KLEINE KIELER
BEITRÄGE ZU

Künstlicher Intelligenz

BAND 6

VK:KIWA

KLEINE KIELER BEITRÄGE ZU

Künstlicher Intelligenz

Pädagogische und fachdidaktische Perspektiven
auf technologische Entwicklungen

BAND 6

Herausgeben von
Kirsten Schindler, Nicolaus Wilder, Virtuelles Kompetenzzentrum:
Künstliche Intelligenz in Bildung, Wissenschaft & Arbeitswelt (VK:KIWA)

Dominik Freinhofer, Gerlinde Schwabl,
Sandra Breitenberger, Anja Steiner

Das PCRR-Framework: Plan – Create – Review – Reflect

Ein prozessorientiertes Framework für die Zusammenarbeit mit Generativer KI in Bildungskontexten

VK:KIWA

KLEINE KIELER BEITRÄGE ZU KÜNSTLICHER INTELLIGENZ | BAND 6

eISSN: 3053-4488

HERAUSGEGEBEN VON

Prof.in Dr. Kirsten Schindler, Bergische Universität Wuppertal
Dr. Nicolaus Wilder, Christian-Albrechts-Universität zu Kiel
Virtuelles Kompetenzzentrum: Künstliche Intelligenz in Bildung, Wissenschaft & Arbeitswelt
https://www.vkkiwa.de/

AUTOR*INNEN

Dominik Freinhofer
Universität Graz, IDea_Lab – das Interdisziplinäre Digitale Labor der Universität Graz, dominik.freinhofer@uni-graz.at

Gerlinde Schwabl
Pädagogische Hochschule Tirol, Institut für Berufspädagogik,
Fachstelle für Medienbildung und Digitalisierung, gerlinde.schwabl@ph-tirol.ac.at

Sandra Breitenberger
Pädagogische Hochschule Oberösterreich, Institut für Berufspädagogik,
sandra.breitenberger@ph-ooe.at

Anja Steiner
Pädagogische Hochschule Tirol, Institut für Berufspädagogik, anja.steiner@ph-tirol.ac.at

Bibliographische Information der Deutschen Nationalbibliothek

Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie. Detaillierte bibliographische Daten sind über https://dnb.d-nb.de abrufbar.

Das Werk ist auf dem Open-Access-Publikationsserver MACAU der Universitätsbibliothek Kiel (https://macau.uni-kiel.de) frei verfügbar: https://doi.org/10.38071/2026-00444-2.

Kiel, 2026
Selbstverlag des Virtuellen Kompetenzzentrums für Künstliche Intelligenz und Wissenschaftliches Arbeiten (VK:KIWA) im Hosting-Service der UB Kiel.

Titelbild: Erstellt mit Recraft (https://www.recraft.ai/).

Geleitwort zur Einordnung

Die technologische Entwicklung der letzten Jahre verändert zeitgleich und substanziell die Idee davon, was Lernende im Umgang mit KI benötigen – und wie wir (als Lehrende) sie dafür vorbereiten sollen. Stand kurz nach der Veröffentlichung von ChatGPT (und dann schnell anderer KI-Chatbots) zunächst das Prompting oder auch Prompt Engineering im Fokus, so hat sich der Akzent inzwischen deutlich verschoben: weg von der Vorstellung, es komme vor allem auf „die eine gute Eingabe" oder auf ein Repertoire an Prompt-Tricks an, und hin zu der Frage, wie Arbeitsprozesse, Kontexte und Qualitätsroutinen gestaltet sein müssen, damit KI-Systeme sinnvoll, nachvollziehbar und verantwortbar genutzt werden können.

Zu dieser Verschiebung trägt auch der technische Fortschritt selbst bei. Neuere Modellgenerationen – einschließlich sogenannter Reasoning-Modelle – sind in vieler Hinsicht robuster gegenüber weniger elaborierten Eingaben, als es frühere Modelle waren. Das heißt jedoch nicht, dass die Anforderungen geringer werden. Sie verändern sich: Entscheidend wird weniger, ob ein Prompt „raffiniert" ist, sondern ob Aufgabenstellung, Rahmenbedingungen, Materialgrundlage und Prüfschritte hinreichend präzise sind; ob Annahmen transparent gemacht werden; ob Zwischenergebnisse geprüft, überarbeitet und begründet werden; und ob die Nutzung der Systeme insgesamt in ein reflektiertes Vorgehen eingebettet ist.

An diese veränderte Entwicklung knüpft das hier vorliegende Rahmenmodell (Framework) an. In seiner revidierten Fassung lässt es sich – im weitesten Sinne – als eine Form von Context Engineering verstehen: nicht als bloßes „Mehr Kontext", sondern als bewusste Gestaltung eines Arbeitsumfelds, in dem Kontext entsteht, strukturiert, aktualisiert und überprüft wird. Gemeint ist dabei ausdrücklich ein breites Spektrum von Kontextdimensionen – von User-Prompt und Systemprompt über Retrieval-gestützte Kontexte (RAG) und kurz- bzw. längerfristige Erinnerungs-/Speichermechanismen bis hin zu Werkzeugen, die in die Bearbeitung eingebunden werden. Im Zentrum steht damit nicht die einzelne Eingabe, sondern ein komplexes Arbeitssetting, in dem das eigene Tun systematisch

geplant, entworfen, überarbeitet und reflektiert wird. Die Interaktion mit KI wird so als iterativer Prozess modelliert, der an unterschiedlichen Stellen unterschiedliche Kompetenzen verlangt – und genau dadurch in Lehr-/Lernszenarien adressierbar wird.

Das PCRR-Framework (Plan – Create – Review – Reflect) macht diese Prozesslogik explizit. Es bietet eine klare Struktur, die zugleich offen genug ist, um sehr unterschiedliche Aufgabenformate zu tragen: von der Erkundung eines Themas über das Erstellen erster Entwürfe bis hin zur Überarbeitung unter Kriterien, zur Fehleranalyse und zur Reflexion der eigenen Entscheidungen. Alle Arbeitsschritte sind entsprechend der Kompetenzen der Lernenden skalierbar, für Lernkontexte dokumentierbar und damit – im Prinzip – auch transparent bewertbar: Nicht nur das Produkt zählt, sondern der Weg dorthin, inklusive der Entscheidungen, Alternativen, Korrekturen und Begründungen.

Mit der vorliegenden Publikation legen die Autor*innen Dominik Freinhofer, Gerlinde Schwabl, Sandra Breitenberger und Anja Steiner eine überarbeitete Fassung des ursprünglichen Frameworks vor. Die Revision macht sichtbar, dass Frameworks dieser Art nicht als „fertige Modelle" verstanden werden sollten, sondern als Vorschläge, die sich in der Praxis bewähren, irritieren, nachgeschärft und weiterentwickelt werden müssen. Das ist kein Nebenaspekt, sondern gehört zur Sache: Wenn sich die technologischen Bedingungen und Nutzungskontexte so dynamisch verändern, dann muss auch ein didaktisches Rahmenmodell in der Lage sein, auf neue Möglichkeiten, neue Grenzen und neue Risiken zu reagieren. Entscheidend ist dabei weniger, ob man dem neuesten Modell hinterherläuft, sondern ob die didaktische Struktur trägt.

Bislang ist das Modell zwar erprobt; belastbare, systematische Evaluationen in größerem Maßstab stehen jedoch – soweit ersichtlich – noch aus. Gerade deshalb ist die Veröffentlichung in der KiKI-Reihe mehr als reine Dokumentation: Sie kann dazu beitragen, die weitere Erprobung zu verbreitern, Vergleichs- und Evaluationsdesigns anzustoßen und zugleich die technologische Weiterentwicklung im Blick zu behalten.

Kirsten Schindler & Nicolaus Wilder

Inhalt

1 Einleitung

Die Jahre 2024 und 2025 haben einen Wendepunkt in der Operationalisierung von *Large Language Models* (LLMs) markiert. Die Disziplin, die in der Frühphase (ca. 2018–2023) der Generativen Künstlichen Intelligenz (GenKI) als ›Prompt Engineering‹ bekannt wurde, hat eine fundamentale Transformation durchlaufen. Wissenschaftliche Untersuchungen belegen, dass sich dieses Feld in die umfassendere, systemtheoretisch fundierte Disziplin des *Context Engineering* aufgelöst hat (Mei et al., 2025).

Dieser Wandel ist primär technologisch getrieben. Während frühere Sprachmodelle (z. B. GPT-3) hochsensibel auf kleinste Formulierungsänderungen reagierten (Holtzman et al., 2022; Zhao et al., 2021), automatisieren moderne *Reasoning-Modelle* (z. B. GPT-5.2-Thinking oder Claude-Opus-4.5) heute viele dieser Schritte. Diese Systeme verfügen über die Fähigkeit zur internen *Chain of Thought* (Gedankenkette): Sie zerlegen komplexe Aufgaben erst intern in logische Zwischenschritte und wägen verschiedene Lösungswege ab, bevor sie eine endgültige Antwort ausgeben (Wei et al., 2023; Zhang et al., 2022).

Zudem hat sich das Verständnis von KI-Architekturen gewandelt: Man agiert heute seltener mit isolierten Modellen, sondern vermehrt mit *Compound AI Systems* (J. Chen et al., 2025). In dieser Architektur fungiert das Sprachmodell lediglich als ›Motor‹, während das Gesamtsystem – die ›Karosserie‹ – durch die Integration von Werkzeugen wie Websuche, Datei-Uploads oder Code-Interpretern erst seine volle Leistungsfähigkeit entfaltet (Buck, 2025; J. Chen et al., 2025). Context Engineering beschreibt dabei die Kunst, diese Zusammenarbeit zwischen Mensch und Maschine sowie den gesamten ›Information Payload‹, den das System während der Inferenz (dem Moment der Antwortgenerierung) berücksichtigt, systematisch zu optimieren (Mei et al., 2025).

Trotz dieser technischen Fortschritte bleibt die menschliche Steuerung per Prompt jedoch oft fehleranfällig und ›brüchig‹ *(prompt brittleness)* (Ceron et al., 2024; Verma et al., 2024). Nutzer*innen ohne spezifische KI-Expertise neigen zu opportunistischen *Trial-and-Error-Strategien*, welche die

Kapazitäten moderner Systeme kaum ausschöpfen (Kaddour et al., 2023; Zamfirescu-Pereira et al., 2023). In der pädagogischen Praxis stellt diese Komplexität eine erhebliche Herausforderung dar. Obwohl Prompting heute als neue digitale Kernkompetenz gehandelt wird (Korzynski et al., 2023), belegen Untersuchungen zur *AI Literacy*, dass insbesondere Lernende häufig in unproduktive Handlungsmuster verfallen (Knoth et al., 2024). Ohne ein tieferes Verständnis der Systemlogik bleiben Anfragen oft oberflächlich, was die Gefahr von Halluzinationen – faktisch falschen, aber plausibel klingenden Ausgaben – signifikant erhöht.

Um KI effektiv und ethisch verantwortungsvoll in Lernprozesse zu integrieren, bedarf es daher mehr als bloßer technischer ›Tricks‹. Es erfordert eine Informationskompetenz, die kritisches Denken und die Fähigkeit zur Evaluation von KI-Outputs ins Zentrum rückt (Lo, 2023). Gutes Prompten wird dabei zunehmend als Spiegelbild ›guten Denkens‹ verstanden (Sawetz, 2025): Die Formulierung eines hochwertigen Inputs ist eine Externalisierung kognitiver Prozesse; sie erfordert eine systematische Problemzerlegung und einen bewussten Perspektivenwechsel. Hier wird deutlich, warum ein didaktischer Rahmen für Bildungskontexte unerlässlich ist: Lehrende benötigen Instrumente, um den Grad der KI-Nutzung transparent zu machen und ethisch zu bewerten, während Lernende eine Struktur brauchen, die sie sicher durch den komplexen Prozess der Mensch-KI-Kollaboration führt (Perkins et al., 2024; Roe et al., 2025).

An dieser Schnittstelle setzt das von Freinhofer et al. (2025) entwickelte *PCRR-Framework* (Plan, Create, Review, Reflect) an. Es schließt die Lücke zwischen technologischer Potenz und pädagogischer Praxis, indem es die Interaktion mit KI nicht als punktuelle Eingabe, sondern als ganzheitlichen, iterativen Arbeitsprozess begreift:

- *Plan:* Die strategische Vorbereitung, in der Ziele präzisiert, Werkzeuge ausgewählt und Kontextquellen zusammengestellt werden.
- *Create:* Die operative Phase der Interaktion, gestaltet als dialogischer Prozess zwischen Mensch und Maschine.

- *Review:* Die kontinuierliche Qualitätskontrolle hinsichtlich fachlicher Korrektheit, logischer Konsistenz und möglicher Biases.
- *Reflect:* Die metakognitive Einordnung des Prozesses, die den individuellen Kompetenzgewinn sichert.

Das PCRR-Framework dient somit als methodische Brücke: Es stellt sicher, dass die Nutzung von KI in der Bildung nicht zur bloßen Delegation von Aufgaben verkommt, sondern als Katalysator für vertieftes Denken und nachhaltigen Kompetenzaufbau fungiert.

Ziel der vorliegenden Publikation ist es, dieses Framework gegenüber der ersten Version (Freinhofer et al., 2025) theoretisch durch das Konzept des Context Engineering zu fundieren, die vier Phasen differenzierter auszuarbeiten und das Framework systematisch im wissenschaftlichen Diskurs zu verorten. Dazu gliedert sie sich in vier Teile. Zunächst wird das theoretische Fundament gelegt: Der Abschnitt zu *Prompt Engineering* und *Context Engineering* zeichnet die Entwicklung von der isolierten Eingabegestaltung hin zur umfassenden Kontextoptimierung nach und konzeptualisiert *Context Engineering* anhand von sechs Dimensionen. Darauf aufbauend wird das PCRR-Framework in seiner aktualisierten Version vorgestellt – von der Grundkonzeption über die detaillierte Ausarbeitung der vier Phasen *Plan, Create, Review* und *Reflect* bis hin zu Überlegungen zur kontextspezifischen Differenzierung. Der dritte Teil verortet das Framework im wissenschaftlichen Diskurs: Die Diskussion grenzt PCRR von bestehenden Ansätzen ab und arbeitet dessen spezifischen Beitrag heraus, während die Limitationen die Grenzen der vorliegenden Arbeit transparent machen. Abschließend fasst das Fazit die zentralen Erkenntnisse zusammen und skizziert die Forschungsagenda für die Weiterentwicklung des Frameworks.

2 Prompt Engineering und Context Engineering

Prompt Engineering (in weiterer Folge auch kurz ›Prompting‹ oder ›Prompten‹ genannt) bezeichnet die Kunst der gezielten Eingabegestaltung für sprachgesteuerte KI-Systeme. Bei Sprachmodellen (z. B. GPT-5.2, Gemini-3.0 oder Claude-Opus-4.5) und den darauf aufbauenden KI-Chatbots (z. B. ChatGPT, Copilot, Perplexity) steht die inhaltliche und sprachliche Formulierung der Eingabe im Zentrum. Der KI-Entwickler Anthropic definiert diese Gestaltung folgendermaßen:

> Prompt engineering is the craft of structuring instructions to get better outputs from AI models. It's how you phrase queries, specify style, provide context, and guide the model's behavior to achieve your goals. (Anthropic, 2025a)

Nicht alle KI-Tools werden jedoch durch Prompts gesteuert – manche, wie der KI-Übersetzer DeepL, operieren über Funktionsauswahl und Parametereinstellungen ohne explizite sprachliche Instruktion.

Als die ersten Sprachmodelle und KI-Chatbots entwickelt wurden, stellte effektives Prompt Engineering die Schlüsselkompetenz im Umgang mit diesen Technologien dar:

> In the early days of engineering with LLMs, prompting was the biggest component of AI engineering work, as the majority of use cases outside of everyday chat interactions required prompts optimized for one-shot classification or text generation tasks. As the term implies, the primary focus of prompt engineering is how to write effective prompts, particularly system prompts. (Anthropic, 2025b)

Effektives Prompt Engineering war in der frühen Modellentwicklung unter anderem deshalb so zentral, weil anfängliche Chatbots fast ausschließlich aus dem Sprachmodell bestanden. Die Unterscheidung zwischen Sprachmodellen und KI-Chatbots ist hierbei zentral: Während Sprachmodelle primär Texte verarbeiten und generieren, erweitern moderne Chatbots diese Grundfunktionalität durch zusätzliche Systeme. Diese verfügen über

Funktionen, die weit über reine Textverarbeitung hinausgehen: *Grounding*[1] für Echtzeitinformationen, *Retrieval-Augmented Generation* (RAG)[2] für externe Wissensquellen, über den einzelnen Chat hinausgehende Kontextinformationen, eine Konsole für das Ausführen von Programmcode, die Canvas-Funktion zum kollaborativen Schreiben, Konnektoren mit anderen Websites und Apps, agentische Fähigkeiten und vieles mehr. Dies ist anhand der offiziellen *Release Notes* von ChatGPT gut zu beobachten: Während das zugrunde liegende Sprachmodell von November 2022 (GPT-3.5) bis Dezember 2025 (GPT-5.2) immer besser wurde, waren die Fortschritte in den Chatbot-Funktionen noch größer. Dazu zählten unter anderem die Websuche (05/23), der Code Interpreter und die Custom Instructions (07/23), die Stimm- und Bild-Unterstützung (09/23), die Custom GPTs (11/23), die ersten Memory-Funktionen (02/24), Canvas und Projekte (12/24), Scheduled Tasks (01/25) und immer mehr werdende Connectors zu anderen Apps seit Anfang 2025 (OpenAI, 2026b).

Die Schreibdidaktikerin Isabella Buck vergleicht treffend das Sprachmodell mit dem Motor und den Chatbot mit der Karosserie – eine Analogie, die die Komplexitätssteigerung verdeutlicht (Buck, 2025). Erweitert man diese Analogie, so sind die Nutzenden die Fahrenden: Sie bestimmen das Ziel, wählen die Route und korrigieren den Kurs – durch ihre Entwürfe, Prompts und die Überarbeitung der KI-generierten Outputs. Erst dieses Zusammenspiel von Mensch und Maschine führt zum gewünschten Ergebnis.

