FORSCHUNGSBERICHT

Automatisierte Kuratierung frühpädagogischer Wissenskorpora

Systemarchitektur der ECE Wissens-Pipeline v1.0 und explorative Evaluation lokaler RAG-Antworten

Sebastian Götz

ORCID 0009-0003-9200-7794

Kita Digital – Die KI-Insel der Frühpädagogik

www.ki-insel.de | beratung@ki-insel.de

Forschungsbericht Version 1.5
18. Oktober 2026

Fassung mit zehn dokumentierten Vergleichsfällen, lokaler Modell- und Hardwarekonfiguration, qualitativer RAGAS-Zuordnung, TDM-Einordnung und Architekturabbildungen. Literaturverzeichnis nach APA 7.

Softwarestand ECE Wissens-Pipeline v1.0
Der Versionswechsel dieses Berichts bezeichnet keine neue Softwareversion.

Zusammenfassung

Hintergrund: Für Retrieval-Augmented Generation (RAG) und die domänenspezifische Anpassung großer Sprachmodelle in der frühkindlichen Bildung, Betreuung und Erziehung werden thematisch geeignete, nachnutzbare und strukturell aufbereitete Textkorpora benötigt. Neben der Quellenqualität ist relevant, ob Fachfragen auf einem vorhandenen Wissensbestand lokal und ohne laufende Internetverbindung bearbeitet werden können.

Methode: Der Beitrag beschreibt eine vierstufige Architektur zur Open-Access-Akquise, regelbasierten Relevanz- und Lizenzklassifikation, Textbereinigung sowie lokalen modellgestützten Qualitätsprüfung. Wissenschaftliche Primärtexte werden nicht generativ umformuliert. Ergänzend werden zehn dokumentierte Antwortpaare aus AnythingLLM mit lokalem RAG und Gemini 3.1 Pro mit hohem Denkaufwand qualitativ mit sieben Quelldokumenten abgeglichen.

Ergebnisse: Der Suchlauf zum Publikationsjahr 2014 betrachtete 2.262 eindeutige Werke. Die Relevanzheuristik schloss 1.056 aus; von 1.199 anschließend geprüften Werken wurden 921 durch die Lizenz-Policy zurückgestellt oder ausgeschlossen. Von 203 technisch validierten PDF-Dateien gelangten 151 in den Korpus, 28 in die manuelle Prüfung und 24 aus weiteren Gründen nicht in den Korpus. Der erweiterte Antwortvergleich zeigt bei mehreren studienspezifischen Fragen einen genaueren Quellenbezug der RAG-Ausgaben. Zugleich treten dort Auslassungen trotz vorhandener Volltextinformationen, unbelegte Ergänzungen und eine fehlerhafte Quellenidentifikation auf. Gemini liefert teilweise unmittelbar nutzbare Vorschläge, verwechselt in einzelnen Fällen jedoch Studien oder theoretische Modelle. Beide Systeme nennen bei einer Kontrollfrage keinen erfundenen Prozentwert.

Schlussfolgerung: Die Pipeline dokumentiert einen Ansatz zur vorsichtigen Kuratierung fachlicher Wissensbestände. Nach vollständiger lokaler Einrichtung können Aufbereitung vorhandener Dokumente, Embeddings, Retrieval und Antwortgenerierung ohne Internet erfolgen. Neue Literaturakquise, Downloads und externe Aktualisierungen bleiben netzabhängig. Die zehn Fälle bilden eine explorative Fallserie; sie belegen weder eine allgemeine Modellüberlegenheit noch eine unabhängige Validierung der Pipeline. Ein kontrollierter Offline-Test wurde nicht protokolliert.

Schlagwörter: Frühpädagogik; Early Childhood Education and Care; Retrieval-Augmented Generation; lokale KI; Offline-Betrieb; Open Access; Forschungsdatenkuratierung.

Projektartefakte und Bezugsadressen

Software-Architektur: https://doi.org/10.5281/zenodo.23182511
Forschungsdatensatz: https://huggingface.co/datasets/Earlychildhoodeducation/ECE-Research-Corpus
EleMo-V3 (Modellkarte): https://doi.org/10.57967/hf/10808

Grafische Übersicht

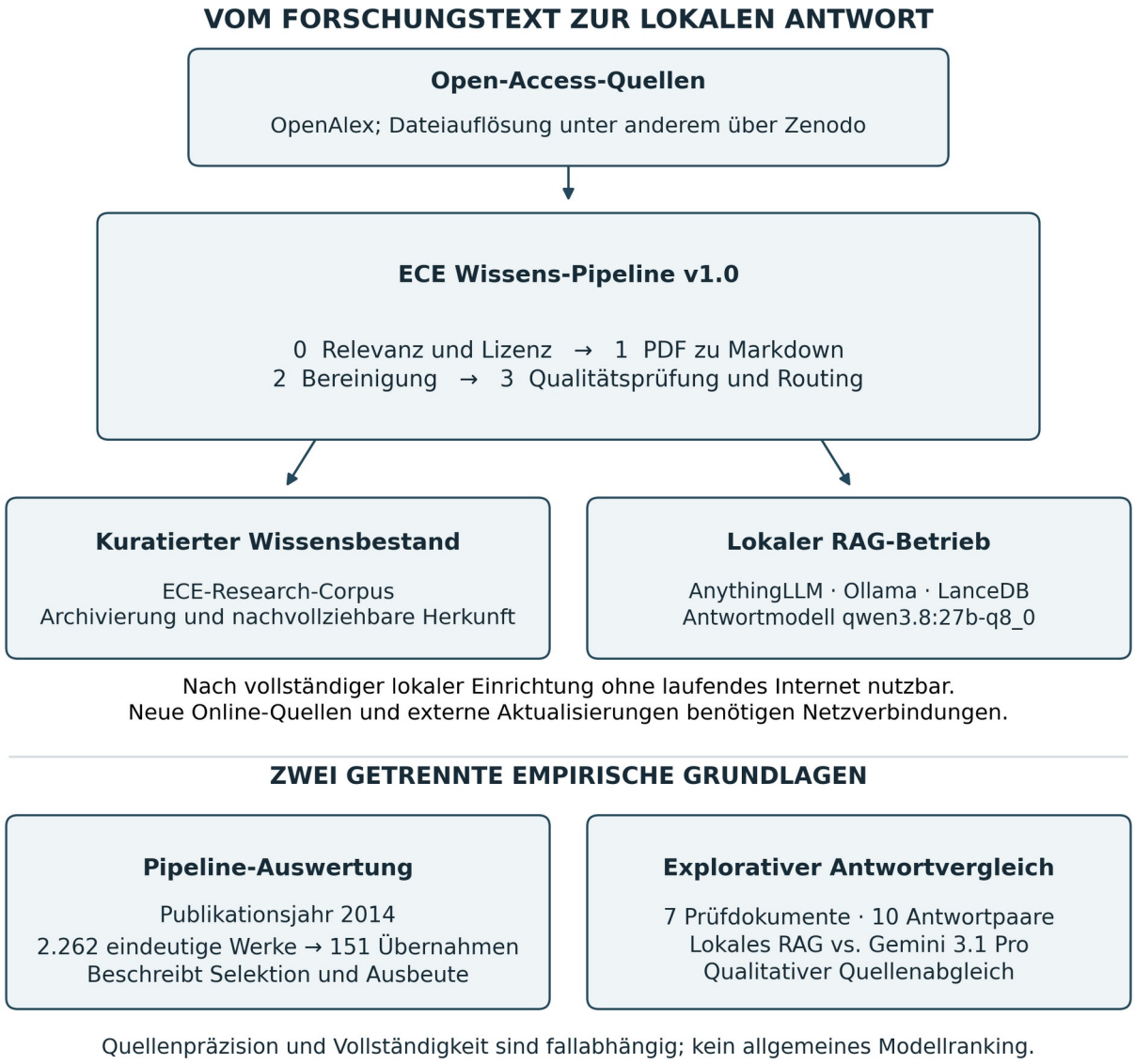


Abbildung 1. Gesamtüberblick über Architektur und Untersuchungsgrundlagen. Die 151 Übernahmen betreffen ausschließlich den Suchlauf zum Publikationsjahr 2014. Die zehn dokumentierten Antwortpaare beruhen auf einem gesonderten Prüfbestand aus sieben Dokumenten.

Die Abbildungen verdichten die dokumentierte Architektur und die explorativen Befunde. Nach vollständiger Einrichtung sind Verarbeitung vorhandener Dokumente und RAG-Antworten bei vollständig lokalen Komponenten ohne laufende Internetverbindung möglich; ein gesonderter Netztrennungstest ist nicht protokolliert.

Abstract

Background: Retrieval-augmented generation (RAG) in early childhood education and care requires reusable, domain-relevant and structurally prepared research texts. Local operation on an existing corpus is an additional architectural objective.

Methods: The paper describes a four-stage pipeline for open-access discovery, rule-based relevance and license screening, deterministic text cleaning and local model-based quality assessment. Ten documented response pairs from a local AnythingLLM RAG system and Gemini 3.1 Pro with high reasoning effort were qualitatively checked against seven source documents.

Results: A bounded search of publications from 2014 examined 2,262 unique works. Of 203 technically valid PDFs, 151 entered the corpus, 28 were routed to manual review and 24 were not included for other reasons. Several RAG responses were more specific to the requested studies, but also omitted available information, added unsupported claims or failed to identify a retrieved source. Gemini provided useful practical suggestions in some cases but confused studies or theoretical frameworks in others. Neither system invented a percentage in the false-premise control question.

Conclusion: The architecture supports offline processing and question answering once the required software, models and documents are available locally and all inference dependencies are local. Literature discovery and external updates require connectivity. The ten-case series is exploratory and does not establish general model superiority or independently validated pipeline quality. No controlled network-disconnection test was documented.

1 Einleitung und Problemstellung

Die Integration generativer Künstlicher Intelligenz in pädagogische Arbeitsfelder stellt besondere Anforderungen an die Nachvollziehbarkeit und Kontextangemessenheit zugrunde liegender Wissensbestände. In der frühkindlichen Bildung, Betreuung und Erziehung (FBBE; international Early Childhood Education and Care, ECEC) sind Fragen der Entwicklungsbegleitung, Beobachtung, Inklusion, des Kinderschutzes und der Zusammenarbeit mit Familien eng mit dem institutionellen Setting verbunden. Ohne geeignete Wissensgrundlagen besteht das Risiko, dass Sprachmodelle Konzepte aus schulischen oder hochschulischen Kontexten unpassend auf die Altersgruppe von null bis sechs Jahren übertragen. Der ECE-Research-Corpus wird vor diesem Hintergrund als domänenspezifische Dateninfrastruktur entwickelt; seine pädagogische Wirksamkeit ist damit noch nicht nachgewiesen.

Retrieval-Augmented Generation verbindet die Antwortgenerierung mit dem Abruf externer Textbelege, die dem Sprachmodell als zusätzlicher Kontext bereitgestellt werden (Lewis et al., 2020). Im hier untersuchten Aufbau bilden lokal gespeicherte Forschungsdokumente diese Wissensgrundlage.

Bei der Nutzung allgemeiner Sprachmodelle und unkuratierter Dokumentbestände für frühpädagogische Fragen sind insbesondere drei mögliche Probleme zu berücksichtigen:

Semantischer Domänendrift: Sprachmodelle können schul- oder hochschuldidaktische Begriffe und Praktiken in Antworten zu frühpädagogischen Settings übernehmen. Häufigkeit und Ausmaß dieses Effekts werden in der vorliegenden Studie nicht systematisch gemessen.

Retrieval-Verzerrung (Chunk Pollution) in RAG-Systemen: Werden PDF-Dokumente ohne strukturelle Prüfung in Vektordatenbanken überführt, können Kopfzeilen, Verlagsinformationen und Literaturverzeichnisse als wenig aussagekräftige Chunks gespeichert und abgerufen werden. Die Pipeline soll solche Artefakte reduzieren; der Effekt auf Retrieval-Kennzahlen wird hier nicht separat quantifiziert.

Urheber- und Lizenzrechtsrisiken: Frei zugängliche Publikationen sind nicht zwangsläufig unter einheitlichen Bedingungen weiterverbreitbar. Die Nutzung nicht-kommerzieller Lizenzen, von

Bearbeitungsverboten und unklar dokumentierten Rechten ist deshalb anhand des jeweiligen Nutzungszwecks und der konkreten Lizenzbedingungen zu prüfen.

Im Sinne von Data-Centric AI richtet die Arbeit den Blick auf die Auswahl, Dokumentation und Aufbereitung von Daten als Voraussetzung für die Nutzung spezialisierter Sprachmodelle und RAG-Systeme. Vorgestellt werden die Architektur der ECE Wissens-Pipeline v1.0 und der daraus entstehende ECE-Research-Corpus (Abbildung 2). Untersucht wird, welche Selektions- und Ausbeutemuster das Verfahren an einer begrenzten Suchstichprobe des Publikationsjahres 2014 erzeugt und wie sich mögliche Anwendungsunterschiede in einer explorativen Fallserie mit zehn dokumentierten Fragen zeigen. EleMo-V3 (Götz, 2026b) beruht auf einer separaten pädagogischen Datenbasis; die Verbindung dieser Trainingsbasis mit dem ECE-Research-Corpus ist für eine spätere Modellgeneration vorgesehen.

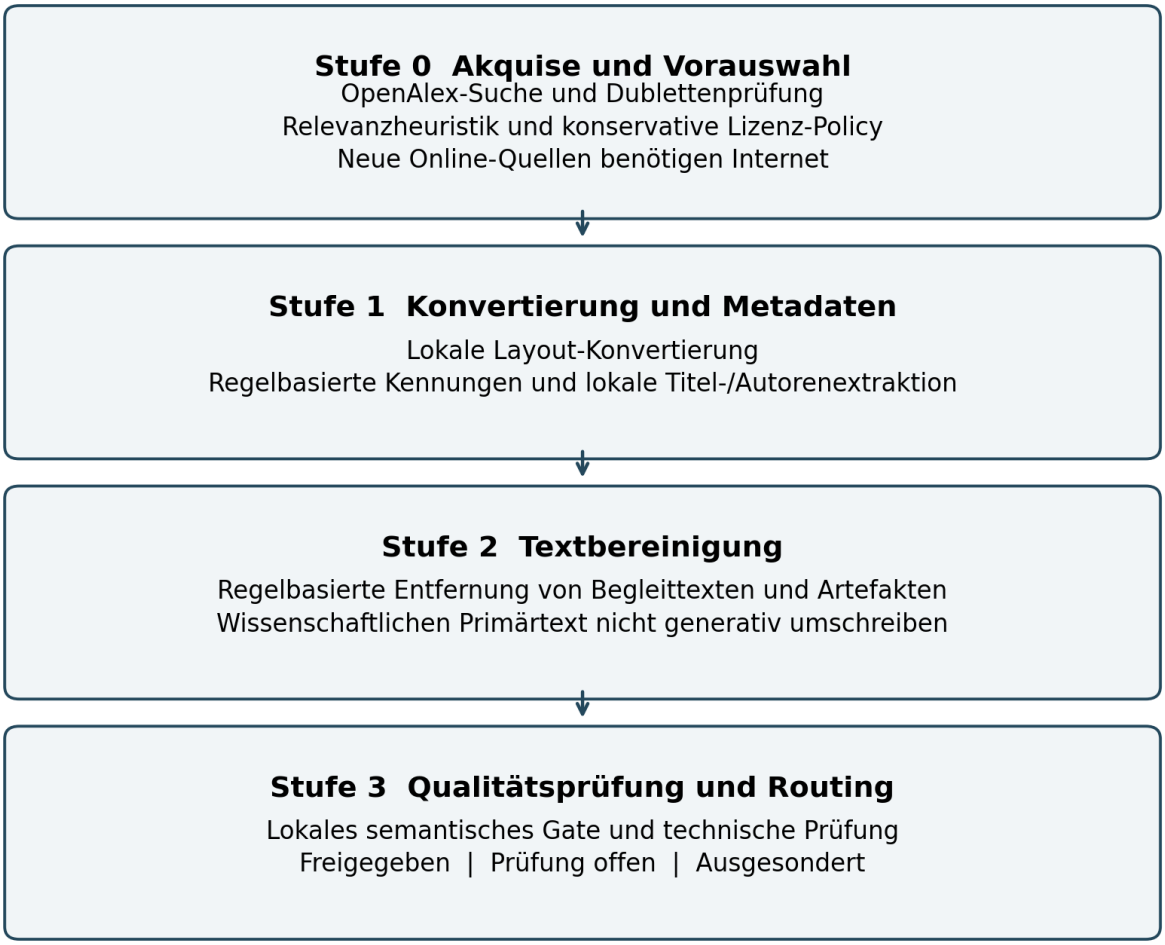


Abbildung 2. Vierstufige Pipeline mit netzabhängiger Akquise und lokal ausführbarer Verarbeitung vorhandener Dokumente. Schematische Darstellung; Konfigurationsdetails siehe Abschnitt 2.

2 Methodik und formale Systemarchitektur der ECE Wissens-Pipeline v1.0

Die ECE Wissens-Pipeline v1.0 ist als deterministisch-hybride Python-Architektur konzipiert, die über das zentrale Steuerungsmodul „Frühpädagogik Bibliothekar.py“ orchestriert wird (Abbildung 2). Als Gestaltungsprinzip gilt, den wissenschaftlichen Primärtext nicht durch ein generatives Modell zusammenzufassen oder umzuschreiben. Sprachmodelle (qwen3.8:27b) werden zur strukturierten Metadatenextraktion und Bewertung bei einer eingestellten Temperatur von T = 0 eingesetzt. Die

eigentlichen Texttransformationen erfolgen regelbasiert; auch deterministische Verfahren können jedoch Textteile fehlerhaft entfernen, weshalb eine Qualitätskontrolle erforderlich bleibt.

Das Modell qwen3.8:27b wurde wegen seiner deutsch- und englischsprachigen Verarbeitung, der Unterstützung strukturierter JSON-Ausgaben und der lokalen Ausführbarkeit ausgewählt. Ein systematischer Modellvergleich war nicht Gegenstand dieses Beitrags.

2.1 Orchestrierung, Preflight-Validierung und Identitätshierarchie

Um inkonsistente Verarbeitungszustände bei mehrstündigen Batch-Läufen zu verhindern, erzwingt der zentrale Orchestrator vor dem Start eines Komplettworkflows einen mehrstufigen Preflight-Check:

- **Umgebungs- und Komponentenprüfung:** Verifikation der vier Kernmodule, der Python- Laufzeitpakete (requests, ollama, reportlab) im aktiven Interpreter sowie der Binärdatei marker_single.
- **Dateisystem- und Registerintegrität:** Schreibtest mittels temporärer Prüfdateien (.kita_digital_write_test.tmp) in allen sechs Pipeline-Verzeichnissen, atomarer Lese-/Schreibtest des zentralen JSON-Registers (_pipeline_verarbeitet.json) und Prüfung des freien Festplattenspeichers (Warnschwelle unter 5,0 GB).
- **LLM- und Datenbank-Verfügbarkeit:** Abfrage des lokalen Ollama-Endpunkts (http://127.0.0.1:11434/api/tags) auf Vorhandensein des Modells qwen3.8:27b sowie Prüfung der laufenden DEVONthink-Instanz (Kita Digital) via AppleScript (osascript).

Über alle vier Verarbeitungsstufen hinweg wird jede Publikation in einem zentralen Register geführt. Da bibliografische Metadaten in Open-Access-Repositorien unvollständig sein können, implementiert die Pipeline eine strikte vierstufige Identitätshierarchie zur Dublettenabwehr:

$$ID(Dokument) = coalesce(OpenAlex\text{-}ID,\ DOI_{normalisiert},\ SHA\text{-}256(PDF),\ Dateiname_{normalisiert})$$

Ein schwächeres Identitätsmerkmal wird nur dann zur Schlüsselbildung herangezogen, wenn alle übergeordneten Merkmale fehlen. Sobald eine Quelle in Stufe 1 mit einer höheren Kennung (z. B. einer nachträglich erkannten OpenAlex-ID) angereichert wird, migriert Stufe 3 automatisch bestehende Registereinträge unter Erhalt sämtlicher Integrations- und Archivstatusfelder.

Abbildung 3 zeigt die Reihenfolge der Schlüsselbildung. Eine spätere stärkere Kennung führt zur beschriebenen Registermigration, nicht zur parallelen Anlage eines zweiten Dokumenteintrags.



Abbildung 3. Vierstufige Identitätshierarchie. Die Pfeile bedeuten ausschließlich den Rückgriff bei fehlender höherer Kennung; ein vorhandener, aber nicht passender höherer Schlüssel rechtfertigt keinen Wechsel zu einem schwächeren Merkmal.

2.2 Stufe 0: Mehrdimensionale Relevanzheuristik und Lizenz-Ampel

Das Akquisemodul („0. Datenbeschaffung OpenAccess.py“) durchsucht die OpenAlex-API mit dem Filter open_access.is_oa:true über 13 definierte Suchanfragen (neun deutschsprachige, vier englischsprachige)

innerhalb eines gewählten Publikationszeitraums. Die vom Dienst als abstract_inverted_index gelieferten Abstracts werden anhand der gespeicherten Wortpositionen rekonstruiert. Danach werden Titel (T), Abstract (A), Autoren-Keywords (K) und OpenAlex-Topics (P; einschließlich subfield und field) normalisiert. OpenAlex stellt hierfür einen offenen wissenschaftlichen Wissensgraphen mit API-Zugang bereit (Priem et al., 2022).

Formales Relevanz-Scoring

Ein Vorfilter verwirft generische Strukturüberschriften ($T_{norm} \in G$), etwa „Literaturverzeichnis“, „Methodik“, „Editorial“ oder „Inhaltsverzeichnis“. Für die übrigen Werke berechnet die Pipeline zwei Relevanzwerte: den Core-Score S_core und den Broad-Score S_broad.

Für jedes Suchmuster x wird nur der höchstgewichtete Trefferort berücksichtigt (Titel T vor Abstract A vor Keywords/Topics $K \cup P$). Die Funktion w(x; gT, gA, gM) gibt entsprechend das Gewicht des zuerst zutreffenden Feldes zurück:

$$\begin{aligned} w(x;\ g_T,\ g_A,\ g_M) &= g_T \quad \textit{falls } x \in T \\ &= g_A \quad \textit{falls } x \notin T \textit{ und } x \in A \\ &= g_M \quad \textit{falls } x \notin (T \cup A) \textit{ und } x \in (K \cup P) \\ &= 0 \quad \textit{andernfalls} \end{aligned}$$

Expliziter Frühpädagogik-Modus (S_core)

Das System prüft drei Mengen von Kernmustern:

- **M_stem:** 11 Wortstamm-Einträge (sechs frühpädagogische Kernstämme einschließlich Umlautvarianten, z. B. frühpädagog* / fruehpaedagog*, kindheitspädagog*, elementarpädagog*, krippenpädagog*, vorschulpädagog*, kindertages*).
- **M_phrase_stem:** acht zweigliedrige Wortstamm-Tupel (vier semantische Begriffspaare einschließlich Umlautvarianten, z. B. „früh“ und „bildung“), erkannt über einen Token-Distanz- Algorithmus (max_gap = 1). Dies erweitert die Erkennung um Flexionsformen und eingeschobene Wörter.
- **M_core:** 35 exakte Fachphrasen (elf deutsch- und 24 englischsprachige Bezeichnungen wie kindertagesbetreuung, kinderkrippe, early childhood education and care, preschool teacher); ergänzt um vier unterstützende Begriffe M_sup (early childhood, early years, preschool, pre school). Der Gesamtscore ergibt sich aus:

$$S_{core} = \sum_{x \in (M_{stem} \cup M_{phrase_stem} \cup M_{core})} w(x;\ 6,\ 4,\ 1) + \sum_{y \in M_{sup}} w(y;\ 2,\ 1,\ 1)$$

Breiter Kita-Kontext-Modus (S_broad)

Für Studien mit spezifischem Elementarbereichsbezug, die keinen expliziten frühpädagogischen Kernbegriff führen, werden drei Dimensionen ausgewertet: Kinder-/Familienbezug (C_child), pädagogischer Kontext (C_ped) und spezifische Kita-Fachthemen (C_topic; mehr als 100 Terme, unter anderem Sprachförderung, Inklusion, Kinderschutz, Transition, Interaktionsqualität, Professionalisierung, Motorik, Gesundheit und Medienpädagogik).

$$S_{broad} = \sum_{x \in C_{child}} w(x;\ 3,\ 2,\ 1) + \sum_{y \in C_{ped}} w(y;\ 3,\ 2,\ 1) + \sum_{z \in C_{topic}} w(z;\ 3,\ 2,\ 1)$$

Negativkontext-Sperre und Entscheidungsregel

N bezeichnet Ausschlussbegriffe aus Primar- und Sekundarstufe, Hochschule, Lehramtsausbildung, Erwachsenenbildung und Geragogik (NEGATIVE_CONTEXT_TERMS). Ein Negativtreffer im Titel blockiert die Aufnahme (B_neg = wahr), sofern der Titel nicht zugleich ein explizites Kita-Setting (E_kita = wahr) nennt.

Die positive Entscheidungsregel lautet: Relevant genau dann, wenn B_neg = falsch und mindestens eine Bedingung erfüllt ist: (a) S_core ≥ 4 mit explizitem Kernmuster in Titel oder Abstract; oder (b) S_broad ≥ 7 mit Kinder-/Familien- und Fachthemenbezug in Titel oder Abstract sowie pädagogischem Kontext in Titel, Abstract, Keywords oder Topics.

Die Regel reduziert Aufnahmen, die allein auf breiten Metadaten-Tags beruhen sollen. Ob dabei einschlägige Arbeiten ausgeschlossen werden, muss durch eine unabhängige Stichprobenprüfung untersucht werden.

Fuzzy-Titel-Deduplizierung und Zenodo-Priorisierung

Zur Erkennung möglicher Preprint-Dubletten werden Titel mittels NFKD-Unicode-Normalisierung, Entfernung diakritischer Zeichen und ß→ss-Konvertierung verglichen. Bei Titeln ab 50 Zeichen und einem Verhältnis kürzerer zu längerer Titellänge von mindestens 0,88 wird der SequenceMatcher eingesetzt; ab Similarity ≥ 0,97 wird ein Werk als potenzielles Titel-Duplikat verworfen.

Bei Zenodo-Datensätzen löst ein spezialisierter Resolver hinterlegte Dateien über die Zenodo-API auf und berechnet einen Dateiprioritätswert P_Zenodo. Damit sollen Hauptmanuskripte von Fragebögen und Anhängen unterschieden werden:

$$P_{Zenodo} = 100 \cdot I(\text{positiver Dateinamensbegriff}) - 100 \cdot I(\text{negativer Dateinamensbegriff}) + min(\text{Dateigröße in MiB}, 50)$$

Dabei ist I eine Indikatorfunktion: I(Bedingung) = 1, wenn die Bedingung erfüllt ist, andernfalls 0.

Konservative Lizenz-Ampel (classify_license)

Die Pipeline klassifiziert gefundene PDF-Quellen anhand maschinenlesbarer Lizenzmetadaten in drei operative Kategorien (Abbildung 4). Diese Einordnung ist eine konservative Auswahlregel, keine rechtsgutachterliche Prüfung:

- **Grün (green):** Public Domain, CC0, CC BY und CC BY-SA. Bei Zenodo-Records mit mehreren Lizenzangaben müssen sämtliche Einträge grün sein.
- **Gelb (yellow):** Fehlende oder widersprüchliche Lizenzdaten, unspezifische Open-Access-Labels (z. B. other-oa, publisher-specific-oa), Bearbeitungsverbote (CC BY-ND), benutzerdefinierte Lizenzen sowie nicht ohne Weiteres zuzuordnende Software-Lizenzen (MIT, Apache, GPL, BSD).
- **Rot (red):** Unter der gewählten konservativen Policy unter anderem CC BY-NC sowie proprietäre Vorbehalte („All rights reserved“). Die Zuordnung bedeutet nicht, dass jede wissenschaftliche Verarbeitung solcher Werke unzulässig wäre. Ausschließlich grüne Quellen werden mit einer Rate-Limiting-Pause (DOWNLOAD_DELAY_SECONDS = 2.0) heruntergeladen, auf HTTP-Status 200, die %PDF-Signatur in den ersten 2.048 Bytes sowie eine Mindestgröße von 1.000 Bytes geprüft und unmittelbar per SHA-256 gehasht.

