

Qualität in der Wissenschaft

Wessen Gedanke ist das?

**Augmentierte Reflexivität und post-individuelle Autorschaft
durch Zusammenarbeit mit generativer KI**

Elena Wilhelm

Zeitschrift: Qualität in der Wissenschaft (QiW)

Jahrgang: 20 (2+3)

Seiten: 45-51

Verlag: UniversitätsVerlagWebler

Ort: Bielefeld

2+3 | 2026

Impressum / Verlagsanschrift

UVW UniversitätsVerlagWebler GmbH
Reepeweg 5, 33617 Bielefeld



Für weitere Informationen besuchen Sie unsere Website:
universitaetsverlagwebler.de

Oder wenden Sie sich direkt an uns:
E-Mail: info@universitaetsverlagwebler.de
Telefon: 0521/ 923 610-0

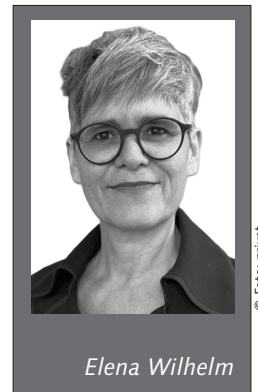
UniversitätsVerlagWebler
Der Fachverlag für Hochschultemen



Elena Wilhelm¹

Wessen Gedanke ist das?

Augmentierte Reflexivität und post-individuelle Autorschaft² durch Zusammenarbeit mit generativer KI



Elena Wilhelm

Discussions of artificial intelligence in higher education have so far revolved around tools, competencies, and regulation. This paper argues that such a perspective overlooks a deeper shift: generative AI changes not only what higher education institutions do, but how thinking itself takes place within them. Drawing on a governance development process at a Swiss university of the arts, the paper reconstructs how reflection, concept formation, and argumentation change through iterative collaboration between humans and generative AI. From this reconstruction, two concepts are developed: augmented reflexivity and post-individual authorship. Both describe a new quality of epistemic practice without negating individual responsibility for judgments and decisions. The paper itself forms part of this practice: its genesis is made transparent in the postscript and serves as an exemplary instance of the concepts developed.

1. Einleitung: Die eigentliche Verschiebung bleibt unsichtbar

Die Diskussion über generative künstliche Intelligenz (generative KI) in Hochschulen bewegt sich derzeit in vertrauten Bahnen: Es geht vorwiegend um Tools, Kompetenzen und ethische Leitlinien (vgl. Kasneci et al. 2023; Zawacki-Richter et al. 2019). Diese Perspektiven sind notwendig. Aber sie verfehlen den Kern der Transformation. Denn sie behandeln generative KI als Werkzeug. Was dabei aus dem Blick gerät, ist grundlegender: Generative KI verändert nicht nur, wie wir arbeiten, sondern wie wir denken. Und im produktivsten Fall entsteht dabei etwas, das weder Mensch noch Maschine allein hätte erzeugen können.

Der Beitrag argumentiert aus einer organisations- und wissenssoziologischen Perspektive. Sein Erkenntnisinteresse gilt der Frage, wie sich epistemische Praxis verändert, wenn Reflexion, Strukturierung und Begriffsbildung in der Interaktion zwischen Mensch und generativen KI-Systemen entstehen. Pädagogische, psychologische und informatische Forschung untersucht derzeit empirisch, wie generative KI Lernen und Kompetenzentwicklung beeinflusst – etwa durch kognitive Entlastung, aber auch durch Komplexitätsreduktion und mögliche Kompetenzverluste. Diese Befunde sind für Hochschulen bedeutsam; sie liegen jedoch ausserhalb des hier verfolgten Erkenntnisinteresses.

Ich schreibe diesen Beitrag aus einer transparent gemachten Perspektive: als Hochschulentwicklerin an einer Kunsthochschule, die generative KI nicht von aussen betrachtet, sondern täglich in strategischen, konzeptionellen und analytischen Prozessen einsetzt. Was sich dabei zeigt, lässt sich nicht auf Effizienzgewinne reduzieren. Generative KI

ist in dieser Praxis nicht nur ein epistemisches Werkzeug, sondern eine epistemische Mitbedingung geworden – im Sinne einer verteilten kognitiven Umgebung, in der Denken und Werkzeug untrennbar verflochten sind (Clark & Chalmers 1998; Clark 2008). Sie verändert, wie Probleme strukturiert, Argumente entwickelt, Perspektiven verschoben und Entscheidungen vorbereitet werden.

Der vorliegende Beitrag versteht diese Beobachtung nicht als Einzelfall, sondern als Hinweis auf eine grundlegendere Verschiebung. Er geht von der These aus, dass in der Interaktion zwischen Mensch und generativer KI eine neue Form der Denkpraxis entsteht. Eine, in der Reflexion, Strukturierung und Generierung nicht mehr eindeutig einer Instanz zugeschrieben werden können. Die Transformation liegt somit nicht primär auf der Ebene der Effizienz, sondern auf der Ebene der Epistemologie.

Der Beitrag entfaltet sich in vier Schritten: Er entwickelt zunächst einen konzeptionellen Rahmen, der generative KI nicht als blosses Werkzeug, sondern als epistemische Mitbedingung fasst. Er rekonstruiert dann exemplarisch einen konkreten Entwicklungsprozess aus der Hochschulpraxis – die Erarbeitung eines Governance-Rahmenwerks an der Zürcher Hochschule der Künste (ZHdK). Darauf aufbauend werden zwei Begriffe vorgeschlagen, die diese Form der Zu-

¹ Der Beitrag ist in Zusammenarbeit mit Claude (Anthropic) und ChatGPT (OpenAI) entstanden.

