partnerschaft und kognitiver Substitution messbare Konsequenzen für die Gedächtnisleistung, die Kompetenzentwicklung und die langfristige akademische Leistung hat.

Die entscheidende Frage für die Hochschulbildung lautet daher nicht, ob KI-Nutzung zulässig sein sollte, sondern wie die Grenze zwischen legitimer Nutzung und unzulässiger Substitution eigener Denkleistung bestimmt, kommuniziert und durchgesetzt werden kann. Diese Frage führt unmittelbar zur Krise der Prüfungsvalidität, der Teil II der vorliegenden Untersuchung eingehend nachgeht.

2.4. Der epistemische Bruch

Die vorangegangenen Abschnitte haben die technischen und operationalen Eigenschaften von LLMs dargelegt. Dieser Abschnitt zieht daraus die Konsequenzen für die epistemologischen Grundlagen der Hochschulbildung. Das Argument lautet, dass generative KI keine inkrementelle Herausforderung darstellt, sondern eine didaktische Zäsur – eine Diskontinuität, die Annahmen in Frage stellt, auf denen die gegenwärtige akademische Bildungspraxis beruht.

Jahrhundertlang haben akademische Institutionen auf einer impliziten epistemologischen Annahme operiert: dass die Qualität eines schriftlichen Produkts als verlässliche Evidenz für die kognitiven Prozesse dient, die es hervorgebracht haben. Eine gut strukturierte Arbeit demonstriert analytisches Denken; ein differenziertes Argument offenbart konzeptuelles Verständnis; angemessener Umgang mit Quellen zeigt Recherchekompetenz. Diese Annahme ermöglichte Prüfung im großen Maßstab: Lehrende konnten Produkte bewerten und berechtigt auf den zugrundeliegenden Prozess schließen.

Generative KI untergräbt diesen Schluss. Ein Text, der jedes formale Kriterium akademischer Qualität erfüllt – kohärente Struktur, differenzierte Argumentation, angemessene Zitation –, kann nun ohne jeden korrespondierenden kognitiven Prozess auf Seiten der einreichenden Person produziert werden. Das Produkt ist epistemisch entleert: Es trägt keine verlässliche Information mehr über das Verständnis der Person, die es vorlegt. Dies ist keine Frage einzelner Studierender, die wirkungsvoller täuschen. Es ist die Auflösung der evidentiellen Beziehung zwischen Produkt und Prozess, auf der das gesamte System akademischer Prüfung gründet.

Prüfungsgültigkeit setzt voraus, dass gemessen wird, was gemessen werden soll. Wenn Hochschulen studentische Leistungen bewerten – sei es in Haus- und Abschlussarbeiten, in Übungsaufgaben, Referaten oder Portfolios –, beanspruchen sie, Verständnis, analytische Fähigkeit und Domänenkompetenz zu messen. Dieser Anspruch setzt jedoch eine Bedingung voraus, die institutionell kaum thematisiert wird: dass der dem Produkt vorausgehende Lernprozess tatsächlich stattgefunden hat. Kompetenz entsteht nicht im Prüfungsmoment, sondern in den asynchronen Lernphasen, in denen Studierende ohne direkte Aufsicht arbeiten – bei der Lektüre, der Bearbeitung

Teil I: Grundlagen – KI als epistemische Disruption

von Übungsaufgaben, der Vor- und Nachbereitung von Lehrveranstaltungen. Wenn generative KI diese Phasen substituiert, ist das schriftliche Produkt nicht mehr deshalb epistemisch entleert, weil im Prüfungsmoment getäuscht wurde, sondern weil der Kompetenzerwerb, den das Produkt anzeigen soll, zu keinem Zeitpunkt stattgefunden hat.

Wenn diese evidentielle Verbindung gekappt ist – wenn Produkte ohne die kognitiven Prozesse generiert werden können, die sie anzeigen sollen –, verliert die Prüfung ihre Gültigkeit im technischen Sinne. Kapitel 10 entwickelt dieses Argument ausführlich. Der hier zu etablierende Punkt ist: Was der Hochschulbildung gegenübersteht, ist nicht primär ein Problem der akademischen Integrität oder didaktischer Anpassung. Es ist eine Gültigkeitskrise – die traditionellen Mechanismen zur Zertifizierung von Kompetenz zertifizieren nicht mehr, was sie zu zertifizieren beanspruchen.