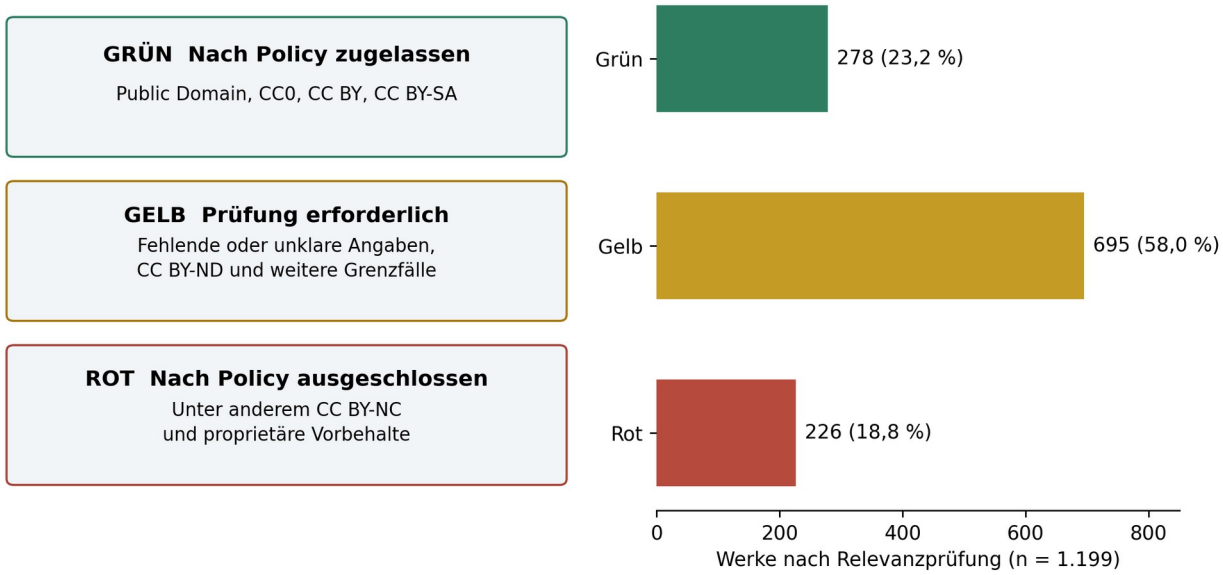


Abbildung 4. Operative Lizenz-Policy und Verteilung der 1.199 als relevant klassifizierten Werke. Die Kategorien sind Auswahlentscheidungen und keine rechtlichen Garantien.

2.3 Stufe 1: Layout-Konvertierung und deterministisch-hybride Metadatenextraktion

In Stufe 1 (1. PDF Umwandeln in Markdown.py) steuert ein ThreadPoolExecutor (PIPELINE_WORKERS = 3) mit dedizierten Semaphoren die rechenintensive Layout-Analyse via marker_single (MAX_PARALLEL_MARKER = 2) und die lokale LLM-Inferenz (MAX_PARALLEL_OLLAMA = 3).

Um die für akademische Korpora wichtige Provenienz-Dokumentation zu unterstützen, ist die Metadatengenerierung strikt zweigeteilt:

- **Deterministische Extraktion (Python):** Publikationsjahr und OpenAlex-ID werden regulär aus dem Dateinamen gelesen. DOIs werden über den regulären Ausdruck `10\.\d{4,9}/[^\s<>"\]\[()]+` ausschließlich aus den ersten 20.000 Zeichen des Dokuments extrahiert, um das Risiko zu reduzieren, dass zitierte DOIs aus dem Literaturverzeichnis am Dokumentende fälschlich als Dokument-DOI erfasst werden. Werden mehrere DOIs gefunden, trennt ein Klassifikator bekannte Repository-DOIs (Präfixe 10.25656/ für peDOCS und 10.5281/zenodo.) als repository_doi vom primären Verlags-DOI (doi) und speichert etwaige weitere Kennungen unter weitere_dois. Lizenzdaten werden ausschließlich aus dem verifizierten Sidecar bzw. Lizenzregister aus Stufe 0 übernommen und niemals aus dem PDF-Text geraten.
- **Semantische Extraktion (qwen3.8:27b):** Das lokale Sprachmodell erhält lediglich die ersten 5.000 Zeichen des Markdown-Rohtextes (OLLAMA_NUM_CTX = 4096, OLLAMA_NUM_PREDICT = 200, temperature = 0, think = False) und ist über Structured Outputs (HEADER_SCHEMA) technisch darauf beschränkt, ein JSON-Objekt mit genau zwei Feldern (titel, autor) zurückzugeben. Kann kein Titel sicher erkannt werden („Unbekannt“), greift ein deterministischer Dateinamen- Fallback. Der Haupttext bleibt in Stufe 1 vollständig unverändert.

2.4 Stufe 2: Regelbasierte Bereinigung zur Vermeidung von Chunk-Pollution

Stufe 2 (2. Markdown verbessern.py) bereinigt den Markdown-Körper deterministisch von nicht-wissenschaftlichen Begleittexten:

- **Konvertierungs- und Layout-Artefakte:** Vollständige Entfernung von HTML-Seitenankern (), lokalen Bildreferenzen (_page_…), leeren Bildtags und internen #page-Links.

- **Repository- und Verlags-Frontmatter:** Algorithmische Erkennung und Entfernung von Self- Archive-Vorspännen, mehrteiligen peDOCS-Deckblättern (Quellenangabe, Kooperationshinweis, Nutzungsbedingungen, Kontaktblock), DJI-Werbeblöcken, Impressums-Sektionen, Article- History-Blöcken sowie 15 Kategorien typischer Verlagsnavigationszeilen (NOISE_LINE_PATTERNS).
- **Thesis-Frontmatter-Reduktion:** Bei Qualifikationsarbeiten (Bachelor-/Masterarbeiten, Dissertationen, Forschungsberichten) prüft das System den Bereich vor dem Abstract (maximal bis Zeile 600) auf neun spezifische Muster (THESIS_SIGNAL_PATTERNS, z. B. submitted to, partial fulfillment, under the supervision of, begleitende lehrperson). Treffen mindestens zwei Signale zu, wird der gesamte administrative Vorspann vor dem Abstract entfernt und der aus dem YAML gesicherte Haupttitel als # {yaml_title} vorangestellt.
- **Deduplizierung von Running Headers und Titelblöcken:** Wiederholte Kopfzeilen (mindestens zwei Vorkommen mit Fachbegriffen wie conference, proceedings, journal oder volume und maximal 220 Zeichen) sowie doppelte initiale Titel-/Autorenblöcke (Titelähnlichkeit mindestens 0,94 innerhalb der ersten 40 Zeilen vor Abstract/Einleitung) werden bereinigt.
- **Selektive Literaturverzeichnis-Entfernung mit Anhang-Schutz:** Literaturverzeichnisse (REFERENCE_NAMES in Deutsch und Englisch) werden sowohl über die Markdown-Überschriftenhierarchie als auch über eine Plain-Text-Suche in den hinteren 60 % des Dokuments identifiziert und entfernt. Folgt nach dem Literaturverzeichnis jedoch ein wissenschaftlicher Anhang (Appendix, Anhang, Annex), erkennt der Parser diese Überschrift und erhält den Anhang vollständig.
- **Konservative Satzreparatur und DOI-Korrektur:** Getrennte Fließtextzeilen werden verbunden, wenn die Folgezeile mit einem Kleinbuchstaben ([a-zäöüß]) beginnt, die vorherige Zeile kein Satzendezeichen ([.!?;:]) aufweist und keine Tabellen-, Listen-, Überschriften- oder Fußnotenmarkierung vorliegt. Abgeschnittene Haupt-DOIs im YAML werden nur korrigiert, wenn repository_doi eine eindeutige, um mindestens vier Zeichen längere Verlängerung desselben DOI-Stamms enthält.

2.5 Stufe 3: Semantisches RAG-Gate, 3-Wege-Routing und Archivierung

Das Abschlussmodul (3. Markdown Überprüfer.py) führt zunächst eine finale RAG-Optimierung durch (NFC-Unicode-Normalisierung, Entfernung unsichtbarer Steuerzeichen, isolierter Seitenzahlen, reiner URL-/E-Mail-Zeilen, Umwandlung externer Links in reinen Sichttext sowie Löschung exakt identischer Textabsätze ab 180 Zeichen Länge).

Anschließend bewertet qwen3.8:27b (OLLAMA_NUM_CTX = 8192, temperature = 0) anhand der ersten 12.000 Textzeichen und bis zu 40 extrahierter Überschriften die Quelle über das Schema RAG_ASSESSMENT_SCHEMA. Die Ergebnisse werden direkt in den YAML-Header zurückgeschrieben (rag_relevanz, rag_wissenschaftlichkeit, rag_dokumenttyp, rag_zielbezug, rag_themen). Ist der Autor im YAML noch als „Unbekannt“ vermerkt, wird er nur dann durch den vom Modell vorgeschlagenen autor_kandidat ersetzt, wenn das Modell seine Sicherheit explizit als autor_sicherheit == "hoch" einstuft.

Parallel prüft der deterministische Validator check_file die YAML-Pflichtfelder, die gewählte Lizenz-Policy (lizenz_ampel == „gruen“), die formale DOI-Integrität, die Mindesttextlänge (Warnung unter 1.500 Zeichen, Fehler unter 500 Zeichen) und verbliebene Artefakte. Zur Reduktion von Fehlalarmen bei Thesis-Frontmatter unterscheidet das System schwache und starke administrative Signale; nur kombinierte Signale lösen eine Warnung aus.

Darauf basierend erfolgt ein transaktionales 3-Wege-Routing:

- **Stufe 3.1 (freigegeben / OK):** Fehler- und warnungsfreie Dokumente mit hoher Relevanz werden freigegeben, in die deduplizierte Import-Warteschlange 4. AnythingLLM – NEU kopiert und via AppleScript in die DEVONthink-Datenbank Kita Digital importiert – einerseits als 3-Datei- Archivpaket (Original.pdf, RAG-Version.md, Original.metadata.json) unter /Akademische Quellen/{Jahr}/{Kurztitel – OpenAlex-ID} und andererseits als flache Kopie in /Akademische Quellen/RAG Markdowns Gesamt. Bei

späteren Titeländerungen ersetzt der Synchronisationsalgorithmus bestehende Einträge in RAG Markdowns Gesamt anhand der stabilen Kennung erst nach erfolgreichem Neuimport, sodass kein Datenverlust eintreten kann.

- **Stufe 3.2 (pruefung_offen):** Dokumente mit technischen Warnungen, formalen Fehlern oder mittlerer Relevanz (RAG_RELEVANCE_MEDIUM) werden zur manuellen Kuratierung ausgesondert. Mittels reportlab generiert das System automatisch einen strukturierten PDF- Prüfbericht (_Pruefbericht_Dateien_mit_Fehlern.pdf), der jeden technischen Fehlercode in eine verständliche Diagnose und Handlungsanweisung übersetzt.
- **Stufe 3.3 (rejected / DAUERHAFT_AUSSORTIERT):** Dokumente mit RAG_RELEVANCE_LOW oder RAG_SCIENTIFIC_QUALITY_LOW werden im YAML mit Aussortiergrund und Datum markiert und als komplettes Dateipaket (Markdown, Original-PDF und .metadata.json) mit Best-Effort- Rollback-Schutz dauerhaft in das Negativarchiv verschoben. Schlägt die Paketzuordnung fehl (etwa weil die Original-PDF fehlt), greift ein Sicherheits-Fallback nach Stufe 3.2.

Abbildung 5 fasst die drei Zielzustände des Gates zusammen. Sie zeigt die berichteten Zuordnungsregeln; die Behandlung gleichzeitig zutreffender Fehler- und Ablehnungsbedingungen lässt sich ohne Prüfung des ausführbaren Routings nicht weiter präzisieren.

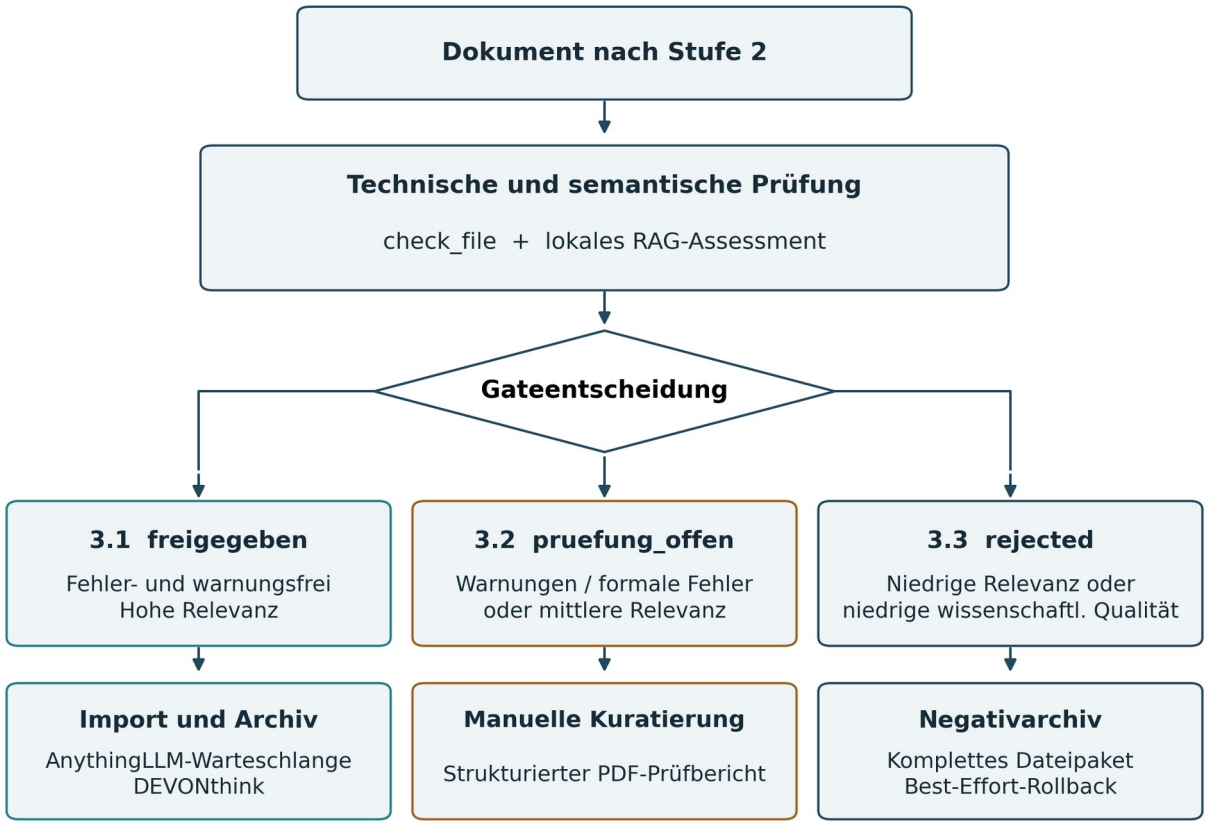


Abbildung 5. Schematischer Entscheidungsbaum des 3-Wege-Routings. Die drei Äste sind Zielzustände, keine zeitliche Folge. Bei gescheiterter Paketzuordnung für das Negativarchiv greift der dokumentierte Fallback nach 3.2; eine Prioritätsregel für kollidierende Prüfbefunde wird hier nicht ergänzt.

2.6 Lokaler Betrieb nach der Einrichtung

Ein wesentliches Architekturmerkmal ist die Trennung zwischen netzabhängiger Beschaffung und lokalem Wissensbetrieb. Nach der Installation aller benötigten Programme, dem Download der Modellgewichte und der Bereitstellung der Dokumente kann die Verarbeitung des vorhandenen Bestands ohne laufende Internetverbindung erfolgen. Dazu gehören die Konvertierung und Bereinigung bereits gespeicherter PDFs,

die lokale Metadaten- und Qualitätsprüfung sowie der anschließende RAG-Betrieb. Eingesetzt wurde ein MacBook Pro mit Apple M3 Max und 128 GB gemeinsamem Arbeitsspeicher (Unified Memory). Ollama führte die KI-Modelle lokal aus; AnythingLLM diente als Chatoberfläche mit RAG, DEVONthink der privaten Dokumentenarchivierung. Diese Ausstattung beschreibt die untersuchte Arbeitsumgebung und keine ermittelte Mindestanforderung.

Für eine vollständig lokale Antwortkette müssen neben dem generierenden Sprachmodell auch das Embedding-Modell für Dokumente und Anfragen, die Vektordatenbank sowie gegebenenfalls ein Reranker lokal verfügbar sein. Websuche, externe Agentenwerkzeuge und Cloud-Fallbacks dürfen nicht Bestandteil dieser Antwortkette sein. Die AnythingLLM-Dokumentation beschreibt lokalen Offline-Betrieb für die Desktop-Anwendung und unterscheidet lokale von Cloud-Embedding-Anbietern (Mintplex Labs, o. J.-a, 2025). Ollama dokumentiert einen lokalen Betriebsmodus mit deaktivierten Cloud-Funktionen (Ollama, o. J.). Die konkrete Anbieterwahl bleibt daher Teil der Systemkonfiguration.

Tabelle 1. Netzabhängigkeiten der Verarbeitungsschritte.

Schritt	Offline möglich	Voraussetzung oder Grenze
Installation und erstmalige Modellbereitstellung	Nach lokaler Bereitstellung	Downloads benötigen Internet; vorbereitete Installationsdateien und Gewichte können auch lokal übertragen werden.
Literatursuche und neue PDF-Downloads	Nein, bei Online-Quellen	OpenAlex, Zenodo und externe Verlagsserver werden über das Internet angesprochen.
Stufen 1 bis 3 für vorhandene Dateien	Ja, bei lokaler Konfiguration	Konverter, OCR- und Sprachmodelle sowie benötigte Ressourcen müssen bereits lokal vorhanden sein.
Embedding und Indexierung	Ja, bei lokalem Anbieter	Gilt auch für das Embedding neuer Nutzerfragen; ein Cloud-Embedder würde eine Netzabhängigkeit erzeugen.
Retrieval und Antwortgenerierung	Ja, bei lokaler Konfiguration	Lokale Vektordatenbank, lokales Antwortmodell und gegebenenfalls lokaler Reranker. Keine externen Werkzeuge.
Veröffentlichung und Aktualisierung	Für externe Dienste nein	Hugging-Face-Uploads, Online-Synchronisation sowie Software- und Modelldownloads sind netzabhängig.

In dieser Konfiguration werden Fachfragen, abgerufene Textstellen und Antworten für die eigentliche Inferenz lokal verarbeitet. Eine Nutzung ist dadurch auch bei fehlender Internetverbindung möglich; sie bleibt auf den lokal vorliegenden Wissensstand begrenzt. Neue Publikationen gelangen erst nach einer erneuten Akquise oder einem lokalen Dateiimport in den Bestand. Für ausgehende Verbindungen sind zusätzlich optionale Telemetrie, Updateprüfungen und Synchronisationsdienste zu berücksichtigen. Die Offline-Fähigkeit ist eine technische Betriebsoption und ersetzt keine organisatorische oder rechtliche Bewertung der jeweiligen Nutzung.

Der lokale Aufbau ist durch die dokumentierte Hardware- und Softwarekonfiguration beschrieben. Ein kontrollierter Kaltstart mit getrennter Internetverbindung, protokollierter Netzwerksperre und überprüften Anbieter-Einstellungen wurde nicht aufgezeichnet. Ein solcher Test sollte in einem Folgelauf auch Anfrage-Embedding, Quellenabruf und Antwortgenerierung erfassen. Systematische Messungen zu Antwortlatenz, Token-Durchsatz und Speicherauslastung liegen nicht vor. Für einen produktiven Offline-Einsatz sind die lokale Ausstattung und die Zahl gleichzeitiger Nutzungen mit den konkret gewählten Modellen und Dokumentbeständen zu bemessen; aus dieser Einzelnutzung folgt keine allgemeine Aussage zu akzeptablen Antwortzeiten.

3 Empirische Fallstudie und explorative Anwendungsevaluation

Um das Selektionsverhalten und die Ausbeute (Yield Rate) zu beschreiben, wurde ein begrenzter Such- und Verarbeitungslauf mit Publikationen aus dem Jahr 2014 (1. Januar bis 31. Dezember) ausgewertet. Pro Suchvektor wurden maximal 500 Treffer betrachtet (MAX_RESULTS_PER_QUERY = 500), und die Zahl neuer Downloads war auf 250 begrenzt (MAX_NEW_DOWNLOADS = 250). Die Auswertung bildet damit keine vollständige Erhebung aller einschlägigen Werke des Publikationsjahres ab.

3.1 Verteilung der Suchvektoren und Deduplizierung

Tabelle 2 zeigt die von OpenAlex angezeigten Suchtreffer, die im Lauf betrachteten Datensätze und die nach suchlaufübergreifender OpenAlex-ID-Deduplizierung verbleibenden eindeutigen Werke.

Tabelle 2. Abfragestatistik zum Publikationsjahr 2014. * Durch das Limit von 500 betrachteten Treffern gedeckelt.

Nr.	Suchvektor	Sprache	OpenAlex	Betrachtet	Neu	Redundanz
1	Frühpädagogik	DE	37	37	37	0,0 %
2	Kindheitspädagogik	DE	1	1	1	0,0 %
3	frühkindliche Bildung	DE	118	118	103	12,7 %
4	frühe Bildung	DE	854	500*	442	11,6 %
5	Kindertagesbetreuung	DE	70	70	34	51,4 %
6	Kindertageseinrichtung	DE	61	61	27	55,7 %
7	Kita Pädagogik	DE	65	65	28	56,9 %
8	Kinderkrippe Pädagogik	DE	7	7	4	42,9 %
9	Kindergarten Pädagogik	DE	146	146	87	40,4 %
10	early childhood education	EN	35.874	500*	498	0,4 %
11	early childhood education and care	EN	22.937	500*	270	46,0 %
12	early years education	EN	135.288	500*	295	41,0 %
13	preschool education	EN	11.191	500*	436	12,8 %
Σ	13 Suchvektoren	DE / EN	206.649	3.005	2.262	24,7 %

Die Suchstichprobe weist zwei für die Interpretation wesentliche Eigenschaften auf:

- **Sprachraum- und Rankingeffekte:** Acht der neun deutschsprachigen Suchanfragen blieben unterhalb des Limits von 500; bei mehreren englischsprachigen Suchanfragen wurden nur die ersten 500 Ergebnisse berücksichtigt. Die in OpenAlex angezeigten Trefferzahlen sind nicht mit der vollständig untersuchten Grundgesamtheit gleichzusetzen. Im Produktivbetrieb wurden höhere Limits konfiguriert; diese Werte sind nicht Bestandteil der hier ausgewerteten Stichprobe.
- **Suchüberlappung:** Von 3.005 betrachteten Treffern waren 743 bereits über einen vorherigen Suchvektor erfasst (24,7 % Redundanz). Die frühe ID-Deduplizierung vermeidet Mehrfachverarbeitung innerhalb des dokumentierten Laufs.

3.2 Mehrstufige Selektionskaskade (Attritions- und Qualitäts-Funnel)

Tabelle 3 und Abbildung 6 dokumentieren die algorithmische Selektionskaskade vom Ausgangsbestand von 2.262 eindeutigen OpenAlex-Werken bis zu 151 in den Korpus übernommenen Dokumenten.

Tabelle 3. Selektions- und Verarbeitungsfunnel. Prozentwerte der Vorstufe beziehen sich jeweils auf die zugehörige Eingangsstufe.

Phase oder Entscheidung	Werke	Anteil an Gesamt	Anteil an Eingangsstufe
Eindeutige Werke	2.262	100,0 %	100,0 %
Thematisch ausgeschlossen (darunter 5 generische Titel)	−1.056	46,7 %	46,7 %
Weitere Dubletten	−7	0,3 %	0,3 %
Eingang in die Lizenzprüfung	1.199	53,0 %	53,0 %
Lizenz gelb	−695	30,7 %	58,0 %
Lizenz rot	−226	10,0 %	18,8 %
Nach Lizenz-Policy zugelassen	278	12,3 %	23,2 %
Technischer Drop-out	−75	3,3 %	27,0 %
Technisch valide PDF-Downloads	203	9,0 %	73,0 %
Manuelle Prüfung	−28	1,2 %	13,8 %
Weitere Nichtaufnahmen	−24	1,1 %	11,8 %
Finaler Korpusbestand	151	6,7 %	74,4 %

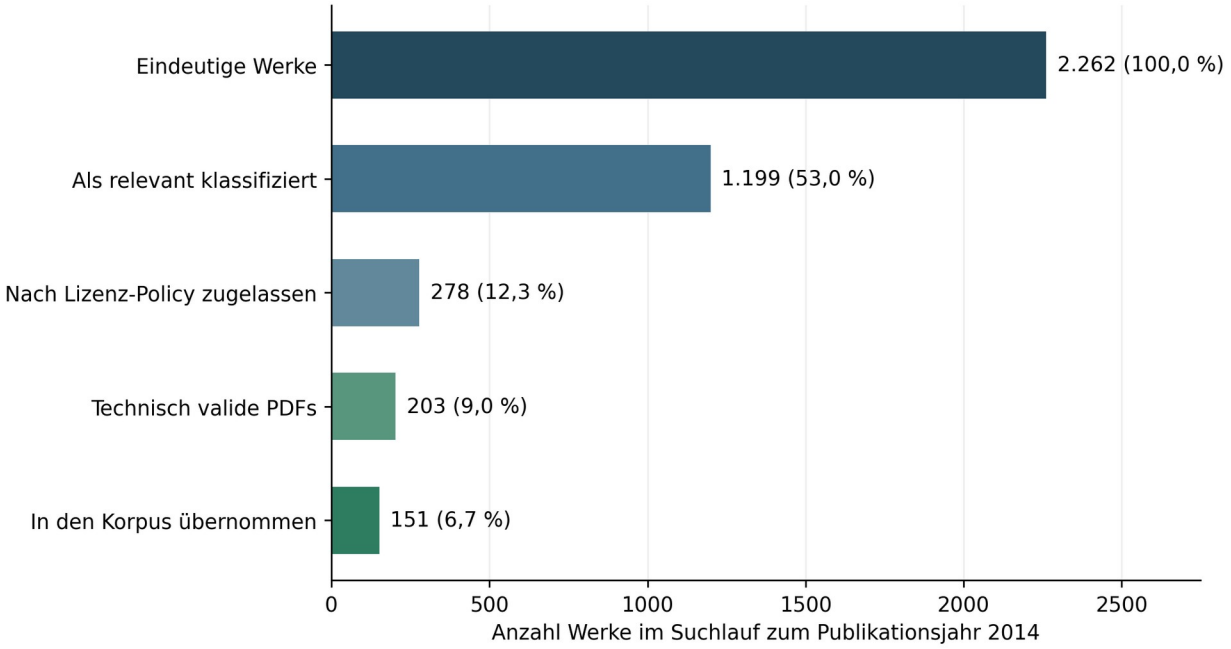


Abbildung 6. Ausbeute der Suchstichprobe 2014. Von 203 PDFs gingen 28 in die manuelle Prüfung; 24 wurden aus weiteren Gründen nicht übernommen.

Wissenschaftliche Interpretation der empirischen Befunde

Aus dem dokumentierten Lauf lassen sich folgende deskriptive Befunde ableiten:

1. Algorithmische thematische Aussonderung (46,7 %): Der regelbasierte Filter ordnete 1.056 von 2.262 eindeutigen Werken als nicht ausreichend einschlägig ein. Dies beschreibt die Selektionsentscheidung der Pipeline, nicht eine unabhängig bestätigte Quote tatsächlich fachfremder Publikationen.

2. Konservative Lizenzklassifikation (76,8 %): Von 1.199 in der Relevanzprüfung verbliebenen Werken erhielten 695 (58,0 %) eine gelbe und 226 (18,8 %) eine rote Klassifikation. 278 Werke (23,2 %) erfüllten die für diesen Lauf festgelegten grünen Lizenzkriterien. Die Verteilung verdeutlicht vor allem, wie stark die gewählte Lizenz-Policy die Datenabdeckung einschränkt; sie beweist keine juristische Fehlerfreiheit.

3. Weiterverarbeitung und manuelle Prüfung: Von 203 erfolgreich heruntergeladenen, technisch validierten PDFs wurden 28 aufgrund mittlerer thematischer Relevanz (RAG_RELEVANCE_MEDIUM) dem Prüfpfad 3.2 zugeordnet. Weitere 24 wurden wegen technischer Filter, Textlänge oder niedriger Relevanz-/Qualitätsbewertung nicht übernommen. Die verbleibenden 151 Dokumente wurden als Fachbibliothek 2014 in den ECE-Research-Corpus importiert. Die inhaltliche Korrektheit dieser Einzelfallentscheidungen wurde nicht unabhängig gemessen.

3.3 Explorative Fallserie mit zehn Fachfragen

3.3.1 Material und Vergleichsbedingungen

Die explorative Fallserie umfasst zehn dokumentierte Vergleichspaare. Das Protokoll RAGvsGemini3.1pro.docx enthält für jede Frage eine Ausgabe von AnythingLLM mit lokalem RAG und eine Ausgabe von Gemini 3.1 Pro mit der Einstellung „Denkaufwand hoch“. Die sieben zur Fragenentwicklung verwendeten Markdowns waren im lokalen RAG-Bestand enthalten. Dieser Bestand ist nicht mit der 2014er-Kohorte aus Abschnitt 3.1 und 3.2 gleichzusetzen.