Wurde ›Prompt Engineer‹ anfangs noch als Zukunftsberuf verkündet und als sehr gut bezahlter Job gehandelt (Harwell, 2023; Stern, 2023), verlor Prompt Engineering zwischen 2023 und 2025 aufgrund dieser rapiden Verbesserungen in der Modellqualität und der Chatbot-Funktionalität

1 Als Grounding wird die Verankerung von KI-Ausgaben in verifizierbaren, externen Quellen bezeichnet. Technisch realisiert durch Echtzeit-Websuche oder Quellenverweise, reduziert Grounding das Risiko von Halluzinationen und erhöht die faktische Zuverlässigkeit (Thoppilan et al., 2022).

2 *RAG (Retrieval-Augmented Generation)* bezeichnet Verfahren, bei denen ein Sprachmodell vor der Antwortgenerierung auf externe Informationsquellen zugreift – etwa durch Websuche, Dokumenten-Upload oder Datenbankanbindung. Dies ermöglicht den Zugriff auf aktuelle oder spezialisierte Informationen, die nicht im Modelltraining enthalten waren, und reduziert Halluzinationen (Lewis et al., 2021; Shuster et al., 2021).

schrittweise an (relativer) Bedeutung. Viele Schlagzeilen in der LLM-Welt lauteten: »Prompt Engineering is dead.« (Genkina, 2024; Gupta, 2025; Hauben, 2025) An dessen Stelle soll das sogenannte *Context Engineering* treten.

2.1 Die Entstehung des Context Engineering

Diese rasanten Entwicklungen im LLM-Bereich führten zur Entstehung des Begriffs *Context Engineering*. 2025 prägte der KI-Entwickler Walden Yan (Mitbegründer von Cognition AI) den Begriff zunächst im Kontext der KI-Agenten-Implementation (Yan, 2025). Shopify-CEO Tobi Lütke erweiterte die Definition zu »the art of providing all the context for the task to be plausibly solvable by the LLM« (Lütke, 2025), während der Softwareentwickler Dexter Horthy Prompt Engineering explizit als Teilbereich des Context Engineering einordnete (Horthy, 2025). Auch die Entwickler von Sprachmodellen und KI-Chatbots nahmen diese Entwicklung ernst. So beschreibt der Claude-Entwickler Anthropic Context, Prompt Engineering und Context Engineering folgendermaßen:

> **Context** refers to the set of tokens included when sampling from a large-language model (LLM). The engineering problem at hand is optimizing the utility of those tokens against the inherent constraints of LLMs in order to consistently achieve a desired outcome. [...] At Anthropic, we view context engineering as the natural progression of prompt engineering. Prompt engineering refers to methods for writing and organizing LLM instructions for optimal outcomes [...]. **Context engineering** refers to the set of strategies for curating and maintaining the optimal set of tokens (information) during LLM inference, including all the other information that may land there outside of the prompts. (Anthropic, 2025b, Hervorhebung im Original)

Google definiert Context Engineering als »[t]he process of dynamically assembling and managing information within an LLM's context window to enable stateful, intelligent agents.« (Milam & Gulli, 2025, S. 6) Google sieht Context Engineering als Evolution des Prompt Engineering:

> Context Engineering represents an evolution from traditional **Prompt Engineering**. Prompt engineering focuses on crafting optimal, often static, system instructions. Conversely, **Context Engineering** addresses the entire payload, dynamically constructing a state-aware prompt based on the user, conversation history, and external data. It involves strategically selecting, summarizing, and injecting different types of information to maximize relevance while minimizing noise. External systems—such as RAG databases, session stores, and memory managers—manage much of this context. The agent framework must orchestrate these systems to retrieve and assemble context into the final prompt. (Milam & Gulli, 2025, S. 7, Hervorhebung im Original)

Diese Diskussion, die zunächst primär in sozialen Medien und Entwicklerblogs[3] stattfand, etabliert sich zunehmend auch in der Wissenschaft. Mei et al. (2025) definierten Context Engineering in ihrem Survey als eine formale Disziplin, die über das reine Design von Instruktionen *(Prompt Design)* hinausgeht und die systematische Optimierung des gesamten ›Information Payloads‹ für LLMs umfasst. Die Arbeit liefert eine essenzielle theoretische Basis, indem sie Context Engineering in eine Taxonomie aus drei funktionalen Kernkomponenten zerlegt: (1) *Context Retrieval* und *Generation* zur Akquise externen Wissens, (2) *Context Processing* zur Strukturierung und Selbstreflexion der Informationen sowie (3) *Context Management*, welches die effiziente Verwaltung von Gedächtnishierarchien und Kompressionsverfahren adressiert. Durch diese Einordnung wird Context Engineering als das übergeordnete Architekturprinzip etabliert, das die Brücke zwischen rohen Daten und der Modellinferenz schlägt, während Prompting als spezifisches Werkzeug innerhalb dieses Rahmens fungiert.

Die Bedeutung von Prompt Engineering relativiert sich dadurch, wird aber keineswegs obsolet. Die Kritik, dass bessere Modelle und geringere Prompt-Sensitivität das Prompt Engineering überflüssig machen würden, unterschätzt die Rolle präziser Kommunikation. Unabhängig davon, wie

3 Da sich das Feld des Context Engineering dynamisch entwickelt und wesentliche Diskurse außerhalb traditioneller Publikationsformate stattfinden, beziehen wir neben wissenschaftlichen Surveys auch Primärquellen aus der Entwickler-Community (Blogartikel, technische Dokumentationen) ein.

intelligent, kompetent oder erfahren mehrere Menschen sind, die zusammenarbeiten – ohne effektive Kommunikation können sie keine gemeinsamen Ziele erreichen. Sprache bleibt das zentrale Medium, um nicht nur Menschen, sondern auch KI-Systemen Ziele, Methoden und relevanten Kontext zu vermitteln.

Ähnlich sieht es der Softwareentwickler Simon Willison, der folgende Anforderungen für effektives Prompten stellt:

> First, you need **really great communication skills**. [...] You need to be able to methodically apply a version of **the scientific method** to your work. Figuring out what works and what doesn't [...] is a huge challenge. [...] The more experience I get with prompting the more areas of expertise I add to my own model of the ideal prompt engineer. Those areas of expertise currently include: **Linguistics** [...,] **Deep learning and associated computer science** [...,] **Human psychology** [...,] **Philosophy** [...,] **Art history** [..., and] **Computer security**. (Willison, 2023, Hervorhebung im Original)

Neben dem Prompt Engineering umfasst der größere Bereich des Context Engineering weitere Bestandteile, die im nächsten Abschnitt behandelt werden.

2.2 Die Dimensionen des Context Engineering

Die wissenschaftliche Systematisierung des Context Engineering befindet sich noch im Entstehen, wobei verschiedene Taxonomien unterschiedliche Aspekte betonen. Mei et al. (2025) unterscheiden in ihrem Survey sechs Input-Komponenten, die bei jeder LLM-Inferenz zusammenwirken: Systeminstruktionen (c_{instr}), externes Wissen (c_{know}), Tool-Definitionen (c_{tools}), persistentes Gedächtnis (c_{mem}), dynamischer Status (c_{state}) und die aktuelle Nutzeranfrage (c_{query}). Milam und Gulli (2025) strukturieren Context Engineering entlang zweier Hauptpfeiler – *Sessions* für den kurzfristigen Gesprächskontext und *Memory* für persistente Informationen –, ergänzt durch Payload-Komponenten wie Systeminstruktionen, Tool-Definitionen und externes Wissen. Horthy (2025) visualisiert Context Engineering als Venn-

Diagramm, in dem sich sechs Bereiche überlappen: Prompt Engineering, RAG, State/History, Memory, Structured Outputs und Tools.

Eine Synthese dieser Ansätze offenbart konvergierende Kernkategorien bei gleichzeitiger terminologischer Heterogenität. Tabelle 1 stellt die Zuordnungen dar.

Dimension dieser Arbeit	*Mei et al., 2025*	*Milam & Gulli, 2025*	*Horthy, 2025*
User Prompt	c_{query}	User Prompt	Prompt Engineering
Systemprompt	c_{instr}	System Instructions	(Teil von Prompt Eng.)
RAG	c_{know}	External Knowledge	RAG
Short-Term Memory	c_{state}	Sessions, History	State/History
Long-Term Memory	c_{mem}	Memory	Memory
Tools	c_{tools}	Tool Definitions	Tool calls

Tabelle 1: Synthese der Context-Engineering-Taxonomien

Basierend auf dieser Synthese konzeptualisieren wir Context Engineering durch sechs Dimensionen. Diese überlappen sich teilweise, was kein konzeptioneller Mangel ist, sondern die vernetzte Natur moderner KI-Interaktion widerspiegelt (siehe Abb. 1). Die konkrete Anordnung der Überlappungen in der Abbildung ist dabei illustrativ zu verstehen – in der Praxis beeinflussen sich alle Dimensionen wechselseitig.

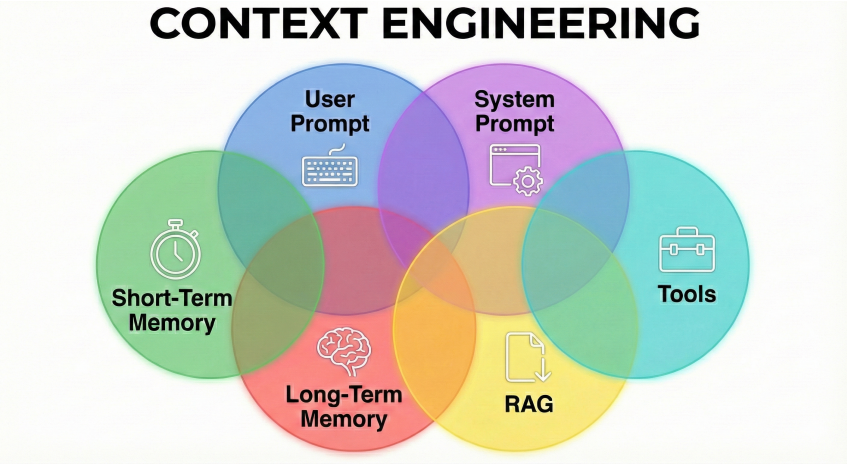

Abbildung 1: Context Engineering. Die Grafik wurde mit Googles Bildmodell ›Nano Banana Pro‹ (›Gemini 3 Pro Image‹) erstellt.

User Prompt: Die Eingabe in einen KI-Chatbot stellt die direkte sprachliche Schnittstelle zwischen Mensch und Maschine dar. Dieser Prompt ist der Anstoß für die Zusammenarbeit und die Generation eines Outputs. Zentral sind dabei vor allem die inhaltliche Formulierung der Instruktion und die logische bzw. syntaktische Strukturierung der Eingabe(n). Dies umfasst etablierte Techniken des Prompt Engineering, wie z. B. das *Persona-Prompting* (Zuweisung einer Rolle), dessen Wirksamkeit allerdings modell- und aufgabenabhängig variiert und möglicherweise ausgedient hat (Basil et al., 2025), *Few-Shot-Prompting* (Bereitstellung von Lösungsbeispielen) sowie der Einsatz von Delimitern (Trennzeichen wie ### oder XML-Tags), um Instruktionstext und Quellmaterial für das Modell eindeutig voneinander abzugrenzen (Schulhoff et al., 2025; White et al., 2023).

Systemprompt: Der Systemprompt bildet den normativen Rahmen der Interaktion. Er umfasst übergeordnete Instruktionen, die das Verhalten des Modells persistent steuern – von funktionalen Regeln bis hin zu stilistischen Vorgaben wie Tonalität oder Kommunikationsstil (Mei et al., 2025; Milam & Gulli, 2025). Systemprompts entstehen dabei auf mehreren Ebenen: Unternehmen wie OpenAI oder Anthropic definieren grundlegende Verhaltensrichtlinien, Plattformen wie Perplexity ergänzen anwendungsspezifische Instruktionen, Anwendungsentwickler*innen konfigurieren spezialisierte

Assistenten (z. B. Custom GPTs oder Gems) und Endnutzer*innen können über *Custom Instructions* eigene Präferenzen hinterlegen. Diese geschichtete Architektur führt dazu, dass der effektive Systemprompt einer Interaktion aus mehreren, hierarchisch angeordneten Quellen zusammengesetzt wird – eine Struktur, die sich metaphorisch als *Prompt-Pyramide* beschreiben lässt.

Retrieval-Augmented Generation (RAG): RAG erweitert die Modell-Wissensbasis durch externe Quellen. Da ein Sprachmodell auf einem statischen (sprich: veralteten), nicht allumfassenden Trainingsdatensatz basiert, ermöglicht RAG den Zugriff auf Informationen, die nicht im Modelltraining vorkamen bzw. die dem Modell nicht mehr zur Verfügung stehen. Dies umfasst in der Praxis die Web-Recherche für Echtzeitinformationen (auch bekannt als *Grounding*), den Upload von Dokumenten (z. B. PDFs oder Bildern) oder die Anbindung an externe Datenbanken (Mei et al., 2025; Milam & Gulli, 2025).

Short-Term Memory: Diese Dimension umfasst den unmittelbaren Gesprächskontext. Es handelt sich um den chronologischen Verlauf einer Konversation: bisherige Eingaben, Outputs und Zwischenergebnisse, die im aktuellen Kontextfenster verfügbar sind. Bei längeren Konversationen werden ältere Teile des Gesprächs aus dem Kontextfenster verdrängt (*Sliding Window*), wodurch frühere Informationen für das Modell nicht mehr zugänglich sind. Diese Flüchtigkeit unterscheidet das Short-Term Memory grundlegend vom persistenten Long-Term Memory (Mei et al., 2025; Milam & Gulli, 2025).

Long-Term Memory: Im Gegensatz zum flüchtigen Short-Term Memory ermöglicht das Long-Term Memory die Persistenz von Informationen über einzelne Konversationen hinaus. Milam und Gulli (2025) beschreiben dies als ›organisierten Aktenschrank‹, in dem wichtige Informationen über mehrere Sitzungen hinweg erfasst und konsolidiert werden. Dies umfasst vom Chatbot gespeicherte Nutzerpräferenzen, Wissensstände oder projektbezogene Informationen, die automatisch oder auf Anfrage in neue Sitzungen übernommen werden. Auch nutzerseitige Stilpräferenzen (›Ich möchte, dass du mich duzt.‹) werden hier konfiguriert (Mei et al., 2025; Milam & Gulli, 2025).

Tools: Die Tool-Dimension umfasst die Fähigkeit von Chatbots, externe Funktionen zu nutzen. Neben den nativen Fähigkeiten (z. B. Internetsuche, Canvas-Feature) können Chatbots durch Schnittstellen (*Application Programming Interface, API*) auch externe Programme steuern. Wesentliche Beispiele im Bildungskontext sind Code-Interpreter für Berechnungen und Datenvisualisierungen, Browser-Module für Faktenverifikation oder Protokolle zur Verbindung mit anderen Programmen (z. B. das *Model Context Protocol, MCP*) (Mei et al., 2025; Milam & Gulli, 2025).

Ergänzend zu diesen sechs Dimensionen ist *Context Management* als übergeordnetes Orchestrierungsprinzip zu verstehen (Mei et al., 2025). Da Sprachmodelle nur eine begrenzte Anzahl an Token verarbeiten können (das sogenannte Kontextfenster), müssen relevante Informationen priorisiert, irrelevante gefiltert und bei Bedarf komprimiert werden. Context Management beschreibt somit nicht eine weitere Input-Komponente, sondern die strategische Verwaltung der sechs Dimensionen angesichts technischer Beschränkungen.

Die sechs Dimensionen operieren nicht isoliert, sondern beeinflussen sich wechselseitig. Für die didaktische Arbeit lassen sie sich durch drei übergeordnete Perspektiven strukturieren:

- *Prozessarchitektur – wie der Workflow organisiert wird:* Diese Perspektive betrifft die strukturelle Organisation der Interaktion: Welche Tools werden aktiviert? Wie wird das Kontextfenster verwaltet? Wie werden Outputs iterativ verfeinert?
- *Inhaltliche Dimension – was eingegeben wird:* Diese Perspektive umfasst die konkreten Informationen, die dem Modell zur Verfügung gestellt werden – User Prompts, RAG-Inhalte, Systemprompts und Memory-Informationen.
- *Sprachlich-technische Umsetzung – wie formuliert wird:* Diese Perspektive betrifft die konkrete Ausgestaltung der Prompts: Präzision bedeutet hier nicht stilistische Perfektion, sondern unmissverständliche Kommunikation der Erwartungen – einschließlich Strukturierung, Formatierung und Explizitheit.

2.3 Implikationen für den Bildungsbereich

Die Vielschichtigkeit des Context Engineering verdeutlicht, warum die aktuelle KI-Nutzung weit über einzelne Prompts hinausgeht. Nutzende formulieren nicht nur Anfragen, sondern orchestrieren komplexe Arbeitsprozesse – sie laden Primärquellen hoch, aktivieren Analysewerkzeuge, managen Kontextfenster über multiple Sitzungen und evaluieren kritisch die generierten Outputs. Im Bildungskontext etwa bedeutet dies, dass Lernende diese Prozesse bewusst und kompetent gestalten können müssen.

Diese Komplexität macht die Bedeutung strukturierter Frameworks wie PCRR deutlich. Während Context Engineering die verfügbaren Komponenten und deren Zusammenspiel beschreibt, bietet PCRR einen didaktischen Rahmen für deren zielgerichteten Einsatz. Das Framework adressiert systematisch, wie die verschiedenen Context-Engineering-Dimensionen in den Phasen Plan, Create, Review und Reflect produktiv genutzt werden können.

Trotz aller technischen Möglichkeiten bleiben fundamentale Herausforderungen: Halluzinationen des Modells, Output-Variabilität und die Gefahr unkritischer Technikabhängigkeit. Context Engineering ist kein Automatismus, sondern erfordert weiterhin fachliche Expertise, kritisches Denken und verantwortungsvollen Umgang.

In der weiteren Arbeit verwenden wir ›Prompting‹ als etablierten Begriff, verstehen darunter aber explizit das gesamte Spektrum des Context Engineering. Prompting fungiert somit als *pars pro toto* für die holistische Gestaltung der Mensch-KI-Kollaboration. Diese umfassende Perspektive ist grundlegend für das PCRR-Framework und die damit verbundenen Lehr-Lernprozesse.