² Im vorliegenden Beitrag wird durchgängig die Schreibweise mit Doppelpunkt für geschlechtergerechte Sprache verwendet (z.B. Autor:innen, Praktiker:in). Eine Ausnahme bildet der Begriff «Autorschaft», der angesichts seiner Häufigkeit im Text aus Gründen der Lesbarkeit in der ungeänderten Form belassen wird.

sammenarbeit fassen: augmentierte Reflexivität und post-individuelle Autorschaft. Abschliessend wird diskutiert, welche Konsequenzen sich daraus für die Hochschulentwicklung ergeben. Die Entstehung dieses Beitrags ist dabei nicht nur Kontext, sondern Teil seines Arguments.

2. Die falsche Rahmung: KI als Tool

Die aktuelle Debatte über generative KI in Hochschulen folgt mehrheitlich einem impliziten Grundmodell: Generative KI wird als Instrument verstanden, das in bestehende Prozesse integriert werden muss. Entsprechend konzentrieren sich Forschung und Praxis auf Fragen der Anwendung, der Kompetenzentwicklung und der Regulierung (vgl. Zawacki-Richter et al. 2019; Kasneci et al. 2023). Diese Rahmung ist nicht falsch, aber unzureichend. Denn sie setzt voraus, dass die grundlegenden Formen wissenschaftlichen Denkens stabil bleiben. Genau diese Annahme ist jedoch fraglich.

Bereits frühe Arbeiten zur erweiterten Kognition (Clark & Chalmers 1998; Clark 2008) sowie zur verteilten und soziomateriellen Wissensproduktion (Hutchins 1995; Latour 2005; Knorr Cetina 1999) zeigen, dass kognitive Prozesse nicht auf das Individuum beschränkt sind. Sie entfalten sich in Interaktionen mit Artefakten, Praktiken und sozialen Strukturen. Aktuelle Arbeiten zu generativer KI deuten in eine ähnliche Richtung: Generative KI wirkt nicht nur als Werkzeug, sondern als aktiver Bestandteil kognitiver Prozesse (Bender et al. 2021; Floridi & Chiriatti 2020; Mollick 2024).

Die Folge ist eine strukturelle Blindstelle: Während intensiv darüber diskutiert wird, wie generative KI genutzt werden soll, bleibt weitgehend unbeachtet, wie sich die Formen des Denkens selbst verändern – und welche neuen Formen der Wissensproduktion dabei entstehen (vgl. Valloir 2024; Williamson 2024).

Diese Perspektive blendet die komplexen und auch problematischen Herstellungsbedingungen generativer KI nicht aus. Grosse Sprachmodelle erzeugen ihre Ausgaben auf der Grundlage statistischer Wahrscheinlichkeiten über Trainingsdaten; sie können fehlerhafte oder erfundene Inhalte produzieren, sie nutzen Wissensbestände realer Autor:innen ohne deren Anerkennung, Bewilligung und Vergütung und werfen Fragen der Validität, der Transparenz und des Datenschutzes auf (Bender et al. 2021). Diese Fragen müssen gelöst werden, sind aber nicht Gegenstand des vorliegenden Beitrags, dessen Fokus auf der Veränderung der epistemischen Praxis selbst liegt.

3. Rekonstruktion einer Praxis: Das Governance-Projekt an der ZHdK

3.1 Ausgangslage: Handlungsdruck, nicht Spielraum

Der Ausgangspunkt war kein offener Denkraum, sondern ein hoch strukturierter Handlungsdruck. Es bestand ein klarer institutioneller Auftrag: die Governance der ZHdK als kohärente Entscheidungsarchitektur neu zu fassen. Gleichzeitig bestanden reale Probleme: unklare Entscheidungswege, Spannungen zwischen departementaler Auto-

nomie und institutioneller Gesamtverantwortung sowie ineffiziente oder überfrachtete Gremienstrukturen.

Der Prozess war somit nicht explorativ im luftleeren Raum, sondern gerichtet, verantwortungsgebunden und entscheidungsrelevant. Diese Ausgangslage ist entscheidend für das Verständnis der Zusammenarbeit mit generativer KI: Sie fand nicht unter Bedingungen unverbindlicher Exploration statt, sondern unter Bedingungen institutioneller Verantwortung. Parallel zur KI-gestützten Konzeptarbeit verlief ein kontinuierlicher Prozess mündlicher und schriftlicher Rückkopplung mit der Hochschulleitung und der Hochschulversammlung – Validierungsrunden, die den konzeptionellen Rahmen immer wieder mit institutioneller Realität konfrontierten.

Der Einsatz generativer KI erfolgte dabei innerhalb klarer Leitplanken: In die Systeme wurden weder Personendaten noch vertrauliche interne Dokumente eingegeben; Gegenstand der Interaktion waren konzeptionelle Entwürfe, Strukturvorschläge und generische Problembeschreibungen.

3.2 Funktionen und Phasen der Zusammenarbeit: Von der Nutzung zur epistemischen Koproduktion

Die Zusammenarbeit mit generativer KI begann nicht am Anfang des Denkprozesses. Zentrale Problemstellungen, erste konzeptionelle Linien und die eigene Rolle waren bereits geklärt. Generative KI wurde zunächst als Strukturierungsinstanz und Resonanzraum hinzugezogen.

In einer ersten Phase war die Arbeitsteilung klar: Inhalt und Richtung lagen bei der Autorin, Formulierung und Verdichtung bei der KI. In einer zweiten Phase verschob sich diese Logik: Konzeptionelle Setzungen – etwa Governance-Prinzipien oder eine Gremien-Typologie – wurden durch die generative KI expliziert, systematisiert und in übergreifende Strukturen überführt.

In einer dritten Phase wurde die generative KI gezielt als Gegenüber eingesetzt: als Differenzverstärker, als Generator alternativer Ordnungen, als Mittel zur Prüfung der Tragfähigkeit von Argumenten.