Die Fragen wurden vor den Vergleichsläufen mit generativer KI aus den sieben Dokumenten entwickelt. Sie prüfen Faktenabruf, Verständnis von Studiendesigns, Praxisübertragung, quellenübergreifende Ableitungen und den Umgang mit einer falschen Fragevoraussetzung. Die Auswahl ist bewusst korpusbezogen und keine repräsentative Stichprobe frühpädagogischer Alltagsfragen. Die Kontrollfrage 10 unterstellt ein zwölfwöchiges Coachingprogramm, das die bezeichnete Studie nicht untersucht.

Tabelle 4. Dokumentationsstand des Systemvergleichs mit zehn Antwortpaaren.

Merkmal	Dokumentierter Stand
Fälle	Zehn Fragen mit je einer dokumentierten RAG- und Gemini-Antwort; Laufzeiten und Wiederholungen nicht separat protokolliert.
Lokale Anwendung	AnythingLLM 1.18.0 mit Ollama auf MacBook Pro (M3 Max, 128 GB); sieben Prüfdokumente im lokalen RAG-Bestand.
Lokales Antwortmodell	qwen3.8:27b-q8_0; Quantisierung Q8_0; konfiguriertes Kontextfenster 16.384 Tokens. Nachträglich dokumentiert; kein Modell-Tag je Fall im Protokollheader.
Retrieval	Embedding-Modell, LanceDB, Chunking, Snippet-Limit und Schwellenwert nachträglich dokumentiert (Tabelle 5); zeitgleiche Kontext- und Indexprotokolle fehlen.
Cloud-Modell	Gemini 3.1 Pro, Einstellung „Denkaufwand hoch“; genaue technische Modellkennung nicht protokolliert.
Cloud-Zusatzfunktionen	Websuche, Grounding, Dateizugriff, Personalisierung und Chat-Historie nicht dokumentiert.
Vergleichskontrolle	Keine dokumentierte Bedingung mit demselben Modell ohne Retrieval; keine Wiederholungsmessungen oder verblindete Fachbewertung.
Offline-Nachweis	Lokale Architektur und Konfiguration dokumentiert; kein gesondertes Testprotokoll bei getrennter Internetverbindung.

Die Ausgaben sind keine kontrollierte Ablationsstudie. AnythingLLM bezeichnet die Anwendung; als lokales Antwortmodell war in Ollama qwen3.8:27b-q8_0 mit Q8_0-Quantisierung und einem konfigurierten Kontextfenster von 16.384 Tokens eingestellt. Davon zu unterscheiden sind das Embedding-Modell qwen3-embedding:8b und das in Abschnitt 2 beschriebene Pipeline-Gate mit eigenständigen Laufparametern. Die nachträglich dokumentierte Konfiguration und der Systemprompt stehen in Abschnitt 3.3.1.1. Das Antwortprotokoll speichert das Modell-Tag nicht je Fall im Header; ein exportierter Korpus- und Indexstand, Retrieval-Rohprotokolle und wiederholte Läufe liegen nicht vor.

Für Gemini sind Websuche beziehungsweise Grounding, Dateizugriff und der Zustand der Chat-Historie nicht protokolliert. In Fall 1 erscheint die isolierte Bezeichnung „My Teaching Cupboard“; Fall 10 enthält Bezüge zur Kita Digital Akademie und zu einem Buchprojekt. Auch die RAG-Ausgaben enthalten teilweise projektbezogene Zusätze. Solche Textspuren belegen keine bestimmte Werkzeugeinstellung, sprechen aber für die Berücksichtigung möglicher zusätzlicher Kontexteinflüsse. Die Cloud-Bedingung wird daher nicht als nachgewiesen retrievalfreie Baseline bezeichnet (Abschnitt 5.5).

3.3.1.1 Lokale RAG-Konfiguration

Die Einstellungen in Tabelle 5 wurden nachträglich aus den lokalen Systemeinstellungen und der Ollama-Laufzeitanzeige dokumentiert. Ein zeitgleich gespeicherter Konfigurationsexport zum Zeitpunkt der Testläufe liegt nicht vor. Das Modell-Tag, die Quantisierung und das konfigurierte Kontextfenster sind damit benannt; der exakte Index- und Laufzustand jedes Antwortpaars lässt sich jedoch nicht rückwirkend rekonstruieren. Der Systemprompt ist vollständig in Anhang C wiedergegeben.

Tabelle 5. Lokale RAG-Konfiguration. Nachträgliche Dokumentation der Systemeinstellungen und Ollama-Laufzeitanzeige; kein zeitgleicher Export der Testläufe.

Parameter	Dokumentierte Einstellung und Einordnung
Anwendung	AnythingLLM 1.18.0; lokale Chatoberfläche mit RAG.
Vektordatenbank	LanceDB.
Lokales Antwortmodell	qwen3.8:27b-q8_0; Quantisierung Q8_0; angezeigte Modell-ID 8f5fb6b71ea0.
Kontextfenster des Antwortmodells	16.384 Tokens (konfigurierter Laufzeitwert). Keine Messung der tatsächlich belegten Kontextlänge.
Ollama-Statusanzeige	30 GB; 100% GPU; llamacpp; Forever. Statuswerte der nachträglichen Laufzeitanzeige, keine Leistungsmessung.
Embedding-Modell	qwen3-embedding:8b; Modellkennung für Embeddings, nicht für die Antwortgenerierung.
Maximale Kontext-Snippets	4; Obergrenze der abgerufenen Snippets für den Antwortkontext.
Dokumentenähnlichkeitsschwelle	Niedrig; angegebener Wert 0,25. Zur Richtung des Vergleichs siehe Erläuterung im Text.
Max embedding chunk length	2048; gemeldeter Grenzwert der Embedding-Konfiguration. Keine Angabe des nativen Token-Kontextfensters des Modells.
Embedding batch size	32; betrifft die Bündelung der Embedding-Verarbeitung, nicht die Zahl der Retrieval-Treffer.
Textchunk-Größe	1500 Zeichen gemäß Einheit der AnythingLLM-Textsplitter-Dokumentation.
Textchunk-Überlappung	250 Zeichen gemäß Einheit der AnythingLLM-Textsplitter-Dokumentation.
Systemprompt	Vollständiger Wortlaut in Anhang C; die Ersetzung von {date} beim damaligen Lauf ist nicht dokumentiert.
Reranking und weitere Kontexte	Suchpräferenz, Reranker, Dokument-Pinning, zusätzliche Anhänge und Chat-Historie nicht protokolliert.
Hardware	MacBook Pro mit Apple M3 Max und 128 GB gemeinsamem Arbeitsspeicher (Unified Memory).
Lokale Modelllaufzeit	Ollama; angezeigtes Inferenz-Backend „llamacpp“ (llama.cpp). Genaue Ollama-Version nicht protokolliert.
Private Archivierung	DEVONthink für die private Dokumentenablage; von der RAG-Vektordatenbank LanceDB zu unterscheiden.
Noch offene Laufparameter	Generierungsparameter (z. B. Temperatur); macOS-, Ollama- und DEVONthink-Versionen. Keine systematischen Laufzeit- oder Speichermessungen.

Die Statusanzeige „30 GB“ wird als ausgegebener Größenwert dokumentiert, nicht als gemessener Spitzenbedarf von RAM oder GPU-Speicher. „100% GPU“ bezeichnet die angezeigte Modellplatzierung beziehungsweise GPU-Auslagerung und keine gemessene GPU-Auslastung. „Forever“ ist die angezeigte Vorhaltezeit des geladenen Modells. Keine dieser Angaben ersetzt Messungen zu Latenz, Durchsatz oder Ressourcenbedarf.

Abbildung 7 trennt die vorgelagerte Indexierung vom Anfragepfad. Die Markierungen L1 bis L5 bezeichnen Stellen, an denen für einen Folgelauf Diagnoseprotokolle benötigt werden; sie lokalisieren keinen der beobachteten Fehler bereits kausal.

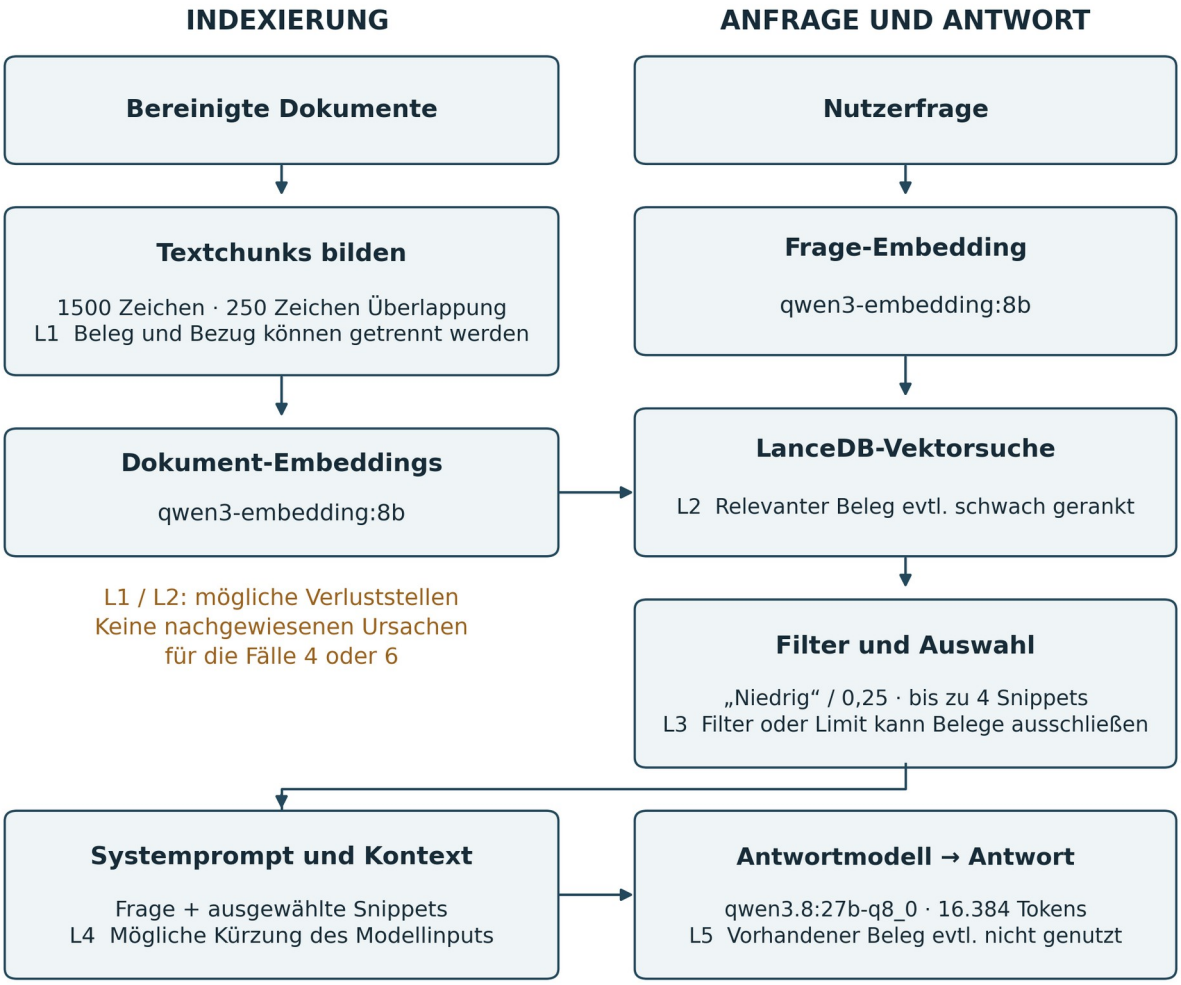


Abbildung 7. Konzeptioneller Retrieval-Pfad mit möglichen Verluststellen. Chunking erfolgt bei der Indexierung. Das Antwortmodell nutzt Q8_0-Quantisierung und ein konfiguriertes Kontextfenster von 16.384 Tokens. Filterrichtung, tatsächliche Rangfolge, zusätzlicher Kontext und Reranking sind nicht historisch protokolliert. Die Verlustmarkierungen sind Hypothesen (Abschnitt 5.7).

AnythingLLM beschreibt für den Textsplitter eine zeichenbasierte Chunk-Größe und Überlappung sowie eine rekursive Trennung entlang von Absatz-, Zeilen- und Wortgrenzen. Änderungen gelten erst für anschließend eingebettete Dokumente; vorhandene Dokumente müssen für einen Vergleich neu eingebettet werden (Mintplex Labs, o. J.-b). Die Werte 1500 und 250 sind deshalb nicht als Tokenzahlen zu lesen. Der separat gemeldete Wert 2048 wird als Konfigurationswert des Embedders dokumentiert; seine konkrete Umsetzung und eine etwaige zusätzliche Trunkierung sind ohne Anbieterexport und Eingabeprotokoll nicht nachgewiesen.

Für die Ähnlichkeitsschwelle wurden das UI-Label „Niedrig“ und der Wert 0,25 dokumentiert. Die ursprüngliche Konfigurationsnotiz bezeichnet den Vergleich als „gleich oder kleiner .25“. Die Herstellerdokumentation erläutert dagegen einen Mindestscore, unterhalb dessen Treffer herausgefiltert werden (Mintplex Labs, o. J.-c). Der Operator ≤ wird deshalb nicht als verifizierte Implementierungsregel übernommen. Für eine exakte Replikation sind Score-Definition, Filterrichtung und Suchpräferenz der eingesetzten Version zu exportieren.

Der Systemprompt verlangt ausdrücklich eine Begründung aus dem lokalen Kontext und den Verzicht auf Vermutungen. Das kann vorsichtige, aber unvollständige Antworten begünstigen, wenn relevante Belege

nicht im sichtbaren Kontext liegen. Seine projektbezogenen Vorgaben zu Zero-Cloud und Praxisentlastung bieten außerdem eine mögliche Erklärung für thematische Zusätze in den Antworten. Beides bleibt ohne kontrollierten Promptvergleich eine Hypothese. Der Prompt ist eine Verhaltensanweisung und kein Nachweis einer technisch abgeschotteten oder rechtlich geprüften Verarbeitung.

3.3.2 Verfahren des Quellenabgleichs

Die qualitative Auswertung erfolgte KI-gestützt anhand des Antwortprotokolls und der sieben Markdowns. Die Antworten wurden nachträglich mit den Quellen abgeglichen; eine unabhängige menschliche Fachbewertung oder verblindete Mehrfachkodierung erfolgte nicht. Geprüft wurden fachliche Richtigkeit, Abdeckung der Frage, Nachvollziehbarkeit der Quellenzuordnung, frühpädagogische Passung und Umgang mit Evidenzgrenzen. Die Matrix dokumentiert konkrete Befunde und vergibt weder Punktwerte noch eine Rangfolge der Modelle.

Aussagen werden danach unterschieden, ob sie im bereitgestellten Text gestützt sind, dem Text widersprechen, gegenüber der Frage unvollständig bleiben oder mit den verfügbaren Unterlagen nicht überprüft werden können. Eine fehlende Quellenangabe macht eine fachlich richtige Aussage nicht automatisch falsch. Umgekehrt begründen Literaturangaben oder präzise Zahlen allein noch keine Belegkorrektheit. Eigene Praxisvorschläge werden gesondert von angeblich empirisch belegten Effekten betrachtet.

Für die RAG-Ausgaben ist zwischen vorhandenem Korpuswissen und dem tatsächlich an das Modell übergebenen Kontext zu unterscheiden. Mehrere Antworten erklären, bestimmte Abschnitte seien im Kontext nicht enthalten, obwohl sie in den bereitgestellten Markdowns vorliegen. Ohne Retrieval-Logs lässt sich nicht entscheiden, ob die Ursache im Index, im Abruf, in einer Kontextkürzung oder in der Verarbeitung durch das Antwortmodell liegt. Diese Trennung von Kontextqualität, Quellentreue und Antwortqualität entspricht auch der mehrdimensionalen Betrachtung in RAGAs (Es et al., 2024).

Für die Heatmap in Abbildung 8 werden die konkret begründeten Befunde nachträglich in vier Kategorien codiert: Q für Quellen- oder Modellverwechslung, A für ausgelassene erfragte Kernangaben, S für fehlerhafte Sach- oder Methodenangaben und Z für unbelegte Ergänzungen oder Überdehnungen. Pro Antwort und Kategorie wird nur das Vorliegen eines Befunds markiert; mehrere Befunde derselben Kategorie werden nicht mehrfach gezählt. Unterschiedliche Kategorien können gleichzeitig vorliegen. Die Codierbegründungen stehen in Tabelle A1. Die Darstellung ist eine redaktionelle Verdichtung des Quellenabgleichs, keine vorab registrierte oder unabhängig validierte Fehlerstatistik.

Tabelle 6. Vergleichsmatrix der zehn dokumentierten Fälle. Gegenübergestellt werden AnythingLLM mit lokalem RAG und Gemini 3.1 Pro mit hohem Denkaufwand.

Fall und Aufgabe	AnythingLLM mit RAG	Gemini 3.1 Pro	Quellenabgleich und Einordnung
1 Spielmerkmale	Drei induktive Merkmale sowie n = 165 und korrekte Häufigkeiten genannt.	Gemini benennt die drei Merkmale und die deduktiven Kategorien korrekt.	Beide treffen den Kern. RAG liefert zusätzliche überprüfbare Detailangaben; eine allgemeine Überlegenheit folgt daraus nicht.
2 Spiel und Lernziele	Passende Studie und Grenze zur Wirksamkeitsprüfung; teilweise zu kategorische Praxisfolgerungen.	Gemini nennt Howard und Fotokarten statt Lannen et al. und Interviews.	Quellen- und Methodenverwechslung bei Gemini. RAG näher am Zieltext, aber mit Überdehnung des vorsichtigen Befunds.

Fall und Aufgabe	AnythingLLM mit RAG	Gemini 3.1 Pro	Quellenabgleich und Einordnung
3 Leadership und Erschöpfung	Direkter und indirekter Effekt richtig; Limitationen unvollständig, einzelne Deutungen fehlerhaft.	Gemini überzeichnet den indirekten Effekt und beschreibt Zeitpunkt sowie Befragte falsch.	RAG bei Kernstatistik genauer. Keine der beiden Antworten bildet das Design und seine Grenzen vollständig ab.
4 Zeitkontingente	62 % und 2,06 Stunden richtig; 34 % der Träger fehlen; Ebenen teilweise vermischt.	Gemini nennt 2,06 Stunden; 62 % und 34 % fehlen.	Beide unvollständig. Der Mittelwert gilt nur bei vorhandenen Zeitkontingenten; Einrichtung und Träger getrennt halten.
5 Lernort-kooperation	Interdependenz, gemeinsame Planung und Qualifizierung quellenbezogen dargestellt.	Gemini beschreibt Abstimmung, geteilte Verantwortung und Curriculumsbezug.	Beide erfassen die Kernaussage. Die Expertise begründet Konzepte, keine gemessene Wirksamkeit der Vorschläge.
6 Kontinuierliche Fortbildung	Abstrakte Rahmenbedingungen; konkrete Formate als im Kontext fehlend bezeichnet.	Gemini nennt Beratung, Hospitation, Qualitätszirkel und Coaching mit Praxisbeispielen.	Gemini beantwortet die offene Praxisfrage konkreter. Relevante Formate sind im Prüfbestand vorhanden; weitere RAG-Quellen hier nicht überprüfbar.
7 Merdeka-Curriculum	Vier richtige Dimensionen; ergänzende Deutschlandaussagen teils unbelegt, Rechtsverweis unpassend.	Gemini verwendet ein anderes Vier-Dimensionen-Modell.	RAG trifft den Studienrahmen. Zusatzbehauptungen bleiben prüfpflichtig; Gemini verfehlt die zentrale Frage.
8 Moralisches Pflicht-bewusstsein	Kritische Einordnung; n und Situationen fehlen; Zielmetadaten als Altersangabe verwendet; themenfremde Persona-Zusätze.	Gemini kritisiert den Schluss auf Angeborenheit; konkrete Stichprobe und Situationen fehlen.	Beide nur teilweise studienspezifisch. Tatsächlich 16 und 4 Kinder im Alter von 3 bis 6 Jahren; zwei Stiftsituationen. Prompt-Einfluss bei RAG: Abschnitt 5.5.
9 Organisatorische Synthese	Befunde und Vorschläge getrennt; zusätzliche, hier nicht verfügbare Quellen genannt.	Gemini trennt formal Befund und Vorschlag; einzelne Wirkungsbehauptungen unbelegt.	Beide bieten brauchbare Ansätze. Ein empirisch geprüfter Entlastungseffekt des Maßnahmenpakets wird nicht belegt.
10 Falsche Voraussetzung	Kein Prozentwert; vorhandene Studie trotz passender OpenAlex-ID als fehlend bezeichnet.	Gemini weist das Programm zurück; erweitert die Aussage über die Zielstudie hinaus.	Beide vermeiden die erfundene Zahl. RAG zeigt einen Identifikationsfehler; Geminis weitergehende Literaturbehauptungen bleiben ungeprüft.

3.3.3 Zentrale Beobachtungen

Studienidentität und konkrete Befunde: In Frage 1 benennen beide Systeme die drei induktiven Spielmerkmale zutreffend; die RAG-Ausgabe ergänzt korrekte Häufigkeiten und Bezugsgrößen. Bei Frage 2 bleibt RAG bei Lannen et al. (2026), während Gemini die Arbeit Justine Howard zuschreibt und Fotokarten als Methode nennt. Der bereitgestellte Text von Lannen et al. beschreibt dagegen Interviews und begründet ausdrücklich den Verzicht auf visuelle Vorgaben. In Frage 7 nennt RAG die von Kristina et al. (2026) im Merdeka-Text untersuchten Dimensionen Plan, Ploy, Pattern und Perspective; Gemini ersetzt sie durch Kommunikation, Ressourcen, Disposition und bürokratische Struktur. Beide Fälle zeigen, warum

thematische Plausibilität nicht mit einer korrekten Antwort auf eine bestimmte Studie gleichgesetzt werden darf.

Genauigkeit und Vollständigkeit: Bei Frage 3 zur Studie von Oppermann et al. (2026) gibt RAG den kleinen indirekten Gesamteffekt von β = −0,06 und den nicht signifikanten direkten Zusammenhang korrekt wieder. Gemini beschreibt einen starken indirekten Pfad und behauptet eine Erhebung aller Konstrukte bei denselben Fachkräften zu einem Zeitpunkt. Tatsächlich wurden Leitungen 2017 zu Führungspraktiken und Fachkräfte 2018 unter anderem zu Teamklima und Erschöpfung befragt. Auch RAG bleibt unvollständig: Es meldet fehlende Limitationen und behandelt strukturelle Einflussgrößen teilweise als nicht modelliert, obwohl der Text entsprechende Kontrollvariablen aufführt. Ein standardisierter Regressionskoeffizient darf zudem nicht unmittelbar als erklärte Varianz interpretiert werden.

Abbildung 8 macht die in Tabelle 6 beschriebenen Muster für die zehn dokumentierten Fälle sichtbar. Eine helle Zelle bedeutet lediglich, dass für diese Kategorie kein konkreter Befund codiert wurde; sie bescheinigt keine fehlerfreie Antwort.

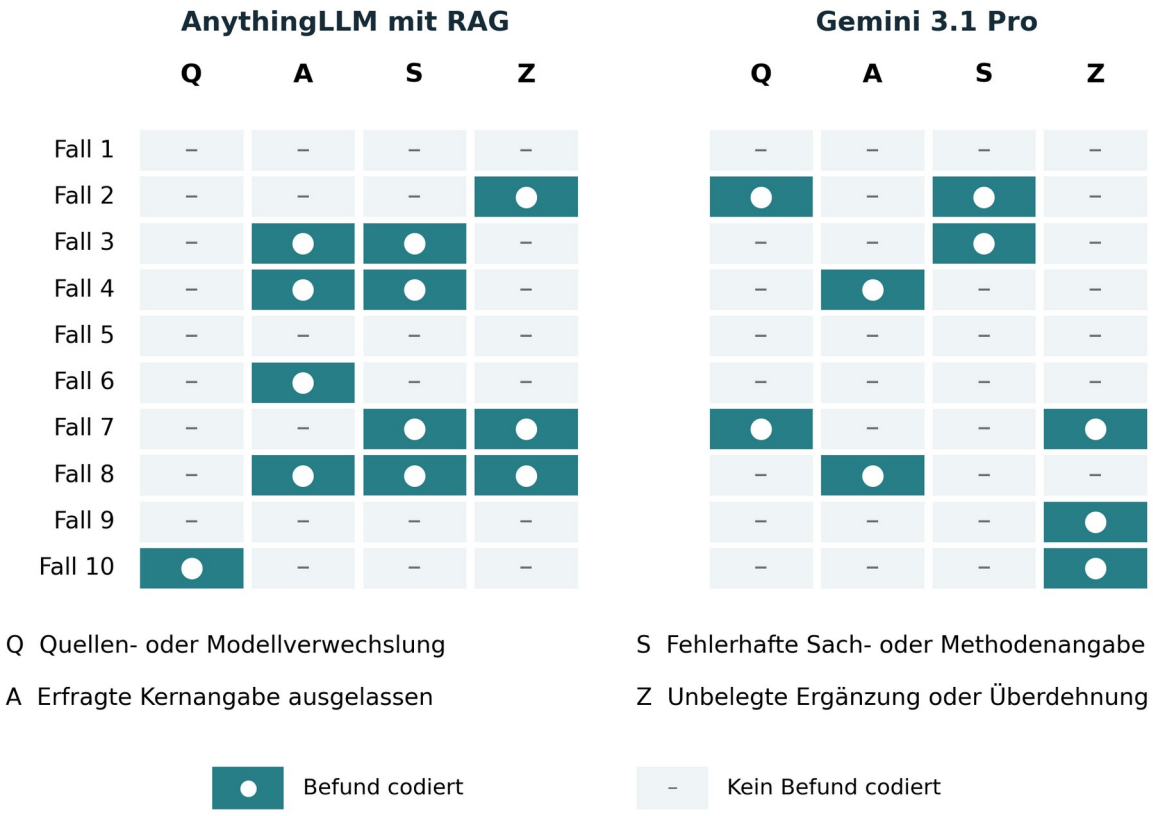


Abbildung 8. Kategoriale Heatmap der zehn dokumentierten Antwortpaare. Dunkel markiert einen codierten Befund, hell keinen konkret codierten Befund. Die Kategorien sind keine RAGAS-Scores oder Schweregrade und werden nicht zu einem Gesamtfehlerwert addiert. Prüfgrundlage und Einzelfallbegründungen: Anhang A und Tabelle A1.

Korrekte Kernstatistiken bleiben als positive Befunde in Tabelle 6 beschrieben. Für Kontrollfrage 10 nennen beide Systeme keinen erfundenen Prozentwert; daraus wird kein allgemeiner Nullbefund zu erfundenen Studienergebnissen abgeleitet. Die zusätzlich sichtbaren Auslassungen in Fall 3 und 8 zeigen, warum eine Zählung allein der Beispiele 4 und 6 die dokumentierten Befunde nicht vollständig abbilden würde.

Bei Frage 4 zur Studie von Fuchs-Rechlin et al. (2026) nennt RAG die 62 % der ausbildenden Einrichtungen mit geregelten Zeitkontingenten und den bedingten Mittelwert von 2,06 Wochenstunden; Gemini nennt den Mittelwert ebenfalls richtig. Beide lassen jedoch die verlangte Kennzahl von 34 % auf Trägerebene aus. Die RAG-Ausgabe überträgt außerdem einen Befund zum Ausbildungsformat von der Einrichtungs- auf die Trägerebene. Der Datenbestand enthält die gesuchten Angaben, dennoch werden sie nicht vollständig und durchgehend bezugsgruppengerecht genutzt.

Praxisnutzen und Quellenbegrenzung: Bei Frage 5 zur Expertise von Freis und Mierendorff (2026) treffen beide Systeme die zentrale Forderung nach gemeinsamer Ausbildungsverantwortung. Bei der offenen Fortbildungsfrage 6 bietet Gemini mehr konkrete Formate. Die dort vorgeschlagenen Zeitdauern sind als Gestaltungsvorschläge zu lesen und nicht als nachgewiesene Dosierungen. RAG erklärt konkrete Formate für im Kontext nicht belegt und verweist auf weitere Arbeiten, die nicht zu den sieben bereitgestellten Prüfdokumenten gehören. Das Kapitel von Tewari et al. (2026) enthält jedoch Abschnitte zu Coaching, Mentoring und professionellen Lerngemeinschaften. Hier zeigt sich eine Lücke zwischen verfügbarem Wissen und nutzbarer Antwort.