3 Das PCRR-Framework

3.1 Grundkonzeption und Zielsetzung

Das von Freinhofer et al. (2025) entwickelte PCRR-Framework – welches aus den Phasen *Plan, Create, Review* und *Reflect* besteht – strukturiert die Kollaboration zwischen Menschen und KI-Systemen als ganzheitlichen Prozess. Ursprünglich für die Berufsbildung Sekundarstufe konzipiert, adressiert Version 2 nun explizit alle Bildungsstufen und Lehr-/Lernkontexte – von der Sekundarstufe I über die Sekundarstufe II bis zur Hochschulbildung – und ist prinzipiell auch auf außerschulische Lern- und Arbeitskontexte übertragbar.

Aufbauend auf dem erweiterten Verständnis von Prompting als Context Engineering beschreibt das Framework einen Kollaborationsprozess, der bereits vor dem Formulieren des eigentlichen Prompts beginnt und sich nach dem Erhalt einer KI-generierten Ausgabe fortsetzt.[4]

Das Framework verfolgt drei zentrale Ziele:

- *Strukturierte Vermittlung:* Lehrende erhalten einen systematischen Rahmen zur Integration Generativer KI in Lernprozesse und zur Vermittlung von KI-Kollaborationskompetenzen.
- *Effektive Nutzung:* Lernende werden befähigt, verantwortungsvoll und zielgerichtet mit KI-Systemen zu arbeiten.
- *Transparente Beurteilung:* Der dokumentierte Prozess ermöglicht die nachvollziehbare Bewertung der Kollaborationsleistung.

4 Während Prompting für die Nutzung von Bildmodellen und anderen modalen Spielformen Generativer KI ebenso relevant ist, beschränkt sich das PCRR-Framework vorerst auf die Nutzung von Sprachmodellen bzw. KI-Chatbots.

3.2 Die vier Phasen im Überblick

Das PCRR-Framework gliedert sich in vier aufeinander aufbauende und rekursive Phasen – Plan, Create, Review und Reflect –, die jedoch nicht linear, sondern iterativ und verzahnt zu verstehen sind (siehe Abb. 2).

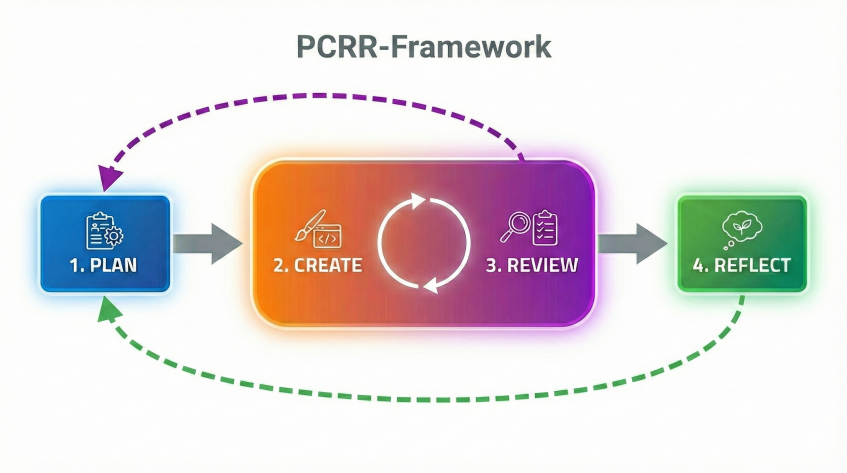

Abbildung 2: Das PCRR-Framework. Die Grafik wurde mit Googles Bildmodell ›Nano Banana Pro‹ (›Gemini 3 Pro Image‹) erstellt.

Diese Phasen beschreiben den gesamten Prozess der KI-Kollaboration, samt Vor- und Nachbereitung seitens der Nutzer*innen – im Bildungskontext also die Lernenden.

- *Plan:* Die strategische Vorbereitungsphase, in der Ziele definiert, Tools sowie deren Funktionen ausgewählt werden und die Kollaborationsstrategie festgelegt wird. Die unterschiedlichen Dimensionen des Context Engineering sind hier bereits mitzudenken: Welche Informationen und Quellen (RAG) sind erforderlich? Wie beeinflussen bestehende Systemprompts das Verhalten des Modells, und sollten eigene Instruktionen konfiguriert werden?[5] Sind für die

5 Ein Systemprompt enthält möglicherweise die Instruktion, dass das Modell ein hilfreicher und freundlicher Assistent sein soll. Die Interpretation dieser Instruktion kann dazu führen, dass der Chatbot den Nutzer*innen nicht widerspricht, auch wenn dies angebracht wäre – ein Phänomen, das in der Forschung als

Aufgabe Modelle mit größerem Kontextfenster (Short-Term Memory) notwendig? Über welche Tools muss ein Chatbot verfügen, um für diese Aufgabe in Frage zu kommen?

- *Create:* Die operative Kollaborationsphase, in der die eigentliche Mensch-KI-Interaktion stattfindet. Alle Context-Engineering-Dimensionen kommen hier zum Tragen: User Prompts werden formuliert, RAG-Quellen eingebunden und Tools aktiviert. Die Phase ist dabei nicht auf das Verfassen von Prompts und das Aktivieren von Funktionen beschränkt. Sie umfasst auch alle weiteren Teilleistungen, die vor und nach dem Prompten auftreten: von der Erstellung eines eigenen Entwurfs über die Bereitstellung von Feedback bis hin zum Überarbeiten des KI-generierten Outputs.
- *Review:* Die kontinuierliche Evaluationsphase, die sich in Mikro-, Meso- und Makro-Reviews unterteilt. Mikro-Reviews finden nach jedem einzelnen Prompt statt – schnelle Plausibilitätschecks sowie direkte Anpassungen. Meso-Reviews finden nach dem Abschluss von Teilaufgaben statt, um den Erfolg des bisherigen Ansatzes zu überprüfen. Das Makro-Review am Ende bewertet das Gesamtergebnis systematisch nach epistemischen, qualitativen, ethischen und rechtlichen Kriterien.
- *Reflect:* Die abschließende Metaebene, auf der der gesamte Prozess, die Lernerfahrung und die Ergebnisqualität reflektiert werden. Diese Erkenntnisse fließen in zukünftige Durchläufe ein.

Abb. 2 verdeutlicht dabei die nicht-lineare Struktur: Create und Review bilden eine integrierte Einheit mit kontinuierlichen Mikro-Iterationen (dargestellt durch den Farb-Gradient und die zirkulären Pfeile). Die (gestrichelten) Rücksprünge zeigen die Makro-Iterationen – von Review zurück zu Plan bei grundlegenden Strategieänderungen. Diese Architektur spiegelt reale KI-Kollaborationsprozesse wider: Selten führt ein einzelner Prompt zum gewünschten Ergebnis. Stattdessen entwickelt sich die Lösung durch iterative

›Sycophancy‹ bezeichnet wird: die Tendenz von Sprachmodellen, Nutzenden übermäßig zuzustimmen oder zu schmeicheln, oft auf Kosten faktischer Korrektheit (Malmqvist, 2024). Nutzer*innen könnten dem entgegenwirken, indem sie eigene Instruktionen definieren, die das Modell zu kritischem Feedback anhalten.

Verfeinerung, wobei jede Phase unterschiedliche Context-Engineering-Dimensionen aktiviert.

Ein zentrales Element des PCRR-Frameworks ist die systematische Dokumentation aller Phasen. Diese umfasst:

- *Planungsentscheidungen:* Begründung der Tool-Wahl, Strategiedefinition, Zielformulierung …
- *Kollaborationsverlauf:* Verwendete Prompts, aktivierte Funktionen, Zwischenergebnisse …
- *Evaluationskriterien:* Angewandte Prüfschritte, identifizierte Probleme, Lösungsansätze …
- *Reflexionserkenntnisse:* Prozessbewertung, Lernerfahrungen, Optimierungspotenziale …

Die Dokumentationstiefe kann dabei an Erfahrungsstufe und Lernziel angepasst werden – von strukturierten Vorlagen für Einsteiger*innen bis zu offenen Reflexionsformaten für Fortgeschrittene. Diese Transparenz dient nicht nur der Leistungsbeurteilung, sondern fördert metakognitives Bewusstsein für den KI-Kollaborationsprozess.

Im Folgenden werden die vier Phasen detailliert ausgearbeitet.

3.2.1 Plan

Die Plan-Phase bildet das konzeptuelle Fundament jeder KI-gestützten Arbeitsaufgabe. Hier werden nicht nur Ziele definiert und Tools ausgewählt, sondern alle (relevanten) Dimensionen des Context Engineering strategisch vorbedacht. Diese Phase determiniert maßgeblich die Qualität des gesamten Kollaborationsprozesses.

Die folgenden Leitfragen strukturieren die Plan-Phase. Ihre konkrete Auswahl und Gewichtung variiert je nach Aufgabentyp, Fachbereich, Vorwissen und Lernziel. Für Einsteiger*innen empfiehlt sich die Bearbeitung der grundlegenden Fragen, während Fortgeschrittene gezielt relevante Aspekte vertiefen können.

Zieldefinition und Aufgabenanalyse: Das übergeordnete Ziel muss präzise formuliert und in operative Teilziele zerlegt werden. Dabei hilft die kontrastive Betrachtung: Wie würde ich ohne KI vorgehen? Welchen Mehrwert könnte KI bieten? Diese Gegenüberstellung schärft das Verständnis für sinnvolle Einsatzszenarien. In MINT-Kontexten könnte dies die Automatisierung von Berechnungen betreffen, in geisteswissenschaftlichen Kontexten eher die Erschließung großer Textkorpora.

Angemessenheitsprüfung des KI-Einsatzes: Nicht jede Aufgabe profitiert von KI-Unterstützung. Die technische Machbarkeit ist nur eine Dimension – rechtliche Rahmenbedingungen (z. B. Datenschutz und Urheberrecht), ethische Überlegungen (z. B. Diskriminierung und Fairness) und didaktische Sinnhaftigkeit müssen gleichwertig geprüft werden. Potenzielle Probleme sollten antizipiert und Gegenstrategien entwickelt werden. Beispiel: Bei der Analyse hochsensibler Daten muss vorab geklärt werden, ob lokale Alternativen zu Cloud-basierten Chatbots existieren.

Context-Engineering-Strategie: Alle sechs Dimensionen des Context Engineering manifestieren sich bereits in der Planungsphase:

- *User Prompt:* Welche Prompting-Techniken sind für diese Aufgabe besonders geeignet?
- *Systemprompt:* Sollten Custom Instructions für konsistente Outputs definiert werden? Haben die Systemprompts der Entwickler einen potenziellen Einfluss?
- *RAG:* Welche externen Quellen oder Dokumente müssen eingebunden werden?
- *Short-Term Memory:* Reicht das Kontextfenster für die geplante Aufgabe? Wie strukturiere ich die Sitzung(en)? Wann sind neue Chats sinnvoll?
- *Long-Term Memory:* Gibt es gespeicherte Präferenzen, die relevant sind?
- *Tools:* Welche Chatbots kommen in Frage? Welche spezifischen Funktionen werden benötigt?

Ausmaß und Art der KI-Kollaboration bzw. Sicherstellung der Eigenleistung: Je nach Aufgabe, eigener Kompetenz und KI-Kompetenz kann die Gestaltung der Mensch-KI-Zusammenarbeit unterschiedlich ausfallen. Dies betrifft sowohl die quantitative Zusammenarbeit (wie soll das Verhältnis von Eigenleistung und KI-Leistung ausfallen) als auch die qualitative Zusammenarbeit (welche Aufgaben soll eher der Mensch ausführen, welche können an die Maschine abgegeben werden und welche werden in einem Wechselspiel umgesetzt). Hier existieren bereits verschiedene konzeptuelle Rahmen, die als Inspiration dienen können. Drei davon seien hier genannt.

- Die *AI Assessment Scale* (AIAS) nach Perkins et al. (2024) definiert fünf Stufen der KI-Integration in Prüfungsformaten, von ›No AI‹ bis ›AI Exploration‹. Das Framework wurde seither weiterentwickelt und für verschiedene Kontexte adaptiert (Furze et al., 2024; Perkins et al., 2025a, 2025b; Roe et al., 2024, 2025). Für die Plan-Phase nimmt Level 2 (›AI-Assisted Ideation‹) eine besondere Relevanz ein, da bei fortgeschrittenen Lernenden auch die Planungsphase mit KI-Unterstützung durchgeführt werden kann: So kann die KI bei der Ideengenerierung, Tool-Auswahl und Herangehensweise unterstützen, ohne die intellektuelle Eigenleistung zu ersetzen.
- Dell'Acqua et al. (2023) unterscheiden in der KI-Nutzung zwischen *Centaur* (sequenzielle Arbeitsteilung zwischen Mensch und KI) und *Cyborg* (integrierte, verschränkte Kollaboration). Randazzo et al. (2025) erweitern diese Typologie um einen dritten Modus, den *Self-Automator*, bei dem Nutzende den gesamten Problemlösungsprozess an die KI delegieren. Die Studie zeigt, dass diese Modi unterschiedliche Kompetenzentwicklungen fördern: Centaurs vertiefen ihre Domänenexpertise, Cyborgs entwickeln vor allem KI-bezogene Fähigkeiten, während Self-Automators weder das eine noch das andere ausbauen. Für Bildungskontexte ergibt sich daraus ein Dilemma: Der Cyborg-Modus, der für erfahrene Nutzende häufig effektiver ist, setzt Fach-, Medien- und KI-Kompetenz voraus und lässt sich daher in klassischen Schulungsformaten nur schwer vermitteln. Lehrende greifen deshalb oft auf den Centaur-Modus zurück – weil er durch explizite Regeln lehr- und lernbar ist –, obwohl langfristig der Cyborg-Modus angestrebt werden sollte (Geyer,

2026). Diese Spannung zwischen Lehrbarkeit und Zielmodus ist bei der didaktischen Gestaltung von KI-Kollaborationen zu berücksichtigen.

- Schließlich kategorisieren Limburg und Buck (2024) KI-Nutzung nach ihrer Funktion für das menschliche Denken: *Ersatz* (KI übernimmt komplett), *Entlastung* (KI übernimmt *lower-order concerns*[6], damit sich die Nutzer*innen auf *higher-order concerns*[7] fokussieren können), *Unterstützung* (KI unterstützt die Lernenden, z. B. zur Behebung von Schreibblockaden) oder *Erweiterung* (KI ermöglicht neue Denkansätze). Diese Systematik unterstützt die bewusste Entscheidung über Art und Umfang der KI-Integration.

Weitere konzeptuelle Rahmen wie das *GPT-Modell* von Steinhoff & Lehnen (2025), welches drei Rollenkonstellationen zwischen KI und Mensch (*Ghost/Client*, *Partner/Explorer* und *Tutor/Learner*), ergänzen diese Perspektiven um die Frage der Rollenverteilung in der Mensch-KI-Interaktion.

Nicht alle Leitfragen sind für jede Aufgabe gleichermaßen relevant. Als Orientierung gilt: Die Zieldefinition und die Angemessenheitsprüfung des KI-Einsatzes bilden den unverzichtbaren Kern jeder Plan-Phase. Die Context-Engineering-Strategie gewinnt mit zunehmender Aufgabenkomplexität und KI-Erfahrung an Bedeutung – für Einsteiger*innen genügt zunächst die Frage nach dem geeigneten Tool und den benötigten Quellen. Die explizite Festlegung des Kollaborationsausmaßes (etwa nach AIAS-Stufen) ist vor allem in Prüfungskontexten erforderlich, kann bei formativen Übungen aber entfallen. Lehrende sollten für ihre spezifischen Lernsituationen

6 Bei *lower-order concerns* handelt es sich um grundlegende, meist bereits routinisierte kognitive Tätigkeiten mit geringem Abstraktionsgrad, z. B. Rechtschreibung, Grammatik, Formatierung, einfache Informationssuche oder Wiederholung. Diese erfordern wenig konzeptionelle oder reflektierende Verarbeitung von den Lernenden.

7 Bei *higher-order concerns* handelt es sich um komplexe, höherstufige kognitive Tätigkeiten mit hohem Abstraktions- und Reflexionsgrad, z. B. Analyse, Bewertung, Synthese, Argumentation, kreative Problemlösung und metakognitives Denken. Diese sind wesentlich für Verstehen, Lernen und Erkenntnisgewinn seitens der Lernenden.

entscheiden, welche Planungselemente vorgegeben, welche optional angeboten und welche bewusst weggelassen werden.

Um die Eigenleistung der Lernenden sicherzustellen, muss in der Plan-Phase dokumentiert werden, wie und wo genuin menschliche Beiträge in den Prozess einfließen. Dies betrifft neben dem finalen Produkt auch den gesamten Arbeitsprozess: Kritische Evaluation, kreative Problemlösung, fachliche Kontextualisierung und ethische Reflexion bleiben menschliche Domänen. Die Dokumentationstiefe orientiert sich am Kompetenzniveau und Lernziel:

- *Basisniveau:* Kernfragen mit Stichpunkten beantworten.
- *Fortgeschrittenes Niveau:* Ausgewählte Aspekte vertiefen, Entscheidungen begründen.
- *Expert*innenniveau:* Fokus auf kritische Entscheidungspunkte und Innovationen.

Die praktische Umsetzung der Plan-Phase kann verschiedene Formen annehmen:

Variante 1 – Analoge Planung: Für den Erwerb von Grundkompetenzen (sowohl fachbezogene als auch KI-Kompetenzen) oder bei sensiblen Aufgaben erfolgt die Planung ohne KI-Unterstützung. Die Dokumentation wird handschriftlich oder digital, aber ohne KI-Tools, erstellt. Dies sichert die vollständige Eigenleistung und fördert tiefes Aufgabenverständnis.