Quer zu dieser zeitlichen Abfolge liegt eine funktionale Ordnung, die den Charakter der Koproduktion genauer trifft als jede Phasenlogik. Generative KI übernahm im Prozess sieben unterscheidbare epistemische Funktionen:

- Strukturierung (Ordnen von Material und Argumentationslinien),
- Resonanz (Zurückspiegeln von Entwürfen zur Prüfung ihrer Kohärenz),
- Verdichtung (sprachliche Präzisierung bereits gesetzter Inhalte),
- Explikation (Überführung impliziten Wissens in systematische Form),
- Differenzverstärkung (Zuspitzung von Unterscheidungen und Sichtbarmachung von Spannungen),
- Alternativengenerierung (Erzeugung konkurrierender Ordnungen) sowie
- Tragfähigkeitsprüfung (Testen von Argumenten gegen Einwände).

Die Phasen unterscheiden sich nicht durch die Verfügbarkeit dieser Funktionen, sondern durch deren Gewichtung: Dominierten zunächst Strukturierung, Resonanz und Verdichtung, verlagerte sich das Gewicht zunehmend auf Explikation, Differenzverstärkung, Alternativengenerierung und Prüfung.

Der Übergang von der Nutzung zur epistemischen Koproduktion lässt sich damit präziser bestimmen: Er vollzieht sich in dem Maß, in dem die reproduktiven Funktionen durch die produktiven abgelöst werden.

Ein zweites generatives KI-System wurde in kritischen Momenten hinzugezogen – nicht als Ersatz, sondern als zusätzliche Perspektive. Es fungierte als externe Reflexionsinstanz auf einer höheren Abstraktionsebene und trug dazu bei, konzeptionelle Knoten zu lösen. Diese funktionale Differenzierung zwischen Systemen war nicht geplant, sondern entstand aus der Praxis.

Ein Beispiel mag dies illustrieren. Als sich die Arbeit an der Frage festgefahren hatte, wie departementale Autonomie und gesamtinstitutionelle Steuerung begrifflich ins Verhältnis zu setzen seien, wurde das zweite System mit dem erreichten Denkstand konfrontiert – verbunden mit der Aufforderung, die Architektur des Arguments zu kritisieren, nicht seine Formulierung. Die Rückmeldung verschob die Ebene der Fragestellung: Nicht die Balance zwischen Autonomie und Steuerung sei zu bestimmen, sondern die Logik, nach der beides begründet wird. Aus dieser Verschiebung entstand die Unterscheidung von Steuerungs- und Ermöglichungslogik, die später zu einem tragenden Element des Governance-Rahmens wurde.

3.3 Momente der Emergenz: Explikation impliziten Wissens

Charakteristisch für den Prozess waren Momente, in denen Formulierungen oder Strukturen entstanden, die unmittelbar als adäquat erkannt wurden, obwohl sie zuvor nicht explizit formuliert worden waren. Ein Beispiel hierfür ist die systematische Ausarbeitung einer Gremientypologie. Das zugrunde liegende Verständnis war bereits vorhanden, wurde jedoch erst im Wechselspiel mit der generativen KI präzise gefasst. Diese Momente sind analytisch zentral. Sie zeigen, dass hier kein einfaches „Generieren“ stattfindet, sondern eine Explikation impliziten Wissens durch Interaktion. Ein Prozess, den Polanyi (1966) mit dem Satz „we can know more than we can tell“ grundlegend beschrieben hat. Das Ergebnis ist weder rein individuell noch rein maschinell erzeugt, sondern ein Produkt des Prozesses.

Wie voraussetzungsvoll diese Explikation ist, zeigt die Entstehung der Typologie der Gremien. Die Unterscheidung verschiedener Gremientypen war im Projekt lange implizit mitgeführt worden – in Arbeitsnotizen, in der Kritik an überfrachteten Strukturen, in der wiederkehrenden Frage, welches Hochschulgremium eigentlich wofür legitimiert sei. Erst die Aufforderung an die generative KI, aus den vorliegenden Arbeitstexten eine systematische Ordnung vorzuschlagen, führte in verschiedenen Iterationen zu einer Typologie, die Gremien nach ihrer Funktion unter-

schied und unmittelbar als adäquat erkannt wurde. Sie war weder von der Autorin diktiert noch von der generativen KI erfunden: Sie explizierte ein Verständnis, das im Material angelegt war.

3.4 Grenzen der Koproduktion: Urteil und Verantwortung

So ausgeprägt die Koproduktion war, so klar blieb ihre Grenze. Problemdefinition, normative Setzungen und kontextuelle Entscheidungen lagen durchgehend bei der Autorin. Was ist das Problem? Welche Balance zwischen Autonomie und Steuerung ist angemessen? Was ist institutionell tragfähig? Dies sind Entscheidungen, die nicht delegiert werden können. Sie erfordern Kontextwissen, Verantwortung und Legitimation.

Hinzu kommt eine Dimension, die im Prozess ebenso zentral war wie die KI-Arbeit: die wiederholten mündlichen und schriftlichen Reflexions- und Validierungsrunden mit der Hochschulleitung und der Hochschulversammlung. Diese Schleifen waren keine nachgelagerte Absegnung, sondern ein eigenständiger epistemischer Prozess. Einer, der nicht nur prüft, sondern Wissen erzeugt: durch die Konfrontation mit institutioneller Erfahrung, organisationaler Praxis und kollektiver Urteilsbildung. KI konnte diesen Prozess vorbereiten und nachbereiten. Ersetzen konnte sie ihn nicht. Hier liegt die entscheidende Differenz: Die generative KI erweitert den Denkraum, aber sie entscheidet nicht. Und sie legitimiert nicht.