Unbelegte Ergänzungen bleiben auch mit RAG möglich. In Frage 7 enthält die RAG-Ausgabe neben dem korrekten Kernrahmen pauschale Deutschlandvergleiche und einen unpassenden Verweis auf § 18 SGB VIII als Fachberatungsgrundlage. Die Norm behandelt die Unterstützung bei Personensorge und Umgangsrecht. Solche Ergänzungen sind vom korrekt abgerufenen Studienbefund zu trennen. Frage 8 zur Untersuchung von Ushkovska und Gulevska (2026) zeigt zudem eine Verwechslung des Metadatenfelds rag_zielbezug „0–6“ mit der tatsächlichen Stichprobe von drei- bis sechsjährigen Kindern. Beide Antworten kritisieren den Schluss auf angeborenes Pflichtbewusstsein, nennen aber weder die Gruppengrößen 16 und 4 noch rekonstruieren sie die beiden Stiftsituationen hinreichend genau.

Umgang mit falschen Voraussetzungen: In Frage 10 zur Studie von Oppermann et al. (2026) erfindet keines der Systeme einen prozentualen Interventionseffekt. Die RAG-Antwort bestreitet jedoch gleichzeitig die Präsenz der gesuchten Studie und führt deren OpenAlex-ID W7165033882 samt passendem Studiendesign auf. Sie vermeidet damit die erfundene Zahl, scheitert aber an der Quellenidentifikation. Gemini erkennt das fehlende Interventionsdesign, weitet die Zurückweisung jedoch auf die gesamte Literatur der Autorinnen aus und ergänzt weitere, hier nicht geprüfte Literatur- und Projektbezüge. Eine angemessene Antwort hätte sich auf die identifizierte Studie beschränkt.

Die Fallserie belegt somit unterschiedliche Stärken und Fehlerarten der dokumentierten Ausgaben. Sie erlaubt keine allgemeine Aussage, dass RAG oder Gemini besser sei, und isoliert auch keinen kausalen Effekt der Korpusbereinigung. Die lokalen Betriebseigenschaften aus Abschnitt 2.6 sind getrennt von der inhaltlichen Antwortqualität zu beurteilen.

Anschluss an RAGAS: Die Befunde aus Tabelle 6 lassen sich den Evaluationsdimensionen des RAGAS-Frameworks zuordnen (Es et al., 2024; Ragas, o. J.-a, o. J.-b, o. J.-c). Tabelle 7 übersetzt die qualitative Fehlerbeschreibung in prüfbare Evaluationsfragen. Sie enthält keine berechneten RAGAS-Werte. Insbesondere indizieren Auslassungen in Fall 4 und 6 zunächst unvollständige Antworten; ein Defizit im Context Recall ist erst belegt, wenn die nötigen Informationen im tatsächlich abgerufenen Kontext fehlen.

Tabelle 7. Qualitative Zuordnung zu RAGAS-Dimensionen. Hypothesen und Datenanforderungen; keine berechneten Metrikwerte.

Dimension	Bezug zu Tabelle 6	Zulässige Einordnung	Für die Messung erforderlich
Context Recall Kontext-Abdeckung	Fall 4: 34 % der Träger fehlen. Fall 6: Coaching, Mentoring und Lerngemeinschaften bleiben ungenutzt.	Mögliche Abruflücke, falls die nötigen Belege im übergebenen Kontext fehlen. Bei vorhandenem Beleg liegt eine nachgelagerte Nutzungslücke vor.	Fachlich geprüfte Referenzantwort oder Referenzstellen sowie tatsächlich abgerufene Kontexte.
Faithfulness Treue zum Kontext	Gemini: Quellenverwechslung in Fall 2 und 7. RAG: unbelegte Ergänzungen etwa in Fall 7 und 8.	Prüfkandidaten für ungestützte Aussagen. Ein Fehler gegenüber dem Zieltext ist erst bei bekanntem Modellkontext als Faithfulness-Defizit einzuordnen.	Vollständige Antwort und tatsächlich übergebener Kontext; bei Gemini Kenntnis des Dokument- und Grounding-Zugangs.

Dimension	Bezug zu Tabelle 6	Zulässige Einordnung	Für die Messung erforderlich
Context Precision Präzision und Rangfolge	Fall 6: Verweise auf andere Arbeiten. Fall 8: Übernahme des Metadatenwerts „0–6“.	Möglicherweise ungeeignete oder ablenkende Treffer. Antworttexte zeigen weder Anzahl noch Rang irrelevanter Chunks; „Chunk Pollution“ ist nicht nachgewiesen.	Geordnete Trefferliste, Chunk-Inhalte und Relevanzurteile nach der gewählten Metrikvariante.

Context Recall untersucht die Abdeckung relevanter Referenzinformationen durch den Abruf. Context Precision bewertet die Relevanz der abgerufenen Chunks unter Berücksichtigung ihrer Rangfolge; die genaue Operationalisierung hängt von der gewählten Variante ab. Faithfulness prüft, ob Antwortaussagen aus dem bereitgestellten Kontext ableitbar sind, und ist deshalb von sachlicher Richtigkeit gegenüber einer externen Referenz zu unterscheiden. Die Quellenverwechslungen bei Gemini sind im vorliegenden Quellenabgleich belegbar, aber ohne dessen Eingabekontext kein gemessener Faithfulness-Mangel. Ebenso rechtfertigen einzelne korrekte RAG-Antworten keinen global hohen Faithfulness-Wert. Eine spätere quantitative Auswertung muss RAGAS-Version, Metrikvarianten, Referenzen, Bewertungsmodell und Prompts festhalten. Die hier verwendeten Begriffsabgrenzungen folgen den aktuellen Metrikbeschreibungen (Ragas, o. J.-a, o. J.-b, o. J.-c).

4 Datenmodell des ECE-Research-Corpus und Orientierung an FAIR

Die aus Stufe 3.1 hervorgehenden Dokumente sowie kuratierte historische Werkbestände werden im Repository Earlychildhoodeducation/ECE-Research-Corpus in komprimierten Apache-Parquet- Dateien bereitgestellt (mit automatischer Unterstützung für Git-LFS/Xet-Versionierung; siehe Abbildung 9). Das ECE-Research-Corpus ist als dynamisches, fortlaufend kuratiertes Repositorium („Living Corpus“) auf Hugging Face konzipiert. Während die vorliegende Studie zur methodischen Validierung exemplarisch die Kohorte des Publikationsjahres 2014 (N = 151 Kerndokumente) heranzieht, wird der Korpus durch wiederkehrende, automatisierte Pipeline-Läufe kontinuierlich um aktuelle Open-Access-Publikationen sowie historische Bestände erweitert. Die Architektur der Wissens-Pipeline v1.0 gewährleistet dabei durch persistente Identifikatoren (OpenAlex-ID, DOI, SHA-256) und das zentrale Transaktionsregister, dass Inkrementalität und Deduplizierung auch über Jahre hinweg sichergestellt sind.

Das Datenmodell orientiert sich an den vier FAIR-Prinzipien (Wilkinson et al., 2016; Abbildung 9). Der Grad der tatsächlichen FAIR-Erfüllung wurde nicht durch ein externes Audit geprüft:

- **Findable (Auffindbar):** Jedes Dokument im Korpus ist über global eindeutige, persistente Identifikatoren (openalex_id, doi, repository_doi) sowie über einen kryptografischen pdf_sha256-Hash der Ursprungsdatei und eine eindeutige dokument_id adressierbar. Die reichhaltigen Metadaten (titel, autor, jahr, kategorie, rag_zielbezug, rag_themen) ermöglichen granulare Abfragen.
- **Accessible (Zugänglich):** Der ECE-Research-Corpus ist öffentlich über den Hugging Face Hub und HTTPS zugänglich. Die Softwarearchivierung und die Modellkarte von EleMo-V3 sind über ihre DOI-Adressen erreichbar. Die Bezugsadressen sind in Abschnitt 7 zusammengeführt. Für die Replikation eines einzelnen Laufs ist zusätzlich der jeweils verwendete Versions- beziehungsweise Indexstand festzuhalten.
- **Interoperable (Interoperabel):** Durch die Verwendung von UTF-8-kodiertem, NFC- normalisiertem Markdown mit standardisiertem YAML-Frontmatter in den Einzeldateien sowie Apache Parquet im Aggregat ist der Korpus unmittelbar kompatibel mit gängigen LLM- und RAG-Frameworks (u. a. Hugging Face Datasets, LangChain, LlamaIndex, AnythingLLM, Unsloth und MLX).
- **Reusable (Wiederverwendbar):** Die Datensätze dokumentieren ihre Lizenz-Provenienz (lizenz, lizenz_original, lizenz_ampel = „gruen“, lizenz_url, lizenz_quelle, download_url). Die Metadaten

erleichtern die Prüfung der Nachnutzungsbedingungen, ersetzen aber nicht die Prüfung der Lizenz- und Rechtekette für jedes enthaltene Werk.

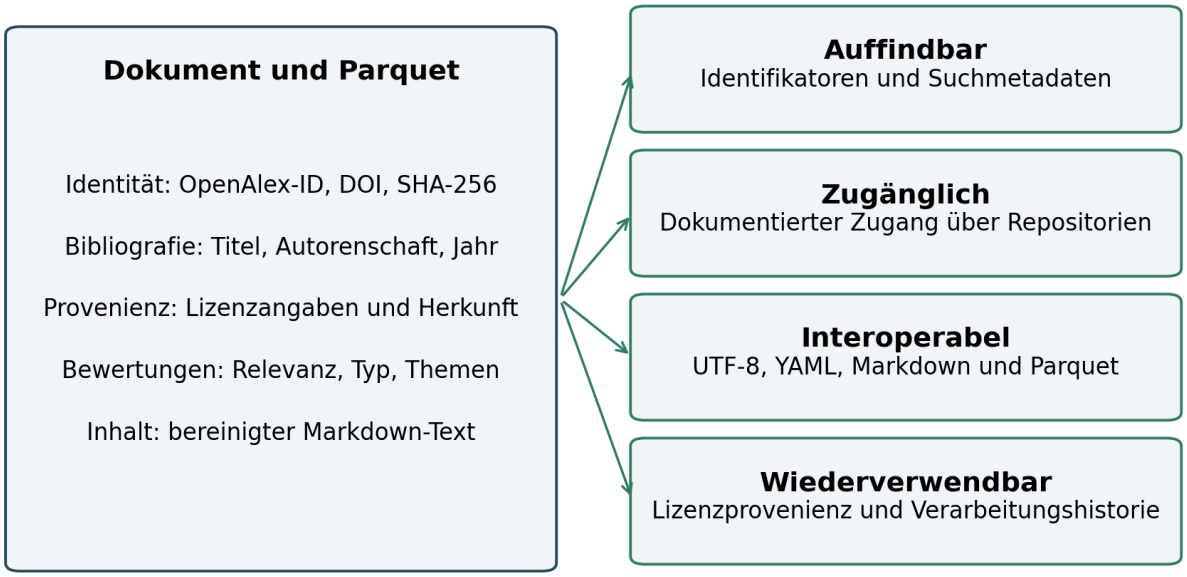


Abbildung 9. Datenmodell und Zuordnung zu den FAIR-Prinzipien. Die Architektur orientiert sich an FAIR; ein externes Audit der Erfüllung liegt nicht vor.

5 Methodenkritik, Limitationen und Threats to Validity

Für eine kritische Einordnung der Systemarchitektur und des Datensatzes sind die folgenden methodischen Limitationen (Threats to Validity) zu berücksichtigen:

5.1 Konservatives Lizenz-Gate und begrenzte Datenabdeckung

In der Stichprobe wurden 58,0 % der nach dem thematischen Filter verbliebenen Open-Access- Werke in die gelbe Kategorie eingeordnet, meist wegen fehlender, unklarer oder unter der gewählten Policy nicht automatisch freigegebener Lizenzinformationen. Darunter können Arbeiten sein, deren konkrete Weiterverwendung unter bestimmten Bedingungen zulässig wäre. Die Pipeline priorisiert dokumentierte Freigabekriterien gegenüber maximaler Abdeckung.

Eine grüne Klassifikation ist daher als regelbasierte Risikoentscheidung zu verstehen, nicht als „100%-ige Legal Precision“ oder Garantie vollständiger Rechtssicherheit. Für eine tatsächliche Weiterverbreitung müssen unter anderem die Rechte des konkreten Werkes, etwaige Drittinhalte, der Anwendungsfall und mögliche ShareAlike-Pflichten berücksichtigt werden. Der konservative Filter kann für andere Forschungskontexte angepasst werden; die rechtliche Bewertung kann jedoch nicht allein aus einem automatischen Metadatenfeld abgeleitet werden.

Zusammen wurden 921 der 1199 lizenzgeprüften Werke (76,8 %) zurückgestellt oder ausgeschlossen. Diese Quote beschreibt die Reichweite der gewählten Policy und nicht den Anteil rechtlich unzulässiger TDM-Nutzungen.

Text- und Data-Mining als alternative Nutzungsgrundlage: Die Lizenz-Policy ist von gesetzlichen Schranken zu unterscheiden. Art. 3 der DSM-Richtlinie (EU) 2019/790 betrifft TDM für wissenschaftliche Forschung durch berechtigte Forschungsorganisationen und Kulturerbe-Einrichtungen bei rechtmäßigem Zugang. Die institutionellen Voraussetzungen ergeben sich insbesondere aus Art. 2 Nr. 1; allein ein nicht kommerzielles

Projekt genügt nicht. Der Opt-out-Vorbehalt aus Art. 4 gilt für diese Schranke nicht. Art. 4 ermöglicht dagegen unter seinen Voraussetzungen auch kommerzielles TDM, sofern rechtmäßiger Zugang besteht und die Rechte nicht wirksam ausdrücklich vorbehalten wurden; bei öffentlich online verfügbaren Inhalten sieht die Richtlinie dafür insbesondere maschinenlesbare Vorbehalte vor (Richtlinie [EU] 2019/790, 2019, Art. 2–4).

Im deutschen Recht sind insbesondere §§ 44b und 60d UrhG einschlägig. § 60d Abs. 3 Nr. 2 erfasst auch einzelne Forschende, sofern sie nicht kommerzielle Zwecke verfolgen. § 44b setzt für allgemeines TDM das Fehlen eines wirksamen Nutzungsvorbehalts voraus; bei online zugänglichen Werken muss dieser maschinenlesbar sein. Eine TDM-Befugnis umfasst jedoch nicht automatisch die freie Veröffentlichung eines Volltextkorpus. § 60d Abs. 4 begrenzt den dort geregelten Zugang auf bestimmte Forschungs- und Überprüfungszwecke. Ob die konkreten Verarbeitungsschritte und Nutzungsbedingungen unter die jeweilige Schranke fallen, wäre gesondert zu prüfen (UrhG, §§ 44b, 60d).

Damit könnte ein gesonderter interner Analysebestand unter passenden Voraussetzungen über die Lizenz-Policy hinausgehen; die zusätzliche Ausbeute wurde nicht ermittelt. Die hier berichtete Policy bleibt auf dokumentierte Nachnutzungsgrundlagen ausgerichtet. Ihre grüne Kategorie umfasst Public Domain, CC0, CC BY und CC BY-SA und ist daher nicht auf CC BY und CC0 beschränkt. Diese Auswahl erleichtert die Prüfung einer späteren Weitergabe, ersetzt aber weder werkbezogene Rechteprüfungen noch die Erfüllung von Namensnennungs- und gegebenenfalls ShareAlike-Pflichten. Das FAIR-Prinzip „Reusable“ verlangt nachvollziehbare Nutzungsbedingungen und bedeutet keine pauschale Freigabe sämtlicher Texte.

5.2 API-Ranking und Trunkierungseffekte

Für hochfrequente englische Suchbegriffe (beispielsweise early years education mit über 135.000 in OpenAlex angezeigten Treffern für 2014) wurde nur ein kleiner gerankter Ausschnitt berücksichtigt. Die Auswahl ist daher von OpenAlex-Ranking und Suchformulierung abhängig und nicht repräsentativ für alle Publikationen des Jahrgangs. Eine spätere Erhöhung des Limits im Produktivbetrieb ändert diese Einschränkung des untersuchten Laufs nicht.

5.3 OCR-Rauschen bei historischen Primärquellen

Die OCR historischer Originaldrucke in Fraktur oder beschädigtem Satzbild kann Erkennungsfehler erzeugen, etwa beim langen „ſ“. Weil die Pipeline wissenschaftliche Texte nicht generativ umschreibt, bleiben solche Fehler gegebenenfalls erhalten. Ihr tatsächlicher Einfluss auf Retrieval und Embeddings wurde in dieser Arbeit nicht gemessen.

5.4 Abhängigkeit von der Klassifikationsgüte des lokalen Sprachmodells

Das semantische Qualitäts-Gate verwendet das lokal ausgeführte Modell qwen3.8:27b. Einstellungen wie temperature = 0 und die Begrenzung auf strukturierte JSON-Ausgaben verhindern nicht zuverlässig Fehlklassifikationen. Grenzfälle, beispielsweise Studien zur Transition Kita–Grundschule, können unterschiedlich bewertet werden. Das Routing mittlerer Relevanz in den Prüfpfad 3.2 schafft eine Möglichkeit zur menschlichen Kontrolle, ersetzt jedoch keine unabhängige Prüfung der automatischen Entscheidungen.

5.5 Grenzen der Güteprüfung und des Systemvergleichs

Eine unabhängige, fachlich annotierte Referenzstichprobe für die regelbasierte Relevanzheuristik, den semantischen Klassifikator und die Textbereinigung liegt für den berichteten Lauf nicht vor. Deshalb können weder Precision und Recall der Selektion noch die Fehlerquote der Texttransformation angegeben werden. Die Ergebnisse der 2014er-Kohorte beschreiben die algorithmisch dokumentierte Ausbeute.

Die Fallserie umfasst zehn vollständig dokumentierte Antwortpaare. Fragen, Modelle und Informationszugang bilden keine kontrollierten Vergleichsbedingungen. Die Fragen wurden anhand bestimmter Korpusdokumente entwickelt und begünstigen damit erwartbar den Zugang zu diesen Dokumenten. Ihre Bearbeitung misst vor allem die Nutzung ausgewählter Quellen und nicht die allgemeine pädagogische Kompetenz eines Modells. Wiederholungsstabilität, Interrater-Reliabilität und inferenzstatistische Unterschiede wurden nicht untersucht.

Die qualitative Auswertung wurde KI-gestützt erarbeitet und ist weder unabhängig noch verblindet. Ein fester Quellenbestand für die Prüfung ersetzt kein annotiertes Referenzset mit unabhängig geprüften Musterantworten. Trotz der nachträglich ergänzten Einstellungen fehlen zeitgleiche Kontext- und Konfigurationsprotokolle. Dies begrenzt die Zuordnung der beobachteten Fehler zu Retrieval, Generierung, Metadaten, Prompt oder Chat-Historie. Der Vergleich mit einem anderen Cloud-Modell isoliert keinen Effekt des Retrievals oder der Korpuskuratierung.

Einfluss von Systemprompt und Personalisierung: Fall 8 zeigt eine besonders deutliche thematische Übertragung aus dem lokalen Systemprompt. Die Frage betrifft die empirische Prüfung von Kants kategorischem Imperativ im Vorschulalter. Die RAG-Antwort ergänzt dennoch Lerngeschichten nach Margaret Carr, DSGVO-bezogene Anonymisierung und eine Zero-Cloud-Notiz-App. Diese Themen entsprechen den Persona-Vorgaben in Anhang C, tragen aber nicht unmittelbar zur Prüfung der erfragten Stichprobe und Situationen bei.

Das Muster lässt sich als Instruction Over-Triggering interpretieren: Themen des Systemprompts werden über den Bedarf der konkreten Frage hinaus aktiviert. Der Ausdruck „Systemprompt-Leakage“ bezeichnet hier diesen thematischen Einfluss und keine Offenlegung eines verborgenen Systemprompts. Die Übereinstimmung ist ein beobachtbarer Befund; sie beweist ohne kontrollierten Vergleich weder die alleinige Ursache noch einen bestimmten internen Mechanismus. Neben dem Systemprompt kommen etwa Chat-Historie oder zusätzliche Retrieval-Kontexte als Einflussquellen in Betracht.

Auch die Cloud-Ausgaben zeigen mögliche Personalisierungseinflüsse, insbesondere die Projektbezüge in Fall 10. Das Fragment „My Teaching Cupboard“ in Fall 1 und die Dialog-Rückfragen belegen für sich genommen weder Websuche noch gespeicherte Nutzerdaten. Für einen Folgelauf sollten frische Sitzungen, dokumentierte Personalisierungseinstellungen und ein neutraler Systemprompt verwendet werden. Ein gezielter Vergleich des lokalen Persona-Prompts mit einer neutralen Fassung bei identischer Frage und identischem Kontext könnte den Beitrag der Promptvorgaben zur thematischen Abweichung in Fall 8 untersuchen.

5.6 Auffälligkeiten in den bereitgestellten Quelldateien

Der Quellenabgleich zeigt auch Grenzen der vorgelagerten Aufbereitung. Im YAML-Header des Kapitels Professional Development in Early Childhood Education von Tewari et al. (2026) steht die DOI 10.1080/09575146.2013.763770. Diese Kennung gehört zum anderen Artikel Continuing professional development in early childhood education in New Zealand (Cherrington & Thornton, 2013). Die Beschränkung der DOI-Suche auf den Dokumentanfang verhindert somit nicht grundsätzlich die Übernahme einer zitierten Fremd-DOI, insbesondere bei kurzen Texten. Der Verlagsnachweis ordnet das Kapitel stattdessen Band 1 von Global Perspectives on ICT for Development, Seiten 22–24, und der DOI 10.5281/zenodo.20773195 zu.

Der Merdeka-Markdown enthält weiterhin das Redaktionsteam und Verlagsnavigation. Sein YAML-Autorenfeld nennt nur Eva Kristina, während der Text fünf Autoren aufführt. In der Spielstudie ist außerdem die extrahierte Tabelle zum Kodierrahmen in Teilen abgeschnitten. Ohne Abgleich mit den jeweiligen Original-PDFs lässt sich der Entstehungsschritt dieser Textverluste nicht sicher bestimmen. Die Auffälligkeiten sind Einzelbefunde und keine geschätzte Fehlerquote der gesamten Pipeline.

Für die Weiterentwicklung ergeben sich daraus Prüfaufgaben: Dokument-DOIs sollten gegen Titel und Autorenschaft validiert, lange administrative Vorspänne zuverlässig abgegrenzt und Metadaten mit dem Volltext abgeglichen werden. Tabellen und zentrale Methodenabschnitte benötigen eine Integritätsprüfung. Für RAG sollten Dokumentidentität und bibliografische Metadaten zusammen mit jedem relevanten Textabschnitt übergeben werden. Diese Maßnahmen sind Vorschläge aus der Fehleranalyse; ihre Umsetzung und Wirksamkeit wurden in dieser Fassung nicht getestet.

5.7 Diagnose von Retrieval-Lücken und Under-generation

Der in Abbildung 7 dargestellte Pfad ordnet die folgenden Hypothesen den Stufen Indexierung, Auswahl, Kontextaufbereitung und Antwortgenerierung zu. Diese Trennung ermöglicht eine gezielte Fehlersuche, sobald die tatsächlich verarbeiteten Chunks und Modellinputs vorliegen.

Die Auslassungen in Fall 4 und 6 sind Ausgangspunkte einer technischen Diagnose, keine bereits lokalisierten Retrieval-Fehler. Im bereitgestellten Praxisanleitungs-Markdown steht die 34-%-Angabe vollständig in einem Fließtextabsatz. Im Kapitel von Tewari et al. sind Coaching und Mentoring sowie Professional Learning Communities ausdrücklich als Formate gegliedert. Ein verlorener Seitenwechsel oder eine zerstörte Tabelle ist für diese Informationen somit nicht als Ursache belegt. Ob dieselben Passagen vollständig indexiert und dem Modell übergeben wurden, bleibt offen.

Hypothese zu Chunking und Auswahl: Bei 1500 Zeichen Chunk-Größe, 250 Zeichen Überlappung und höchstens vier Kontext-Snippets können Zahlen, Bezugsgruppen und erläuternde Nachbarabsätze auf unterschiedliche Treffer verteilt sein. Ein relevanter Chunk könnte außerhalb der ausgewählten Treffer liegen oder am Ähnlichkeitsfilter scheitern. Umgekehrt können überlappende Treffer aus derselben Passage die inhaltliche Breite der vier Snippets verringern. Diese Möglichkeiten wären anhand der gespeicherten Chunk-Grenzen, Rangplätze und Scores zu prüfen. Die nachträglich angegebenen Einstellungen belegen keine der Ursachen für einen konkreten Fall.

Hypothese zur Anfrageformulierung: In Fall 6 könnte die deutschsprachige Frage nach konkreten Fortbildungsformaten abstrakte Beiträge höher gewichten als die englischen Formatüberschriften bei Tewari et al. Die Antwort verweist auf ein Angebots-Nutzungs-Modell; daraus folgt nicht, dass die gesuchten Formate nur dort oder nur theoretisch behandelt werden. Eine mögliche sprachliche oder begriffliche Abruflücke lässt sich mit kontrollierter Anfrageerweiterung um „coaching“, „mentoring“ und „professional learning communities“ prüfen. Liegen die Belege bereits im Modellkontext, sind stattdessen Kontextkürzung, Antwortplanung und die vorsichtige Promptvorgabe als Erklärungen zu untersuchen.

Für einen Folgelauf sollte zunächst die relevante Referenzpassage im Index nachgewiesen und anschließend ihr Weg durch Kandidatensuche, Filter, Rangfolge und endgültigen Modellkontext protokolliert werden. Eine zusätzliche Antwortgenerierung mit fest vorgegebenen Referenzpassagen trennt Probleme des Abrufs von Problemen der Kontextnutzung. Darauf aufbauend lassen sich Snippet-Limit, Chunking und Reranking jeweils einzeln variieren. Änderungen am Chunking erfordern einen neu aufgebauten Index; ein bloßer Wechsel der Anzeigeeinstellung verändert vorhandene Chunks nicht (Mintplex Labs, o. J.-b).

Als Erweiterungen kommen hybride Suche aus lexikalischem BM25 und Vektorsuche sowie mehrstufiges Retrieval in Betracht. BM25 könnte exakte Begriffe wie „Träger“ und „Zeitkontingente“ ergänzend gewichten; eine unbekannte Prozentzahl wird dadurch nicht automatisch gefunden. Ein agentisch gesteuerter zweiter Abruf könnte unzureichend beantwortete Teilfragen gezielt nachrecherchieren. Beides sind zu prüfende Entwicklungsvorschläge und keine in dieser Fallserie getesteten Verbesserungen. Für den Offline-Betrieb müssten auch diese Such-, Steuerungs- und gegebenenfalls Reranking-Komponenten lokal laufen. Der Vergleich sollte neben Quellenabdeckung und Faktentreue auch Latenz und Ressourcenbedarf erfassen.

6 Fazit und Ausblick

Die ECE Wissens-Pipeline v1.0 beschreibt einen technisch nachvollziehbaren Weg, domänenspezifische Forschungstexte auszuwählen, zu konvertieren und für RAG-Anwendungen aufzubereiten. Die Fallstudie des Publikationsjahrs 2014 dokumentiert eine mehrstufige Selektionskaskade bis zu 151 übernommenen Werken. Die Ergebnisse belegen den Verarbeitungsablauf und die Ausbeute; eine unabhängige Bestätigung sämtlicher Relevanz-, Lizenz- und Qualitätsentscheidungen steht aus.

Die zehn dokumentierten Antwortpaare machen den Anwendungsnutzen konkreter und zeigen zugleich Grenzen. Quellenbezogenes Retrieval unterstützt in mehreren Fällen die korrekte Wiedergabe spezifischer Studien. Es verhindert jedoch weder Auslassungen noch unbelegte Ergänzungen und gewährleistet nicht, dass das Modell eine abgerufene Quelle richtig identifiziert. Gemini beantwortet einzelne offene Praxisfragen konkreter, zeigt bei mehreren gezielt studienspezifischen Aufgaben aber Quellen- oder Modellverwechslungen. Die Fallserie erlaubt eine qualitative Beschreibung dieser Muster; sie begründet kein allgemeines Modellranking.

Der lokale Betrieb ist ein eigenständiger Beitrag der Architektur. Nach vollständiger Einrichtung kann das System vorhandene Dokumente aufbereiten und Fachfragen mittels lokaler Embeddings, lokaler Vektorsuche und lokalem Sprachmodell ohne laufende Internetverbindung beantworten. Die Akquise neuer Online-Literatur und externe Aktualisierungen bleiben davon getrennt. Diese Betriebsform ermöglicht eine Nutzung des vorhandenen Wissensbestands unabhängig von Cloud-Inferenzdiensten, ohne für sich genommen eine höhere fachliche Antwortqualität zu garantieren.