Variante 2 – KI-assistierte Planung: Entsprechend AIAS-Level 2 kann ein Chatbot bei der Planung unterstützen. Wie dies konkret aussehen kann, darf und/oder soll, orientiert sich an den Rahmenbedingungen und muss von der Lehrperson vorgegeben werden. Ein sehr simpler Prompt könnte folgendermaßen aussehen:

Beispiel-Prompt: Planungsphase für eine Hausarbeit

```
Ich plane eine Hausarbeit zum Thema [X]. Hilf mir bei
der Planungsphase:
Welche KI-Tools eignen sich für Literaturrecherche,
Gliederung und Schreibprozess?
Welche Funktionen sollte ich aktivieren/deaktivieren?
```

```
Wo sind menschliche Eigenleistungen unverzichtbar?
Welche ethischen/rechtlichen Aspekte muss ich beachten?
```

Das Gespräch mit der KI kann in diesem Fall auch als Chatexport in die Dokumentation integriert werden.

Die Plan-Phase generiert somit konkrete Vorgaben für die darauffolgenden Phasen: Create (welche Tools, welche Prompting-Strategien), Review (welche Prüfkriterien, welche Dokumentation) und Reflect (welche Aspekte besonders reflektiert werden sollen). Sie ist damit nicht nur Startpunkt, sondern Referenzrahmen für den gesamten PCRR-Zyklus. Sie macht den Denk-, Lern- und Arbeitsprozess transparent und damit bewertbar – unabhängig vom späteren Endprodukt.

3.2.2 Create

Die Create-Phase bildet das operative Zentrum des PCRR-Frameworks. Hier wird die in der Plan-Phase entwickelte Strategie in konkrete Mensch-KI-Interaktion umgesetzt. Diese Phase umfasst dabei nicht nur die direkte Arbeit mit dem KI-System, sondern auch alle vorbereitenden und nachbearbeitenden Tätigkeiten, die für eine erfolgreiche Kollaboration notwendig sind.

Die Create-Phase beginnt bereits vor der ersten Prompt-Eingabe – etwa wenn Lernende einen eigenen Textentwurf erstellen, der später als Grundlage für die KI-Bearbeitung dient, oder wenn sie relevante Quellen für den Upload zusammenstellen. Sie endet auch nicht mit der KI-generierten Antwort, sondern erstreckt sich auf die kritische Durchsicht, Überarbeitung und Integration der Ergebnisse in die eigene Arbeit.

Diese erweiterte Perspektive ist wichtig, weil sie die Create-Phase als genuinen Kollaborationsprozess begreift. Es handelt sich nicht um eine Abfolge isolierter Prompt-Response-Paare, sondern um einen iterativen Dialog, in dem beide Partner – Mensch und KI – spezifische Beiträge leisten. Der Mensch bringt Fachwissen, Kontextverständnis und kritisches Urteilsvermögen ein. Die KI bietet Verarbeitungskapazität, Mustererkennung und Zugriff auf breite Wissensbereiche.

Drei miteinander verbundene Dimensionen prägen die Create-Phase:

- *Prozessarchitektur* – wie die Kollaboration strukturiert wird: Diese Dimension umfasst alle Entscheidungen zur Gestaltung des Arbeitsprozesses: Welche Modelle und Funktionen werden aktiviert? Wie werden komplexe Aufgaben segmentiert? In welchen Iterationszyklen wird gearbeitet? Die Wahl zwischen *Reasoning*- oder Standard-Modellen, zwischen sequenzieller oder integrierter Bearbeitung (*Centaur vs. Cyborg*) prägt fundamental die Ergebnisqualität. Studierende, die beispielsweise ihre Bachelorarbeiten strukturieren, profitieren von einer mehrstufigen Herangehensweise: eigene Kernthesen entwickeln → mit KI diskutieren → Gegenargumente explorieren → spezifische Aspekte vertiefen. Dieser iterative Prozess führt zu substanzielleren Ergebnissen als der Versuch, direkt eine vollständige Gliederung generieren zu lassen.

- *Inhaltliche Fundierung* – was eingebracht wird: Die Substanz und Qualität der bereitgestellten Informationen determinieren die Güte der KI-Outputs. Diese Dimension umfasst (a) die eigenen Vorarbeiten (z. B. Braindumps, Skizzen oder Rohentwürfe), die der KI als Grundlage dienen, (b) Kontextinformationen zur jeweiligen Situation, (c) Bereitstellung zusätzlicher Inhalte (z. B. durch verlinkte Websites oder hochgeladene Dateien) sowie (d) die eigene Fachexpertise, die notwendig ist, um relevante Informationen auszuwählen, sinnvolle Fragen zu formulieren und später die Outputs kritisch zu evaluieren.

- *Sprachlich-technische Umsetzung* – wie formuliert wird: Diese Dimension betrifft die konkrete Ausgestaltung der Prompts. Präzision bedeutet hier nicht stilistische Perfektion, sondern unmissverständliche Kommunikation der Erwartungen. Zentrale Aspekte dabei sind (a) die sprachliche Formulierung der Prompts (z. B. in ganzen Sätzen, möglichst ohne grammatikalische Fehler), (b) die Strukturierung von Informationen (z. B. das Anordnen essenzieller Informationen am Anfang und Ende des Prompts oder die Verwendung von Delimitern zur Trennung verschiedener Prompt-Bestandteile) sowie (c) Explizitheit ohne Interpretationsspielraum (also die klare Benennung von Anforderungen, Detailtiefe, Format oder Umfang).

Diese drei Dimensionen operieren nicht isoliert, sondern beeinflussen sich wechselseitig: Die gewählte Prozessarchitektur bestimmt, welche Inhalte wann relevant werden. Die verfügbaren Inhalte beeinflussen die sprachliche Gestaltung. Die sprachlich-technischen Möglichkeiten wirken auf die Prozessgestaltung zurück. Erst ihr Zusammenspiel ermöglicht effektive KI-Kollaboration.

Alle sechs Dimensionen des Context Engineering werden in der Create-Phase praktisch relevant: Die User Prompts bilden die primäre Interaktionsform zwischen Nutzenden und Chatbots. Der Systemprompt kann vorab über Custom Instructions definiert werden, um konsistente Outputs über mehrere Interaktionen zu gewährleisten. RAG erweitert kontinuierlich die verfügbare Informationsbasis durch hochgeladene Dokumente oder referenzierte Webseiten. Das Short-Term Memory muss bewusst verwaltet werden – irrelevante Informationen können das Kontextfenster belasten und die Outputqualität mindern; strategische Entscheidungen über Sitzungsstrukturen und Informationsfluss sind hier essenziell. Long-Term Memory kann bei wiederkehrenden Aufgaben relevante Präferenzen bereitstellen, einschließlich nutzerseitig definierter Stilpräferenzen. Die Tool-Dimension wird durch gezielte Aktivierung spezifischer Funktionen (wie Bildgenerierung, Deep Research, Canvas oder Code-Interpreter) genutzt.

Die Herausforderung für die Praxis besteht darin, entsprechende Prompting-Techniken zu wählen. Die Literatur zählt hier bereits dutzende unterschiedliche Techniken (Schulhoff et al., 2025). Da das PCRR-Framework universal eingesetzt werden soll und die Mensch-Maschine-Kollaboration sehr individuell ausfallen kann, gibt es keine pauschalen Anleitungen für die Create-Phase. Idealerweise wird diese an unterschiedliche Disziplinen, Institutionen oder Lernsituationen angepasst und die Lernenden in die entsprechende Nutzung eingeschult. Dabei sollte effektives Scaffolding betrieben werden, um die Umsetzung von den Lernenden erwarten zu können. White et al. (2023) haben sechs verschiedene *Prompting Patterns* identifiziert, aus denen man für den eigenen Bereich spezifische Prompting-Techniken ableiten kann: *Input Semantics, Output Customization, Error Identification, Prompt Improvement, Interaction* und *Context Control*.

Trotz der Kontextabhängigkeit jeder konkreten Anwendung haben wir in der eigenen Lehr- und Weiterbildungspraxis sieben Prinzipien identifiziert, die sich über verschiedene Bildungskontexte und Aufgabentypen hinweg als durchgängig wertvoll erweisen. Die Auswahl orientiert sich dabei bewusst an der Vermittelbarkeit: Während viele der in der Literatur dokumentierten Techniken ein hohes Maß an technischem Verständnis voraussetzen und sich in klassischen Lehr-Lern-Settings nur schwer einführen lassen, sind die folgenden Prinzipien für Lernende leicht nachvollziehbar und unmittelbar umsetzbar. Sie konvergieren mit den oben genannten Systematisierungen, priorisieren jedoch didaktische Zugänglichkeit: sprachliche Präzision, Kontextualisierung, beispielbasiertes Arbeiten, externe Ressourcenintegration, dialogische Optimierung, Fehlertoleranz und Aufgabensegmentierung.

Sprachliche Präzision: Vollständige Sätze mit eindeutigen Formulierungen reduzieren den Interpretationsspielraum. Fachbegriffe sollten konsistent verwendet, Abkürzungen ausgeschrieben oder erklärt werden. Die Investition in klare Formulierungen zahlt sich durch bessere Outputs und weniger Nacharbeit aus.

Kontextualisierung: Die systematische Bereitstellung relevanter Kontextinformationen verbessert die Angemessenheit der Outputs signifikant. Dabei könnten sich die Lernenden an den W-Fragen (Wer, Was, Wen, Wann, Wo, Wie, Warum, Wozu) orientieren. Implizite Annahmen müssen explizit gemacht und Rahmenbedingungen und Einschränkungen klar benannt werden, um keinen unerwünschten Interpretationsspielraum zuzulassen.

Beispielbasiertes Arbeiten (Few-Shot Prompting): Wenn ein Sprachmodell konkrete Strukturen, Stile oder Formate übernehmen soll, sollte dies durch konkrete Beispiele vorgegeben werden, anstatt abstrakte Beschreibungen zu verwenden. Dies gilt vor allem für standardisierte Aufgaben, domänenspezifische Konventionen oder persönliche Kommunikationsweisen. Um dem Verdacht der KI-Nutzung zu entgehen, könnten Lernende beispielsweise folgendermaßen vorgehen: ›Im Anhang findest du drei meiner vergangenen Englischhausübungen. Erzeuge nun einen Text für den hochgeladenen Arbeitsauftrag, in welchem du mein Sprachniveau, meinen Schreibstil und meine typischen Fehler übernimmst.‹

Externe Ressourcenintegration (RAG): Die Einbindung relevanter Dokumente oder Webseiten erweitert die Informationsbasis erheblich. Die Verfügbarkeit und korrekte Verarbeitung sollte jedoch verifiziert werden, da unter anderem technische Beschränkungen (z. B. eine robots.txt-Datei, die den Zugriff auf Websites durch Webcrawler regelt) den Zugriff verhindern können.

Dialogische Optimierung: Erfolgreiche KI-Kollaboration ist iterativ. Rückfragen sollten eingefordert und beantwortet, Zwischenergebnisse evaluiert und Verbesserungswünsche kommuniziert werden. Dieser dialogische Prozess führt zu qualitativ besseren Ergebnissen als der Versuch, alle Anforderungen in einem einzelnen Prompt zu spezifizieren. Chatbots können auch als Prompt Engineers fungieren, welche effektive Prompts erstellen oder vorgegebene Prompts optimieren, bevor sie eingesetzt werden.

Fehlertoleranz: Aktuelle Chatbot-Systeme haben die Tendenz, schnell plausibel klingende Antworten zu erzeugen, weil sie Nutzer*innen mit wenig Aufwand zufriedenstellen wollen. Forscher*innen von OpenAI schreiben dazu:

> language models hallucinate because the training and evaluation procedures reward guessing over acknowledging uncertainty [...]. [H]allucinations persist due to the way most evaluations are graded – language models are optimized to be good test-takers, and guessing when uncertain improves test performance. (Kalai et al., 2025)

Diese Voreiligkeit führt häufig zu vermeidbaren Halluzinationen. Eine Möglichkeit, dadurch entstehende Fehlinformationen zu verringern, besteht darin, den Chatbot explizit dazu aufzufordern, Unsicherheit zu äußern, z. B. durch einen Prompt wie ›Markiere unsichere Informationen in deiner Ausgabe, für die du keinen Beleg findest‹ oder ›Falls du auf den Link aufgrund einer robots.txt-Datei nicht zugreifen kannst, sag mir das ehrlich, anstatt Informationen zu erfinden.‹ Diese Technik ist auch als *Escape Hatch* bekannt, weil sie der KI einen Fluchtweg bietet. Aktuelle Modelle integrieren jedoch zunehmend Mechanismen zur Unsicherheitskommunikation in ihre Systemprompts, was eine entsprechende Ergänzung im Nutzer*innen-Prompt mittelfristig obsolet machen könnte. Aktuell kann eine explizite Aufforderung zur Markierung unsicherer Informationen in bestimmten Kontexten – etwa

bei faktenintensiven Aufgaben – dennoch weiterhin die Outputqualität verbessern.

Aufgabensegmentierung (Prompt Chaining): Komplexe Aufgaben sollten in validierbare Teilschritte zerlegt werden, deren Ergebnisse aufeinander aufbauen können. Diese Segmentierung ermöglicht auch bessere Qualitätskontrolle: Hat sich im ersten Zwischenergebnis ein Fehler eingeschlichen, kann dieser bereinigt werden, bevor zum zweiten Schritt übergegangen wird.

Daneben gibt es viele weitere Techniken, die je nach Modell, Eingabe und Aufgabenart relevant sein könnten, wie die *Chain-of-Thought-Technik* (z. B. indem man ›Think step by step‹ am Ende des Prompts ergänzt), um die Outputqualität von Non-Reasoning-Modellen zu erhöhen, oder das Verwenden von sogenannten *Delimitern* (Sonderzeichen wie Hashtags oder XML-Tags), um Instruktion und Inhalt oder unterschiedliche Inhalte im selben Prompt zu trennen.

Für die vertiefte Auseinandersetzung mit spezifischen Prompting-Techniken verweisen wir auf etablierte Systematisierungen wie den Prompt Report (Schulhoff et al., 2025), die empirisch fundierten Prompting Science Reports von Meincke et al. (2025a, 2025b, 2025c) sowie Basil et al. (2025), die Übersichten von White et al. (2023), B. Chen et al. (2025), Singh et al. (2024) und Vatsal und Dubey (2024) und die Prompting Guides der einzelnen Sprachmodell-Entwickler (Anthropic, 2025a; Google, 2025; OpenAI, 2026a).

Wie auch die Plan-Phase sollte die Create-Phase dokumentiert werden, um den Prozess transparent und bewertbar zu gestalten. Art und Umfang der Dokumentation orientieren sich am Verwendungszweck und müssen für unterschiedliche Kontexte adaptiert werden: Für Übungsaufgaben genügt oft die Speicherung der Hauptprompts und des finalen Outputs. Für benotete Arbeiten sollte der Prozess nachvollziehbar dokumentiert werden – verwendete Prompts mit Begründungen, wichtige Iterationsschritte, menschliche Vor- und Nachbearbeitung. Für Forschungskontexte könnte eine vollständige Dokumentation inklusive aller Zwischenschritte und verworfener Ansätze angebracht sein. Die Dokumentation dient nicht nur der Transparenz und Bewertbarkeit, sondern unterstützt auch die Reflexion und kontinuierliche Verbesserung der eigenen KI-Kollaborationskompetenz.

Der Übergang zur Review-Phase ist dabei kein klarer Schritt. Vielmehr ist die Create-Phase über kontinuierliche Evaluationsprozesse eng mit der Review-Phase verzahnt.

3.2.3 Review

Die Review-Phase erfüllt mehrere kritische Funktionen im PCRR-Framework. Sie sichert die fachliche Korrektheit der Ergebnisse, überprüft die Zielerreichung, validiert den menschlichen Eigenanteil und gewährleistet die ethische und rechtliche Unbedenklichkeit der generierten Inhalte. Darüber hinaus entwickeln Lernende durch systematisches Review kritische Evaluationskompetenzen, die für verantwortungsvolle KI-Nutzung unerlässlich sind.

Sie ist jedoch keine eigenständige Phase oder singuläre Aktivität, sondern ein kontinuierlicher Evaluationsprozess, der die gesamte KI-Kollaboration begleitet. Sie ist somit mit der Create-Phase verzahnt und manifestiert sich auf drei unterschiedlichen Ebenen – Mikro, Meso und Makro –, die jeweils spezifische Funktionen erfüllen und unterschiedliche Konsequenzen nach sich ziehen.

Die strukturierte Anwendung der drei Review-Ebenen fördert die Entwicklung kritischer Evaluationskompetenz. Anfänger*innen beginnen beispielsweise mit expliziten Checklisten und entwickeln durch wiederholte Anwendung intuitive Evaluationsmuster. Dieser Lernprozess ist essenziell für verantwortungsvolle KI-Nutzung.

Lehrende sollten diesen Kompetenzaufbau durch differenziertes Feedback unterstützen. Die Qualität der Reviews – nicht nur die der Endergebnisse – sollte in die Bewertung einfließen. Besonders die Fähigkeit, in Meso-Reviews strategische Entscheidungen zu treffen und zu begründen, zeigt fortgeschrittene KI-Kollaborationskompetenz.

3.2.4 Review-Dimensionen

Unabhängig von der Review-Ebene orientiert sich die Evaluation an fünf zentralen Prüfdimensionen (siehe Abb. 3). Diese bilden das analytische

Raster, das auf Mikro-, Meso- und Makro-Ebene mit unterschiedlicher Intensität und Systematik angewendet wird.

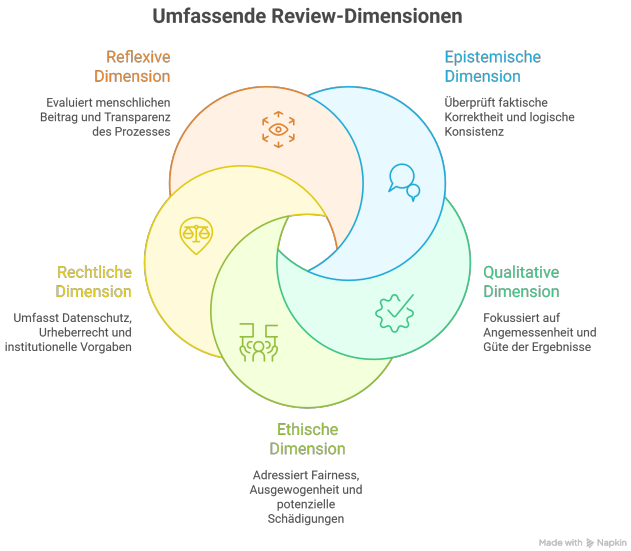

Abbildung 3: Die Review-Dimensionen. Die Grafik wurde mit Napkin AI erstellt.