Die bisherige institutionelle Reaktion hat sich im Wesentlichen auf Detektion konzentriert: KI-generierten Inhalt zu identifizieren, um die Integrität traditioneller Prüfungsformate zu wahren. Diese Reaktion ist verständlich, greift jedoch zu kurz. Selbst eine vollständige Detektion – die technisch nicht realisierbar bleibt – würde die epistemologische Ausgangslage nicht wiederherstellen. Die entscheidende Frage ist nicht allein, ob Studierende KI eingesetzt haben, sondern was die Existenz dieser Technologie für die Schlussfolgerungsstruktur der Prüfung bedeutet. Wenn Produkte ohne Prozess generiert werden können, stellt die Detektion einzelner Fälle kein Vertrauen in die nicht detektierten Fälle wieder her. Die Indikatorbeziehung ist kategorial gebrochen – nicht nur in markierten Ausnahmefällen.

Wenn das asynchron erstellte Produkt nicht länger als verlässliche Evidenz für den kognitiven Prozess dienen kann – was tritt an seine Stelle? Diese Frage führt zu dem Konzept, das für das normative Argument dieses Buches zentral ist: epistemische Handlungsfähigkeit – die Fähigkeit und Disposition, Wissensbehauptungen selbst zu entwickeln, zu begründen und Verantwortung dafür zu übernehmen. In einer Situation, in der KI jede textuelle Ausgabe produzieren kann, ist das unterscheidende Merkmal gebildeter Personen nicht, was sie produzieren, sondern wie sie sich zu Wissen verhalten: ob sie ihre Behauptungen begründen können, Evidenz prüfen, die Grenzen ihres eigenen Verständnisses erkennen und intellektuelle Verantwortung für ihre Positionen übernehmen.

Kapitel 11 entwickelt epistemische Handlungsfähigkeit als zentrales Bildungsziel für das KI-Zeitalter. Die Grundlage für dieses Argument wird hier gelegt: Wenn traditionelle Indikatoren ihre Aussagekraft verlieren, wenn Produkte nicht länger verlässlich Prozesse anzeigen, verschiebt sich die Bildungsaufgabe vom Erzeugen bewertbarer Ausgaben zur Entwicklung der kognitiven Fähigkeiten und epistemischen Dispositionen, die genuines Verständnis ausmachen. Dieser Wandel ist nicht optional – er wird durch die technologische Realität erzwungen, die dieses Kapitel beschrieben hat.

2.5. Akademische Wissensproduktion

Die technologischen Eigenschaften von LLMs, wie in den vorangegangenen Abschnitten dargelegt, tragen Konsequenzen, die über unmittelbare Anwendungsfragen hinausreichen. Sie berühren die grundlegenden Prinzipien der Curriculum- und Prüfungsgestaltung innerhalb akademischer Institutionen und damit das Selbstverständnis akademischer Bildung insgesamt.

Die in den vorangegangenen Abschnitten entwickelte Analyse zwingt Hochschulen zu grundlegender Reflexion. Diese muss allerdings an einem früheren Punkt ansetzen, als es die Prüfungspraxis nahelegt. Akademisches Lernen vollzieht sich zu einem wesentlichen Teil in Phasen, die strukturell der Beobachtung entzogen sind: in der eigenständigen Lektüre, der Bearbeitung von Übungsaufgaben, der Vor- und Nachbereitung von Lehrveranstaltungen, der Auseinandersetzung mit Fachliteratur. Diese asynchronen Lernphasen sind nicht peripher, sondern konstitutiv: In ihnen entstehen die kognitiven Strukturen, die Kompetenz ausmachen. Generative KI ist in diesen Phasen maximal wirksam – gerade weil keine Aufsicht stattfindet und keine unmittelbare Konsequenz die Delegation sichtbar macht. Die Delegation geschieht nicht im Moment der Abgabe, sondern kontinuierlich, in den Stunden des scheinbar eigenständigen Arbeitens.