Als nächste Evaluationsschritte bieten sich ein Vergleich desselben Antwortmodells mit und ohne Retrieval sowie mit neutralem und themenspezifischem Systemprompt, ein eingefrorener Korpus- und Indexstand, vollständige Kontextprotokolle, wiederholte Läufe und eine unabhängige Fachbewertung an. Die dokumentierte RAG-Konfiguration und die qualitative RAGAS-Zuordnung ermöglichen gezieltere Prüfungen der in Abschnitt 5.7 formulierten Hypothesen. Eine spätere automatisierte Bewertung kann mit lokal bereitgestellten Bewertungsmodellen konzipiert werden; in dieser Fassung wurde sie nicht durchgeführt. Ergänzend sind Textintegrität, Metadatenzuordnung und Betrieb bei getrennter Internetverbindung zu prüfen. Die Verbindung des ECE-Research-Corpus mit der eigenständigen Trainingsbasis von EleMo-V3 bleibt eine Perspektive für künftige Forschungsprototypen.

7 Daten- und Code-Verfügbarkeit, Rechte und Interessenkonflikt

Software und Quellcode: Die ECE Wissens-Pipeline v1.0 ist unter https://doi.org/10.5281/zenodo.23182511 öffentlich archiviert (Götz, 2026a). Maßgeblich ist die mit dem konkreten Softwarestand bereitgestellte Lizenz.

Forschungsdaten: Der ECE-Research-Corpus ist unter https://huggingface.co/datasets/Earlychildhoodeducation/ECE-Research-Corpus öffentlich zugänglich. Die Kennzeichnung des Repositories mit CC BY 4.0 ist keine pauschale Neulizenzierung fremder Volltexte. Für jedes enthaltene Werk gelten dessen individuelle Rechte und gegebenenfalls ShareAlike-Bedingungen; die Rechtekette ist vor einer Weitergabe von Volltexten werkbezogen zu prüfen.

Sprachmodell: EleMo-V3 ist über die öffentliche Hugging-Face-Modellkarte dokumentiert (Götz, 2026b). Dort wird die DOI https://doi.org/10.57967/hf/10808 ausgewiesen; die Modellkarte ist unter https://huggingface.co/Earlychildhoodeducation/EleMo-V3 erreichbar. Das Modell beruht auf einer eigenständigen pädagogischen Kerndatenbasis. Die Entwicklung einer kombinierten Folgegeneration ist geplant und nicht Ergebnis dieser Studie.

Vergleichsmaterial: Das Protokoll mit zehn Antwortpaaren ist in Anhang B wiedergegeben. Die sieben Prüfdokumente sind bibliografisch und über OpenAlex-IDs nachgewiesen; Prüfsummen der Markdown-Dateien stehen in Anhang A. Die RAG-Einstellungen wurden nachträglich dokumentiert (Tabelle 5); der Systemprompt steht in Anhang C. Zeitgleiche Konfigurations-, Index- und Retrieval-Protokolle liegen nicht vor.

KI-Unterstützung: Die zehn Fragen wurden mithilfe generativer KI aus den sieben Quellen entwickelt. Auch die Textüberarbeitung, der Quellenabgleich und die Erstellung der Dokumentfassungen erfolgten KI-gestützt. Die inhaltliche Verantwortung liegt beim Autor; eine unabhängige Evaluation durch externe Fachgutachterinnen oder Fachgutachter erfolgte nicht.

Interessenkonflikt: Ich bin Gründer der Initiative Kita Digital und Entwickler der beschriebenen Pipeline sowie des zugehörigen Datensatzes. Durchführung und Auswertung liegen damit in derselben Hand; die explorative Evaluation erfolgte weder unabhängig noch verblindet.

8 Literaturverzeichnis

Das Verzeichnis enthält die im Bericht verwendeten Fachquellen, Rechtsgrundlagen, Herstellerdokumentationen und Projektartefakte. Bei dynamischer Softwaredokumentation ist der Abrufstand angegeben. Literaturangaben innerhalb der dokumentierten Modellausgaben in Anhang B sind Untersuchungsgegenstand und werden nicht automatisch als bestätigte Quellen übernommen.

Cherrington, S., & Thornton, K. (2013). Continuing professional development in early childhood education in New Zealand. *Early Years*, *33*(2), 119–132. https://doi.org/10.1080/09575146.2013.763770

Es, S., James, J., Espinosa Anke, L., & Schockaert, S. (2024). RAGAs: Automated evaluation of retrieval augmented generation. In N. Aletras & O. De Clercq (Hrsg.), *Proceedings of the 18th Conference of the European Chapter of the Association for Computational Linguistics: System Demonstrations* (S. 150–158). Association for Computational Linguistics. https://doi.org/10.18653/v1/2024.eacl-demo.16

Freis, M., & Mierendorff, J. (2026). *Praxisanleitung am Lernort (Berufs-)Praxis: Forschungsergebnisse und methodisch-didaktische Konzepte des Praxislernens in der Frühpädagogik* (WiFF Expertisen, Band 61). Deutsches Jugendinstitut. https://doi.org/10.36189/wiff22026

Fuchs-Rechlin, K., Barbarino, B., Nachtigall, C., & Pilchowski, A. (2026). *Praxisanleitung in der Frühen Bildung: Auf dem Weg zur Professionalisierung?* (WiFF Studien kompakt, Band 7). Deutsches Jugendinstitut. https://doi.org/10.36189/wiff32026

Götz, S. (2026a). *ECE Wissens-Pipeline: Automatisierte Kuratierung domänenspezifischer Open-Access-Literatur für RAG-Systeme in der Frühpädagogik* (Version 1.0) [Software]. Zenodo. https://doi.org/10.5281/zenodo.23182511

Götz, S. (2026b). *EleMo-V3 (Elementary Pedagogical Model)* [Großes Sprachmodell]. Hugging Face. https://doi.org/10.57967/hf/10808

Kristina, E., Munawar, M., Khairani, C., Siraj, S., & Najmuddin, N. (2026). Principals strategies for Merdeka curriculum implementation in early childhood education. *Academia Open*, *11*(1). https://doi.org/10.21070/acopen.11.2026.13388

Lannen, P., Pritz, L. N., Duss, I., Ruedisueli, C., & Wustmann Seiler, C. (2026). Play through children's eyes—The subjective meaning of play to young children. *Early Childhood Education Journal*. Vorab-Onlinepublikation. https://doi.org/10.1007/s10643-026-02359-6

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. (2020). Retrieval-augmented generation for knowledge-intensive NLP tasks. In H. Larochelle, M. Ranzato, R. Hadsell, M. F. Balcan, & H. Lin (Hrsg.), *Advances in neural information processing systems* (Band 33, S. 9459–9474). Curran Associates. https://papers.nips.cc/paper/2020/hash/6b493230205f780e1bc26945df7481e5-Abstract.html

Mintplex Labs. (o. J.-a). *Embedding models*. Abgerufen am 17. Oktober 2026, von https://docs.anythingllm.com/features/embedding-models

Mintplex Labs. (o. J.-b). *Text splitting & chunking*. Abgerufen am 17. Oktober 2026, von https://docs.anythingllm.com/setup/embedder-configuration/text-splitting

Mintplex Labs. (o. J.-c). *Using documents in AnythingLLM*. Abgerufen am 17. Oktober 2026, von https://docs.anythingllm.com/chatting-with-documents/introduction

Mintplex Labs. (2025, 14. Juli). *AnythingLLM Desktop app privacy policy*. https://docs.anythingllm.com/installation-desktop/privacy

Ollama. (o. J.). *FAQ*. Abgerufen am 17. Oktober 2026, von https://docs.ollama.com/faq

Oppermann, E., Hummel, T. G., & Anders, Y. (2026). Psychische Gesundheit in der Kita: Die Rolle von Leadership und Teamklima für das Belastungserleben frühpädagogischer Fachkräfte. *Zeitschrift für Erziehungswissenschaft*, *29*(2), 533–551. https://doi.org/10.1007/s11618-025-01353-9

Priem, J., Piwowar, H., & Orr, R. (2022). *OpenAlex: A fully-open index of scholarly works, authors, venues, institutions, and concepts*. arXiv. https://doi.org/10.48550/arXiv.2205.01833

Ragas. (o. J.-a). *Context precision*. Abgerufen am 17. Oktober 2026, von https://docs.ragas.io/en/stable/concepts/metrics/available_metrics/context_precision/

Ragas. (o. J.-b). *Context recall*. Abgerufen am 17. Oktober 2026, von https://docs.ragas.io/en/stable/concepts/metrics/available_metrics/context_recall/

Ragas. (o. J.-c). *Faithfulness*. Abgerufen am 17. Oktober 2026, von https://docs.ragas.io/en/stable/concepts/metrics/available_metrics/faithfulness/

Richtlinie (EU) 2019/790 des Europäischen Parlaments und des Rates vom 17. April 2019 über das Urheberrecht und die verwandten Schutzrechte im digitalen Binnenmarkt und zur Änderung der Richtlinien 96/9/EG und 2001/29/EG, Amtsblatt der Europäischen Union L 130, 92–125 (2019). https://eur-lex.europa.eu/eli/dir/2019/790/oj

Sozialgesetzbuch (SGB) Achtes Buch (VIII) – Kinder- und Jugendhilfe, § 18, BGBl. I S. 1163 (1990; geltende Fassung). https://www.gesetze-im-internet.de/sgb_8/__18.html

Tewari, S., Rai, A., & Anuradha. (2026). Professional development in early childhood education. In *Global perspectives on ICT for development* (Band 1, S. 22–24). ZenToks Publication. https://doi.org/10.5281/zenodo.20773195

Urheberrechtsgesetz [UrhG], §§ 44b und 60d, BGBl. I S. 1273 (1965; geltende Fassung). https://www.gesetze-im-internet.de/urhg/BJNR012730965.html

Ushkovska, A., & Gulevska, V. (2026). Possibilities for the implementation of Immanuel Kant's categorical imperative in preschool education. *New Horizons – Journal of Student Research*, *3*(1), 54–66. https://doi.org/10.20544/nhjsr.2026.644

Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., Bonino da Silva Santos, L., Bourne, P. E., Bouwman, J., Brookes, A. J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C. T., Finkers, R., . . . Mons, B. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, *3*, Artikel 160018. https://doi.org/10.1038/sdata.2016.18

Anhang A Prüfgrundlage der zehn Fälle

Die folgenden Notizen dokumentieren den Erwartungshorizont und den Quellenabgleich. Sie sind keine unabhängigen Fachratings. Die genannten Abschnitte beziehen sich auf die sieben Markdowns des Prüfbestands; ergänzende Quellen aus dem größeren RAG-Bestand wurden nur beurteilt, soweit sie im Prüfmaterial verfügbar waren.

Fall 1

Prüfquelle: Lannen et al. (2026), Results und Data Analysis

Erwartungshorizont: Drei induktive Kategorien; 165 differenzierte Antworten aus 327 Interviews, 186 Kategoriezuordnungen; soziale Interaktionen 44, Objektbezug 35, Kinderangelegenheit 12.

Abgleich: RAG nennt Kategorien und Häufigkeiten passend. Gemini trifft die Kategorien ebenfalls. Die isolierte Bezeichnung „My Teaching Cupboard“ in der Cloud-Ausgabe ist kein rekonstruierbarer Primärbeleg. Beide Antworten verdichten die kindliche Perspektive teilweise stärker, als die deskriptiven Häufigkeiten allein erlauben.

Fall 2

Prüfquelle: Lannen et al. (2026), Procedure and Participants und Practical Implications

Erwartungshorizont: Offene verbale Interviews; Wahlfreiheit, Eigeninitiative und unterstützende Erwachsenenrollen; keine Interventionsstudie.

Abgleich: RAG erkennt die fehlende Kausalevidenz, formuliert aber kategorisch, die beschriebene Praxis werde nicht als Spiel erfahren. Der Text sagt vorsichtiger, die Benennung als spielerisch garantiere diese Erfahrung nicht. Gemini verwechselt die Zielstudie mit Arbeiten von Howard und beschreibt Fotokarten statt der hier gewählten Interviews.

Fall 3

Prüfquelle: Oppermann et al. (2026), Abschnitte 2.1, 2.2, 2.3, 3.2 und 4.2

Erwartungshorizont: 1.394 Fachkräfte und 204 Leitungen; Leitungsbefragung 2017, Fachkräftebefragung 2018; β indirekt = −0,06, p = 0,023; direkter Zusammenhang nicht signifikant.

Abgleich: RAG gibt die Kernkoeffizienten richtig wieder, nutzt die vorhandenen Limitationen aber nicht vollständig. Strukturelle Variablen wurden kontrolliert; β ist kein R^2. Gemini behauptet fälschlich einen einzigen Erhebungszeitpunkt und dieselben Fachkräfte als Auskunftspersonen für alle Konstrukte. Das korrelative Design erlaubt trotz zeitversetzter Erhebung keine kausale Identifikation.

Fall 4

Prüfquelle: Fuchs-Rechlin et al. (2026), Einrichtungen zeigen großes Engagement und Rahmenbedingungen unzureichend

Erwartungshorizont: 62 % der ausbildenden Einrichtungen mit verbindlichen Zeitkontingenten; bei vorhandenen Kontingenten MW 2,06 Stunden; 34 % der Träger mit verbindlicher Regelung.

Abgleich: RAG nennt die Einrichtungswerte, lässt aber 34 % aus und überträgt Angaben zum Ausbildungsformat auf die Trägerebene. Aus zwei Mittelwerten darf zudem keine gesicherte individuelle 1:1-Zuordnung abgeleitet werden. Gemini nennt 2,06 Stunden, lässt beide zentralen Anteilswerte aus und beschreibt Trägerart und Regression verkürzt.

Fall 5

Prüfquelle: Freis und Mierendorff (2026), Abschnitte 3.2, 3.3 und 4

Erwartungshorizont: Interdependenz der Lernorte, gemeinsame Kompetenzziele und Ausbildungsplanung, Weiterbildung der Anleitenden und Aushandlung der Verantwortung.

Abgleich: Beide Antworten erfassen die Kernaussage. RAG formuliert mehrere Aufgaben unmittelbar entlang des Textes; Gemini fasst konzeptionell zusammen. Aussagen zur Qualitätssteigerung sind als begründete Konzeption zu lesen, nicht als im Beitrag experimentell gemessener Effekt.

Fall 6

Prüfquelle: Tewari et al. (2026), Abschnitte 3 und 4; ergänzend Freis und Mierendorff (2026)

Erwartungshorizont: Kontextbezug, aktives Lernen, kontinuierliche Begleitung, Coaching und Mentoring sowie professionelle Lerngemeinschaften sind im Prüfbestand benannt.

Abgleich: RAG verweist auf zusätzliche Arbeiten und ein Angebots-Nutzungs-Modell, deren tatsächliche Kontextgrundlage hier nicht vollständig vorliegt. Die Aussage fehlender konkreter Formate ist auf den behaupteten Abrufkontext beschränkt, trifft aber nicht auf den bereitgestellten Prüfbestand zu. Gemini ist aufgabenbezogen konkreter; Zeitdauern und Wirkungsversprechen sind nicht als nachgewiesene Studienergebnisse abgesichert.

Fall 7

Prüfquelle: Kristina et al. (2026), Abstract, Metode sowie Hasil dan Pembahasaan A bis D

Erwartungshorizont: Plan, Ploy, Pattern, Perspective; qualitative Untersuchung in drei PAUD-Einrichtungen mit drei Leitungen und sechs Lehrkräften.

Abgleich: RAG nennt die richtigen Dimensionen und kennzeichnet Beispiele teilweise als Ableitungen. Der Deutschlandvergleich enthält unbelegte Verallgemeinerungen und den sachfremden §-18-Verweis. Gemini nennt ein anderes Implementierungsmodell und ergänzt nicht nachgewiesene Beispiele. Die genannte Zahl deutscher Einrichtungen wurde nicht als Ergebnis der Zielstudie gewertet.

Fall 8

Prüfquelle: Ushkovska und Gulevska (2026), Methodology, Empirical Examination und Discussion

Erwartungshorizont: Zwei Gruppen mit 16 und 4 Kindern im Alter von 3 bis 6 Jahren; jeweils fehlt ein Stift, worauf ein Kind seinen anbietet. Keine Kontrollgruppe und kein Nachweis von Angeborenheit.

Abgleich: Beide Antworten thematisieren Sozialisation und alternative Motive, rekonstruieren den empirischen Teil aber unzureichend. RAG verwechselt den Metadaten-Zielbezug mit dem tatsächlichen Alter und fügt einen nicht belegten „Hofman-Effekt“ sowie projektbezogene Ratschläge zu Carr, Anonymisierung und einer Zero-Cloud-App hinzu. Die thematische Übereinstimmung mit dem Systemprompt wird in Abschnitt 5.5 diskutiert. Geminis pauschale entwicklungspsychologische Erklärungen sind keine Messbefunde dieser Studie.

Fall 9

Prüfquelle: Fuchs-Rechlin et al. (2026), Kurz & knapp; Freis und Mierendorff (2026), Abschnitt 4; Oppermann et al. (2026), Abschnitt 4.3

Erwartungshorizont: Organisatorische Absicherung, Teamverantwortung, Qualifizierung und Lernortkooperation; kein geprüfter Gesamteffekt eines konkreten Maßnahmenpakets.

Abgleich: Beide Systeme kennzeichnen Handlungsvorschläge gesondert. RAG nennt zusätzlich Kita-Entwicklung und Kita-Sozialarbeit, die nicht als vollständige Prüfdokumente vorliegen. Gemini formuliert unter „Berichtete Befunde“ unter anderem eine signifikante Belastungszunahme ohne identifizierbaren Beleg. Konkrete Zeitpläne und Entlastungsaussagen bleiben Ableitungen.

Fall 10

Prüfquelle: Oppermann et al. (2026), Abstract sowie Abschnitte 2 und 3

Erwartungshorizont: Die bezeichnete Studie ist eine Mediationsanalyse und untersucht kein zwölfwöchiges Coachingprogramm. Ein entsprechender Prozentwert ist nicht vorhanden.

Abgleich: Beide Systeme nennen keinen erfundenen Wert. RAG bestreitet jedoch die Quellenidentität trotz passender OpenAlex-ID und Stichprobe. Gemini identifiziert das Thema, behauptet aber weitergehend das Nichtvorliegen einer solchen Studie im gesamten Werk der Autorinnen. Weitere von Gemini genannte Publikationen und institutionelle Zuordnungen sind hier nicht vollständig geprüft.

Codierung der Heatmap

Die Codes Q, A, S und Z entsprechen Abbildung 8. Erfasst werden die nachstehend konkretisierten Befunde, nicht sämtliche denkbaren Fehler oder jede ungeprüfte Literaturangabe. Als eigene Vorschläge erkennbare Praxisbeispiele werden nicht allein wegen fehlender Messdaten als Fehler codiert. Eine Quellenverwechslung (Q) kann mit einem zusätzlich konkret falschen Methodendetail (S) zusammentreffen.

Tabelle A1. Fallbezogene Begründungen der Heatmap. „–“ bedeutet kein konkret codierter Befund in den vier Kategorien; weitere Quellen können außerhalb des Prüfbestands liegen.

Fall	AnythingLLM mit RAG	Gemini 3.1 Pro
1	– — Kein konkret codierter Befund; Kernkategorien und Häufigkeiten passend.	– — Kein konkret codierter Befund; Kernkategorien passend.
2	Z — Zu kategorische Folgerung, die beschriebene Praxis werde nicht als Spiel erfahren.	Q / S — Q: Howard statt Lannen et al. S: Fotokarten statt der tatsächlich eingesetzten Interviews.
3	A / S — A: vorhandene Limitationen unvollständig. S: strukturelle Kontrollen und Deutung des Koeffizienten fehlerhaft.	S — Kleinen indirekten Effekt überzeichnet; Erhebungszeitpunkt und Auskunftspersonen falsch beschrieben.
4	A / S — A: 34 % der Träger fehlen. S: Einrichtungs- und Trägerebene bei Angaben zum Ausbildungsformat vermischt.	A — Die Anteilswerte 62 % und 34 % fehlen.
5	– — Kein konkret codierter Befund; konzeptionelle Kernaussage getroffen.	– — Kein konkret codierter Befund; konzeptionelle Kernaussage getroffen.
6	A — Konkrete Fortbildungsformate bleiben ungenutzt, obwohl sie im Prüfbestand stehen.	– — Kein konkret codierter Befund; vorgeschlagene Zeitdauern werden nicht als gemessene Dosierungen gewertet.
7	S / Z — S: § 18 SGB VIII als Fachberatungsgrundlage unpassend. Z: unbelegte pauschale Deutschlandvergleiche.	Q / Z — Q: anderes Vier-Dimensionen-Modell. Z: konkrete örtliche Abläufe und Transferbehauptungen ohne Beleg im Zieltext.
8	A / S / Z — A: Gruppengrößen und Stiftsituationen fehlen. S: 0–6 aus Metadaten statt tatsächlichem Alter 3–6. Z: unbelegter „Hofman-Effekt“.	A — Gruppengrößen und konkrete Stiftsituationen werden nicht hinreichend rekonstruiert.

Fall	AnythingLLM mit RAG	Gemini 3.1 Pro
9	– — Kein konkret codierter Befund; weitere genannte Quellen liegen nicht vollständig zur Prüfung vor.	Z — Signifikante Belastungszunahme als berichteter Befund ohne identifizierbaren Beleg.
10	Q — Präsenz der Zielstudie trotz passender OpenAlex-ID und passendem Design bestritten.	Z — Zurückweisung über die Zielstudie hinaus auf das gesamte Werk der Autorinnen ausgedehnt.

Identität der Prüfdokumente

Die Prüfsumme bezeichnet die für den Quellenabgleich verwendete Markdown-Datei. Sie ist kein Nachweis, dass genau diese Bytefassung im RAG-Index verwendet wurde. Die PDF-Prüfsummen in den YAML-Headern beziehen sich dagegen auf die dort bezeichneten ursprünglichen PDFs.

OpenAlex-ID	SHA-256 der bereitgestellten Markdown-Datei
W7162201820	6e49fc9478945ab1be1ca80ed4d31c2b c7a279dcfa956a8becac68bc4450d570
W7213927762	95b74016d87cb83ac63d6467f25770c9 68908dbeee2aa2db5b5c020b2be8c1fc
W7212201653	b9982fcce6c3515de6bc2a2e2a89b64b 5a441ce0ae3834e9c8ae6716a7bc5d48
W7212215487	b2d344e546c235a95a8b208d91afaa2a c0743112f6a258fb5e7212273f99ce1b
W7165033882	437f3c25e63b538e6684845951665cb5 7240e498fd4edf040d1d282a8cd2924b
W7163050204	be33e54229f0ae9da586dfa93707f29b c4bac0e4e69edd77befbb9c7f9c53e8c
W7165453726	476d9cbb772a9dc823405e4c2758ac92 6eac36edf9c0608c7fb84c1b24c6dacc

Anhang B Fragen und vollständige Antworten

Grundlage ist das Vergleichsprotokoll RAGvsGemini3.1pro.docx. Die zehn Fragen und beide zugehörigen Antworten werden in ihrer ursprünglichen Reihenfolge wiedergegeben. Der Wortlaut einschließlich nicht belegter Behauptungen und Rückfragen bleibt erhalten; Absatz-, Überschriften- und Tabellenformatierung wurden vereinheitlicht. Lediglich ein offensichtlicher Groß-/Kleinschreibungsfehler im Wort „Erklärvariablen“ in Fall 3 wurde berichtigt. Die Aussagen in diesem Anhang sind untersuchte Modellausgaben und keine bestätigten Aussagen des Forschungsberichts. Nicht aufgezeichnete Retrieval-Kontexte, Suchprotokolle oder Quellenlinks wurden nicht rekonstruiert. Der nachträglich dokumentierte lokale Systemprompt steht getrennt in Anhang C.

Fall 1

1. Welche drei zusätzlichen Merkmale von Spiel wurden in „Play Through Children’s Eyes“ aus den Aussagen der Kinder herausgearbeitet, und wie unterscheiden sie sich von den fünf theoretisch vorgegebenen Kategorien?

AnythingLLM mit RAG

Induktive Merkmale von Spiel in „Play Through Children’s Eyes"

Die drei zusätzlich herausgearbeiteten Merkmale

Aus den 165 differenzierten Kinderantworten (186 Kategoriezuordnungen durch Mehrfachcodierung) traten drei induktive Merkmale hervor, die im bestehenden deduktiven Rahmen (Smith & Vollstedt, 1985) nicht ausreichend abgebildet werden konnten:

INDUKTIVES MERKMAL	N	ANTEIL
Soziale Interaktion	44	26,7 %
Objektbezug (object reference)	35	21,2 %
Kinderangelegenheit (children’s business)	12	7,3 %

Alle drei Merkmale wurden als „coherent with a larger framework of child development and with what we know about playfulness, the way children engage in play" bewertet.

Die fünf deduktiv vorgegebenen Kategorien (zum Vergleich)

DEDUKTIVES MERKMAL	N	ANTEIL	HERKUNFT
Positive Affektqualität	26	15,8 %	Krasnor & Pepler (1980)
Flexibilität	20	12,1 %	Krasnor & Pepler (1980)
Intrinsische Motivation	15	9,1 %	Krasnor & Pepler (1980)
Nicht-Literalität	13	7,9 %	Krasnor & Pepler (1980)
Mittel vor Ziel (means before end)	21	12,7 %	Rubin et al. (1983)

Wesentliche Unterscheidung zwischen den beiden Merkmalsgruppen

Perspektive der Operationalisierung Die fünf deduktiven Kategorien wurden von Erwachsenen – Eltern, Lehrkräften, externen Beobachtern, Theoretikern und Praktikern – entwickelt (Barnett, 2013; Glenn et al., 2013). Sie beschreiben Spiel primär über individuelle, innere Qualitäten (Affekt, Motivation, Prozessorientierung, Nicht-Literalität). Die drei

induktiven Merkmale hingegen sind aus den Aussagen der Kinder selbst hervorgegangen, weil deren Antworten „could not be adequately represented by the existing framework" (S. 3).

Schwerpunktverlagerung: relational und referenziell vs. intrapsychisch

Soziale Interaktion war mit 26,7 % das häufigste Merkmal in den Kinderantworten. Kinder verorten Spiel demnach zuerst im Miteinander, während der theoretische Rahmen dies nicht als eigenständige Kategorie vorsieht.

Objektbezug und Kinderangelegenheit betonen die konkrete, materielle und soziale Zuordnung von Spiel („etwas, das man mit Dingen tut" bzw. „etwas, das Kinder tun"), wo die deduktiven Kategorien stärker auf abstrakte, individuelle Prozesse abzielen.

Konsequenz für die Spieldefinition Die Studie folgert ausdrücklich: „children's statements are not completely congruent and therefore cannot be equated with adults' statements." Die drei induktiven Merkmale erweitern das Spielverständnis, ersetzen die deduktiven Kategorien aber nicht – sie komplementieren sie um die dimensionale Ebene, die aus kindlicher Perspektive salient ist, ohne dass der bisherige Erwachsenenrahmen als falsch gilt.

Methodischer Hinweis

Die kombinierte deduktiv-induktive Vorgehensweise diente gezielt dem doppelten Ziel: Die etablierten Kategorien fungierten als sensitising and comparative points of reference; Antworten, die sich nicht in diesen Rahmen fassen ließen, wurden induktiv gruppiert, „allowing additional features salient from the children's perspective to emerge from the material." Damit bleibt der Smith-und-Vollstedt-Rahmen als empirisch geprüfter Ausgangspunkt erhalten, wird aber um kindliche Sichtweisen erweitert.

Gemini 3.1 pro

In der Studie „Play Through Children's Eyes – The Subjective Meaning of Play to Young Children“ (2026) wurden die Beschreibungen von Kindern zum Thema „Spiel“ sowohl deduktiv anhand etablierter Theorie als auch induktiv anhand der direkten Kinderaussagen untersucht.

Die Analyse brachte drei zusätzliche Merkmale (Kategorien) hervor, die aus der Perspektive der Kinder zentral für das Spiel sind:

Children's business (Kinderangelegenheit): Spiel wird von Kindern als ihre eigene Domäne betrachtet, in der sie Autonomie und Kontrolle haben, oft in Abgrenzung zur Welt und den Vorgaben der Erwachsenen.

My Teaching Cupboard

Social interactions (Soziale Interaktionen): Die Anwesenheit und Einbindung von Mitspielern (Peers) ist für Kinder ein entscheidendes Kriterium, um eine Aktivität als „Spiel“ zu bewerten.