Die *epistemische Dimension* prüft die faktische Korrektheit und logische Konsistenz der Inhalte. Stimmen die Aussagen? Sind Argumente schlüssig? Funktionieren bereitgestellte Links? Werden Quellen korrekt wiedergegeben? Besondere Aufmerksamkeit gilt möglichen Halluzinationen – plausibel klingenden, aber falschen Behauptungen. Bei Mikro-Reviews genügt oft eine Plausibilitätsprüfung, während Meso- und Makro-Reviews systematische Verifikation erfordern.

Die *qualitative Dimension* fokussiert die Angemessenheit und Güte der Ergebnisse. Entspricht der Output den gesetzten Standards? Passt die Darstellung zu Kontext und Zielgruppe? Wurden die definierten Ziele erreicht? Diese Dimension gewinnt besonders bei Meso-Reviews an Bedeutung, wenn Teilziele evaluiert werden, und kulminiert im Makro-Review mit der Frage, ob das Gesamtwerk seinen intendierten Zweck erfüllt.

Die *ethische Dimension* adressiert Fairness, Ausgewogenheit und potenzielle Schädigungen. Werden Stereotype reproduziert? Sind marginalisierte Perspektiven angemessen repräsentiert? Könnte die Verwendung der Inhalte problematisch sein? Diese Dimension erfordert besondere Sensibilität, kritische Reflexion und oft einen Perspektivwechsel: Wie würden Betroffene diese Darstellung wahrnehmen?

Die *rechtliche Dimension* umfasst Datenschutz, Urheberrecht und institutionelle Vorgaben zur KI-Nutzung. Wurden geschützte Inhalte reproduziert? Sind Persönlichkeitsrechte gewahrt? Entspricht die KI-Nutzung den geltenden Richtlinien? Verstöße in dieser Dimension können gravierende Konsequenzen haben und sollten daher frühzeitig identifiziert werden.

Die *reflexive Dimension* evaluiert den menschlichen Beitrag und die Transparenz des Prozesses. Ist der menschliche Eigenanteil substanziell und erkennbar? Wurde kritisch mit KI-Outputs umgegangen? Ist der Arbeitsprozess nachvollziehbar dokumentiert? Diese Dimension bereitet den Übergang zur Reflect-Phase vor und stellt sicher, dass das Produkt einer kritischen Überprüfung auf akademische Integrität standhalten würde.

Die Intensität der Prüfung variiert je nach Review-Ebene: Mikro-Reviews setzen auf schnelle Plausibilitätschecks, Meso-Reviews vertiefen die Prüfung bei strategisch wichtigen Teilprodukten, und Makro-Reviews durchlaufen alle Dimensionen methodisch und vollständig.

Mikro-Reviews

Mikro-Reviews finden nach jedem KI-Output statt und dienen der unmittelbaren Qualitätskontrolle und Prozesssteuerung. Sie sind in der Regel kurz und fokussiert, müssen nicht vollständig dokumentiert werden, sollten aber bei kritischen Entscheidungen festgehalten werden. Diese kontinuierlichen Checks ermöglichen agile Anpassungen während des Arbeitsprozesses und verhindern, dass sich Fehler durch mehrere Iterationen ziehen.

Nach jeder KI-Ausgabe erfolgt eine kurze Evaluation. Für Einsteiger*innen empfiehlt sich anfangs eine strukturierte Herangehensweise:

- Plausibilität: Klingt die Antwort schlüssig?
- Relevanz: Adressiert sie die gestellte Frage?
- Qualität: Inwieweit entspricht sie den Erwartungen?
- Aktion: Weitermachen, nachfragen oder neu ansetzen?

Dokumentiert werden dabei nur kritische Punkte: überraschende Erkenntnisse, identifizierte Probleme oder Strategiewechsel.

Die Intensität der Mikro-Reviews variiert situationsabhängig. Ein generierter Übungssatz in einer Fremdsprache erfordert eine andere Prüftiefe als statistische Auswertungen für eine empirische Arbeit. Entscheidend ist, dass Lernende ein Gespür dafür entwickeln, wann oberflächliche Plausibilitätsprüfung ausreicht und wann tiefere Evaluation notwendig ist.

Meso-Reviews

Meso-Reviews markieren den Abschluss von Teilaufgaben oder Arbeitsabschnitten. Nach Fertigstellung eines Kapitels, nach Abschluss einer Datenanalyse oder nach Erstellung eines Konzeptentwurfs erfolgt eine umfassendere Evaluation. Diese mittlere Review-Ebene prüft, ob die bisherige Strategie funktioniert oder ob fundamentale Anpassungen notwendig sind – bis hin zur Rückkehr in die Plan-Phase. Meso-Reviews sind Entscheidungspunkte im Arbeitsprozess. Eine beispielhafte Überprüfung könnte folgendermaßen aussehen:

- Kohärenzprüfung: Fügen sich die Teilergebnisse zu einem stimmigen Ganzen? Gibt es Widersprüche zwischen verschiedenen Abschnitten? Ist ein roter Faden erkennbar?
- Zielerreichung: Wurden die für diesen Abschnitt gesetzten Teilziele erreicht? Entspricht die Qualität den Anforderungen? Sind Anpassungen für die weiteren Abschnitte notwendig?
- Strategieevaluation: Funktioniert der gewählte Ansatz? Sind die genutzten Tools und Prompting-Strategien effektiv? Sollten grundlegende Änderungen vorgenommen werden?

- Entscheidungspunkt: Basierend auf dem Meso-Review wird entschieden: Fortfahren mit der nächsten Teilaufgabe, Überarbeitung des aktuellen Abschnitts oder Rückkehr zur Plan-Phase für grundlegende Neukonzeption.

Ein Beispiel: Nach Fertigstellung des Literaturteils einer Seminararbeit zeigt das Meso-Review, dass die KI primär englischsprachige Quellen einbezogen hat, obwohl deutsche Literatur zentral wäre. Dies macht eine Rückkehr zur Plan-Phase sowie eine Anpassung der Tool-Auswahl (z. B. mit Zugang zu deutschsprachigen Datenbanken) und der Prompting-Strategie erforderlich.

Makro-Reviews

Das Makro-Review markiert den Abschluss der KI-Kollaboration und unterscheidet sich fundamental von den prozessbegleitenden Reviews. Während Mikro- und Meso-Reviews noch Kurskorrekturen ermöglichen, ist das Makro-Review eine finale Evaluation des Endergebnisses. Es erfüllt drei zentrale Funktionen: die systematische Qualitätssicherung des Endergebnisses, die Dokumentation der Create- und Review-Phasen für die Bewertung und die Vorbereitung der Reflect-Phase durch eine abschließende Bestandsaufnahme.

Im Makro-Review werden alle fünf Prüfdimensionen methodisch und vollständig durchlaufen. Der entscheidende Unterschied zu den vorgelagerten Review-Ebenen liegt in der Systematik und im Bezugsrahmen: Nicht einzelne Outputs oder Teilprodukte werden geprüft, sondern das Gesamtwerk. Dies ermöglicht das Erkennen von Mustern, die in Einzelbetrachtungen nicht sichtbar werden – etwa ein durchgängiger Bias, wiederkehrende Argumentationslücken oder Halluzinationen, die sich durch mehrere Iterationen gezogen haben.

Für die Durchführung empfiehlt sich ein strukturiertes Vorgehen: Zunächst wird das komplette Ergebnis zusammenhängend gelesen oder betrachtet, ohne bereits zu evaluieren – dies ermöglicht einen Gesamteindruck. Dann wird jede der fünf Dimensionen systematisch abgearbeitet und Befunde werden dokumentiert. In der Synthese werden die Einzelbefunde zu einer Gesamtbewertung zusammengeführt, Stärken und Schwächen identifiziert.

Abschließend wird ein strukturierter Makro-Review-Bericht erstellt, der alle Prüfergebnisse, die Gesamtbewertung und offene Punkte für die Reflect-Phase festhält.

Diese Dokumentation ermöglicht faire Bewertung durch Lehrende, liefert die Faktenbasis für die Reflect-Phase und dokumentiert die erreichte Produktqualität. Das Makro-Review ist somit eine abschließende Produktkontrolle, die die Grundlage für die prozessorientierte Reflect-Phase schafft.

3.2.5 Reflect

Die Reflect-Phase markiert den Abschluss des PCRR-Zyklus und unterscheidet sich fundamental von der vorangegangenen Review-Phase. Während das Review das Produkt und seine Qualität evaluiert, richtet die Reflect-Phase den Blick auf den Prozess, die Lernerfahrung und den Kompetenzgewinn. Die KI-Kollaboration ist beendet, das Ergebnis liegt vor – nun geht es darum, aus der Erfahrung zu lernen.

Die Reflect-Phase erfüllt mehrere wichtige Funktionen im Lernprozess. Sie fördert metakognitive Kompetenzen durch die bewusste Auseinandersetzung mit dem eigenen Arbeits- und Lernprozess. Sie identifiziert erfolgreiche Strategien und Verbesserungspotenziale für zukünftige Projekte. Sie macht implizites Prozesswissen explizit und damit transferierbar. Und sie dokumentiert den individuellen Lernfortschritt in der KI-Kollaboration.

Dabei verschiebt sich der Fokus von klassischem Feedback – das retrospektiv fragt: ›Wie habe ich diese Aufgabe gemeistert?‹ – hin zu Feedforward, das prospektiv fragt: ›Wie kann ich die nächste Aufgabe besser meistern?‹ (vgl. Höfler, 2025). Diese Zukunftsorientierung ist zentral für dynamische Lernprozesse und wird im PCRR-Framework durch die zyklische Verbindung der Reflect-Phase mit einer neuen Plan-Phase strukturell verankert.

Die Reflect-Phase adressiert drei miteinander verbundene Reflexionsebenen:

- *Prozessreflexion:* Wie verlief die Kollaboration konkret? Diese Ebene fokussiert auf den Ablauf, die gewählten Strategien und deren Effektivität. Welche Phasen des PCRR-Frameworks waren besonders

herausfordernd oder erhellend? Wie gut funktionierte das Zusammenspiel der verschiedenen Context-Engineering-Dimensionen?

- *Erfahrungsreflexion:* Was wurde subjektiv erlebt? Hier geht es um die persönliche Dimension – Frustrationen und Erfolgserlebnisse, Überraschungen und Bestätigungen, Momente der Überforderung und des Flow. Diese emotionale und motivationale Ebene ist wichtig, weil sie zukünftige Einstellungen zur KI-Kollaboration prägt.
- *Kompetenzreflexion:* Was wurde gelernt? Diese Ebene identifiziert den konkreten Kompetenzzuwachs – sowohl in Bezug auf KI-Kollaboration als auch fachlich. Welche neuen Prompting-Techniken wurden erworben? Welches Verständnis für Context Engineering hat sich entwickelt? Aber auch: Welche fachlichen Einsichten entstanden durch die intensive Auseinandersetzung mit dem Thema?

Die Reflect-Phase sollte systematisch angegangen werden, um oberflächliche Allgemeinplätze zu vermeiden. Folgende Leitfragen strukturieren die Reflexion:

Rückblick auf die Plan-Phase: War die initiale Planung realistisch und vollständig? Welche unvorhergesehenen Herausforderungen traten auf? Wie gut funktionierte die Tool-Auswahl und Strategiedefinition? Was würde ich bei der nächsten Planung anders machen?

Analyse der Create-Phase: Welche Prompting-Strategien waren besonders effektiv? Wie entwickelte sich die Interaktion über die Zeit? Wo war die Balance zwischen KI-Nutzung und Eigenleistung optimal? Welche Context-Engineering-Dimensionen waren kritisch für den Erfolg?

Evaluation der Review-Prozesse: Haben die Mikro-Reviews rechtzeitig Probleme identifiziert? Waren die Meso-Reviews hilfreich für strategische Entscheidungen? Hat das Makro-Review alle relevanten Aspekte erfasst? Wie könnte das Review-System optimiert werden?

Gesamtbetrachtung: Entsprach der Aufwand dem Nutzen? War das PCRR-Framework hilfreich oder hinderlich? Wie hat sich mein Verständnis von KI-Kollaboration entwickelt? Was war die wichtigste Erkenntnis aus diesem Projekt?

Ein zentraler Bestandteil der Reflect-Phase ist die explizite Formulierung von *Lessons Learned*:

- *Methodische Erkenntnisse:* Welche Arbeitsweisen haben sich bewährt? Beispiel: ›Die Zerlegung komplexer Aufgaben in Teilschritte mit jeweiligen Meso-Reviews hat die Qualität deutlich verbessert.‹
- *Technische Einsichten:* Welche Tools oder Funktionen waren besonders nützlich? Beispiel: ›Der Code-Interpreter eignet sich hervorragend für Datenvisualisierung, stößt aber bei komplexeren statistischen Analysen an Grenzen.‹
- *Persönliche Entwicklung:* Wie hat sich die eigene Kompetenz entwickelt? Beispiel: ›Meine Fähigkeit, präzise Prompts zu formulieren, hat sich deutlich verbessert, besonders bei der Kontextualisierung.‹
- *Warnungen und Fallstricke:* Was sollte zukünftig vermieden werden? Beispiel: ›Zu lange Create-Sitzungen ohne Meso-Review führen zu inkonsistenten Ergebnissen.‹

KI-unterstützte Reflexion

Nach der eigenständigen Reflexion kann optional eine KI zur Vertiefung eingesetzt werden. Dies sollte jedoch klar als Ergänzung, nicht als Ersatz der eigenen Reflexion gekennzeichnet sein:

1. Eigene Reflexion verfassen (nach Leitfragen oder frei).
2. KI-Feedback einholen durch gezieltes Prompting: ›Ich habe folgenden Reflexionsbericht über meine KI-Kollaboration verfasst: [Reflexion]. Welche blinden Flecken siehst du? Welche Aspekte könnte ich vertiefen? Gibt es Widersprüche in meiner Darstellung?‹
3. Meta-Reflexion des KI-Feedbacks: Kritische Auseinandersetzung mit den KI-Anmerkungen.

Diese dreistufige Struktur nutzt die KI als Reflexionspartner, bewahrt aber die Authentizität der persönlichen Reflexion.

Dokumentationsformen

Die Dokumentation der Reflect-Phase kann verschiedene Formen annehmen, abhängig von Kontext und Kompetenzstand:

- *Strukturierter Reflexionsbericht:* Ein ausformulierter Text, der systematisch die verschiedenen Reflexionsebenen abarbeitet. Diese Form eignet sich besonders für benotete Arbeiten und fördert tiefe Auseinandersetzung.
- *Portfolio-Eintrag:* Eine kürzere, fokussierte Reflexion, die Teil eines größeren Lernportfolios wird. Hier steht der persönliche Lernfortschritt im Vordergrund.
- *Peer-Reflexion:* Austausch mit anderen Lernenden über die jeweiligen Erfahrungen. Dies kann schriftlich oder mündlich erfolgen und bietet zusätzliche Perspektiven.

Die Qualität der Reflexion bemisst sich nicht an der Länge, sondern an Tiefe, Ehrlichkeit und Konkretheit. Eine prägnante, selbstkritische Reflexion mit klaren *Lessons Learned* ist wertvoller als ausschweifende Allgemeinplätze.

Die Reflect-Phase schließt den aktuellen PCRR-Zyklus ab, bildet aber gleichzeitig die Brücke zum nächsten. Die gewonnenen Erkenntnisse fließen in die Plan-Phase zukünftiger Projekte ein. So entsteht eine Aufwärtsspirale kontinuierlicher Kompetenzentwicklung.

Die systematische Reflexion macht aus einzelnen KI-Kollaborationserfahrungen übertragbares Wissen. Sie transformiert *Trial-and-Error* in bewusste Kompetenzentwicklung. Und sie fördert die kritische Distanz, die für verantwortungsvolle KI-Nutzung unerlässlich ist.

4 Differenzierung

Das PCRR-Framework wurde in den vorangegangenen Abschnitten in seiner vollständigen theoretischen Tiefe dargestellt. Diese Vollständigkeit ist beabsichtigt: Als Referenzwerk soll es alle relevanten Aspekte der KI-Kollaboration abdecken, sodass Forschende und Praktizierende je nach Bedarf darauf zugreifen können. Für die praktische Anwendung bedeutet dies jedoch, dass das Framework an spezifische Kontexte angepasst werden muss. Eine unreflektierte Eins-zu-eins-Übertragung – etwa der vollständigen Plan-Phase mit allen Context-Engineering-Dimensionen auf eine erste Unterrichtseinheit mit KI – würde Lernende wie Lehrende überfordern.

Bereits 2024 und 2025 wurden von den Autor*innen erste Praxisversuche durchgeführt (Freinhofer et al., 2025) und praxisrelevante Überlegungen angestellt (Schwabl & Vötsch, 2025). Aktuell führen wir eine Aktionsforschung mit Lehramtsstudierenden durch, die zu einem späteren Zeitpunkt (nach einer zweiten Erhebungswelle) veröffentlicht wird. Diese Erprobungen bestätigten einerseits das Potenzial des PCRR-Ansatzes, offenbarten aber auch zwei Entwicklungsbedarfe: Zum einen fehlte Version 1 ein hinreichend fundiertes theoretisches Fundament – diesem Bedarf begegnet die hier vorgestellte Version 2 durch die Integration des Context-Engineering-Rahmens und die systematische Ausarbeitung der Phasen. Zum anderen zeigte sich, dass der direkte Unterrichtseinsatz klarere Handreichungen, praxisnähere Anleitungen und einfachere Ableitungen erfordert – diese Operationalisierung ist Gegenstand laufender und geplanter Folgepublikationen.

4.1 Differenzierungsvariablen

Die Anpassung des PCRR-Frameworks kann entlang mehrerer Variablen erfolgen, die sich teilweise überschneiden und wechselseitig beeinflussen. Zu den zentralen Variablen zählen das Alter der Lernenden, das unter anderem über rechtliche Zugangsbeschränkungen zu KI-Tools entscheidet; die Disziplin bzw. der Fachbereich, der spezifische Anforderungen an Tools, Output-Formate und Prüfkriterien stellt; die Fachkompetenz der Lernenden, die bestimmt, wie viel KI-Unterstützung sinnvoll und vertretbar ist; die KI-

Kompetenz der Lernenden, die das erforderliche Maß an *Scaffolding* und expliziter Anleitung beeinflusst; die KI-Kompetenz der Lehrenden, die die Lernenden unterschiedlich stark vorbereiten, unterstützen und fördern können; die bisherige Erfahrung mit dem PCRR-Framework selbst, die eine schrittweise Komplexitätssteigerung ermöglicht; sowie die institutionellen Rahmenbedingungen, die oft den praktischen Möglichkeitsraum definieren.