Daraus folgt eine Unterscheidung, die für die Hochschulbildung unausweichlich wird: jene zwischen Produktorientierung und Prozessorientierung. Wenn der Zweck akademischer Arbeit allein im Ergebnis bestünde wäre KI ein legitimes Mittel für dessen Produktion. Wenn der Zweck jedoch im Prozess liegt – im Denken, das durch Schreiben geschieht, im Verständnis, das sich durch die Auseinandersetzung mit Quellen entwickelt, in der Kompetenz, die sich durch anstrengendes Engagement konsolidiert –, dann muss dieser Prozess ins Zentrum zurückgestellt werden.

Diese Unterscheidung erklärt, warum scheinbar vorteilhafte KI-Nutzung aus bildungswissenschaftlicher Perspektive problematisch sein kann. KI kann die Qualität von Ausgaben verbessern – Texte werden sprachlich polierter, Argumente strukturierter, Präsentationen professioneller. Gleichzeitig wird die Wirksamkeit des Lernprozesses nicht gesteigert, sondern möglicherweise beeinträchtigt. Kompetenzerwerb findet nicht statt, wenn die kognitive Anstrengung, die ihn erst ermöglicht, an KI delegiert wird. Die Studierenden erhalten ein verbessertes Produkt und gleichzeitig eine verminderte Bildungserfahrung.

Hochschulbildung steht daher vor einer dualen Herausforderung mit inhärenter Spannung. Sie muss Formate entwickeln, die den kompetenten Umgang mit KI vermitteln – denn diese Kompetenz wird beruflich zunehmend unverzichtbar. Sie muss gleichzeitig sicherstellen, dass grundlegende kognitive und epistemische Fähigkeiten – kritisches Denken, analytisches Begründen, Domänenexpertise, epistemische Verantwortung – nicht durch KI-

Teil I: Grundlagen – KI als epistemische Disruption

ermöglichte Vermeidung kognitiver Anstrengung verkümmern. Teil IV dieses Buches erörtert, wie mit dieser Spannung produktiv umgegangen werden kann.

Diese duale Herausforderung lässt sich in drei Kompetenzdomänen konkretisieren, die an Bedeutung gewinnen, gerade weil KI sie nicht substituieren kann.

Kritisches Denken bezeichnet die Fähigkeit, Information zu bewerten, Argumente zu prüfen, Annahmen zu identifizieren und begründete Urteile zu bilden. KI-Systeme können Information aggregieren, aber nicht beurteilen, welche Information in einem spezifischen Kontext relevant, verlässlich oder handlungsleitend ist. Diese evaluative Fähigkeit bleibt eine genuin menschliche Leistung und gewinnt an Bedeutung, nicht an Bedeutungslosigkeit, je reichlicher Information verfügbar und je unsicherer ihre Herkunft wird.

Domänenspezifische Expertise bezeichnet akkumuliertes Fachwissen in einer Disziplin, das tiefes Kontextverständnis, die Kenntnis theoretischer Debatten, methodischer Standards und disziplinärer Normen einschließt. KI-Systeme können disziplinäre Texte produzieren, die im Bereich des Trainingsdatenstands kompetent wirken. Sie können jedoch keine neuen theoretischen Einsichten entwickeln, methodische Angemessenheit in neuartigen Kontexten beurteilen oder die impliziten Qualitätsnormen anwenden, die Expertinnen und Experten durch jahrelange Fachsozialisation erworben haben. Tiefe Domänenkenntnis bleibt die Grundlage, auf der KI-Outputs überhaupt kritisch bewertet werden können.

Epistemische Verantwortung schließlich bezeichnet die Disposition, für eigene Wissensbehauptungen Rechenschaft zu übernehmen – begründen zu können, warum eine Position vertreten wird, welche Evidenz sie stützt und wo die Grenzen des eigenen Wissens liegen. Diese Disposition ist nicht durch Delegation erlernbar: Sie setzt voraus, dass Lernende tatsächlich durch den Prozess des Begründens, Abwägens und Revidierens gegangen sind. KI kann diesen Prozess ersetzen, aber nicht stellvertretend vollziehen.

Diese drei Kompetenzdomänen werden in den nachfolgenden Kapiteln als analytischer Referenzrahmen verwendet. Sie markieren, was Hochschulbildung auch unter KI-Bedingungen leisten muss – und was durch technologische Erweiterung nicht ersetzt werden kann.