Object reference (Objektbezug): Der Umgang mit konkreten Spielzeugen oder Gegenständen ist aus Kindersicht ein greifbarer, maßgeblicher Bestandteil der Spielhandlung.

Der Unterschied zu den fünf theoretischen Kategorien

Die fünf theoretisch vorgegebenen Kategorien basierten auf etablierten Definitionen der Spielforschung (ursprünglich von Krasnor & Pepler, 1980, und erweitert durch Smith & Vollstedt, 1985). Diese umfassen:

Positive affect (Positive Emotionen / Freude)

Flexibility (Flexibilität)

Intrinsic motivation (Intrinsische Motivation)

Nonliterality (So-tun-als-ob / Nicht-Wörtlichkeit)

Means before end (Zweckfreiheit / Der Weg ist das Ziel)

Der wesentliche Unterschied: Die fünf theoretischen Kategorien sind abstrakte, psychologische Konstrukte, die von Erwachsenen (Forschenden) formuliert wurden, um das Phänomen „Spiel“ von außen zu klassifizieren. Die drei zusätzlichen Merkmale hingegen entstanden induktiv aus der gelebten Realität der Kinder. Sie zeigen, dass Kinder Spiel sehr viel konkreter und kontextbezogener definieren: Für sie entscheidet nicht nur das abstrakte Gefühl oder die Zweckfreiheit, sondern ganz praktisch, womit sie spielen (Objekte), mit wem sie spielen (Soziales) und wer das Sagen hat (Kinderangelegenheit).

Möchtest du mehr über den Einfluss der Erwachsenen beim kindlichen Spiel erfahren?

Fall 2

2. In einer Kita werden fast alle als „Spiel“ bezeichneten Aktivitäten von Erwachsenen geplant und auf ein vorgegebenes Lernergebnis ausgerichtet. Welche Veränderungen lassen sich aus „Play Through Children’s Eyes“ begründen, und welche Aussagen über deren Wirksamkeit erlaubt die Studie nicht?

AnythingLLM mit RAG

„Play Through Children’s Eyes" (Lannen et al., 2026) – Befund und Grenze der Übertragbarkeit

1. Was der beschriebene Kita-Alltag laut Studie problematisiert

Die Studie stellt die in Ihrer Frage geschilderte Konstellation – erwachsenengeplante, an ein vorgegebenes Lernergebnis ausgerichtete Aktivitäten mit dem Label „Spiel" – als zentrales Unterscheidungsproblem heraus (Wustmann Seiler et al., 2024, zitiert in Lannen et al., 2026):

„Labelling an activity as playful does not necessarily mean that children experience it as play, particularly when the activity is strongly directed, evaluated, or organised around a predefined learning outcome."

Das bedeutet: Die Benennung einer Tätigkeit als „Spiel" ist für die Kinderperspektive irrelevant, wenn die Tatsachenstruktur – starke Führung, Bewertung, Festlegung auf ein externes Ziel – demSpielcharakter widerspricht.

Konkrete, aus dem Text begründbare Veränderungen

ASPEKT IN DER STUDIE	KONSEQUENZ FÜR DIE BESCHRIEBENE KITA-PRAXIS
Means before end – Spiel erlaubt die Beschäftigung mit einer Aktivität um ihrer selbst willen, ohne dass der Prozess einem extern bestimmten Zweck untergeordnet wird (Lannen et al., 2026)	Aktivitäten dürfen nicht strukturell auf ein vorab fixiertes Lernergebnis zulaufen; das Mittel (die Beschäftigung) darf nicht zum bloßen Vehikel des Endzwecks degradiert werden.
Onlooker-Rolle – aufmerksame Präsenz und Beobachtung ohne aktive Intervention im laufenden Spiel (Rüdisüli et al., 2023); häufigeres Einnehmen dieser Rolle ist mit späterer Spielfreude der Kinder assoziiert (Rüdisüli et al., 2026)	Pädagog:innen beginnen in der Beobachtungs- und Anwesenheitsrolle; erst nach Verständnis von Spielsituation, Intentionen und Entwicklungsbedarf wird flexibel in eine aktivere Rolle gewechselt, wenn Unterstützung tatsächlich erforderlich ist.
Assoziation: akademisch orientierte Lernumgebungen ↔ niedrigere Spielfreude (Rüdisüli et al., 2026)	Ein überwiegender Anteil stark strukturierter, lernzielgesteuerter Angebote korreliert mit einer Reduktion subjektiv erlebter Spielfreude. Eine Reduktion des „akademischen" Charakters der Angebote ist indiziert.

Zusammengefasst: Die Studie liefert eine theoretische und beschreibende Begründung, warum die in Ihrer Frage geschilderte Praxis – „fast alles geplant, an ein Lernergebnis ausgerichtet, als Spiel bezeichnet" – aus Kindperspektive nicht als Spiel erfahren wird, und wo pädagogisches Handeln ansetzen kann (Beobachtung vor Intervention, Selbstzweckhaftigkeit der Aktivität, flexible Rollenwechsel).

2. Was die Studie nicht erlaubt

Hier ist die Abgrenzung entscheidend, weil die Frage nach „Wirksamkeit" eine andere Evidenzstufe verlangt, als die Studie sie erbringt:

Keine Interventions- / Wirksamkeitsevidenz. Lannen et al. (2026) ist eine qualitative Studie zur subjektiven Bedeutung von Spiel für junge Kinder (Kindperspektive). Sie untersucht, wie Kinder Spiel erleben und woran sie dessen Merkmale aus ihrer Sicht erkennen. Sie ist keine kontrollierte Intervention, die beispielsweise prüft, ob die Umstellung von lernzielorientierten auf freie Spielformen messbare Entwicklungsgewinne (z. B. in Sprach- oder Mathematikkompetenzen) hervorruft.

Keine kausalen Aussagen. Die im Text genannten Zusammenhänge – häufigere Onlooker-Rolle ↔ spätere Spielfreude; akademischere Umgebungen ↔ niedrigere Spielfreude – werden als Assoziationen aus verwandten Studien (Rüdisüli et al., 2023; 2026) referenziert, nicht als kausal-belegte Wirkungsmechanismen aus dieser Studie selbst. Aus der Kita-Praxis lässt sich daher nicht ableiten: „Wenn wir Planungsgrad X auf Y senken, steigt die Spielfreude um Z."

Keine Aussage über den Lernerfolg. Die Studie thematisiert die Unterscheidung Spiel vs. erwachsenengestaltetes spielerisches Lernen, sagt aber nichts über den quantitativen Effekt auf vordefinierte Lernziele (Lesefähigkeit, Zahlenverständnis, soziale Kompetenz usw.) in dem Sinne, dass „weniger Steuerung = weniger Lernerfolg" oder umgekehrt. Das wäre eine völlig andere Frage.

Keine Generalisierung über den hier vorliegenden Befundkreis. Die vorliegenden Kontexte enthalten die Diskussion, die Stärken-/Schwächen-Angabe und die Metadaten. Die vollständige Methodik, die genaue Stichprobenbeschreibung und die detaillierten Kodierkategorien der qualitativen Analyse sind in den mir vorliegenden Auszügen nicht enthalten. Eine abschließende Bewertung der externen Validität für die spezifische Kita-Situation ist daher aus dem vorliegenden Kontext nicht möglich.

Keine Empfehlung eines konkreten Anteils. Die Studie liefert keinen Richtwert wie „mindestens 60 % der Zeit muss freies Spiel sein". Sie begründet, warum die Erfahrungsebene des Kindes zählt, gibt aber kein quantitatives Planungsschema vor.

3. Kurzformel für die Praxis

Die Studie legitimiert, die Frage „Erleben die Kinder das hier als Spiel?" als pädagogisch relevant zu stellen und die Onlooker-Rolle als Erstprinzip zu verankern. Sie legitimiert nicht, daraus eine Wirksamkeitsgarantie für bestimmte Lernergebnisse abzuleiten oder den Planungsgrad auf einen festen prozentualen Wert zu reduzieren, ohne weitere Evidenz.

Quellenbasis: Lannen, P., Pritz, L. N., Duss, I., Rüdisüli, C., & Wustmann Seiler, C. (2026). Play Through Children’s Eyes – The Subjective Meaning of Play to Young Children. OpenAlex W7213927762, CC BY 4.0. Zitierte Teilstudien: Rüdisüli et al. (2023; 2026); Wustmann Seiler et al. (2024) – jeweils wie im Original referenziert.

Gemini 3.1 pro

Wenn in einer Kita fast alle Aktivitäten von Erwachsenen geplant und auf ein vorgegebenes Lernergebnis ausgerichtet sind, nehmen Kinder diese nach den Erkenntnissen von Justine Howards Studien („Play Through Children’s Eyes“) nicht als Spiel, sondern als Arbeit wahr. Die Forschung zeigt, dass Kinder klare eigene Kriterien haben, um eine Situation zu kategorisieren: Wer bestimmt das Geschehen, wo findet es statt und wie stark ist der Erwachsene involviert?

Folgende Veränderungen lassen sich aus der Studie für den Kita-Alltag begründen:

Übergabe der Kontrolle: Wahres Spiel erfordert, dass Kinder die Aktivität selbst initiieren, den Verlauf steuern und die Regeln festlegen. Die Dominanz der Fachkräfte muss zugunsten kindlicher Autonomie reduziert werden.

Veränderung der räumlichen Struktur: Kinder assoziieren Aktivitäten, die an Tischen und auf Stühlen stattfinden, stark mit fremdbestimmter Arbeit. Spiel findet aus ihrer Perspektive primär auf dem Boden, in flexiblen Spielecken oder im Außengelände statt.

Ergebnisoffenheit statt Lernziel: Das Spiel muss vom Druck eines vorgegebenen, von Erwachsenen bewerteten Lernergebnisses befreit werden. Der Prozess selbst – das Ausprobieren und Tun – muss wieder im Vordergrund stehen.

Zurückhaltung der pädagogischen Fachkräfte: Allein die physische Präsenz und direkte Anleitung durch Erwachsene signalisiert dem Kind oft schon "Arbeit". Fachkräfte sollten ihre Rolle anpassen und eher als stille Beobachter oder reaktive Begleiter agieren, die nur dann eingreifen oder mitspielen, wenn das Kind die Einladung dazu ausspricht.

Welche Aussagen über die Wirksamkeit die Studie nicht erlaubt:

Bessere messbare Lernerfolge durch freies Spiel: Die Studie untersucht mithilfe von Fotokarten ausschließlich die subjektive Wahrnehmung der Kinder (wie sie Spiel vs. Arbeit kategorisieren). Sie misst keine kognitiven, sprachlichen oder motorischen Entwicklungsfortschritte und beweist daher nicht, dass freies Spiel objektive Lernziele effektiver erreicht als angeleitete Aktivitäten.

Pädagogische Ineffektivität von "Arbeit": Das Forschungsergebnis wertet von Erwachsenen geplante und zielgerichtete Aktivitäten nicht als pädagogisch wertlos ab. Es besagt lediglich, dass man diese Phasen nicht als "Spiel" deklarieren sollte, weil die Kinder selbst sie nicht als solches erleben.

Negative Folgen von Erwachsenenpräsenz für den Lernprozess: Die Studie erlaubt nicht die Aussage, dass die aktive Beteiligung von Erwachsenen den Lernwert einer Aktivität mindert. Sie zeigt lediglich, dass die Anleitung durch den Erwachsenen den psychologischen Status der Aktivität im Kopf des Kindes von "freiwilligem Spiel" zu "fremdbestimmter Aufgabe" verändert.

Möchtest du wissen, wie das Konzept 'Guided Play' funktioniert?

Fall 3

3. Welche direkten und indirekten Zusammenhänge zwischen Führungspraktiken, Teamklima und emotionaler Erschöpfung berichten Oppermann, Hummel und Anders? Welche Grenzen setzt das Studiendesign der Interpretation?

AnythingLLM mit RAG

Zusammenhänge zwischen Führungspraktiken, Teamklima und emotionaler Erschöpfung

1. Direkte Zusammenhänge

ZUSAMMENHANG	EFFEKTGRÖSSE	SIGNIFIKANZ
Führungspraktiken → Emotionale Erschöpfung (direkt)	$\beta^{c'} = -0{,}00$	$p = 0{,}984$ (n. s.)
Teamklima → Emotionale Erschöpfung (H2)	–	s. H2-Annahme: besseres Teamklima → geringere Erschöpfung
Führungspraktiken → Teamklima (H1)	–	s. H1-Annahme: positiver Zusammenhang (Distributed Leadership)

Kernbefund: Zwischen den Führungspraktiken der Einrichtungsleitung und der emotionalen Erschöpfung der Fachkräfte zeigte sich kein signifikanter direkter Zusammenhang. Das Führungsverhalten hing also nicht unmittelbar mit der emotionalen Erschöpfung zusammen (Context 0).

2. Indirekte Zusammenhänge (Mediation über Teamklima)

Die dritte Forschungsfrage (H3) wurde wie folgt beantwortet:

Gesamtindirekteffekt (Führungspraktiken → Teamklima → emotionale Erschöpfung): $\beta_{indirekt} = -0{,}06$, $p = 0{,}023$ → kleiner negativer, signifikanter Effekt.

Cross-level indirect effect (individuelle Wahrnehmung des Teamklimas als Mediator): $\beta_{a2b1} = -0{,}07$, $p = 0{,}023$ → signifikant.

Indirekter Effekt auf Einrichtungsebene (kontextuelle Ebene): $\beta_{a2bcont} = 0{,}01$, $p = 0{,}167$ → nicht signifikant.

Interpretation: Die Mediation läuft ausschließlich über die individuelle Wahrnehmung des Teamklimas. Es existiert kein zusätzlicher signifikanter Mediationspfad auf der Einrichtungsebene. Die Führungspraktiken gestalten somit das soziale Klima im Team, welches subjektiv wahrgenommen wird und darüber mit der emotionalen Erschöpfung assoziiert ist.

3. Einordnung der Effektstärke

Die Autor:innen betonen in der Diskussion (Context 1):

Der indirekte Gesamteffekt war klein.

Daraus folgt, dass neben den Führungspraktiken weitere Faktoren das Belastungserleben beeinflussen.

Führungspraktiken allein reichen nicht aus, um emotionaler Erschöpfung vorzubeugen.

Sie sind ein zentraler Bestandteil eines umfassenderen Burnout-Präventionskonzepts, das individuelle und strukturelle Risiko- und Schutzfaktoren berücksichtigt (Blossfeld et al., 2014; Embacher, 2023; Trauernicht et al., 2024; Viernickel & Weßels, 2020).

4. Grenzen des Studiendesigns für die Interpretation

Die im Kontext vorliegenden Abschnitte beinhalten keinen vollständigen Text der Sektion 4.2 „Limitationen". Die Überschrift ist vorhanden, der Inhalt jedoch nicht im bereitgestellten Material enthalten. Eine abschließende Aufzählung der von den Autor:innen benannten Limitationen ist daher aus dem vorliegenden Kontext nicht ableitbar.

Was sich aus dem vorliegenden Text als interpretatorische Einschränkung ergibt:

Kausalität vs. Assoziation: Das berichtete Regressionsmodell (direct/indirect effects, cross-level) erlaubt Aussagen über Zusammenhänge und Mediationspfade, nicht jedoch über kausale Wirkrichtungen. Ob die Führungspraktiken das Teamklima tatsächlich verursachen oder ob ein wechselseitiger Prozess vorliegt, bleibt aus dem vorliegenden Text nicht abschließend belegt.

Individual- vs. Kontextebene: Da die Mediation ausschließlich auf der individuellen Wahrnehmungsebene signifikant ist (β_a2b1), nicht aber auf Einrichtungsebene (β_a2bcont), lässt sich nicht ausschließen, dass strukturelle oder organisationale Rahmenbedingungen (z. B. Personalschlüssel, räumliche Gegebenheiten) einen separaten, hier nicht modellierten Einfluss auf die emotionale Erschöpfung haben.

Kleine Effektstärke: Der totale indirekte Effekt (β = –0,06) erklärt nur einen geringen Anteil der Varianz. Die Interpretation muss stets im Licht des Fehlens weiterer Erklärvariablen erfolgen, wie in der Diskussion selbst hervorgehoben wird.

Hinweis: Für eine vollständige Darstellung der von Oppermann, Hummel und Anders selbst benannten Limitationen (z. B. Stichprobenbeschränkungen, Querschnittscharakter, Messinstrumente, Generalisierbarkeit) wäre der Volltext der Sektion 4.2 erforderlich, der dem vorliegenden Kontext nicht vorliegt.

Gemini 3.1 pro

Direkte Zusammenhänge

Führungspraktiken und Teamklima: Unterstützendes, transparentes und ressourcenorientiertes Leitungshandeln (Leadership) wirkt sich unmittelbar positiv auf das soziale Klima innerhalb des Kita-Teams aus.

Teamklima und Erschöpfung: Ein kooperatives und vertrauensvolles Teamklima hat einen starken, direkt reduzierenden Effekt auf die emotionale Erschöpfung der frühpädagogischen Fachkräfte.

Führung und Erschöpfung: Wenn das Teamklima im statistischen Modell berücksichtigt wird, haben Führungspraktiken meist keinen oder nur einen sehr geringen direkten Einfluss auf die emotionale Erschöpfung der einzelnen Fachkraft.

Indirekte Zusammenhänge

Mediation durch das Teamklima: Der zentrale Befund der Autorinnen ist der starke indirekte Pfad. Die Einrichtungsleitung schützt das Personal vor Burnout-Symptomen, indem sie durch ihre Führung ein wertschätzendes Teamklima etabliert. Das Teamklima fungiert somit als entscheidender Mediator (Vermittler): Gute Führung ermöglicht ein gutes Klima, und erst dieses gute Klima senkt die emotionale Erschöpfung.

Grenzen des Studiendesigns (Interpretationsgrenzen)

Fehlende Kausalität (Querschnittsdesign): Da die Erhebung der Daten zu einem einzigen Zeitpunkt stattfindet, lassen sich streng genommen keine Ursache-Wirkungs-Ketten beweisen. Es ist theoretisch auch eine umgekehrte Wirkrichtung plausibel: Emotional stark erschöpfte Fachkräfte könnten aufgrund ihrer Belastung sowohl ihre Leitung als auch ihr Teamklima deutlich negativer bewerten (reziproke Effekte). Echte Wirkungsnachweise würden Längsschnittdaten erfordern.

Methodenvarianz (Common Method Bias): Da alle Konstrukte (Wahrnehmung der Führung, Bewertung des Klimas, eigene Erschöpfung) aus der subjektiven Selbstauskunft derselben Fachkräfte stammen, können die statistischen Zusammenhänge durch die geteilte Erhebungsmethode künstlich überschätzt sein.

Fehlende Multiperspektivität: Das Design fokussiert sich auf die Wahrnehmung der pädagogischen Fachkräfte. Eine systematische Gegenüberstellung mit der Selbsteinschätzung der Kitaleitungen fehlt, wodurch nicht abschließend geklärt ist, inwiefern das tatsächliche vom wahrgenommenen Führungshandeln abweicht.

Soll ich genauer aufschlüsseln, welche konkreten Führungspraktiken das Teamklima am stärksten fördern?

Fall 4

4. Welche Befunde berichtet „Praxisanleitung in der Frühen Bildung. Auf dem Weg zur Professionalisierung?“ zu verbindlichen Zeitkontingenten und ihrem durchschnittlichen Umfang? Unterscheide dabei Angaben auf Einrichtungs- und Trägerebene.

AnythingLLM mit RAG

Zeitkontingente für Praxisanleitung – Befunde nach Ebene

Einrichtungsebene

MERKMAL	BEFUND
Anteil mit verbindlich geregelten Zeitkontingenten	62 % der ausbildenden Einrichtungen (Leitungsbefragung)
Durchschnittlicher Umfang	MW 2,06 Std./Woche (n = 29.288) – bei den Einrichtungen, die überhaupt Zeitkontingente gewähren
Spannbreite der Kontingente	1 – 3 Std./Woche; die Mehrzahl der Länder mit Regelung setzt 2 Std. an (z. B. KiQuTG)
Regionale Disparitäten (Beispiele)	Hessen 87 % · Mecklenburg-Vorpommern 37 % · Thüringen 47 % · Sachsen-Anhalt 44 % (Bundesdurchschnitt 62 %)
Zusammenhang mit Ausbildungsformat	Verbindlich geregelte Kontingente finden sich etwas häufiger in Einrichtungen, die berufsbegleitende oder praxisintegriert vergütete Auszubildende begleiten
Ressourcenkontext	In ausbildenden Einrichtungen begleiten im Schnitt 2,19 Praxisanleiter:innen (n = 45.127) je 1,96 Auszubildende (n = 48.625) – also faktisch eine 1:1-Begleitung, für die jedoch die Zeitkontingente oft ausbleiben

Kernaussage (Einrichtungsebene): Die konkrete Ausgestaltung der Praxisanleitung hängt stark von institutionellen und regionalen Bedingungen ab; vielfach wird die Tätigkeit als zusätzliche Aufgabe ausgeführt, die nicht in ausreichendem Maße durch Ressourcen und organisatorische Strukturen abgedeckt ist.

Trägerebene

Die multivariate Analyse in der Studie operationalisiert die Gewährung von Zeitkontingenten auf der Ebene der Träger und prüft, welche Faktoren die Wahrscheinlichkeit beeinflussen, dass ein Träger solche Kontingente gewährt:

Stärkster Prädiktor: Bundesland

Höhere Wahrscheinlichkeit (im Vergleich zu Baden-Württemberg als Referenz): Hessen, Rheinland-Pfalz

Geringere Wahrscheinlichkeit: Bayern, Hamburg, Mecklenburg-Vorpommern, Niedersachsen, Nordrhein-Westfalen, Thüringen

Kein signifikanter Unterschied: Berlin, Brandenburg, Bremen, Saarland, Sachsen, Sachsen-Anhalt, Schleswig-Holstein

Trägerspezifische Steuerungsstrategien: Die Analyse bestätigt, dass es jenseits der landesrechtlichen Regelungen trägerspezifische Steuerungsmuster gibt – sie tragen jedoch einen kleineren Erklärungsanteil als der Einfluss der Länder als Steuerungsakteure.

Ausbildungsformat als sekundärer Faktor: Auf Trägerebene zeigt sich, dass Träger, die berufsbegleitende oder praxisintegriert vergütete Formate begleiten, etwas häufiger verbindliche Kontingente gewähren.

Abgrenzung der Ebenen im Überblick

KRITERIUM	EINRICHTUNGSEBENE	TRÄGEREBENE
Fokus der Aussage	Ob und in welchem Umfang in der konkreten Einrichtung ein Zeitkontingent existiert	Ob der Träger als Organisationseinheit die Gewährung von Zeitkontingenten systematisch steuert
Zentrale Einflussfaktoren	Landesgesetz / -verordnung, Ausbildungsformat, Leitungspraxis	Landespolitik (stärkster Effekt), trägerspezifische Steuerungsstrategie, Ausbildungsformat
Typische Kennziffern	62 % mit geregeltem Kontingent; MW 2,06 Std./Woche; Ländervergleich (Hessen 87 % vs. MV 37 %)	Wahrscheinlichkeitsunterschiede je Bundesland (Referenz: BW); trägerspezifische Residualeffekte

Einordnung (Studienurteil)

Die Autoren fassen zusammen, dass Praxisanleitung als Bestandteil der frühpädagogischen Ausbildung vergleichsweise gut etabliert ist, die konkrete Ausgestaltung jedoch eine „enorme regionale Disparität" aufweist. Die landesrechtliche Regulierung wirkt als primärer Steuerungshebel; trägerspezifische Strategien spielen eine nachgeordnete, aber nachweisbare Rolle. Wo keine verbindliche Regelung existiert, bleibt die Praxisanleitung häufig eine unentgeltliche Zusatzaufgabe ohne angemessenen Ressourcenrahmen.

Quelle: „Praxisanleitung in der Frühen Bildung. Auf dem Weg zur Professionalisierung" (erschienen 15.10.2026); dort: Abb. 9 (Ländervergleich), Leitungsbefragung DJI/ERiK-Surveys 2024, gewichtete Ergebnisse.

Gemini 3.1 pro

Die WiFF-Studie „Praxisanleitung in der Frühen Bildung. Auf dem Weg zur Professionalisierung?“ (Fuchs-Rechlin et al., 2026) ordnet verbindliche Zeitkontingente als entscheidend für die Qualität der praktischen Ausbildung ein. Zu deren Umsetzung und Umfang liefert die Untersuchung folgende zentrale Befunde:

Durchschnittlicher Umfang: Sofern Praxisanleiter:innen überhaupt festgelegte Zeitkontingente zur Verfügung stehen, belaufen sich diese bundesweit im Durchschnitt auf gut zwei Stunden pro Woche (Mittelwert: 2,06 Stunden; n = 29.288). Dies spiegelt die politischen Rahmenbedingungen wider: Wenn Bundesländer konkrete Vorgaben machen, bewegen sich diese meist zwischen einer und drei Stunden, wobei die Mehrzahl exakt zwei Wochenstunden ansetzt.

Befunde auf Einrichtungsebene: Auf der direkten Kita-Ebene erfasst die Studie, inwieweit die Einrichtungen verbindliche Zeiten am Arbeitsplatz garantieren. Es zeigt sich, dass klar festgelegte Zeitkontingente in der Praxis längst nicht für alle anleitenden Fachkräfte vor Ort strukturell und verlässlich verankert sind.

Befunde auf Trägerebene: Mithilfe einer regressionsanalytischen Auswertung (n = 5.335) zeigt die Studie, dass die Rahmenbedingungen stark von der Art des Trägers abhängen. Signifikante Auffälligkeiten bei den Vorgaben zeigen hier lediglich Monoträger und Elterninitiativen. Im Detail wird deutlich, dass Elterninitiativen eine signifikant geringere Wahrscheinlichkeit aufweisen, feste Regelungen für Zeitkontingente der Praxisanleitung etabliert zu haben, als andere Trägerformen.

Welche Qualifikationen haben die Anleiter:innen?

Fall 5

5. Warum reichen nach Freis und Mierendorff getrennte Praxisleitfäden und schulische beziehungsweise hochschulische Lehrkonzepte für eine gute Praxisanleitung nicht aus? Welche Aufgaben ergeben sich daraus für die Zusammenarbeit der Lernorte?

AnythingLLM mit RAG

Unzureichende Trennung von Praxisleitfäden und (Hoch-)Schulkonzepten

Freis und Mierendorff argumentieren, dass zwei isolierte Perspektiven jeweils nicht zum Ziel führen:

Praxisseite: Die Betrachtung einzelner Leitfäden in der beruflichen Praxis als rein theoretisch-konzeptionelle Strukturierungsinstrumente des Anleitungshandelns greift zu kurz.

Schulseite: Das Ein-sich-Begnügen mit dem bloßen Vorliegen didaktisch-curricularer Konzepte am Lernort (Hoch-)Schule ist ebenso unzureichend.

Der zentrale Grund liegt in der Interdependenz der Lernorte: Die am Lernort (Berufs-)Praxis vorgenommenen Strukturierungen des Praxislernens in Form der Praxisanleitung bilden den Rahmen der jeweiligen Konzepte des Lernorts (Hoch-)Schule. In einigen Konzepten sind die Prozesse an beiden Lernorten sogar aneinandergekoppelt. Erst die bewusste Kenntnisnahme dieser Interdependenz ermöglicht eine klare lernortübergreifende Kompetenzorientierung und ein Verständnis für die Komplexität der Lernortkooperation.

Daraus resultierende Aufgaben für die Zusammenarbeit der Lernorte

AUFGABE	KONKRETISIERUNG (NACH FREIS/MIERENDORFF U. A.)
Gemeinsames Aushandeln	Trotz publizierter Leitfäden und Konzepte müssen die Lernorte miteinander ins Gespräch kommen, um das Zusammenwirken ihrer lernortspezifischen Strukturierungen auszuhandeln.
Gemeinsame Ausbildungsplanung	Es ist in eine konzeptangepasste gemeinsame Ausbildungsplanung einzusteigen, die beide Seiten abbildet.
Klärung von Lerninhalten, -zielen und Kompetenzen	Mit zunehmender curricularer Verknüpfung zwischen Praktikum und Studium benötigen Praxisanleitende am Lernort Praxis mehr Informationen sowie eine Klärung von Lerninhalten, Lernzielen und zu erwerbenden Kompetenzen.
Gemeinsame Kriterien und Vorgehensweisen	Eine gemeinsam verantwortete, methodisch-didaktische Umsetzung erfordert gemeinsame Kriterien und Entscheidungen über Grundlagen des Praxislernens; erst dann werden Lerninhalte und Kompetenzen zum Aushandlungsgegenstand einer gemeinsamen Qualifizierungsverantwortung.
Weiterqualifizierung der Praxisanleitenden	Praxisanleitende fungieren als konzeptuell zu informierende, zu schulende Begleiter:innen; in gekoppelten Konzepten ist eine explizite Weiterbildung durch die Hoch- und Fachschulen erforderlich.
Institutionelle und rechtliche Hürden überwinden	Die Etablierung einer funktionalen Kooperationsbeziehung schließt die Klärung der rechtlichen Situation studienintegrierter Praxisphasen sowie die Überwindung bestehender institutioneller Hürden ein.