Darüber hinaus spielen aufgabenbezogene Faktoren eine Rolle: Die Komplexität und der Typ der Aufgabe – ob Brainstorming, Textproduktion oder Datenanalyse – erfordern unterschiedliche PCRR-Tiefen. Der Bewertungskontext unterscheidet zwischen formativem Üben, bei dem Fehler produktiv genutzt werden können, und summativen Prüfungen, die eine vollständige Dokumentation erfordern. Schließlich variiert der Autonomiegrad: Wie viel Struktur wird durch Lehrende vorgegeben, wie viel gestalten Lernende selbst?

Eine erschöpfende Behandlung aller Variablen würde den Rahmen dieser Publikation sprengen. Im Folgenden werden daher drei Kernvariablen vertieft, die für die praktische Anpassung besonders relevant sind: KI-Kompetenz, Fachkompetenz und institutionelle Rahmenbedingungen. Diese Auswahl bildet sowohl individuelle Voraussetzungen der Lernenden als auch kontextuelle Faktoren ab.

4.2 KI-Kompetenz

Die KI-Kompetenz der Lernenden bestimmt maßgeblich, wie explizit das PCRR-Framework angeleitet werden muss. Für Lernende ohne Vorerfahrung empfiehlt sich ein stark strukturierter Einstieg: Die Plan-Phase kann zunächst auf die Kernfragen der Zieldefinition und Tool-Auswahl reduziert werden, während komplexere Aspekte wie Context Management oder die strategische Nutzung von Systemprompts erst später eingeführt werden. In der Create-Phase sollten anfangs wenige, klar definierte Prompting-Prinzipien vermittelt werden – etwa sprachliche Präzision und Kontextualisierung –, bevor fortgeschrittene Techniken wie *Few-Shot-Prompting* oder *Prompt Chaining* hinzukommen. Die Review-Phase kann zunächst auf Mikro-Reviews mit einfachen Checklisten beschränkt werden; die differenzierte Unterscheidung zwischen Mikro-, Meso- und Makro-Reviews setzt ein

Verständnis iterativer Arbeitsprozesse voraus, das sich erst mit Erfahrung entwickelt.

Mit wachsender KI-Kompetenz verschiebt sich der Fokus: Fortgeschrittene Lernende benötigen weniger explizite Anleitung für einzelne Phasen, profitieren aber von der Reflexion komplexerer Zusammenhänge – etwa der Frage, wie verschiedene Context-Engineering-Dimensionen zusammenwirken oder wann ein Meso-Review einen Rücksprung in die Plan-Phase nahelegt. Die Dokumentationsanforderungen können parallel angepasst werden: Während Anfänger*innen von strukturierten Vorlagen profitieren, die jeden Schritt explizit abfragen, können Fortgeschrittene zu offeneren Reflexionsformaten übergehen.

4.3 Fachkompetenz

Die Fachkompetenz der Lernenden beeinflusst, welche Art und welches Ausmaß der KI-Nutzung didaktisch sinnvoll sind. Dieses Prinzip wurde bereits in der Plan-Phase unter dem Aspekt ›Sicherstellung der Eigenleistung‹ angesprochen: Je weniger Fachkompetenz vorhanden ist, desto wichtiger ist es, dass Lernende zunächst eigene Grundlagen entwickeln, bevor KI-Unterstützung hinzukommt.

Limburg und Buck (2024) unterscheiden zwischen KI als *Ersatz, Entlastung, Unterstützung* und *Erweiterung* menschlichen Denkens. Für Noviz*innen in einem Fachgebiet ist der Ersatz-Modus problematisch: Wer die Grundlagen eines Themas noch nicht beherrscht, kann KI-generierte Outputs nicht angemessen evaluieren – die epistemische Dimension der Review-Phase setzt Fachwissen voraus. Auch Entlastung durch Delegation von Routineaufgaben greift hier zu kurz, da Noviz*innen diese Routinen erst noch entwickeln müssen. Für sie ist daher der Unterstützungs-Modus zentral: KI dient dazu, Verständnisfragen zu klären, Feedback auf eigene Entwürfe zu erhalten oder Denkprozesse zu begleiten – ohne das eigene Denken zu ersetzen. Mit zunehmender Fachkompetenz werden Entlastung und Erweiterung (Exploration neuer Perspektiven) sinnvoller.

Für die Praxis bedeutet dies: Bei fachlich unerfahrenen Lernenden sollte die Create-Phase stärker auf eigene Vorarbeiten setzen – der Rohentwurf vor

dem Prompting gewinnt an Bedeutung. Die Review-Phase erfordert möglicherweise zusätzliche Unterstützung durch Lehrende oder Peers, da die eigenständige Faktenprüfung noch nicht zuverlässig geleistet werden kann. Die Reflect-Phase kann explizit thematisieren, welche fachlichen Erkenntnisse durch die KI-Kollaboration gewonnen wurden – und welche Wissenslücken sie offengelegt hat.

4.4 Institutionelle Rahmenbedingungen

Unabhängig von individuellen Kompetenzen setzen institutionelle Rahmenbedingungen oft die praktischen Grenzen der PCRR-Anwendung. Drei Faktoren sind dabei besonders relevant: verfügbare Zeit, technische Infrastruktur und rechtliche Vorgaben.

Der Zeitfaktor betrifft sowohl Lernende als auch Lehrende. Die vollständige Durchführung aller PCRR-Phasen – einschließlich differenzierter Reviews und ausführlicher Reflexion – erfordert erheblich mehr Zeit als eine Ad-hoc-KI-Nutzung. Für kürzere Unterrichtseinheiten oder zeitlich begrenzte Aufgaben muss das Framework entsprechend reduziert werden. Gleichzeitig erhöht die im Framework geforderte Prozessdokumentation den Korrekturaufwand für Lehrende. Hier sind pragmatische Lösungen gefragt: die selektive Dokumentation einiger weniger, aber kritischer Entscheidungspunkte, Peer-Review-Formate oder eine KI-gestützte Auswertung der Dokumentation.

Die technische Infrastruktur bestimmt, welche Context-Engineering-Dimensionen überhaupt nutzbar sind. Schulen ohne Zugang zu datenschutzkonformen KI-Tools mit erweiterten Funktionen (große Kontextfenster, Websuche, Datei-Upload, Code-Interpreter, Chatexport usw.) können bestimmte Aspekte des Frameworks nicht umsetzen. Die Plan-Phase muss dann realistisch einschätzen, welche Tools tatsächlich verfügbar sind, statt von idealen Bedingungen auszugehen. Auch die Frage der Chancengerechtigkeit stellt sich: Wenn Lernende außerhalb der Institution auf unterschiedlich leistungsfähige Tools zugreifen können, entstehen Ungleichheiten, die bei der Bewertung berücksichtigt werden müssen.

Rechtliche Vorgaben betreffen Altersbeschränkungen für KI-Tools (die meisten erfordern ein Mindestalter von 13 Jahren), Datenschutzbestimmungen (insbesondere bei der Verarbeitung personenbezogener Daten) und institutionelle KI-Richtlinien, die den erlaubten Einsatzrahmen definieren. Diese Vorgaben sind nicht verhandelbar und müssen in der Plan-Phase als Randbedingungen berücksichtigt werden.

5 Diskussion

Das in dieser Arbeit vorgestellte PCRR-Framework verortet sich in einer dynamischen Landschaft bildungstechnologischer Modelle. Um den spezifischen Mehrwert von PCRR für die Bildungspraxis herauszuarbeiten, ist eine differenzierte Abgrenzung zu bestehenden Ansätzen des Prompt Engineering, der Assessment-Gestaltung, der Lehrerprofessionalisierung und der Lerntheorie notwendig. Während Frameworks wie CLEAR (Lo, 2023) oder AIAS (Perkins et al., 2024) spezifische Teilaspekte der KI-Interaktion adressieren (z. B. die Prompting-Komponente respektive die Kollaborations-Komponente), positioniert sich PCRR als methodische Klammer, die technisches Handeln (Prompt und Context Engineering) in einen pädagogisch-didaktischen Gesamtprozess integriert.

Als Gegenüberstellung wurden fünf Modelle ausgewählt: das *CLEAR-Framework* nach Lo (2023), die *AI Assessment Scale* nach Perkins et al. (2024), das *TPACK-Modell*, das *KI-Kompetenzmodell* nach Alles et al. (2025) und der *Experiential Learning Cycle* nach Kolb (1984).

5.1 Vom Prompting zum Prozess: Verhältnis zu Prompt-Strukturierungs-Frameworks

Für die Strukturierung einzelner Prompts haben sich in der Praxis verschiedene Frameworks etabliert, wie z. B. das *CLEAR-Framework* (*Concise, Logical, Explicit, Adaptive, Reflective*; Lo, 2023) oder das *Five S Model* (AI for Education, 2025). Diese Ansätze übersetzen Grundprinzipien effektiver Kommunikation mit Sprachmodellen in einprägsame Akronyme und bieten niedrigschwellige Orientierungshilfen für die direkte Eingabegestaltung. Empirisch validiert sind sie bislang nicht; ihre Verbreitung verdanken sie primär ihrer Praktikabilität.

PCRR operiert auf einer anderen Abstraktionsebene. Es fragt nicht nach der optimalen Formulierung eines einzelnen Prompts, sondern nach der

Einbettung von Prompts in einen didaktisch reflektierten Gesamtprozess. Während Frameworks wie CLEAR Heuristiken für die Create-Phase liefern können, adressiert PCRR die vorgelagerte strategische Planung, die nachgelagerte Evaluation und die metakognitive Reflexion. Die Ansätze sind damit komplementär: Prompt-Strukturierungs-Frameworks können innerhalb der Create-Phase von PCRR als Orientierung dienen, ohne dass PCRR diese Detailebene selbst spezifiziert. PCRR definiert das *Was* und *Warum* der KI-Kollaboration; Frameworks wie CLEAR können das *Wie* der Eingabegestaltung konkretisieren.

5.2 Von der Policy zur Pädagogik: Integration der AI Assessment Scale

Die bereits vorgestellte *AI Assessment Scale* nach Perkins et al. (2024) ist ein vielrezipierter Ansatz zur Klassifizierung von KI-Einsatzszenarien in Aufgaben- und Prüfungsformaten – von ›No AI‹ (Level 1) bis ›AI Exploration‹ (Level 5). Die AIAS ist primär ein Instrument der Ordnungspolitik und des Assessment-Designs; sie definiert, was erlaubt ist und welcher Grad der Nutzung angestrebt wird. Inzwischen liegen auch Implementierungsempfehlungen (Perkins et al., 2025a), kontextspezifische Adaptionen (z. B. Roe et al., 2024, 2025) sowie erste empirische Erprobungen (Furze et al., 2024) vor. Was jedoch weiterhin fehlt, ist eine systematische didaktische Anleitung, wie Lernende die jeweiligen Stufen kompetent ausfüllen können – hier setzt das PCRR-Framework an.

PCRR fungiert hier als das operative Gegenstück zur normativen Ebene des AIAS. In der Plan-Phase des PCRR-Zyklus entscheiden Lehrende oder Lernende unter Rückgriff auf das AIAS, welcher Grad der KI-Unterstützung für die Aufgabe angemessen ist (z. B. AIAS Level 2 ›AI Planning‹ oder Level 4 ›Full AI‹). In den darauffolgenden Phasen (Create, Review) operationalisiert PCRR diese Entscheidung. Während AIAS die Integrität des Assessments sichert, sichert PCRR die Qualität des Lernprozesses innerhalb der gewählten AIAS-Stufe.

5.3 Operationalisierung von Handlungskompetenz: Verhältnis zum TPACK-Modell

Das *TPACK-Modell (Technological Pedagogical Content Knowledge)* beschreibt das notwendige Professionswissen von Lehrkräften (Mishra & Koehler, 2006). In der Ära der Generativen KI wurde dieses Modell erweitert, um Kontextwissen (XK) stärker zu betonen (Mishra et al., 2023) und ethische Dimensionen sowie *Intelligent-TPACK* zu integrieren (Celik, 2023). TPACK bleibt jedoch eine Beschreibung von Wissensstrukturen (Potenzial), nicht von Handlungsprozessen (Performanz).

PCRR kann als die methodische Operationalisierung von *GenAI-TPACK* verstanden werden. Um PCRR im Unterricht erfolgreich zu implementieren, benötigen Lehrkräfte spezifisches *Technological Pedagogical Knowledge* (TPK) – etwa das Wissen darüber, wie Halluzinationen in der Review-Phase didaktisch genutzt werden können. PCRR ist somit das Werkzeug, mit dem Lehrkräfte ihr theoretisches TPACK-Wissen in konkrete Lernarrangements übersetzen. Es transformiert das statische Wissen über Technologie in einen dynamischen Workflow für Lernende.

5.4 Kompetenzziele vs. Kompetenzerwerb: Das Modell nach Alles et al.

Das Kompetenzmodell von Alles et al. (2025) definiert vier zentrale Handlungsfelder für den Umgang mit KI: Verstehen, Anwenden, Reflektieren und Mitgestalten. Es beschreibt die Ziele der KI-Bildung (*Learning Outcomes*). Das Modell von Alles et al. (2025) wird als repräsentatives Beispiel für aktuelle KI-Kompetenzraster herangezogen, da es exemplarisch die normativen Bildungsziele definiert. Diese Auswahl dient dazu, die Funktion des PCRR-Frameworks als methodischen Wegbereiter (*Learning Process*) abzugrenzen, der notwendig ist, um solche abstrakt definierten Kompetenzstufen – vom Verstehen bis zum Mitgestalten – in der Praxis operativ zu erreichen. PCRR beschreibt den Weg (*Learning Process*), auf dem diese Kompetenzen iterativ

erworben werden. PCRR ist somit der didaktische Motor zur Erreichung der von Alles et al. (2025) postulierten Kompetenzstufen.

5.5 Fundierung in der Lerntheorie: Der Experiential Learning Cycle nach Kolb

Das PCRR-Framework weist eine konzeptionelle Verwandtschaft mit dem *Experiential Learning Cycle* (Kolb, 1984; Kolb & Kolb, 2023) auf, ohne dessen Struktur eins zu eins zu replizieren. Beide Modelle teilen die Grundannahme, dass Lernen ein zyklischer Prozess ist, der Handlung und Reflexion verbindet. Die Phasen lassen sich lose parallelisieren: Plan entspricht der vorbereitenden Konzeptualisierung, Create verbindet aktives Experimentieren mit konkreter Erfahrung, Review initiiert die kritische Beobachtung der Ergebnisse, und Reflect vertieft diese Beobachtung auf der Prozessebene.

Kolbs Zyklus aus *Concrete Experience, Reflective Observation, Abstract Conceptualization* und *Active Experimentation* bildet das lerntheoretische Fundament, auf dem PCRR aufbaut und es für das Zeitalter der Mensch-Maschine-Kollaboration spezifiziert:

- *Plan* entspricht der *Abstract Conceptualization*: Die strategische Planung und Theoriebildung vor der Handlung.
- *Create* verbindet *Active Experimentation* und *Concrete Experience*: Das aktive Prompting und das unmittelbare Erleben der KI-Reaktion.
- *Review* initiiert die *Reflective Observation* auf der Produktebene: Die kritische Betrachtung der generierten Outputs durch Mikro-, Meso- und Makro-Reviews.
- *Reflect* vertieft die *Reflective Observation* auf der Prozessebene: Die metakognitive Auseinandersetzung mit dem gesamten Kollaborationsverlauf und den eigenen Lernerfahrungen.

Anders als bei Kolb, wo der Zyklus typischerweise mit der konkreten Erfahrung beginnt, setzt PCRR die strategische Planung an den Anfang. Diese Modifikation trägt der Spezifik der KI-Kollaboration Rechnung: Ohne vorherige Klärung von Zielen, Tools und ethischen Rahmenbedingungen besteht die Gefahr ineffektiver oder problematischer KI-Nutzung. PCRR adaptiert somit

das Grundprinzip des erfahrungsbasierten Lernens für einen Kontext, in dem die ›Erfahrung‹ (die KI-Interaktion) durch bewusste Vorbereitung qualitativ verbessert werden kann.

5.6 Zusammenfassung der Diskussion

Zusammenfassend unterscheidet sich das PCRR-Framework durch seine Ganzheitlichkeit und Prozessorientierung. Es reduziert KI-Kompetenz nicht auf technisches Prompting (wie CLEAR) oder statisches Wissen (wie TPACK), sondern modelliert die Interaktion als zirkulären, hermeneutischen Prozess. Es schließt die Lücke zwischen der normativen Ebene (AIAS, Kompetenzmodelle) und der operativen Ebene (Prompting), indem es Lernenden und Lehrenden eine strukturierte Heuristik an die Hand gibt, um vom *Trial-and-Error* zum systematischen Context Engineering zu gelangen.

6 Limitationen

Obwohl das PCRR-Framework einen umfassenden Ansatz zur Strukturierung der Mensch-KI-Kollaboration bietet und das theoretische Fundament des Context Engineering in didaktische Handlungsschritte übersetzt, unterliegt die vorliegende Arbeit sowie das Framework selbst spezifischen Limitationen. Diese betreffen die empirische Validierung, die praktische Umsetzbarkeit unter Ressourcenbeschränkungen sowie die Dynamik der technologischen Entwicklung.