4. Augmentierte Reflexivität

Der Begriff der augmentierten Reflexivität beschreibt eine Form des Denkens, in der Reflexion durch die Interaktion mit generativen KI-Systemen systematisch erweitert wird. Reflexivität gilt als zentrales Merkmal wissenschaftlicher Praxis (Schön 1983; Giddens 1991). In der Zusammenarbeit mit generativer KI verändert sich jedoch ihre Qualität grundlegend. Drei Dimensionen sind zentral: Explizierung (implizite Annahmen werden sichtbar gemacht); Beschleunigung (Denkzyklen werden verdichtet) sowie Erweiterung (Alternativen werden systematisch generiert).

Reflexion wird damit zu einem dialogischen, iterativen Prozess. Sie ist nicht mehr ausschliesslich intern, sondern entsteht im Wechselspiel mit einem System, das keine eigene Intentionalität besitzt (Floridi & Chiriatti 2020; Vallor 2024), aber dennoch strukturierend wirkt. Augmentierte Reflexivität bezeichnet somit eine neue Form epistemischer Praxis: schneller, expliziter, relationaler.

Der Begriff wird in diesem Beitrag neu vorgeschlagen; ein etablierter Terminus gleichen Namens findet sich, soweit ersichtlich, weder in der deutsch- noch in der englischsprachigen Literatur. Seine beiden Bestandteile schliessen gleichwohl an je eigene Traditionen an. Das Attribut „augmentiert“ nimmt das Forschungsprogramm der Intelligenzverstärkung auf, das Licklider (1960) mit der Idee einer Mensch-Maschine-Symbiose und Engelbart (1962) mit dem Konzept des „augmenting human intellect“ begründet haben: Rechner sollen menschliches Denken nicht

ersetzen, sondern erweitern. Der Begriff der Reflexivität entstammt der im Folgenden aufgenommenen sozialwissenschaftlichen Tradition. Neu ist die Verbindung beider Linien und ihre Pointe: Augmentiert wird hier nicht die Rechen- oder Gedächtnisleistung, sondern der reflexive Akt.

Damit ist jedoch erst die operative Oberfläche beschrieben. Analytisch weiterführend ist die Frage, worin sich augmentierte Reflexivität von den etablierten Reflexivitätsbegriffen der Sozialwissenschaften unterscheidet. Schön (1983) fasst Reflexion als nachträgliches oder handlungsbegleitendes Nachdenken einer Praktiker:in über ihr eigenes Tun; Giddens (1991) und Beck (1996) beschreiben eine institutionelle Reflexivität, in der modernes Wissen fortlaufend auf sich selbst zurückwirkt. Beide Modelle setzen ein reflektierendes Subjekt voraus, dem ein Gegenstand – die eigene Praxis, das eigene Wissen – gegenübersteht. Augmentierte Reflexivität verschiebt diese Konstellation: Das reflektierende Gegenüber ist nicht mehr allein die Praktiker:in, die auf sich zurückblickt, sondern eine dialogische Anordnung, in der ein zweites, nicht-intentionales System Annahmen zurückspiegelt, variiert und in ungewohnte Ordnungen überführt. Reflexivität wird dadurch von einer inneren zu einer verteilten Operation.

Diese Verschiebung ist mehr als eine Beschleunigung. Sie verändert die Struktur des reflexiven Aktes selbst. In der klassischen Reflexion ist das Subjekt zugleich Quelle und Prüfinstanz seiner Gedanken; blinde Flecken bleiben schwer zugänglich, weil das Denken sich nicht vollständig überblicken kann. Augmentierte Reflexivität externalisiert einen Teil dieser Prüfung: Implizite Annahmen werden nicht durch Introspektion, sondern durch ihre Konfrontation mit einer maschinell erzeugten Alternative sichtbar. Der epistemische Gewinn liegt nicht darin, dass die generative KI „besser weiss“, sondern darin, dass sie das Denken zwingt, sich zu explizieren, um überhaupt anschlussfähig zu bleiben. Reflexivität wird performativ: Sie realisiert sich im Zwang zur Artikulation.

Gerade darin liegt jedoch eine Ambivalenz, die für die Hochschulentwicklung von zentraler Bedeutung ist. Augmentierte Reflexivität ist nicht per se erkenntnisförderlich. Generative Systeme neigen dazu, die Prämissen ihrer Nutzer:innen zu bestätigen, sprachlich zu glätten und auf statistisch erwartbare Formulierungen zu konvergieren (Bender et al. 2021). Wo dies ungeprüft bleibt, kehrt sich die Augmentierung in ihr Gegenteil: Statt Annahmen sichtbar zu machen, verstärkt sie diese; statt Alternativen zu eröffnen, homogenisiert sie das Denken auf ein plausibles Mittelmaß. Augmentierte Reflexivität kann mithin sowohl erweiternd als auch verengend wirken. Die entscheidende Differenz liegt nicht im Werkzeug, sondern in der Art seiner Nutzung.

Aus dieser Ambivalenz lässt sich ein produktives Gestaltungsprinzip ableiten, das über die bloße Diagnose hinausweist: das Prinzip der reflexiven Reibung. Erkenntnisförderlich ist die Interaktion mit generativer KI nicht dort, wo sie Zustimmung erzeugt, sondern dort, wo sie kontrollierten Widerstand leistet – wo das System nicht gebeten wird, einen Gedanken auszuformulieren, sondern ihn zu

bestreiten, seine Architektur zu kritisieren oder ihn in eine fremde Ordnung zu übersetzen. Das in Abschnitt 3.2 beschriebene Vorgehen, ein zweites System gezielt zur Kritik der Argumentarchitektur einzusetzen, ist ein Beispiel für solche produktiv gemachte Reibung. Augmentierte Reflexivität ist demnach kein Zustand, der sich mit dem Zugang zu leistungsfähigen Systemen von selbst einstellt, sondern eine Praxis, die bewusst gegen die Bestätigungsneigung der Systeme organisiert werden muss.