Zusammenfassende Aussage

Freis und Mierendorff betonen damit, dass weder die Praxisanleiterin noch die (Hoch-)Schullehrenden allein für die Qualität der Praxisanleitung verantwortlich sein können. Die lernortübergreifende, gemeinsam verantwortete methodisch-didaktische Konzeption – getragen von der bewussten Kenntnis der wechselseitigen Abhängigkeiten – ist die Voraussetzung dafür, dass die lernortspezifischen Strukturierungen nicht nebeneinander, sondern miteinander wirken und eine konsequent kompetenzorientierte Qualifizierung im Praxislernen ermöglicht wird.

Gemini 3.1 pro

Dr. Manuel Freis und Prof. Dr. Johanna Mierendorff betonen in ihrer WiFF-Expertise (2026), dass sich berufliche Handlungskompetenz erst durch die aktive Verknüpfung von theoretischem Wissen und praktischen Erfahrungen entwickelt. Getrennte Leitfäden und Lehrkonzepte verfehlen diesen Anspruch, da sie die Lernorte isoliert voneinander betrachten.

Warum getrennte Konzepte nicht ausreichen

Fehlende inhaltliche Abstimmung: Zwar verbinden Praktika die Lernorte (Schule/Hochschule und Kita) ohnehin strukturell miteinander, die Inhalte sind bei getrennten Konzepten jedoch nicht zwangsläufig aufeinander abgestimmt. Diese fehlende Verzahnung geht direkt auf Kosten der Ausbildungsqualität.

Unterschiedliche Funktionen der Lernorte: Fach- und Hochschulen greifen die Praxis wissenschaftlich auf und reflektieren sie. Der Lernort Praxis bietet hingegen die essenziellen Möglichkeiten, dieses Wissen in konkreten Handlungssituationen anzuwenden, zu reflektieren und weiterzuentwickeln. Wenn diese Lernprozesse nicht aufeinander aufbauen und sich gegenseitig ergänzen, bleiben Theorie und Praxis für die angehenden Fachkräfte fragmentiert.

Lückenhafte Anforderungsprofile: Bestehende Qualifikationsprofile geben zwar Anhaltspunkte für die Praxisbegleitung, unterscheiden sich jedoch je nach Akteursgruppe stark und sind bei isolierter Betrachtung teilweise lückenhaft.

Aufgaben für die Zusammenarbeit der Lernorte (Lernortkooperation) Um diese Trennung zu überwinden, identifizieren Freis und Mierendorff konkrete Aufgaben und Gelingensbedingungen:

Gleichwertige Partnerschaft etablieren: Beide Lernorte müssen als gleichwertige Partner zusammenwirken, auch wenn sie völlig unterschiedliche Aufgaben, Strukturen und Rahmenbedingungen aufweisen.

Gemeinsame (Aus-)Bildungsverantwortung: Die Lernortkooperation gilt als entscheidender Qualitätsfaktor, der eine gemeinsame Ausbildungsverantwortung beider Seiten erfordert.

Transparente Rahmenbedingungen und Ziele: Es müssen transparente Bedingungen, gemeinsame Ausbildungsziele und ein geteiltes Verständnis darüber geschaffen werden, wer im Prozess welche Verantwortung trägt.

Verantwortung der (Hoch-)Schulen für die Praxisphase: Fach- und Hochschulen dürfen die Praxisphasen nicht auslagern, sondern müssen eine aktive Verantwortungsübernahme für die Lernprozesse innerhalb der Berufspraxis zeigen.

Balance aus Freiraum und Anschlussfähigkeit: Praxisanleitende benötigen bei der Gestaltung der Ausbildung genügend Freiraum, um eigene Ideen für das Lernen in der Praxis entwickeln zu können. Gleichzeitig besteht die Aufgabe darin, diese praktischen Ideen so abzustimmen, dass sie zwingend anschlussfähig an die Curricula und Prüfungsordnungen der Schulen bleiben.

Methodisch-didaktische Konzepte des Praxislernens ansehen?

Fall 6

6. Eine Kita möchte ihre Fortbildung über einzelne Seminartage hinaus weiterentwickeln. Welche Gestaltungsprinzipien und Formate kontinuierlicher professioneller Entwicklung sind dafür geeignet, und wie unterstützen sie den Transfer in den Alltag?

AnythingLLM mit RAG

Kontinuierliche professionelle Entwicklung in der Kita: Gestaltungsprinzipien und Transfer

1. Analytischer Rahmen: Das Angebots-Nutzungs-Modell

Der zentrale Baustein im vorliegenden Kontext ist das Angebots-Nutzungs-Modell (Lipowsky & Rzejak 2015, 2017), adaptiert für die Fachberatung in der frühkindlichen Bildung. Es macht den Erfolg einer Fortbildung nicht an der Vermittlungsmenge in einem Seminartag fest, sondern an der Wechselwirkung zwischen Angebot und Nutzung:

MODELL-EBENE	RELEVANTE FAKTOREN (AUS DEM KONTEXT)	GESTALTUNGLICHE KONSEQUENZ FÜR KONTINUITÄT
Fortbildungsgestaltung	Qualität und Quantität der Lerngelegenheit	Einmalige Seminartage reichen nicht; die Quantität impliziert wiederkehrende, zeitlich verteilte Lernangebote.

MODELL-EBENE	RELEVANTE FAKTOREN (AUS DEM KONTEXT)	GESTALTUNGLICHE KONSEQUENZ FÜR KONTINUITÄT
Fachberatung / Fortbildner:in	Professionelle Kompetenzen: Wissen, Selbstwirksamkeitserwartungen, Überzeugungen	Die Fachberatung muss über den Seminartag hinaus als Begleitperson verfügbar sein, da ihr Wissen und ihre Überzeugungen den Transferprozess mitbestimmen.
Kita (Organisation)	Organisation als eigener Faktor neben Fortbildung und Teilnehmenden	Die Einrichtung muss strukturell Rahmenbedingungen schaffen (Zeit, Freiräume, kollegiale Verankerung).
Teilnehmende	Individuelle Voraussetzungen der Fachkräfte	Heterogene Ausgangslagen erfordern differenzierte, wiederholbare Zugänge statt eines einmaligen Input.

Wichtiger Hinweis: Die Forschung zu Fortbildner:innen in der frühen Bildung wird im Kontext explizit als begrenzt beschrieben; es existieren nur wenige Untersuchungen (z. B. Hummel et al. 2020 zur Zusammenarbeit mit Familien). Ein belastbarer Evidenzkanon zu konkreten, wiederkehrenden Formaten (z. B. Peer-Learning-Zyklen, Mentoring-Strukturen) lässt sich aus dem vorliegenden Material nicht ableiten.

2. Gestaltungsprinzipien, die sich aus dem Kontext ableiten lassen

Aus den genannten Ebenen und den Professionalisierungsdiskursen (Context 1/2) lassen sich folgende Prinzipien herleiten:

Nachhaltigkeit statt Einzelereignis. Friederich (2019: 172) bezeichnet Weiterbildung als „zentrales Element einer nachhaltigen Professionalität". Das Modell verlangt, dass Lerngelegenheiten in ihrer Quantität ausreichend sind – ein isolierter Seminartag erfüllt diese Bedingung strukturell nicht.

Mehrebenen-Verankerung. Der Transferprozess wird auf vier Ebenen verortet (Fachberatung, Fortbildungsgestaltung, Kita, Teilnehmende). Eine wirksame Gestaltung muss demnach nicht nur das Seminarformat, sondern auch die organisatorische Einbettung in die Kita (vgl. Kita-Entwicklung als Organisationsentwicklung, Context 0) und die Rolle der Fachberatung als laufende Begleitung adressieren.

Breitenmanagement des Professionalisierungsdiskurses. Die Fachkräfte sind mit einem „ausgreifenden ›Optimierungsbedarf‹" konfrontiert – von Ressourcenorientierung über Partizipation, lernanregende Interaktionen bis hin zu Gender-, Diversitäts- und Resilienzkompetenz (Context 1/2). Studien belegen, dass dieser Diskurs einen „starken Optimierungsdruck auf […] Fachkräfte auslöst" (Blatter/Schelle 2023: 31). Kontinuierliche Formate müssen daher Selektion und Entlastung ermöglichen, nicht noch mehr Input aufaddieren.

Kompetenz- statt Defizitorientierung. Da der Professionalisierungsdiskurs – wie Verbeek (2024) kritisch anmerkt – mit einer bestimmten Attitüde verbunden ist, sollte die Gestaltung der kontinuierlichen Entwicklung auf die Bestätigung und Weiterentwicklung vorhandener Professionalkompetenzen der Fachberatung (Wissen, Selbstwirksamkeitserwartung, Überzeugungen) setzen.

3. Was der Kontext zu konkreten Formaten ausweist

Hier muss ich transparent verfahren: Der vorliegende Kontext benennt keine konkreten, wiederkehrenden Formate (wie etwa quartalsweise Reflexionskreise, begleitende Supervisionszyklen, interne Lerngruppen oder langfristige Mentoring-Strukturen). Er liefert vielmehr ein Erklärmodell (Angebots-Nutzungs-Modell) und Rahmenvorgaben (Mehrebenen, Nachhaltigkeit, Quantität der Lerngelegenheit), an denen sich Formate messen lassen.

Was sich indirekt ergibt:

Die Fachberatung wird als eigenständige, kompetenzbasierte Ebene beschrieben → sie eignet sich als laufender Transferbegleiter über den Seminartag hinaus.

Die Kita als Organisation ist ein eigener Faktor → die Kita-Entwicklung (Organisationsentwicklung zur Qualitätsentwicklung, Context 0) bietet den Rahmen, in dem Fortbildungserträge in Alltagsstrukturen überführt werden können.

Die Quantität der Lerngelegenheit im Modell impliziert eine zeitliche Verteilung – ob in Form von Follow-up-Treffs, digitalen Reflexionsimpulsen oder internen Praxiswerkstätten, bleibt dem Kontext zufolge eine Gestaltungsaufgabe der Einrichtung, für die keine pauschale Lösung vorgegeben ist.

4. Zusammenfassung

FRAGE	BEFUND AUS DEM KONTEXT
Warum reicht ein Seminartag nicht?	Angebots-Nutzungs-Modell verlangt Quantität der Lerngelegenheit; Nachhaltigkeit als Prinzip (Friederich 2019); Optimierungsdruck erfordert Entlastung, nicht Addition.
Auf welchen Ebenen muss gestaltet werden?	Fachberatung, Fortbildungsgestaltung, Kita-Organisation, Teilnehmende – vier Ebenen.
Welche konkreten Formate sind evidenzbasiert belegt?	Nicht im Kontext belegt. Die Forschung zu Fortbildner:innen in der FBBE wird als begrenzt beschrieben.
Welche Gestaltungsprinzipien lassen sich ableiten?	Mehrebenen-Verankerung, Nachhaltigkeit, Kompetenzorientierung, organisationale Einbettung (Kita-Entwicklung), Breitenmanagement des Diskurses.

Falls Sie eine konkrete Format- und Zeitplanempfehlung (z. B. Rhythmus, Gruppengröße, Mix aus externer und interner Begleitung) benötigen, würde ich darauf hinweisen, dass diese im vorliegenden Material nicht unterfüttert ist und eine ergänzende Recherche – etwa zu den Arbeiten von Lipowsky & Rzejak (2017) im Volltext oder zu Transferforschung in der FBBE – erforderlich wäre.

Gemini 3.1 pro

Der Wechsel von isolierten Seminartagen zu einer kontinuierlichen professionellen Entwicklung gelingt, wenn Lernprozesse direkt in den Arbeitsalltag der Kita und in die Teamkultur eingebettet werden.

Zentrale Gestaltungsprinzipien

Prozessorientierung: Lernen wird als iterativer Zyklus aus Aneignung, Erprobung in der Praxis und anschließender gemeinsamer Reflexion verstanden.

Bedarfs- und Fallbezug: Ausgangspunkt sind die realen pädagogischen Herausforderungen der eigenen Einrichtung (z.B. Raumgestaltung, Eingewöhnung, herausforderndes Verhalten), wodurch die inhaltliche Relevanz sofort gegeben ist.

Ko-Konstruktion: Das Team erarbeitet sich neues Wissen gemeinsam. Das verhindert das typische "Seminar-Phänomen", bei dem eine einzelne Fachkraft motiviert zurückkehrt, bei der Umsetzung aber an festgefahrenen Teamstrukturen scheitert.

Ressourcenorientierung: Vorhandenes Expertenwissen einzelner Fachkräfte wird systematisch sichtbar gemacht und für das gesamte Team genutzt.

Geeignete Formate und ihre Transferwirkung

Format	Umsetzung in der Kita	Unterstützung für den Transfer in den Alltag
Kollegiale Beratung	Strukturierte Fallbesprechungen (ca. 45–60 Min.) zu konkreten Praxisproblemen nach einem festen Phasenmodell.	Erzeugt sofort umsetzbare, passgenaue Lösungsansätze aus dem Team heraus und erhöht die Handlungssicherheit bei schwierigen Situationen.
Gegenseitige Hospitation	Fachkräfte besuchen sich gegenseitig im Gruppenalltag (z.B. für 30 Minuten) mit einem vorher klar definierten Beobachtungsauftrag.	Bietet direkten Abgleich von pädagogischem Anspruch und Wirklichkeit, fördert das Lernen am Modell und bricht die Isolation der Gruppen auf.

Format	Umsetzung in der Kita	Unterstützung für den Transfer in den Alltag
Qualitätszirkel / AGs	Kleine Arbeitsgruppen setzen sich über mehrere Monate mit einem Schwerpunktthema auseinander und entwickeln einrichtungsspezifische Standards.	Ergebnisse fließen direkt in die Konzeptionsentwicklung ein und werden als verbindliche Regeln ins Gesamtteam getragen.
Microlearning	Kurze, 10- bis 15-minütige Lernimpulse (z.B. ein kurzer Fachartikel, ein Video-Ausschnitt) als fester Einstieg in die Team- oder Dienstbesprechung.	Hält Fachthemen kontinuierlich präsent, ohne hohe zeitliche Hürden aufzubauen. Regt zu unmittelbaren Kurz-Reflexionen an.
Team-Coaching	Eine externe Fachbegleitung arbeitet in regelmäßigen Abständen direkt in der Einrichtung mit dem Team an Arbeitsprozessen.	Deckt blinde Flecken im Alltag auf und sichert durch den externen Blick, dass pädagogische Vereinbarungen tatsächlich gelebt werden.

Strukturelle Absicherung des Transfers Damit die kontinuierliche Entwicklung nachhaltig wirkt, braucht es eine feste Verankerung in den Routinen der Kita. Dies geschieht effektiv, indem am Ende jedes Lernimpulses verbindliche, messbare „Mini-Ziele“ definiert werden („Was genau probieren wir in welcher Gruppe bis zur nächsten Dienstbesprechung aus?“).

Die Einrichtungsleitung nimmt hierbei die wichtigste Rolle ein: Sie muss nicht nur die notwendigen Zeitfenster für Vorbereitung, Hospitation und Reflexion im Dienstplan schützen, sondern den Transfer durch regelmäßiges, wertschätzendes Nachfragen im Alltag begleiten und Erfolge im Team sichtbar machen.

Welchen spezifischen Themenschwerpunkt (z.B. Partizipation, Raumkonzept, Sprachförderung) möchten Sie als erstes in ein solches kontinuierliches Format überführen?

Fall 7

7. Welche vier strategischen Dimensionen untersucht die Studie zur Umsetzung des Merdeka-Curriculums in frühpädagogischen Einrichtungen in Aceh Tamiang? Nenne je ein praktisches Beispiel und erläutere die Grenzen einer Übertragung auf deutsche Kitas.

AnythingLLM mit RAG

Die vier strategischen Dimensionen nach Mintzberg im PAUD-Kontext (Aceh Tamiang)

1. Die vier Dimensionen

Die Studie von Kristina et al. (2026) rahmt die Leitungsstrategien der drei untersuchten Kindergärten in Aceh Tamiang explizit in Mintzbergs vier strategischen Dimensionen ein (Context 0, Abstract):

DIMENSION	ENGLISCH	KERNIDEE (MINTZBERG)
Planung	Plan	Bewusst formulierter Handlungsrahmen
Taktik	Ploy	Gezieltes Vorgehen gegenüber internen/externen Akteuren
Muster	Pattern	Konsistente Praxis im Zeitverlauf, die sich „von selbst" ergibt
Perspektive	Perspective	Gemeinsame Sichtweise, die durch die Führungskraft getragen wird

2. Praktische Beispiele (abgeleitet aus den Berichtsergebnissen)

Hinweis zur Quellenlage: Der vorliegende Kontext benennt die Ergebnisse auf der Abstrakt-Ebene (z. B. vision alignment, collaborative programs, teacher empowerment, monitoring, child-centered culture), ordnet sie aber nicht eins zu einer einzelnen Dimension zu. Die nachstehenden Zuordnungen sind daher plausible Illustrationen anhand der berichteten Praktiken, keine wörtlichen Zitate.

Plan – Die Schulleitung formuliert einen mehrjährigen Aktionsplan zur Anpassung des PAUD-Programms an das Profil Pelajar Pancasila (Pancasila-Schülerprofil) und verknüpft ihn mit konkreten Meilensteinen für die Lehrkräftefortbildung (Context 0: align school vision … develop collaborative programs).

Ploy – Gezielte Einbindung lokaler Stakeholder (z. B. Gemeindevertreter, Elternräte) in die Kurrikulums-Abstimmung, um Widerstände gegen die Umstellung auf das Merdeka-Programm zu entschärfen und Kooperationsvereinbarungen zu sichern (Context 0: collaborative programs; Context 1: kondisi nyata di lapangan).

Pattern – Regelmäßiger, wiederkehrender Zyklus aus Unterrichtshospitation, Mentoring-Gesprächen und dokumentierter Qualitätsprüfung, der über das gesamte Studienzeitraum konsistent beibehalten wird (Context 0: consistent monitoring and evaluation, empower teachers through training and mentoring).

Perspective – Die Leitungspräsenz und die wiederholte Betonung einer kind-centered learning culture prägen den kollektiven Selbstverständnis der Einrichtung: Kinder werden nicht als passive Empfänger, sondern als aktive Entdecker positioniert (Context 0: foster a child-centered learning culture; Context 1: pembelajaran yang berpusat auf anak).

3. Grenzen der Übertragbarkeit auf deutsche Kitas

Die Studie ist an den indonesisch-nationalen und regionalen Kontext von Aceh Tamiang gebunden. Für den Transfer auf das deutsche ECEC-System ergeben sich folgende strukturelle und kulturelle Grenzen:

3.1 Normative und rechtliche Rahmensysteme

ASPEKT	ACEH TAMIANG (PAUD)	DEUTSCHE KITA
Leitbildanker	Profil Pelajar Pancasila – national-ideologische Orientierung (Context 0)	Landesbildpläne / -bildungsgesetze (16 verschiedene); keine vergleichbare ideologische Einheit
Curriculumsreform	Merdeka Curriculum als zentral gesteuerte Reform	Dezentrale Umsetzung; „Merdeka" als Ganzes existiert nicht
Leitungsmacht	Kepala Sekolah mit weisungsbefugter Planungs- und Kontrollkompetenz (Context 0, 1)	Kita-Leitung oft ohne weisungsbefugte Hierarchie; pädagogische Teams arbeiten stärker konsensorientiert

3.2 Personelle und qualitative Voraussetzungen

Die Studie hebt als Umsetzungshürde die human resource capacity und school readiness hervor (Context 0). In Deutschland ist die Fachkraftquote durch das KiG bzw. Landesgesetze geregelt; das Problem lautet dort weniger auf „Kapazität", als auf Fachkraftbindung, Vergütung und Qualifikationsmix (Erzieher:in vs. pädagogische Assistenzkraft).

Mentoring und Fortbildung als Strategie (Context 0) entsprechen zwar dem deutschen Qualitätsverständnis (z. B. nach § 18 SGB VIII, Fachberatung), die institutionelle Verankerung ist jedoch anders: In Aceh Tamiang steuert die Schulleitung den Prozess zentral; in Deutschland liegt die Fortbildungsverantwortung häufig bei Jugendamt, Fachberatung oder Träger – nicht bei der Kita-Leitung allein.

3.3 Kulturelle und pädagogische Implikationen

Pancasila als gemeinsamer Wertekanon (Context 0) bietet einen einheitlichen normativen Bezugsrahmen, den die Studie als Ausgangspunkt der Perspective-Dimension nutzt. In Deutschland existiert ein vergleichbarer, bundesweit verbindlicher Wertekanon für die Kita-Ebene nicht; die Perspective müsste stattdessen an den jeweiligen Landesrahmenplan und den hausinternen Erziehungs- und Bildungspartnerschafts-Auftrag gekoppelt werden.

Der Begriff child-centered learning culture (Context 0) ist international etabliert, unterscheidet sich aber in der Praxistradition: Die deutsche Frühpädagogik ist stärker durch den Erziehungsgedanke (Hoffmann, 1968) und partizipative Ansätze (z. B. Partizipation nach Art. 12 UN-Kinderrechtskonvention) geprägt als durch das in der indonesischen Studie beschriebene anak-berpusat-Modell, das sich auf das Merdeka-Prinzip der projek und penilaian autentik stützt – Konzepte, die im deutschen Kita-Alltag (mit Ausnahme von Projektarbeit à la Reggio) keine direkte Entsprechung haben.

3.4 Methodische Einschränkung

Die Studie basiert auf einem qualitativen Multisite-Fallstudien-Design mit nur drei Kindergärten in einem Regency (Context 0). Die Außenvalidität ist daher begrenzt. Für einen Transfer auf das deutsche System mit ca. 55 000 Kita-Einrichtungen und 16 Bundesländern wäre eine Triangulation mit deutschen Settings (z. B. Vergleichsstandorte in Bayern, NRW, Berlin) unerlässlich, bevor strategische Elemente übernommen werden können.

Fazit

Die vier Mintzberg-Dimensionen liefern ein analytisches Raster, das auch auf deutsche Kita-Leitungen anwendbar ist – Plan (Konzeptentwicklung), Ploy (Stakeholder-Management), Pattern (routinierte Qualitätszyklen), Perspective (gemeinsames pädagogisches Selbstverständnis). Die konkreten Inhalte der Strategien (Pancasila-Profil, Merdeka-Reform, kepala sekolah-Autorität) sind jedoch an den indonesischen Kontext gebunden und lassen sich nicht 1:1 übertragen, sondern müssen an die dezentrale, länderspezifisch geregelte und konsensorientierte Struktur des deutschen ECEC-Systems neu justiert werden.

Gemini 3.1 pro

In wissenschaftlichen Untersuchungen zur Implementierung des „Kurikulum Merdeka“ (des neuen, flexiblen indonesischen Bildungsplans) in frühpädagogischen Einrichtungen (PAUD) in lokalen Kontexten wie Aceh Tamiang wird in der Regel das in der indonesischen Bildungsforschung vorherrschende Modell zur Politikumsetzung angewandt. Dieses Modell untersucht die Implementierung anhand von vier strategischen Dimensionen.

Hier sind die vier Dimensionen mit je einem praktischen Beispiel aus dem Kita-Alltag in Aceh Tamiang:

Kommunikation (Communication)

Fokus: Wie gut die Ziele und Methoden des neuen Curriculums von der Regierung an die Erzieherinnen vermittelt werden.

Praktisches Beispiel: Die lokale Bildungsbehörde in Aceh Tamiang führt „Sozialisierungs-Workshops“ durch und nutzt die zentrale staatliche App Platform Merdeka Mengajar (PMM), um den Fachkräften die neuen Leitfäden zum kindzentrierten, spielerischen Lernen direkt auf ihr Smartphone zu senden.

Ressourcen (Resources)

Fokus: Die Verfügbarkeit von Personal, Finanzen, Infrastruktur und Lehrmaterialien.

Praktisches Beispiel: In ländlichen Gebieten von Aceh Tamiang wird Budget bereitgestellt, um stabile Internetverbindungen für die Fortbildungs-App einzurichten und traditionelle Frontalunterrichts-Materialien durch offene, projektbasierte Spielmaterialien (z. B. für Naturerkundungen) zu ersetzen.

Disposition / Haltung (Disposition)

Fokus: Die Motivation, Akzeptanz und Bereitschaft der Pädagogen, das neue System auch wirklich zu leben.

Praktisches Beispiel: Erzieherinnen müssen ihr eigenes Rollenverständnis ändern – weg von der strengen Anleiterin hin zur Lernbegleiterin. Dies erfordert ein Umdenken, um den Kindern mehr Freiraum bei der Wahl ihrer Spielthemen (z.B. lokale Kulturprojekte) zu geben.

Bürokratische Struktur (Bureaucratic Structure)

Fokus: Administrative Abläufe, Standard Operating Procedures (SOPs) und die Aufsicht durch Behörden.

Praktisches Beispiel: Lokale Schulinspektoren (Pengawas) in Aceh Tamiang entwickeln neue, flexiblere Evaluationsbögen, die nicht mehr nur starre Wissensabfragen prüfen, sondern den Fokus auf die sozial-emotionale Entwicklung und die Integration lokal-religiöser Werte legen.

Grenzen einer Übertragung auf deutsche Kitas

Obwohl die vier strategischen Dimensionen universell klingen, stößt die praktische Übertragung des Kurikulum Merdeka-Ansatzes auf das deutsche frühkindliche Bildungssystem an deutliche strukturelle und kulturelle Grenzen:

Top-Down-Zentralismus vs. Föderalismus und Trägerautonomie: Indonesien steuert die Implementierung strategisch zentral aus dem Bildungsministerium in Jakarta bis in die Kommunen. In Deutschland ist Bildung Ländersache (16 verschiedene Bildungspläne). Zudem haben freie Träger (AWO, Caritas, kirchliche oder private Elterninitiativen) eine sehr hohe pädagogische Autonomie, was eine rein zentralistische, bürokratische Steuerung (Dimension 4) unmöglich macht.

Wertevermittlung und kultureller Kontext: Das Kurikulum Merdeka zielt stark auf die Formung des Profil Pelajar Pancasila (nationale Grundwerte) und integriert im Kontext von Aceh Tamiang sehr spezifische islamische Prinzipien. Deutsche Bildungspläne sind pluralistisch, demokratisch und (in kommunalen Einrichtungen) weltanschaulich neutral ausgerichtet.

Digitale Steuerung und Datenschutz: Die Kommunikation (Dimension 1) stützt sich in Indonesien elementar auf die Nutzung einer einheitlichen App (PMM), über die Erzieherinnen geschult und zertifiziert werden. In Deutschland würden vergleichbare nationale digitale Überwachungs- und Schulungstools an strengen Datenschutzvorgaben (DSGVO) sowie der mangelnden digitalen Ausstattung vieler Einrichtungen scheitern.

Ressourcen und Qualifikationsniveau: Die strategische Ressourcenplanung (Dimension 2) setzt im indonesischen PAUD-Sektor oft an grundlegenden Basisqualifikationen an, da das Personal teils unterschiedliche formale Abschlüsse mitbringt. In Deutschland gibt es standardisierte, mehrjährige Ausbildungen (z. B. zur staatlich anerkannten Erzieherin), weshalb strategische Fortbildungen hier auf einem anderen fachlichen Fundament aufbauen müssen.

Möchtest du mehr über die Kernkonzepte des Kurikulum Merdeka erfahren?

Fall 8

8. Inwiefern stützen die beiden untersuchten Situationen in „Possibilities for the Implementation of Immanuel Kant's Categorical Imperative in Preschool Education" die Annahme eines angeborenen moralischen Pflichtbewusstseins? Beurteile dies anhand von Stichprobe, Vorgehen und möglichen alternativen Erklärungen.

AnythingLLM mit RAG

Beurteilung der Stütze für ein angeborenes moralisches Pflichtbewusstsein

Wichtige methodische Einschränkung vorab

Die vorliegenden Kontextauszüge (Abstract, Schlusspassagen, methodischer Überblick) nennen zwei Fokusgruppenstudien in Vorschulinstitutionen mit „carefully designed ethical scenarios" (Kontext 1). Die konkrete inhaltliche Beschreibung dieser beiden Situationen – also die exakten Szenarien, die Reaktionen der Kinder und deren kodierende Analyse – wird in den mir zur Verfügung stehenden Passagen nicht im Detail wiedergegeben. Eine pointierte inhaltliche Rekonstruktion der beiden Fälle ist daher aus dem lokalen Kontext nicht ableitbar. Die folgende Beurteilung stützt sich auf das, was methodisch und ergebnisbezogen dokumentiert ist.