6.1 Empirische Validierungslücke

Die vorliegende Publikation ist primär konzeptioneller Natur. Während das Framework auf etablierten Theorien (Kolb, 1984; Mishra et al., 2023) und aktuellen *Best Practices* des Prompt Engineering aufbaut, steht der breite empirische Nachweis seiner Wirksamkeit noch aus. Die in der Differenzierung erwähnten Pilotversuche (Freinhofer et al., 2025; Schwabl & Vötsch, 2025) liefern erste positive Indikatoren, stellen jedoch noch keine statistisch signifikante Evidenz für den Lernerfolg dar. Es bleibt zu untersuchen, ob die Anwendung des PCRR-Frameworks tatsächlich zu einer messbaren Steigerung der fachlichen oder der KI-bezogenen Kompetenz führt, oder ob die intensive Beschäftigung mit dem Prozess (Plan, Review) zu einer kognitiven Überlastung (*Cognitive Load*) führt, die den fachlichen Lernzuwachs in bestimmten Kontexten hemmen könnte.

6.2 Komplexität und Implementierungshürden

Das PCRR-Framework fordert von Lernenden und Lehrenden ein hohes Maß an metakognitiver Disziplin. Die Forderung, nicht nur Ergebnisse zu produzieren, sondern den gesamten Prozess kleinteilig zu planen (Plan), iterativ zu gestalten (Create) und auf drei Ebenen zu evaluieren (Review), steht im Konflikt zur oft opportunistischen Nutzungspraxis (*Trial-and-Error*), die Lernende intuitiv bevorzugen (vgl. Zamfirescu-Pereira et al., 2023). Für die

Unterrichtspraxis ergibt sich daraus ein Skalierungsproblem: Die im Framework geforderte detaillierte Dokumentation des Prozesses (Prompts, Reflexionen, Meso-Reviews) erhöht den Korrekturaufwand für Lehrende massiv. Wenn die Prozessbewertung mehr Zeit in Anspruch nimmt als die Bewertung des Endprodukts, droht das Framework an der ökonomischen Realität des Schul- und Hochschulalltags zu scheitern. Zukünftige Iterationen des Modells müssen daher Wege aufzeigen, wie diese Dokumentation effizienter – möglicherweise selbst KI-gestützt – erfolgen kann.

6.3 Technologische Abhängigkeit und Chancengerechtigkeit

Das Framework setzt implizit den Zugang zu leistungsfähigen KI-Modellen voraus, die über entsprechende Context-Engineering-Funktionen (großes Kontextfenster, Datei-Uploads, Reasoning-Fähigkeiten) verfügen. Diese Funktionen sind derzeit (Stand Januar 2026) häufig an kostenpflichtige Abonnements gebunden (z. B. ChatGPT Plus, Gemini Pro oder Claude Pro). Darüber hinaus sind diese Chatbots für den pädagogischen Gebrauch aufgrund von Datenschutzbedenken nicht geeignet. Datenschutzkonforme KI-Tools stehen vielen Lernenden nicht ausreichend zur Verfügung. Dies wirft die Frage der Chancengerechtigkeit auf (*Digital Divide*). Wenn das PCRR-Framework nur mit High-End-Modellen (*Compound AI Systems*) sein volles Potenzial entfaltet, werden Lernende benachteiligt, die auf kostenlose, eingeschränkte Basis-Modelle angewiesen sind. In diesen Fällen funktionieren Strategien wie RAG oder komplexes Reasoning oft unzureichend, was zu Frustration führen kann. Das Framework muss daher künftig stärker differenzieren, welche PCRR-Schritte auch mit ›Low-Tech‹-Zugängen effektiv durchführbar sind.

6.4 Das Dilemma der Zielgruppen-Breite

Eine wesentliche Herausforderung betrifft die breite Adressierung des PCRR-Frameworks, das von der Sekundarstufe bis zur Hochschulbildung Anwendung finden soll. Kılınç (2024) kritisiert in Bezug auf die AI Assessment Scale (AIAS), dass eine fehlende Differenzierung zwischen *K-12* (Primar- und

Sekundarstufe) und *Higher Education* (Hochschulbereich) die pädagogische Präzision verwässert. Diese Kritik lässt sich unmittelbar auf PCRR übertragen: Die kognitiven Voraussetzungen für die Plan-Phase (strategische Antizipation) und die Reflect-Phase (Metakognition) unterscheiden sich zwischen 15-jährigen Schüler*innen und 25-jährigen Master-Studierenden fundamental. Der Anspruch des Frameworks, als universelles Modell zu fungieren (›One size fits all‹), birgt die Gefahr, spezifische didaktische Erfordernisse der jeweiligen Bildungsstufen zu nivellieren und damit an Schärfe zu verlieren.

6.5 Volatilität der Technologie: Vom Prompting zu Agenten

Das PCRR-Framework in seiner jetzigen Form ist stark auf die Interaktion mit Chatbots ausgerichtet, bei denen der Mensch der steuernde Akteur bleibt (*Human-in-the-loop*). Die rasante Entwicklung hin zu *Agentic AI* – also Systemen, die autonom Ziele verfolgen, Werkzeuge nutzen und sich selbst korrigieren – könnte Teile des Frameworks obsolet machen. Wenn zukünftige Modelle die Phasen Plan (Strategieentwicklung) und Review (Fehlerkorrektur) zunehmend autonom und zuverlässig übernehmen, verschiebt sich die notwendige menschliche Kompetenz drastisch: Weg vom detaillierten Context Engineering hin zur bloßen Zielvorgabe und finalen Abnahme. Das Framework läuft Gefahr, Kompetenzen zu trainieren (z. B. manuelles *Prompt Chaining*), die durch die technische Evolution automatisiert werden. Eine kontinuierliche Anpassung des Modells an den Grad der KI-Autonomie ist daher unabdingbar.

6.6 Komplexität der Erfolgsmessung

Ein ungelöstes theoretisches Problem stellt die objektive Evaluation der PCRR-Ergebnisse dar. Singh et al. (2024) illustrieren die Vielfalt technischer Metriken zur Bewertung von LLM-Outputs (z. B. BLEU, ROUGE oder BERT-Score). Diese quantitativen Maße greifen jedoch für pädagogische Kontexte zu kurz, da PCRR nicht die Textqualität, sondern die Kompetenzentwicklung und den Denkprozess in den Fokus rückt. Es fehlt derzeit an validierten

Instrumenten, um die Qualität einer ›Planung‹ oder ›Reflexion‹ im PCRR-Zyklus standardisiert zu messen. Anders als bei rein technischen Benchmarks ist die Bewertung der kollaborativen Leistung zwischen Mensch und Maschine hochgradig subjektiv und kontextabhängig, was die Vergleichbarkeit von Lernergebnissen erschwert.

6.7 Strukturelle Verortung der Ethik

Im Vergleich zu Celiks (2023) Konzept des Intelligent-TPACK, welches ethische Überlegungen als integrativen Bestandteil professionellen Lehrerwissens durchgängig mitdenkt, weist PCRR eine strukturelle Einschränkung auf. Ethische Aspekte werden zwar in der Plan-Phase unter ›Angemessenheitsprüfung‹ angesprochen und in der Review-Phase als eigenständige Prüfdimension operationalisiert, doch eine systematische Durchdringung aller Phasen erfolgt nicht.

Diese Einschränkung ist teilweise der Komplexität des Feldes geschuldet: KI-Ethik im Bildungsbereich befindet sich in einer formativen Phase, in der verbindliche Orientierungen weitgehend fehlen. Rechtliche Rahmenbedingungen (Datenschutz, Urheberrecht, KI-Verordnung) variieren zwischen Ländern und entwickeln sich dynamisch. Institutionelle Richtlinien – sofern vorhanden – unterscheiden sich erheblich zwischen Schultypen, Bundesländern und Hochschulen. Das PCRR-Framework kann diese Heterogenität nicht auflösen, sondern nur darauf verweisen, dass ethische und rechtliche Prüfungen kontextspezifisch erfolgen müssen.

Für zukünftige Versionen des Frameworks ist eine stärkere Integration ethischer Reflexion angestrebt, sobald sich belastbare Orientierungsrahmen etabliert haben. Bis dahin wird Lehrenden empfohlen, die ethische Dimension nicht als nachgelagerten Prüfschritt zu behandeln, sondern bereits in der Plan-Phase explizit zu thematisieren – etwa durch die Leitfrage: ›Welche ethischen Risiken (Bias, Datenschutz, Abhängigkeit) könnten bei dieser Aufgabe auftreten, und wie gehe ich damit um?‹

Trotz dieser Limitationen stellt das PCRR-Framework einen notwendigen Schritt dar, um die ›Black Box‹ der KI-Nutzung in bildungsrelevanten Kontexten zu öffnen. Die aufgezeigten Grenzen sind als Arbeitsaufträge für die

weitere forschende und praktische Auseinandersetzung zu verstehen. Sie unterstreichen die Notwendigkeit, KI-Didaktik nicht als statisches Regelwerk, sondern als agile Disziplin zu begreifen, die sich mit ihrem Gegenstand ständig weiterentwickeln muss.

7 Fazit und Ausblick

Die vorliegende Publikation verfolgt das Ziel, das PCRR-Framework theoretisch zu fundieren und angesichts der rasanten Entwicklungen im Bereich der Sprachmodelle und KI-Chatbots zu aktualisieren. Die zentralen Beiträge dieser Arbeit lassen sich in drei Bereichen zusammenfassen.

Erstens wurde mit dem Konzept des *Context Engineering* ein theoretisches Fundament etabliert, das über das traditionelle Prompt Engineering hinausgeht. Die Konzeptualisierung in sechs Dimensionen – *User Prompt, Systemprompt, RAG, Short-Term Memory, Long-Term Memory* und *Tools* – bildet die Komplexität moderner KI-Interaktion ab und verdeutlicht, warum isolierte Prompting-Tipps der Realität heutiger *Compound AI Systems* nicht mehr gerecht werden. Diese Rahmung ermöglicht es, die Mensch-KI-Kollaboration als orchestrierten Prozess zu verstehen, der strategische Planung, technische Kompetenz und kritische Evaluation erfordert.

Zweitens wurde das PCRR-Framework selbst substanziell weiterentwickelt. Die ursprünglich für die Berufsbildung konzipierte Erstversion wurde zu einem bildungsstufenübergreifenden Modell erweitert. Die vier Phasen – *Plan, Create, Review, Reflect* – wurden theoretisch vertieft und mit den Dimensionen des Context Engineering verknüpft. Neben der stärkeren Rückkopplung zwischen den Create- und Review-Phasen erfuhr letztere auch eine differenzierte Ausarbeitung: Die Unterscheidung zwischen *Mikro-Reviews* (nach jedem Output), *Meso-Reviews* (nach Teilaufgaben) und *Makro-Reviews* (finale Evaluation) bildet reale Arbeitsprozesse ab und überwindet die naive Vorstellung eines einmaligen Faktenchecks am Ende. Die klare Trennung zwischen produktbezogenem Review und prozessbezogenem Reflect schärft zudem das Verständnis dafür, dass KI-Kollaboration sowohl Ergebnis- als auch Lernprozessqualität erfordert.

Drittens wurde das PCRR-Framework in der bestehenden Landschaft bildungstechnologischer Modelle verortet. Die Diskussion zeigte, dass sich das PCRR-Framework komplementär zu anderen Modellen des KI-Einsatzes in der pädagogischen Praxis verhält. Es positioniert sich als methodische

Klammer, die technisches Handeln in einen pädagogisch-didaktischen Gesamtprozess integriert.

7.1 Implikationen für die Praxis

Für Lehrende bietet das PCRR-Framework einen strukturierten Rahmen, um KI-Kollaboration systematisch in Lernprozesse zu integrieren. Es ermöglicht die transparente Kommunikation von Erwartungen – welche Phasen sollen durchlaufen werden, welche Dokumentation wird erwartet, welche Kriterien gelten für die Bewertung. Gleichzeitig macht das Framework deutlich, dass effektive KI-Nutzung Kompetenzen erfordert, die explizit vermittelt und schrittweise aufgebaut werden müssen. Die im Differenzierungskapitel skizzierten Anpassungsmöglichkeiten zeigen, dass das Framework nicht als starres Regelwerk, sondern als adaptiver Orientierungsrahmen zu verstehen ist.

Für Lernende strukturiert das PCRR-Framework einen Prozess, der andernfalls oft unsystematisch und oberflächlich verläuft. Die explizite Planungsphase fördert strategisches Denken, die differenzierten Review-Ebenen schulen kritische Evaluation, und die Reflexionsphase sichert den Transfer der Lernerfahrungen. Dabei geht es nicht darum, jeden KI-Einsatz mit maximaler Dokumentationstiefe zu begleiten, sondern darum, ein Bewusstsein für die Komplexität der Mensch-KI-Kollaboration zu entwickeln, das auch in weniger strukturierten Kontexten wirksam bleibt.

7.2 Forschungsagenda

Die in den Limitationen identifizierten Herausforderungen definieren zugleich die Forschungsagenda für die Weiterentwicklung des PCRR-Frameworks. Drei Stränge sind dabei vorrangig.

Der erste Strang betrifft die praktische Operationalisierung. Die theoretische Vollständigkeit des hier vorgestellten Frameworks muss für spezifische Kontexte übersetzt werden. Aktuell wird an einer Dokumentationsvorlage für Lehrende und Lernende in der Sekundarstufe gearbeitet, die das Framework für den Unterrichtseinsatz handhabbar macht. Diese Vorlage wird

Gegenstand einer eigenständigen Publikation sein, die konkrete Materialien und Anwendungsbeispiele bereitstellt.

Der zweite Strang umfasst die empirische Validierung. Die konzeptionelle Plausibilität des PCRR-Frameworks muss durch empirische Forschung gestützt werden. Eine laufende Aktionsforschung in der Berufsbildung untersucht, wie das Framework in realen Unterrichtssituationen funktioniert, welche Anpassungen sich als notwendig erweisen und welche Effekte auf Lernprozesse und -ergebnisse beobachtbar sind. Diese Ergebnisse werden wichtige Erkenntnisse für die Weiterentwicklung des Frameworks liefern und gleichzeitig zur noch spärlichen empirischen Basis der KI-Didaktik beitragen.

Der dritte Strang adressiert die Assessment-Dimension. Die im Framework geforderte Prozessdokumentation wirft die Frage auf, wie diese systematisch bewertet werden kann. Die Entwicklung eines Bewertungsrasters, das sowohl die Prozess- als auch die Produktqualität erfasst, ist Gegenstand einer geplanten Folgepublikation. Dieses Raster soll Lehrenden konkrete Kriterien an die Hand geben und gleichzeitig die Fairness und Vergleichbarkeit von Bewertungen sicherstellen.

Darüber hinaus bleiben grundlegendere Forschungsfragen offen. Wie verändert sich das PCRR-Framework, wenn KI-Systeme zunehmend agentisch agieren und Teile der Plan- und Review-Phase autonom übernehmen? Welche Instrumente eignen sich, um die Qualität von KI-Kollaborationsprozessen valide zu messen? Wie kann Ethik nicht nur als Prüfdimension in der Review-Phase, sondern als durchgängige Haltung im gesamten Prozess verankert werden? Diese Fragen verweisen auf die Notwendigkeit, KI-Didaktik als dynamische Disziplin zu begreifen, die sich mit ihrem Gegenstand kontinuierlich weiterentwickelt.

Die vorliegende Arbeit ist ein offener Vorschlag, der von der kritischen Auseinandersetzung durch andere Forschende und Praktizierende profitieren kann. Empirische Erprobungen in unterschiedlichen Bildungskontexten, theoretische Weiterentwicklungen einzelner Komponenten oder kritische Revisionen des Gesamtansatzes sind ausdrücklich erwünscht. Wir laden dazu ein, das Framework aufzugreifen, zu adaptieren und – wo nötig – zu widersprechen.

7.3 Schluss

Das PCRR-Framework versteht sich als Beitrag zu einer verantwortungsvollen Integration Generativer KI in Bildungskontexten. Es basiert auf der Überzeugung, dass KI-Nutzung weder unkritisch delegiert noch pauschal abgelehnt werden sollte, sondern als Kompetenz begriffen werden muss, die systematisch entwickelt werden kann. Die vier Phasen – *Plan, Create, Review, Reflect* – bieten dafür einen strukturierten Rahmen, der technische, kognitive und metakognitive Anforderungen integriert.

Die hier vorgestellte Version 2 ist dabei kein Endpunkt, sondern ein Zwischenstand. Die technologische Entwicklung wird weitere Anpassungen erfordern, die empirische Forschung wird Stärken und Schwächen des Ansatzes offenlegen, und die praktische Erprobung wird zeigen, wo theoretische Eleganz an pragmatische Grenzen stößt. Das PCRR-Framework ist somit selbst einem iterativen Prozess unterworfen – ganz im Sinne der Logik, die es für die KI-Kollaboration propagiert. Alle bisherigen und künftigen Materialien rund um das PCRR-Framework (von Publikationen bis hin zu Unterrichtsmaterielien) werden auf der Website https://pcrr.info/ zur Verfügung gestellt.

7.4 Hinweise

*CrediT Autor*innenschaft Beitragserklärung*

- *Dominik Freinhofer,* Universität Graz, IDea_Lab – das Interdisziplinäre Digitale Labor der Universität Graz: Conceptualization, Writing – Original Draft, Writing – Review & Editing, Supervision, Project Administration.
- *Gerlinde Schwabl*, Pädagogische Hochschule Tirol, Institut für Berufspädagogik, Fach stelle Medienbildung & Digitalisierung, Lehrgangsleitung ›KI im IT-Unterricht der Berufsbildung‹: Conceptualization, Writing– Review & Editing, Project Administration.
- *Sandra Breitenberger*, Pädagogische Hochschule Oberösterreich, Institut für Berufspädagogik: Conceptualization, Writing – Review & Editing, Project Administration.

- *Anja Steiner,* Pädagogische Hochschule Tirol, Institut für Berufspädagogik: Conceptualization, Writing – Review & Editing, Project Administration. Erklärung von Generativer KI und KI-gestützten Technologien im Schreibprozess.

Erklärung von Generativer KI und KI-gestützten Technologien im Schreibprozess

Alle Ideen stammen von den menschlichen Autor*innen dieses Papers. KI-Systeme wurden eingesetzt, um die Qualität der Publikation zu verbessern– nicht, um menschliches Denken zu ersetzen, sondern um dieses zu entlasten, zu unterstützen und zu erweitern (vgl. Limburg & Buck, 2024).