Diese Einsicht hat unmittelbare Konsequenzen für die Hochschulentwicklung, und hier liegt der eigentliche zukunftsweisende Ertrag des Begriffs. Wenn augmentierte Reflexivität eine erlernbare und gestaltbare Praxis ist, dann wird sie zu einer institutionellen Fähigkeit, nicht nur zu einer individuellen Kompetenz. Hochschulen stehen damit vor einer neuen Formatierungsaufgabe: Sie müssen nicht primär regeln, ob generative KI genutzt werden darf, sondern Bedingungen schaffen, unter denen ihre Nutzung reflexionsförderlich statt reflexionersetzend wirkt. Das betrifft die Gestaltung von Prüfungs- und Betreuungsformaten ebenso wie die Frage, welche Formen des Denkens in Forschung, Lehre und Selbstverwaltung institutionell honoriert werden. Eine Hochschule, die augmentierte Reflexivität ernst nimmt, wird Reibung, Widerspruch und die Explikation von Annahmen nicht als Störung, sondern als Qualitätsmerkmal epistemischer Prozesse behandeln – und sie entsprechend in ihren Verfahren verankern.

So verstanden ist augmentierte Reflexivität kein blosses Effizienzversprechen, sondern eine normative Kategorie: Sie benennt eine Qualität des Denkens, die unter Bedingungen generativer KI weder garantiert noch ausgeschlossen ist, sondern hergestellt werden muss. Der Begriff verweist damit über die vorliegende Rekonstruktion hinaus auf ein Forschungs- und Gestaltungsprogramm, das die Bedingungen ihrer Herstellung – didaktisch, organisational und kulturell – zum Gegenstand macht.

5. Post-individuelle Autorschaft

Augmentierte Reflexivität beschreibt, wie unter Bedingungen generativer KI gedacht wird. Damit stellt sich unmittelbar eine Folgefrage: Wem ist das so Gedachte zuzurechnen? Wenn Reflexion zu einer verteilten Operation wird, verliert auch die Zuschreibung ihrer Ergebnisse ihre Selbstverständlichkeit. Der Gedanke, der in einer dialogischen Anordnung entsteht, hat keinen eindeutigen Ursprungsort mehr. Diese Folgefrage führt vom ersten zum zweiten Begriff dieses Beitrags.

Im Folgenden wird post-individuelle Autorschaft als spezifische epistemologische Kategorie für die Genese von Gedanken in Mensch-KI-Konstellationen ausgearbeitet. Sie beschreibt eine Verschiebung in der Zuschreibung von geistiger Leistung, die in bestehenden Theorien zwar angelegt, aber für die Mensch-KI-Interaktion noch nicht systematisch gefasst ist.

Der Begriff wird dabei ausdrücklich als epistemologische Kategorie vorgeschlagen. Er beansprucht weder eine urheberrechtliche Neubestimmung von Autorschaft noch eine

bibliometrische Operationalisierung – etwa im Sinne prozentualer Anteile menschlicher und maschineller Beiträge. Im Begutachtungsverfahren dieses Beitrags wurde genau dieser Einwand erhoben: Post-individuelle Autorschaft sei nicht operationalisierbar – weder rechtlich noch für die Identifikation von Plagiaten, die Feststellung wissenschaftlichen Fehlverhaltens oder die Zurechnung von Forschungsleistung, etwa nach dem Muster, vierzig Prozent eines Textes stammten von der generativen KI und sechzig Prozent von der Autorin. Der Einwand ist ernst zu nehmen, benennt aber exakt das, worum es dem Begriff nicht geht. Eine solche Quantifizierung würde gerade verfehlen, was der Begriff beschreibt: einen Prozess, dessen Beiträge sich nicht additiv zerlegen lassen. Für Fragen wissenschaftlicher Integrität, der Zurechnung von Forschungsleistungen oder der Plagiatsprüfung bleiben deshalb die etablierten Kategorien individueller Verantwortung maßgeblich. Wie die geforderte Transparenz hergestellt werden kann, zeigt exemplarisch der Nachtrag dieses Beitrags. Dass der Einwand im Peer-Review-Verfahren formuliert wurde, hat dabei über seinen Inhalt hinaus Bedeutung: Die Begutachtung bildet eine weitere reflexive Schleife, in der die dialogisch entwickelten Konzepte an der Urteilskraft der Fachgemeinschaft geprüft und geschärft werden. Der Beitrag durchläuft damit genau jene Form kollegialer Validierung, deren Unersetzbarkeit er behauptet.

Klassische Modelle gehen von individueller Autorschaft aus, auch in kollaborativen Kontexten. Post-individuelle Autorschaft betrifft damit die Genese eines Gedankens, nicht die Zurechenbarkeit von Verantwortung: Wer für einen Text entsteht, bleibt eindeutig bestimmbar, auch wenn nicht mehr eindeutig bestimmbar ist, wer ihn im engeren Sinne gedacht hat. Verschieben wird nicht die Zurechenbarkeit, sondern deren Bezugspunkt – von der alleinigen Urheberschaft eines Ergebnisses hin zur bewussten Steuerung eines Prozesses.

In der Interaktion mit generativer KI wird diese Zuschreibung unscharf. Begriffe, Argumentationslinien und Strukturen entstehen im Wechselspiel. Sie sind weder vollständig intendiert noch vollständig generiert.

Die Frage lautet damit nicht mehr ausschliesslich: Wer verantwortet diesen Text? Sie lautet zunächst: Wessen Gedanke ist das? Post-individuelle Autorschaft ist in diesem Sinne keine Theorie des Textes, sondern eine Theorie der Entstehung wissenschaftlicher Gedanken unter Bedingungen generativer KI.