Allgemeine Herangehensweise: Primär wurde der KI-Chatbot Claude (Anthropic) in der Projekt Funktion verwendet, wobei die referenzierten Open-Access-Publikationen hochgeladen und ein umfangreicher Projekt-Prompt definiert wurde. Als Modellvariante diente vorwiegend Claude Opus-4.5, ergänzt durch Funktionen wie erweitertes Nachdenken, Websuche und Recherche. Für spezifische Aufgaben kam zusätzlich Gemini-3.0 (Google) zum Einsatz, z.B. wenn ein größeres Kontextfenster nötig war. Die Übersetzung ins Englische erfolgte mit dem KI-Chatbot Claude (Anthropic) und wurde vom Hauptautor, der als Englischlehrer ausgebildet ist, überprüft und überarbeitet.

Funktionen der KI im Schreibprozess: Die Autor*innen verfassten das Grundgerüst selbst. KI-Chatbots dienten als kritische Sparringspartner zur Identifikation von Stärken und Schwächen, zur sprachlichen Optimierung (akademische Ausdrucksweise, roter Faden, einheitlicher Stil), zum Verfassen von Übergängen ohne inhaltlichen Mehrwert sowie zur Ausformulierung von Notizen zu Kapiteln– stets mit manueller Prüfung auf unerwünschte Ergänzungen.

Spezifische Einsatzbereiche: Die Einleitung entstand mit Gemini unter Nutzung der Deep Research-Funktion, wobei die KI bei der Synthese technischer Fachbegriffe und der Verzahnung von Forschungsansätzen unterstützte. Theoretische Diskussion und Limitationen wurden ebenfalls mit Gemini entwickelt– hier erwies sich die fehlende ›Betriebsblindheit‹ des Modells als besonders hilfreich. Fazit und Abstract wurden mit Claude erarbeitet. Ein weiteres Element war der systematische Quellenabgleich: Claude

und Gemini analysierten die verwendeten Publikationen und identifizierten relevante Anknüpfungspunkte für das Framework. Die Autor*innen behielten dabei stets die Hoheit über die konzeptuelle Einordnung– die Modelle lieferten Optionen, die Autor*innen entschieden, welche Quellen als inspirierend versus strukturgebend positioniert werden sollten. Einzelne Bereiche– wie das dreistufige Review-System und die Praxisbeispiele entstanden im Dialog mit Claude und wurden gemeinsam ausgearbeitet.

Charakter der Zusammenarbeit: Die Entwicklung erfolgte in einem iterativen Ko-Konstruktionsprozess. Die Autor*innen legten konzeptuelle Grundzüge vor, die Modelle erstellten Entwürfe, und die Autor*innen überarbeiteten diese– oft mit dem Hinweis, dass Formulierungen zu abstrakt, zu prätentiös oder zu umgangssprachlich geraten waren. Die Modelle wurden explizit angewiesen, bei Unklarheiten nachzufragen und Widerspruch einzubringen. Die finale Textverantwortung verbleibt bei den Autor*innen.

Einordnung nach referenzierten Frameworks: Nach der AI Assessment Scale (Perkins et al., 2024) bewegte sich der KI-Einsatz primär auf den Ebenen ›AI Planning‹ und ›AI Collaboration‹, einzelne Teilaufgaben auf der Ebene ›Full AI‹. Erklärung über Interessenskonflikte Die Autor*innen erklären, dass keine Interessenkonflikte im Zusammenhang mit dieser Veröffentlichung bestehen.

Erklärung über Interessenskonflikte

Die Autor*innen erklären, dass keine Interessenkonflikte im Zusammenhang mit dieser Veröffentlichung bestehen.

Literaturverzeichnis

AI for Education. (2025). *Prompt framework for educators: The five "S" model*. Abgerufen am 6. Januar 2026 von https://www.aiforeducation.io/ai-resources/the-five-s-model

Alles, S., Falck, J., Flick, M., & Schulz, R. (2025). *KI-Kompetenzen für Lehrende und Lernende: Aus der Praxis für die Praxis – eine adaptierbare Basis*. https://doi.org/10.5281/zenodo.15001018

Anthropic. (2025a). *Best practices for prompt engineering*. Abgerufen am 12. Januar 2026 von https://web.archive.org/web/20251219031008/https://claude.com/blog/best-practices-for-prompt-engineering

Anthropic. (2025b). *Effective context engineering for AI agents*. Abgerufen am 6. Januar 2026 von https://www.anthropic.com/engineering/effective-context-engineering-for-ai-agents

Basil, S., Shapiro, I., Shapiro, D., Mollick, E., Mollick, L., & Meincke, L. (2025). *Prompting science report 4: Playing pretend: Expert personas don't improve factual accuracy*. https://doi.org/10.48550/arXiv.2512.05858

Buck, I. (2025). *Wissenschaftliches Schreiben mit KI*. UTB. https://doi.org/10.36198/9783838563657

Celik, I. (2023). Towards intelligent-TPACK: An empirical study on teachers' professional knowledge to ethically integrate artificial intelligence (AI)-based tools into education. *Computers in Human Behavior, 138*, Article 107468. https://doi.org/10.1016/j.chb.2022.107468

Ceron, T., Falk, N., Barić, A., Nikolaev, D., & Padó, S. (2024). *Beyond prompt brittleness: Evaluating the reliability and consistency of political worldviews in LLMs*. https://doi.org/10.48550/arXiv.2402.17649

Chen, B., Zhang, Z., Langrené, N., & Zhu, S. (2025). Unleashing the potential of prompt engineering for large language models. *Patterns, 6*(6), Article 101260. https://doi.org/10.1016/j.patter.2025.101260

Chen, J., Ye, J., & Wang, G. (2025). *From standalone LLMs to integrated intelligence: A survey of compound AI systems*. https://doi.org/10.48550/arXiv.2506.04565

Dell'Acqua, F., McFowland III, E., Mollick, E., Lifshitz-Assaf, H., Kellogg, K. C., Rajendran, S., Krayer, L., Candelon, F., & Lakhani, K. R. (2023). *Navigating the jagged technological frontier: Field experimental evidence of the effects of AI on knowledge worker productivity and quality*. https://doi.org/10.2139/ssrn.4573321

Freinhofer, D., Schwabl, G., Aichinger, S., Breitenberger, S., Steindl, S., & Hechenberger, T. (2025). Prompten nach Plan: Das PCRR-Framework als pädagogisches Werkzeug für den Einsatz von Künstlicher Intelligenz. *Medienimpulse, 63*(1). https://doi.org/10.21243/mi-01-25-26

Furze, L., Perkins, M., Roe, J., & MacVaugh, J. (2024). The AI assessment scale (AIAS) in action: A pilot implementation of GenAI-supported assessment. *Australasian Journal of Educational Technology, 40*(4), 38–55. https://doi.org/10.14742/ajet.9434

Genkina, D. (2024, 6. März). *AI prompt engineering is dead: Long live AI prompt engineering*. https://spectrum.ieee.org/prompt-engineering-is-dead

Geyer, B. (2026, 5. Januar). *Bin ich eine Betrügerin, weil ich mit KI schreibe? Zwischen Autorinnenschaft, Kontrolle und Ko-KI-Kreation*. https://barbarageyer.substack.com/p/bin-ich-eine-betrugerin-weil-ich

Google. (2025). *Strategien für Prompt-Design*. Abgerufen am 6. Januar 2026 von https://ai.google.dev/gemini-api/docs/prompting-strategies

Gupta, M. (2025, 7. Juli). *Prompt engineering is dead: Prompt engineering is obsolete – now what?* https://medium.com/data-science-in-your-pocket/prompt-engineering-is-dead-debb01e9720e

Harwell, D. (2023, 25. Februar). *Tech's hottest new job: AI whisperer. No coding required*. *The Washington Post*. Abgerufen am 12. Januar 2026 von https://www.washingtonpost.com/technology/2023/02/25/prompt-engineers-techs-next-big-job/

Hauben, F. (2025, 29. Juni). *Prompt engineering is dead. Long live context engineering*. https://buildaiwithai.substack.com/p/prompt-engineering-is-dead-long-live

Höfler, E. (2025). Von Feedback zu Feedforward: Eine begriffliche Neuorientierung im Sinne von Futures Literacy. *R&E-SOURCE, 12*(4), 64–77. https://doi.org/10.53349/resource.2025.i4.a1492

Holtzman, A., West, P., Shwartz, V., Choi, Y., & Zettlemoyer, L. (2022). *Surface form competition: Why the highest probability answer isn't always right*. https://doi.org/10.48550/arXiv.2104.08315

Horthy, D. (2025). *12-factor-agents: Factor 03 – own your context window*. Abgerufen am 6. Januar 2026 von https://github.com/humanlayer/12-factor-agents/blob/main/content/factor03-own-your-context-window.md

Kaddour, J., Harris, J., Mozes, M., Bradley, H., Raileanu, R., & McHardy, R. (2023). *Challenges and applications of large language models*. https://doi.org/10.48550/arXiv.2307.10169

Kalai, A. T., Nachum, O., Vempala, S. S., & Zhang, E. (2025). *Why language models hallucinate*. https://doi.org/10.48550/arXiv.2509.04664

Kılınç, S. (2024). *Comprehensive AI assessment framework: Enhancing educational evaluation with ethical AI integration*. https://doi.org/10.48550/arXiv.2407.16887

Knoth, N., Tolzin, A., Janson, A., & Leimeister, J. M. (2024). AI literacy and its implications for prompt engineering strategies. *Computers and Education: Artificial Intelligence, 6*, Article 100225. https://doi.org/10.1016/j.caeai.2024.100225

Kolb, A. Y., & Kolb, D. A. (2023). *8 things to know about the experiential learning cycle*. Abgerufen am 6. Januar 2026 von https://learningfromexperience.com/books/8-things-to-know-about-the-experiential-learning-cycle/

Kolb, D. A. (1984). *Experiential learning: Experience as the source of learning and development*. Prentice Hall.

Korzynski, P., Mazurek, G., Krzypkowska, P., & Kurasinski, A. (2023). Artificial intelligence prompt engineering as a new digital competence. *Entrepreneurial Business and Economics Review, 11*(3), 25–37. https://doi.org/10.15678/EBER.2023.110302

Lewis, P., Perez, E., Piktus, A., Petroni, F., Karpukhin, V., Goyal, N., Küttler, H., Lewis, M., Yih, W.-t., Rocktäschel, T., Riedel, S., & Kiela, D. (2021). *Retrieval-augmented generation for knowledge-intensive NLP tasks*. https://doi.org/10.48550/arXiv.2005.11401

Limburg, A., & Buck, I. (2024). KI und Kognition im Schreibprozess. *Journal für Schreibwissenschaft, 15*(26), 8–23. https://doi.org/10.3278/JOS2401W

Lo, L. S. (2023). The CLEAR path: A framework for enhancing information literacy through prompt engineering. *The Journal of Academic Librarianship, 49*(4), Article 102720. https://doi.org/10.1016/j.acalib.2023.102720

Lütke, T. (2025). *Context engineering* [Tweet]. Abgerufen am 6. Januar 2026 von https://x.com/tobi/status/1935533422589399127

Malmqvist, L. (2024). Sycophancy in large language models: Causes and mitigations. https://doi.org/10.48550/arXiv.2411.15287

Mei, L., Yao, J., Ge, Y., Wang, Y., Bi, B., Cai, Y., Liu, J., Li, M., Li, Z.-Z., Zhang, D., Zhou, C., Mao, J., Xia, T., Guo, J., & Liu, S. (2025). *A survey of context engineering for large language models*. https://doi.org/10.48550/arXiv.2507.13334

Meincke, L., Mollick, E., Mollick, L., & Shapiro, D. (2025a). *Prompting science report 1: Prompt engineering is complicated and contingent*. https://doi.org/10.48550/arXiv.2503.04818

Meincke, L., Mollick, E., Mollick, L., & Shapiro, D. (2025b). *Prompting science report 2: The decreasing value of chain of thought in prompting*. https://doi.org/10.48550/arXiv.2506.07142

Meincke, L., Mollick, E., Mollick, L., & Shapiro, D. (2025c). *Prompting science report 3: I'll pay you or I'll kill you – but will you care?* https://doi.org/10.48550/arXiv.2508.00614

Milam, K., & Gulli, A. (2025). *Context engineering: Sessions, memory*. Abgerufen am 6. Januar 2026 von https://www.kaggle.com/whitepaper-context-engineering-sessions-and-memory

Mishra, P., & Koehler, M. J. (2006). Technological pedagogical content knowledge. *Teachers College Record, 108*(6), 1017–1054. https://doi.org/10.1111/j.1467-9620.2006.00684.x

Mishra, P., Warr, M., & Islam, R. (2023). TPACK in the age of ChatGPT and generative AI. *Journal of Digital Learning in Teacher Education, 39*(4), 235–251. https://doi.org/10.1080/21532974.2023.2247480

OpenAI. (2026a). *OpenAI cookbook*. Abgerufen am 12. Januar 2026 von https://cookbook.openai.com/

OpenAI. (2026b). *ChatGPT – Release notes*. Abgerufen am 12. Januar 2026 von https://help.openai.com/en/articles/6825453-chatgpt-release-notes

Perkins, M., Furze, L., Roe, J., & MacVaugh, J. (2024). The artificial intelligence assessment scale (AIAS). *Journal of University Teaching and Learning Practice, 21*(6). https://doi.org/10.53761/q3azde36

Perkins, M., Roe, J., & Furze, L. (2025a). How (not) to use the AI assessment scale. *Journal of Applied Learning & Teaching, 8*(2). https://doi.org/10.37074/jalt.2025.8.2.15

Perkins, M., Roe, J., & Furze, L. (2025b). Reimagining the artificial intelligence assessment scale. *Journal of University Teaching and Learning Practice, 22*(7). https://doi.org/10.53761/rrm4y757

Randazzo, S., Lifshitz-Assaf, H., Kellogg, K. C., Dell'Acqua, F., Mollick, E. R., Candelon, F., & Lakhani, K. R. (2025). *Cyborgs, centaurs and self-automators*. https://doi.org/10.2139/ssrn.4921696

Roe, J., Perkins, M., & Furze, L. (2025). Applying the AI assessment scale in writing and translation for EFL. *Artificial Intelligence in Education*, 1–16. https://doi.org/10.1108/AIIE-02-2025-0034

Roe, J., Perkins, M., & Tregubova, Y. (2024). *The EAP-AIAS*. https://doi.org/10.48550/arXiv.2408.01075

Sawetz, J. (2025). *Good prompting = good thinking*. Abgerufen am 6. Januar 2026 von https://www.sawetz.com/publications/

Schulhoff, S., Ilie, M., Balepur, N., Kahadze, K., Liu, A., Si, C., Li, Y., Gupta, A., Han, H., Dulepet, P. S., Vidyadhara, S., Ki, D., Agrawal, S., Pham, C., Kroiz, G., Li, F., Tao, H., Srivastava, A., … Resnik, P. (2025). *The prompt report*. https://doi.org/10.48550/arXiv.2406.06608

Schwabl, G., & Vötsch, M. (2025). Macht Künstliche Intelligenz richtig satt? *HiBiFo – Haushalt in Bildung & Forschung, 14*(4), 41–53. https://doi.org/10.3224/hibifo.v14i4.4

Shuster, K., Poff, S., Chen, M., Kiela, D., & Weston, J. (2021). Retrieval augmentation reduces hallucination in conversation. *Findings of the Association for Computational Linguistics: EMNLP 2021*, 3784–3803. https://doi.org/10.18653/v1/2021.findings-emnlp.320

Singh, A., Ehtesham, A., Gupta, G. K., Chatta, N. K., Kumar, S., & Khoei, T. T. (2024). *Exploring prompt engineering*. https://doi.org/10.48550/arXiv.2410.12843

Steinhoff, T. & Lehnen, K. (2025). Schreiben mit Künstlicher Intelligenz: Das GPT-Modell (Ghost, Partner, Tutor). *Leseräume – Zeitschrift für Literalität in Schule und Forschung*, *12*(11), 1–14.

Stern, J. (2023, 29. November). *Talking to chatbots is now a $200K job*. *The Wall Street Journal*. Abgerufen am 12. Januar 2026 von https://www.wsj.com/tech/ai/talking-to-chatbots-is-now-a-200k-job-so-i-applied-258bd5f0

Thoppilan, R., de Freitas, D., Hall, J., Shazeer, N., Kulshreshtha, A., Cheng, H.-T., Jin, A., Bos, T., Baker, L., Du, Y., Li, Y., Lee, H., Zheng, H. S., Ghafouri, A., Menegali, M., Huang, Y., Krikun, M., Lepikhin, D., Qin, J., … Le, Q. (2022). *LaMDA*. https://doi.org/10.48550/arXiv.2201.08239

Vatsal, S., & Dubey, H. (2024). *A survey of prompt engineering methods*. https://doi.org/10.48550/arXiv.2407.12994

Verma, M., Bhambri, S., & Kambhampati, S. (2024). *On the brittle foundations of ReAct prompting*. https://doi.org/10.48550/arXiv.2405.13966

Wei, J., Wang, X., Schuurmans, D., Bosma, M., Ichter, B., Xia, F., Chi, E., Le, Q., & Zhou, D. (2023). *Chain-of-thought prompting elicits reasoning*. https://doi.org/10.48550/arXiv.2201.11903

White, J., Fu, Q., Hays, S., Sandborn, M., Olea, C., Gilbert, H., Elnashar, A., Spencer-Smith, J., & Schmidt, D. C. (2023). *A prompt pattern catalog*. https://doi.org/10.48550/arXiv.2302.11382

Willison, S. (2023, 21. Februar). *In defense of prompt engineering*. https://simonwillison.net/2023/Feb/21/in-defense-of-prompt-engineering/

Yan, W. (2025, 12. Juni). *Don't build multi-agents*. https://cognition.ai/blog/dont-build-multi-agents

Zamfirescu-Pereira, J. D., Wong, R. Y., Hartmann, B., & Yang, Q. (2023). Why Johnny can't prompt. In A. Schmidt et al. (Hrsg.), *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (S. 1–21). ACM. https://doi.org/10.1145/3544548.3581388

Zhang, Z., Zhang, A., Li, M., & Smola, A. (2022). *Automatic chain-of-thought prompting*. https://doi.org/10.48550/arXiv.2210.03493

Zhao, T. Z., Wallace, E., Feng, S., Klein, D., & Singh, S. (2021). *Calibrate before use*. https://doi.org/10.48550/arXiv.2102.09690