Diese Beobachtung berührt eine Debatte, die in den Geistes- und Kulturwissenschaften seit Jahrzehnten geführt wird. Barthes (1968) hat mit dem „Tod des Autors“ die Vorstellung verabschiedet, ein Text sei Ausdruck eines dahinterstehenden Autor-Subjekts, und die Einheit des Textes in seine Lektüre verlegt. Foucault hat in „Was ist ein Autor?“ (1969) gezeigt, dass Autorschaft keine biographische Tatsache ist, sondern eine diskursive Funktion, ein Konstrukt, das Texten nachträglich Einheitlichkeit, Herkunft und Verantwortlichkeit verleiht. Diese Perspektive ist erhellend, trifft aber den hier beschriebenen Sachverhalt nur teilweise. Denn Foucault fragt nach der Funktion

des Autornamens im Diskurs. Der vorliegende Beitrag fragt nach dem Produktionsprozess selbst. Die Verschiebung liegt damit nicht auf der Ebene der Zuschreibung durch Andere, sondern auf der Ebene der Entstehung.

Von der Gegenwartsdiskussion her ist der Begriff zudem von der jüngst vorgeschlagenen Kategorie der „post-authorship“ (Abdul-Jabbar & Bhatt 2025) abzugrenzen. Sie bezeichnet einen Zustand, in dem das menschliche Subjekt nicht mehr primärer Ursprung von Denken und Ausdruck ist – nicht den Barthes'schen Tod, sondern die Ersetzung des Autors. Post-individuelle Autorschaft formuliert die entgegengesetzte Diagnose: nicht Ablösung, sondern Rekonfiguration. Der Mensch bleibt Träger von Urteil und Verantwortung, während sich der Ursprungsort des Gedankens verteilt. Als Terminus wird post-individuelle Autorschaft, soweit ersichtlich, in diesem Beitrag erstmals vorgeschlagen.

Diese Beobachtung knüpft an Theorien verteilter und relationaler Wissensproduktion an (Latour 2005; Suchman 2007; Barad 2007), erhält jedoch durch die Dichte und Geschwindigkeit der Interaktion mit KI eine neue Qualität. Während verteilte Kognition bislang vor allem in Mensch-zu-Mensch-Kontexten oder in der Interaktion mit stabilen Artefakten beschrieben wurde (Hutchins 1995), erzeugt die Iteration mit generativer KI eine qualitativ neue Form der Ko-Produktion.

Post-individuelle Autorschaft bedeutet nicht den Verlust individueller Autorschaft, sondern deren Transformation: von einer individuellen zu einer prozessualen und relationalen Kategorie. Damit verschiebt sich auch Verantwortung: weg von der Zuschreibung eines Ergebnisses hin zur bewussten Steuerung eines Prozesses.

6. Die Differenz generativer KI-Systeme

„Die KI“ existiert nicht als einheitliche Instanz (Vallor 2024; Mollick 2024). Unterschiedliche Systeme rahmen Probleme unterschiedlich, setzen unterschiedliche Schwerpunkte und eröffnen unterschiedliche Perspektiven. Im beschriebenen Prozess wurde diese Differenz produktiv genutzt: Ein System fungierte als iterativer Denkpartner, ein anderes als externe Reflexionsinstanz. Die Auswahl erfolgte situativ und erfahrungsbasiert. Für die Hochschulentwicklung ergibt sich daraus eine zentrale Kompetenz: nicht die Beherrschung eines Systems, sondern die Steuerung eines pluralen epistemischen Settings.

7. Diskussion: KI als epistemischer Bruch

Die Analyse legt nahe, dass generative KI nicht primär als Werkzeug verstanden werden sollte, sondern als Auslöser eines epistemischen Bruchs, einer Verschiebung, die nicht lediglich neue Inhalte produziert, sondern die Bedingungen des Denkens selbst verändert. Der Begriff knüpft an Bachelards Konzept der „rupture épistémologique“ (Bachelard [1938] 1993) an, meint hier aber keine Zäsur in der Wissenschaftsgeschichte, sondern eine Transformation der epistemischen Praxis auf der Ebene individueller und institutioneller Wissensarbeit.

Die produktivste Nutzung liegt nicht in der Automatisierung, sondern in der Transformation von Denkformen. Strategische Arbeit verschiebt sich: Reflexion beginnt früher, wird iterativer und entkoppelt sich teilweise von sozialen Dynamiken wie Hierarchie oder Konsensdruck (vgl. Mollick 2024).

Für Hochschulen stellt sich damit eine grundlegende Frage: Nicht mehr, wie generative KI reguliert wird, sondern welche Formen des Denkens institutionell hervorgebracht und welche verhindert werden (Vallor 2024; Williamson 2024).

Dagegen liesse sich einwenden, wissenschaftliche Reflexion habe ihre etablierten Orte – Publikationen, Tagungen, den kritischen Diskurs der Fachgemeinschaft – und die dialogische Reflexion mit generativer KI drohe, die mühevollen Arbeit des Lesens und der literaturgestützten Auseinandersetzung zu ersetzen. Die hier rekonstruierte Praxis legt das Gegenteil nahe: Augmentierte Reflexivität ergänzt diese Formen, sie ersetzt sie nicht. Sie bereitet den Eintritt in kollegiale, diskursive und publikationsförmige Validierung vor – ersetzen kann sie ihn nicht, weil ihr fehlt, was diese Formen auszeichnet: Erfahrung, Verantwortung und die Urteilskraft einer wissenschaftlichen Gemeinschaft. Der vorliegende Beitrag selbst folgt dieser Logik: Er stellt die dialogisch entwickelten Konzepte dem Fachdiskurs zur Prüfung.

Eine besondere Schärfe gewinnt diese Frage im Kontext von Kunsthochschulen. Hier ist Autorschaft keine abstrakte akademische Kategorie – sie ist Berufsidentität, Rechtsgrundlage, ästhetisches Qualitätskriterium und institutioneller Kernauftrag zugleich. Misselhorn hat in ihrer philosophischen Analyse argumentiert, dass Autorschaft für die Identität und den Status von Kunstwerken konstitutiv ist, und dass KI eben diese Autorschaft nicht übernehmen kann, weil sie weder intentional handelt noch über eine Persönlichkeit oder Weltsicht verfügt, die sie zum Ausdruck bringen könnte (Misselhorn 2023). Dieser Einwand trifft die Vorstellung autonomer KI-Produktion. Er erledigt jedoch nicht die Frage der Mensch-KI-Koproduktion. Und genau hier liegt die ungelöste Spannung.

Bajohr hat gezeigt, dass die Frage der Autorschaft in Mensch-KI-Konstellationen nicht binär zu verstehen ist, sondern als Spektrum von Distanzgraden zwischen Schreibenden und Text (Bajohr 2024). Primäre, sekundäre und tertiäre Autorschaft unterscheiden sich durch die Tiefe der maschinellen Vermittlung, ohne dass die menschliche Verantwortung dabei verschwindet. Dieses Modell ist auf künstlerische Prozesse übertragbar: Die Frage ist nicht, ob KI „Kunst macht“, sondern wie sich künstlerische Eigenständigkeit unter Bedingungen epistemischer Ko-Produktion neu konstituiert.

Für Kunsthochschulen stellt sich damit eine doppelte Herausforderung. Einerseits die institutionell-strategische, die dieser Beitrag beschreibt. Andererseits die künstlerisch-pädagogische: Wenn Autorschaft relational wird, verän-

dert sich, was Ausbildung leisten soll. Nicht ob KI eingesetzt werden darf, sondern was künstlerische Eigenständigkeit im Zeitalter der Ko-Produktion bedeutet – und wie sie gelehrt, bewertet und entwickelt werden kann. Diese Frage ist offen. Sie muss gestellt werden.

8. Fazit

Die gegenwärtige Debatte über generative KI unterschätzt die Tiefe der Veränderung. Sie fokussiert auf Werkzeuge, während sich die eigentliche Transformation auf der Ebene des Denkens vollzieht. Mit den Konzepten der augmentierten Reflexivität und der post-individuellen Autorschaft wurde ein Vorschlag gemacht, diese Entwicklung zu fassen. Beide verweisen auf eine qualitative Verschiebung: Wissensproduktion wird relational, prozessual und verteilt. Die vorgeschlagenen Begriffe stützen sich dabei auf einen einzelnen, selbst rekonstruierten Fall; ihre Tragfähigkeit über diesen Fall hinaus muss sich in weiterer, davon unabhängiger Forschung noch erweisen. Die Herausforderung für Hochschulen besteht nicht darin, diese Entwicklung zu kontrollieren, sondern darin, sie aktiv zu gestalten und Verantwortung dafür zu übernehmen.

Nachtrag: Methodische Transparenz: Der Text als Beispiel

Dieser Beitrag ist selbst Teil dessen, was er beschreibt. Er entstand in einem iterativen Prozess zwischen drei Akteur:innen: der Autorin sowie zwei KI-Systemen – Claude (Anthropic) und ChatGPT (OpenAI). Erste Skizzen und konzeptionelle Linien wurden vor der KI-Zusammenarbeit entwickelt. Die KI-Systeme wurden hinzugezogen, nachdem der inhaltliche Rahmen bereits gesetzt war. Die Zusammenarbeit verlief in drei Phasen. Zunächst diente ChatGPT als Resonanzraum und Strukturierungsinstanz: Erste Argumentationslinien wurden verdichtet, implizite Annahmen expliziert und eine tragfähige Gliederung entwickelt. In einer zweiten Phase fungierten beide Systeme als Denkkollegen: Die zentralen Begriffe – augmentierte Reflexivität und post-individuelle Autorschaft – wurden in iterativen Schleifen geschärft, Argumentationslinien wechselseitig überprüft und der Praxisbericht (Abschnitt 3) auf der Grundlage von Rohmaterial der Autorin entwickelt. In der dritten Phase übernahm Claude die Rolle einer externen Validierungsinstanz: kritische Rückmeldung zu Struktur, Argumentation und Konsistenz, Prüfung der theoretischen Verankerung, Ergänzung fehlender Zitationen und Überarbeitung des Literaturverzeichnisses. Durchgehend lagen alle Richtungsentscheide, inhaltlichen Urteile und normativen Setzungen bei der Autorin; sie gab jeden Abschnitt frei und trägt die Letztverantwortung für den Gesamttext. Die Transparenz dieses Prozesses ist nicht nur methodische Offenlegung. Sie ist Teil der Argumentation: Der Text zeigt nicht nur, was er behauptet. Er ist selbst ein Beispiel dafür. Wissenschaftliche Gedanken entstehen zunehmend in epistemischen Praktiken, deren Verantwortung individuell bleibt, deren Genese jedoch relational geworden ist.

Literaturverzeichnis

- Abdul-Jabbar, Wisam Kh./Bhatt, Ibrar (2025): Post-authorship and AI-Assisted Writing. Singularity, Communication, and Pedagogy in Higher Education, in: Postdigital Science and Education, Jg. 7, Nr. 4, S. 1136-1149. <https://doi.org/10.1007/s42438-025-00586-5>
- Anthropic (2026): Claude (Version Claude Opus 4.8) [Large Language Model], verfügbar unter: <https://claude.ai/>
- Bachelard, Gaston ([1938] 1993): La formation de l'esprit scientifique. Contribution à une psychanalyse de la connaissance objective, Paris: Vrin.
- Bajohr, Hannes (2024): Autorschaft und Künstliche Intelligenz, in: Stephanie Catani (Hrsg.), Handbuch Künstliche Intelligenz und die Künste, Berlin/Boston: De Gruyter, S. 265-280.
- Barad, Karen (2007): Meeting the Universe Halfway. Quantum Physics and the Entanglement of Matter and Meaning, Durham: Duke University Press.
- Barthes, Roland (1968): La mort de l'auteur, in: Manteia, Nr. 5, S. 12-17.
- Beck, Ulrich (1996): Das Zeitalter der Nebenfolgen und die Politisierung der Moderne, in: Ulrich Beck/Anthony Giddens/Scott Lash, Reflexive Modernisierung. Eine Kontroverse, Frankfurt a. M.: Suhrkamp, S. 19-112.
- Bender, Emily M./Gebriu, Timnit/McMillan-Major, Angelina/Shmitchell, Shmargaret (2021): On the Dangers of Stochastic Parrots. Can Language Models Be Too Big?, in: Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency (FACCT '21), New York: ACM, S. 610-623. <https://doi.org/10.1145/3442188.3445922>
- Clark, Andy (2008): Supersizing the Mind. Embodiment, Action, and Cognitive Extension, Oxford: Oxford University Press.
- Clark, Andy/Chalmers, David (1998): The Extended Mind, in: Analysis, Jg. 58, Nr. 1, S. 7-19. <https://doi.org/10.1093/analys/58.1.7>
- Engelbart, Douglas C. (1962): Augmenting Human Intellect. A Conceptual Framework, SRI Summary Report AFOSR-3223, Menlo Park: Stanford Research Institute.
- Floridi, Luciano/Chiratti, Massimo (2020): GPT-3. Its Nature, Scope, Limits, and Consequences, in: Minds and Machines, Jg. 30, Nr. 4, S. 681-694. <https://doi.org/10.1007/s11023-020-09548-1>
- Foucault, Michel (1969): Qu'est-ce qu'un auteur?, in: Bulletin de la Société française de philosophie, Jg. 63, Nr. 3, S. 73-104.
- Giddens, Anthony (1991): Modernity and Self-Identity. Self and Society in the Late Modern Age, Cambridge: Polity Press.
- Hutchins, Edwin (1995): Cognition in the Wild, Cambridge, MA: MIT Press.
- Kasneci, Enkelejd/Seßler, Kathrin/Küchemann, Stefan/Bannert, Maria/Dementieva, Daryna/Fischer, Frank/Gasser, Urs/Groh, Georg/Günemann, Stephan/Hüllermeier, Eyke/Krusche, Stephan/Kutyniok, Gitta/Michaeli, Tilman/Nerdel, Claudia/Pfeffer, Jürgen/Poquet, Oleksandra/Sailer, Michael/Schmidt, Albrecht/Seidel, Tina/Stadler, Matthias/Weller, Jochen/Kuhn, Jochen/Kasneci, Gjergji (2023): ChatGPT for Good? On Opportunities and Challenges of Large Language Models for Education, in: Learning and Individual Differences, Jg. 103, Art. 102274. <https://doi.org/10.1016/j.lindif.2023.102274>
- Knorr Cetina, Karin (1999): Epistemic Cultures. How the Sciences Make Knowledge, Cambridge, MA: Harvard University Press.
- Latour, Bruno (2005): Reassembling the Social. An Introduction to Actor-Network-Theory, Oxford: Oxford University Press.
- Licklider, Joseph C. R. (1960): Man-Computer Symbiosis, in: IRE Transactions on Human Factors in Electronics, Jg. HFE-1, S. 4-11.
- Misselhorn, Catrin (2023): Künstliche Intelligenz – das Ende der Kunst?, Ditzingen: Reclam.
- Mollick, Ethan (2024): Co-Intelligence. Living and Working with AI, New York: Portfolio.
- OpenAI (2026): ChatGPT (Version GPT-5.6 Thinking) [Large Language Model], verfügbar unter: <https://chatgpt.com/>
- Polanyi, Michael (1966): The Tacit Dimension, Garden City, NY: Doubleday.
- Schön, Donald A. (1983): The Reflective Practitioner. How Professionals Think in Action, New York: Basic Books.
- Suchman, Lucy (2007): Human-Machine Reconfigurations. Plans and Situated Actions, 2. Aufl., Cambridge: Cambridge University Press.
- Vallor, Shannon (2024): The AI Mirror. How to Reclaim Our Humanity in an Age of Machine Thinking, Oxford: Oxford University Press.
- Williamson, Ben (2024): The Social Life of AI in Education, in: International Journal of Artificial Intelligence in Education, Jg. 34, S. 97-104. <https://doi.org/10.1007/s40593-023-00342-5>
- Zawacki-Richter, Olaf/Marin, Victoria I./Bond, Melissa/Gouverneur, Franziska (2019): Systematic Review of Research on Artificial Intelligence Applications in Higher Education – Where Are the Educators?, in: International Journal of Educational Technology in Higher Education, Jg. 16, Art. 39. <https://doi.org/10.1186/s41239-019-0171-0>

■ **Elena Wilhelm**, Prof. Dr., Leiterin Hochschulentwicklung, Zürcher Hochschule der Künste
E-Mail: elena.wilhelm@zhdk.ch

Zusammenarbeit mit EBSCOhost

Unsere Zeitschriften sind in den Fachdatenbanken von EBSCO indexiert.

- Hochschulmanagement (HM)
- Das Hochschulwesen (HSW)
- Qualität in der Wissenschaft (QIW)
- Personal- und Organisationsentwicklung (POE)
- Zeitschrift für Beratung und Studium (ZBS)



EBSCOhost ist eine intuitive Online-Rechercheplattform, die von Tausenden Institutionen und Millionen Nutzer:innen weltweit genutzt wird. Mit hochwertigen Datenbanken und Suchfunktionalitäten unterstützt EBSCOhost Wissenschaftler:innen, Forschende und Studierende dabei, die benötigten Informationen schnell und einfach zu finden.