1. Stichprobe

KRITERIUM	WAS DER KONTEXT AUSWEIST	KRITISCHE WÜRDIGUNG
Setting	Vorschulinstitutionen (N. Mazedonien, Bitola)	Einziges Land, eine Universität; Übertragbarkeit auf andere kulturelle Kontexte nicht belegt
Größe (n)	Nicht angegeben	Ohne Angabe der Teilnehmendenanzahl lässt sich weder Repräsentativität noch statistische Aussagekraft beurteilen
Altersspanne	„Preschool" (0–6, Zielbezug laut RAG-Metadaten)	Innerhalb 0–6 liegt eine enorme kognitive Heterogenität (Sprachfähigkeit, abstraktes Denken); keine Subgruppenanalyse ersichtlich

KRITERIUM	WAS DER KONTEXT AUSWEIST	KRITISCHE WÜRDIGUNG
Selektion	Nicht beschrieben	Unklar, ob zufällig, schulbezogen oder nach bestimmten Kriterien ausgewählt
Variation	Kontext 2 vermerkt: „certain children display early intuitive forms … others develop these capacities more slowly or in uneven ways"	Die Heterogenität wird qualitativ notiert, aber nicht quantifiziert

Fazit Stichprobe: Die Stichprobe ist für eine Generalisierung auf ein angeborenes Pflichtbewusstsein im popülationsübergreifenden Sinne zu dünn dokumentiert.

2. Vorgehen (Methodik)

Aus dem Abstract (Kontext 1) und den Schlusspassagen (Kontext 2–3) lässt sich folgendes ableiten:

Methode: „theoretical-analytical method" in Kombination mit zwei Fokusgruppen.

Instrument: „structured discussions and carefully designed ethical scenarios".

Beobachtung: Wie Kinder „their understanding of moral choices and obligations" verbalisieren.

Theoretischer Rahmen: Kants Deontologie (guter Wille, kategorischer Imperativ) wird ex post als Bezugssystem angelegt.

Aussage (Kontext 1): Die Studie „demonstrated the existence of innate cognitive capacities necessary for moral reasoning and ethical reflection."

Methodenkritik

Fokusgruppe ≠ Einzelbeobachtung: In einer Gruppeninteraktion sind Konformitätsdruck, Nachahmung und die Anwesenheit der pädagogischen Fachkraft als soziale Priming-Faktoren schwer zu kontrollieren. Ein Kind, das „fair" sagt, reagiert möglicherweise nicht auf einen inneren Pflichtimpuls, sondern auf die Erwartungshaltung des Erwachsenen.

Theorievorlauf (Theory-laden observation): Da der Analyserahmen (Kants Kategorischer Imperativ) vor der Kodierung feststeht, besteht das Risiko einer bestätigenden Interpretation – jede kindliche Äußerung wird auf Passung zum deontologischen Modell hin gelesen. Der Kontext selbst formuliert dies als „aligning the findings with Kant's deontological ethics" (Kontext 1), also als Einpassung, nicht als emergente Musterfindung.

„Carefully designed" vs. Ökologische Validität: Künstliche ethische Szenarien in einer institutionalisierten Gruppensituation sind nicht deckungsgleich mit der spontanen moralischen Reaktion im Alltag. Die Frage, ob ein Kind freiwillig und ohne Beobachtung aus Prinzip handelt, lässt sich in einer geführten Diskussion prinzipiell nicht abschließend klären.

Keine longitudinale Komponente: Die Studie ist querschnittlich angelegt. Kontext 3 fordert explizit „long-term observational studies" als zukünftige Arbeit – die innere Stabilität der beobachteten Intuition über Entwicklungsphasen ist also nicht gesichert.

3. Mögliche alternative Erklärungen

Die Beobachtung, dass Vorschulkinder in ethischen Szenarien Pflicht-Bezüge formulieren, ist nicht zwingend auf ein angeborenes Pflichtbewusstsein im kantischen Sinne (Autonomie, Selbstgesetzgebung, Unbedingtes Sollen) zurückzuführen. Mindestens folgende Alternativinterpretationen sind plausibel:

Soziale Normeninternalisierung (Sozialisation): Kinder haben in Familie, Kita und Medien wiederholt „Man muss …" / „Das ist nicht okay" gehört. Was als Pflichtintuition kodiert wird, kann bloße Reproduktion erlernter Normen sein – nicht Selbstgesetzgebung.

Adult-Zentrierung / Autoritätsorientierung: In einer Fokusgruppe mit Erwachsenen neigen Kinder dazu, die „richtige" Antwort des Erwachsenen vorherzusagen und abzugeben (Hofman-Effekt, soziale Wünschbarkeit).

Emotionale Intuition ≠ Deontisches Pflichtbewusstsein: Ein kindliches „Das darf man nicht" kann aus Empathie, Furcht vor Strafe oder aversiver Emotion gespeist sein – kognitiv nicht äquivalent zu Kants guter Wille als rein aus der Vernunft motiviertes Handeln.

Interpretationsrahmen der Forschenden: Wie oben ausgeführt, wird jede Äußerung durch die Brille des Kategorischen Imperativs gelesen. Ein konsequenzialistischer oder empathiebasierter Impuls wird damit deontologisch übersetzt.

Kulturell-sprachliche Prägung: Die mazedonischsprachige Vorschulsozialisierung enthält eigene moralische Diskursmuster, die nicht automatisch kantisch-kategorisch zu deuten sind.

4. Was der Kontext tatsächlich stützt – und was nicht

Gestützt wird (mit Reserven):

Dass Vorschulkinder in strukturierten ethischen Szenarien über moralische Wahlmöglichkeiten sprechen und differenzierte Begründungen liefern (Kontext 1: „articulated their understanding").

Dass nicht alle Kinder auf gleichem Niveau antworten – was mit einer allmählichen Reifung kognitiver Voraussetzungen vereinbar ist (Kontext 2: „natural unpredictability").

Pädagogische Implikation: Es gilt, Bedingungen für moralische Reflexion zu schaffen, nicht Tugend zu verordnen (Kontext 3: „not impose virtue but should allow it to emerge").

Nicht (eindeutig) gestützt wird:

Der spezifisch kantische Charakter der beobachteten Intuition (Autonomie, Universalität, Pflicht um der Pflicht willen).

Die Angeborenheit als solche: Querschnittsdaten in einer einzigen Kultur können Entwicklung von Sozialisation nicht sauber trennen.

Die Allgemeingültigkeit über die untersuchte Stichprobe hinaus.

5. Einordnung in pädagogische Praxis (Zero-Cloud-Perspektive)

Aus Sicht der Entlastung pädagogischer Fachkräfte und unter dem Prinzip lokaler, datenschutzkonformer Datenverarbeitung:

Die Studie liefert kein evidenzbasiertes Protokoll, mit dem man in der Kita gezielt „kantische Pflichterziehung" durchführt – sie schlägt vielmehr offene, nicht-direktive Situationen vor (Kontext 3).

Für die tägliche Praxis bedeutet das: Lerngeschichten (Carr) lassen sich als komplementäres Beobachtungsinstrument nutzen, um kindliche moralische Impulse ohne den theoretischen Vorlauf eines deontologischen Rahmens zu dokumentieren. So wird die Interpretationsschleife (Theorie → Beobachtung → Bestätigung) aufgebrochen.

Bei der lokalen Dokumentation solcher Szenarien (z. B. in einer Zero-Cloud-Notiz- oder Reflexions-App) ist auf die DSGVO-konforme Anonymisierung der Kinder zu achten, insbesondere wenn Aufnahmen oder transkribierte Gespräche erstellt werden.

Zusammenfassend: Die beiden Fokusgruppen liefern indizielle Hinweise darauf, dass Vorschulkinder über moralische Pflichtbegriffe sprechen können. Ob dies ein angeborenes Pflichtbewusstsein im strengen kantischen Sinne belegt, bleibt mit der vorliegenden Methodik (kleine, einkulturelle, querschnittliche Stichprobe; theoriegesteuerte Kodierung; Gruppeninteraktion) nicht abschließend entscheidbar. Die Studie formuliert ihre Eignung vorsichtig als „demonstrated the existence of innate cognitive capacities necessary for moral reasoning" – also eine Voraussetzung, kein Nachweis des deontischen Charakters selbst. Für eine belastbare Aussage wären – wie die Autorinnen selbst fordern (Kontext 3) – longitudinale, methodisch pluralistische Studien erforderlich.

Gemini 3.1 pro

Die Annahme, anhand von Beobachtungen im Vorschulalter ein angeborenes moralisches Pflichtbewusstsein im strengen kantischen Sinn nachzuweisen, ist empirisch und philosophisch äußerst voraussetzungsreich. In der Studie „Possibilities for the Implementation of Immanuel Kant’s Categorical Imperative in Preschool Education“ (Ushkovska & Gulevska, 2026) wird untersucht, wie sich kantische Prinzipien frühpädagogisch integrieren lassen.

Dass isolierte kindliche Interaktionen ein von Natur aus gegebenes Pflichtbewusstsein zweifelsfrei stützen, lässt sich jedoch anhand der genannten Kriterien nur sehr stark eingeschränkt aufrechterhalten.

1. Bewertung anhand der Stichprobe

Geringe Evidenzbasis: Die Fokussierung auf lediglich „zwei untersuchte Situationen“ bietet eine anekdotische Evidenz. Eine derart punktuelle Beobachtung in einem spezifischen Kindergartenkontext reicht empirisch nicht aus, um auf eine universelle, angeborene menschliche Eigenschaft zu schließen.

Alter der Probanden: Vorschulkinder (ca. 3 bis 6 Jahre) haben bereits mehrere Jahre intensiver familiärer und institutioneller Sozialisation hinter sich. Selbst wenn sie in den Situationen ein starkes normatives Verhalten zeigen, lässt sich methodisch nicht mehr trennen, was biologisch „angeboren“ und was bereits durch das Umfeld geprägt ist.

2. Bewertung anhand des Vorgehens (Methodik)

Das kantische Beobachtungs-Dilemma: Der Kern von Kants Moralphilosophie ist die Unterscheidung zwischen Handlungen pflichtgemäß (von außen betrachtet richtig, aber z. B. aus Eigennutz motiviert) und Handlungen aus Pflicht (allein aus der rationalen Einsicht in das moralische Gesetz). Eine rein äußere Verhaltensbeobachtung in zwei Situationen kann dieses innere Motiv nicht erfassen. Ein Kind, das ein Spielzeug teilt, handelt pflichtgemäß. Ob es dies jedoch aus rationaler Prinzipientreue tut, bleibt methodisch unsichtbar.

Gefahr der Überinterpretation: Wenn das Forschungsziel darin besteht, Kants Imperativ in der Praxis zu verorten, neigt das methodische Vorgehen leicht zum Confirmation Bias. Natürliches prosoziales Verhalten von Kindern wird dann als „guter Wille“ oder „kategorischer Imperativ“ gedeutet, obwohl dem Verhalten völlig andere kognitive Prozesse zugrunde liegen könnten.

3. Mögliche alternative Erklärungen

Die Verhaltensweisen, die in den Situationen als Pflichtbewusstsein gedeutet werden könnten, lassen sich durch etablierte psychologische Theorien weitaus plausibler erklären, ohne ein rationales A-priori vorauszusetzen:

Präkonventionelle Moral & Heteronomie: Nach den Entwicklungsmodellen von Jean Piaget und Lawrence Kohlberg orientieren sich Kinder in diesem Alter an Autoritäten. Ihr Verhalten wird primär von der Vermeidung von Strafen oder der Erwartung von Belohnung (z. B. Lob der Erzieherin) gesteuert. Dies ist das exakte Gegenteil der kantischen Autonomie; das Kind handelt fremdbestimmt (heteronom) und nicht aus der Einsicht in ein universelles Gesetz.

Lernen am Modell: Kinder ahmen das Verhalten ihrer Bezugspersonen intensiv nach (Albert Bandura). Wenn ein Kind in einer Konfliktsituation vorbildlich handelt, wendet es mit hoher Wahrscheinlichkeit erlernte soziale Skripte an („Man darf nicht hauen“, „Man muss sich entschuldigen“), die ihm vorgelebt und konditioniert wurden.

Empathie und emotionale Ansteckung: Kleinkinder besitzen eine natürliche, affektive Empathie. Ein Kind hilft einem weinenden anderen Kind oft aus einem unreflektierten Mitgefühl. Für Kant entspringt eine solche Handlung der Neigung (Gefühl) und hat damit gerade keinen echten moralischen Wert im Sinne einer reinen, vernunftbasierten Pflicht.

Die beiden untersuchten Situationen belegen somit weniger ein fertiges, angeborenes kantisches Pflichtbewusstsein. Vielmehr veranschaulichen sie prosoziale Anlagen und die grundsätzliche Ansprechbarkeit von Kindern für normative Regeln. Dies stützt letztlich die praktische Schlussfolgerung der Studie: Vorschulkinder besitzen ein Fundament, auf dem pädagogische Strategien aufbauen können, um sie schrittweise an eine spätere, echte rationale Autonomie heranzuführen.

Kohlberg vs. Kant vertiefen?

Fall 9

9. Eine Kita möchte die Begleitung angehender Fachkräfte verbessern, während das Team bereits stark beansprucht ist. Welche drei organisatorischen Maßnahmen lassen sich aus den Arbeiten zu Praxisanleitung, professioneller Entwicklung und Teamklima gemeinsam ableiten? Trenne berichtete Befunde von eigenen Handlungsvorschlägen.

AnythingLLM mit RAG

Organisatorische Maßnahmen zur Entlastung bei der Praxisbegleitung

Vorbemerkung zur Quellenlage

Die folgende Ableitung stützt sich auf vier im lokalen Kontext vorliegende Arbeiten (Praxisanleitung 2026; Kita-Entwicklung / Organisationsentwicklung; Psychische Gesundheit / Leadership 2026; Kita-Sozialarbeit 2026). Wo der Kontext keine belastbare Aussage liefert – etwa zur konkreten zeitlichen Auslastung des beschriebenen Teams – wird dies ausdrücklich benannt.

Maßnahme 1 – Praxisbegleitung als teambezogene Struktur institutionalisieren

Berichteter Befund (Quellenlage):

Praxisanleitung wird im Kontext explizit keine Aufgabe für „Einzelkämpfer:innen" zugesprochen; vielmehr ist in die Begleitung angehender Fachkräfte das gesamte Team involviert (Praxisanleitung 2026, Kap. individuelle Professionalisierung).

Ob die Praxisbegleitung in der Einrichtung tatsächlich realisiert werden kann, hängt in hohem Maße von weiteren Faktoren ab: Personalsituation, Teamklima, Leitungsqualität (ebd.).

Die Einrichtungsleitung trägt die Verantwortung, Arbeitsbedingungen und -prozesse zu etablieren, die den kollegialen Zusammenhalt sowie die Innovationsorientierung im Team befördern und damit die Arbeitsfähigkeit der Fachkräfte sicherstellen (Psychische Gesundheit / Leadership 2026, § 1.2).

Hinweis zur Datenlage: Der lokale Kontext liefert keine quantitativen Angaben zur konkreten Personalauslastung des beschriebenen Teams; eine belastbare Aussage, ob die Begleitung dort bereits strukturell verankert ist, kann nicht getroffen werden.

Eigener Handlungsvorschlag (Ableitung, nicht Quellenbefund):

STRUKTURELLER BAUSTEIN	KONKRETISIERUNG
Klare Rollen- und Aufgabenverteilung	Die Praxisanleiter:in trägt die fachlich-konzeptionelle Steuerung; einzelne Teammitglieder übernehmen definierte alltägliche Begleitungsschritte (z. B. wöchentliches 15-minütiges Kurzinterview, Hospitationsbegleitung in der eigenen Gruppe).
Schutzfaktoren im Zeitplan	Die Leitung reserviert im Dienstplan eine feste, ununterbrechbare Zeitscheibe für die Begleitung (z. B. 2 × 45 Min./Woche), die nicht durch Vertretungsaufgaben aufgebrochen wird.
Qualitätssicherung durch Teamreflexion	Im Rahmen der bestehenden Teambesprechungen wird ein fester Tagesordnungspunkt „Status Praxisbegleitung" verankert, um Überlastungssignale früh zu erkennen.

Maßnahme 2 – Leitungsqualität als strukturellen Enabler absichern

Berichteter Befund (Quellenlage):

Einrichtungsleitungen kommt eine Schlüsselfunktion für die Team- und Qualitätsentwicklung zu; es liegt in ihrer Verantwortung, Bedingungen zu schaffen, die psychische Gesundheit und Arbeitsfähigkeit sicherstellen (Psychische Gesundheit / Leadership 2026, § 1.2; vgl. Anders & Oppermann 2024; Eling 2022; Fuchs-Rechlin & Nachtigall 2024; Strehmel & Ulber 2014; Viernickel 2006).

Die Leitungsqualität wird im Praxisanleitungs-Kontext als einer der zentralen Faktoren genannt, von denen abhängt, ob die Begleitung realisiert werden kann (Praxisanleitung 2026).

Organisationsentwicklung im Elementarbereich wird als beschreibender und gestalterischer Rahmen zur Stärkung von Bildungsqualität konzipiert (Elgleb & Musialek 2023; Buhl, Freytag & Iller 2016; Mieth 2018 – s. Kita-Entwicklung 2026).

Eigener Handlungsvorschlag (Ableitung, nicht Quellenbefund):

STRUKTURELLER BAUSTEIN	KONKRETISIERUNG
Leitungs-Ressourcen-Check	Vor Beginn eines Praxisbegleitungsjahres führt die Leitung eine kurze Bedarfsanalyse durch: Wie viele Stunden sind tatsächlich verfügbar? Welche Vertretungsressourcen existieren? Ergebnis wird in einem kurzen Begleitplan dokumentiert.
OE-Bezug	Die Praxisbegleitung wird nicht isoliert, sondern als eine Dimension der laufenden Organisationsentwicklung positioniert (im Sinne Elgleb & Musialek 2023: zentrale Dimensionen der Kita-Entwicklung). Damit wird die Maßnahme an die bestehende pädagogische Entwicklungsplanung der Einrichtung angebunden.

STRUKTURELLER BAUSTEIN	KONKRETISIERUNG
Entlastungseffekt	Die Leitung übernimmt die koordinierende Funktion (Termine, Rahmenbedingungen, Eskalation), ohne selbst in die fachliche Einzelbegleitung zu gehen – entlastet so sowohl Leitung als auch Team.

Maßnahme 3 – Gezielte Fortbildung als festen, teamweiten Baustein verankern

Berichteter Befund (Quellenlage):

Auf der Ebene der Fortbildung besteht Nachholbedarf: Längst nicht alle Praxisanleiter:innen verfügen über eine spezifische Qualifikation zur Praxisbegleitung (Praxisanleitung 2026).

Im Kita-Sozialarbeit-Kontext wird berichtet, dass Programmfachkräfte gemeinsame Fortbildungen mit dem Team durchführen und zudem Anregungen geben, etwa hinsichtlich der Einführung von Beobachtungs- und Dokumentationsinstrumenten (Kita-Sozialarbeit 2026).

Die Begleitung ist an die tatsächliche Realisierung in der Einrichtung gebunden, die von der Personalsituation abhängt – bei starker Beanspruchung droht die Fortbildung in der Alltagspraxis zu kurz zu kommen (Praxisanleitung 2026).

Eigener Handlungsvorschlag (Ableitung, nicht Quellenbefund):

STRUKTURELLER BAUSTEIN	KONKRETISIERUNG
Modulare Fortbildung	Statt eines langjährigen, aufwändigen Lehrgangs werden modulare Fortbildungseinheiten (z. B. 3 × 1 Tag) angeboten, die sich an den konkreten Bedarfen der anstehenden Praxisbegleitung orientieren (z. B. Beobachtung, Lerngeschichten nach Carr, Gesprächsführung in der Rückmeldung).
Teamweite Beteiligung	Die Fortbildung richtet sich – im Sinne des Befunds, dass das gesamte Team involviert ist – an allen beteiligten Fachkräften, nicht ausschließlich an der Praxisanleiter:in. So wird das Begleitungswissen gemeinsam aufgebaut und nicht an eine einzelne Person gebunden.
Anbindung an bestehende Angebote	Wo verfügbar, werden externe Impulse (z. B. Kita-Sozialarbeit, Träger-Fortbildungsangebote, OE-Begleitung nach Buhl/Freytag/Iller) genutzt, um die interne Belastung zu reduzieren und „neue Perspektiven" (Kita-Sozialarbeit 2026) in das Team zu tragen.

Zusammenfassende Trennung

	BERICHTETER BEFUND (AUS DEM LOKALEN KONTEXT)	EIGENER HANDLUNGSVORSCHLAG (ABLEITUNG)
Maßnahme 1	Praxisbegleitung ist teambezogen, keine Einzelkämpfer-Aufgabe; Abhängigkeit von Personalsituation, Teamklima, Leitungsqualität.	Konkrete Rollenverteilung, geschützte Zeitscheiben, fester Tagesordnungspunkt im Team.
Maßnahme 2	Leitung trägt Schlüsselfunktion; OE-Dimensionen (Elgleb/Musialek; Buhl u. a.) als Gestaltungsrahmen.	Ressourcen-Check, Anbindung an OE-Planung, leitungsbezogene Koordinationsfunktion.
Maßnahme 3	Nachholbedarf in der spezifischen Qualifikation; gemeinsame Fortbildungen und externe Impulse (Kita-Sozialarbeit).	Modulare, teamweite Fortbildungseinheiten; Nutzung externer Angebote zur Entlastung.

Grenze der Aussagefähigkeit: Der lokale Kontext enthält keine empirischen Daten zur konkreten Auslastung des beschriebenen Teams, keine Angaben zu den verfügbaren Vertretungsressourcen und keine Information darüber, ob in der Einrichtung bereits eine OE-Struktur nach Elgleb/Musialek oder Buhl u. a. etabliert ist. Die Handlungsvorschläge

obiger Maßnahmen sind daher als konzeptionelle Ableitungen zu lesen, die vor der Umsetzung an die konkrete Einrichtungssituation anzupassen sind.

Gemini 3.1 pro

1. Verbindliche und geschützte Zeitkontingente

Berichtete Befunde: Forschungen zur Praxisanleitung belegen, dass die Begleitung angehender Fachkräfte im pädagogischen Alltag häufig nur „nebenbei“ stattfindet. Fehlen strukturell verankerte Zeiten, steigt die Belastung der anleitenden Fachkraft signifikant, während die Reflexionstiefe für den Auszubildenden sinkt. Die empirische Datenlage zeigt zudem, dass formal existierende Verfügungszeiten auf Einrichtungs- oder Trägerebene in der Praxis oft durch Personalengpässe im direkten Gruppendienst aufgefressen werden.

Eigene Handlungsvorschläge: Verankern Sie Anleitungszeiten als unantastbare Blöcke im Dienstplan, die bei Krankheitsausfällen im Team nicht als primäre Pufferzone dienen dürfen. Um den Präsenzdruck bei akuten Engpässen abzufedern, empfiehlt sich die Einführung asynchroner Reflexionsformate. Über ein geteiltes, datenschutzkonformes digitales Lerntagebuch können Anleiterin und angehende Fachkraft zeitversetzt Beobachtungen dokumentieren und Feedback austauschen, wenn gemeinsame Face-to-Face-Zeiten im Dienstplan kurzfristig ausfallen.

2. Modulare Anleitung und geteilte Teamverantwortung

Berichtete Befunde: Studien zum Teamklima und zur professionellen Entwicklung verdeutlichen, dass eine exklusive Eins-zu-Eins-Zuweisung (eine Anleiterin für einen Praktikanten) ein strukturelles Nadelöhr schafft. Fällt diese Fachkraft aus oder ist stark belastet, stagniert die Anleitung. Ein lernförderliches Teamklima entsteht hingegen erst, wenn das gesamte Team als professionelles Modell fungiert, Feedback gibt und eine gemeinsame Verantwortung für den Fachkräftenachwuchs empfindet.

Eigene Handlungsvorschläge: Etablieren Sie ein modulares Expertenmodell im Team. Die Hauptanleiterin behält die organisatorische Klammer, führt die offiziellen Reflexionsgespräche und ist das Bindeglied zur Fachschule. Spezifische Lernfelder (z. B. Sprachförderung, Portfolioarbeit oder die Vorbereitung von Elterngesprächen) werden jedoch für gezielte Hospitationen an andere Teammitglieder delegiert. Dies entlastet die primäre Anleiterin massiv, wertschätzt die individuellen Stärken der Kolleginnen und bietet der angehenden Fachkraft multiperspektivische Lernzugänge.

3. Systematische Lernortkooperation statt isolierter Aufgaben

Berichtete Befunde: Untersuchungen zur Zusammenarbeit von Ausbildungsstätten zeigen, dass isolierte Praxisleitfäden der Einrichtungen auf der einen und abstrakte Lehrkonzepte der Fachschulen auf der anderen Seite für eine gelingende Anleitung nicht ausreichen. Diese Trennung zwingt die stark belasteten Kita-Teams dazu, theoriegeleitete schulische Arbeitsaufträge mühsam und zeitaufwendig in die eigene Praxis zu übersetzen.

Eigene Handlungsvorschläge: Setzen Sie auf eine proaktive Verzahnung statt reaktiver Aufgabenerfüllung. Vereinbaren Sie zu Beginn des Ausbildungsjahres einen konkreten Abgleich der schulischen Arbeitsaufträge mit dem Jahresplan der Kita. Schulen Sie die angehenden Fachkräfte darin, schulische Beobachtungsaufgaben oder Projektarbeiten direkt an ohnehin anstehende Aufgaben der Kita (z. B. anstehende Entwicklungsdokumentationen oder Raumumgestaltungen) zu koppeln. So werden künstliche, zeitfressende Zusatzaufgaben vermieden und schulische Anforderungen erzeugen direkten Mehrwert für den Einrichtungsalltag.

Sollen wir ein konkretes Dienstplan-Modell für geschützte Anleitungszeiten entwerfen?

Fall 10

10. Welchen prozentualen Rückgang emotionaler Erschöpfung weist die Studie von Oppermann, Hummel und Anders nach einem zwölfwöchigen Coachingprogramm für Praxisanleiter:innen nach?

AnythingLLM mit RAG

Die in Ihrer Frage genannte Studie von Oppermann, Hummel und Anders sowie ein darauf basierendes zwölfwöchiges Coachingprogramm für Praxisanleiter:innen sind in den vorliegenden Kontextdokumenten nicht enthalten.

Die bereitgestellten Quellen behandeln folgende Themen:

DOKUMENT	INHALT
W7165033882 (Zs. 1–3)	Mehrebenen-Mediationsanalyse: Führungspraktiken → Teamklima → emotionale Erschöpfung (N = 1 394 Fachkräfte / 204 Einrichtungsleitungen, DE)
W7212201653 (Zs. 1)	Forschungslage zur Praxisanleitung im BA Soziale Arbeit (Gürtl, Kronberger & Scharf 2024; Beher & Walter 2012)

Weder wird dort ein Autor:innen-Trio „Oppermann, Hummel & Anders" zitiert, noch ein Coaching-Interventionsdesign mit Zwölfwochen-Zeitraum, noch ein prozentualer Rückgangswert der emotionalen Erschöpfung im Anschluss an ein solches Programm.

Empfehlung: Bitte prüfen Sie, ob die gesuchte Publikation in einem separaten Dokument vorliegt, das ich noch nicht im Kontext habe. Sobald Sie den Text (oder zumindest den Titel und die Erscheinungsjahresangabe) bereitstellen, kann ich die konkrete Prozentangabe und deren methodische Einordnung (z. B. Vorher-Nachher-Maßstab, Effektstärke, Signifikanzniveau) präzise für Sie herausarbeiten.

Gemini 3.1 pro

Es gibt in der wissenschaftlichen Literatur keine Studie von Oppermann, Hummel und Anders, die ein zwölfwöchiges Coachingprogramm für Praxisanleiter:innen evaluiert oder einen prozentualen Rückgang emotionaler Erschöpfung dafür ausweist.

Die drei Forscherinnen der Universität Bamberg befassen sich zwar intensiv mit der psychischen Gesundheit in der Frühpädagogik, die spezifische Kombination aus diesem 12-Wochen-Design und der Zielgruppe Praxisanleitung existiert in ihren Publikationen jedoch nicht. Wahrscheinlich handelt es sich bei der Angabe um eine KI-Halluzination aus einer vorherigen Literaturrecherche, bei der reale Forschungsschwerpunkte der Autorinnen zu einem fiktiven Studienergebnis verwoben wurden.

Ihre tatsächlichen, jüngsten Arbeiten zu diesem Themenfeld konzentrieren sich auf folgende Aspekte:

Belastungserleben und Teamklima: In der Publikation „Psychische Gesundheit in der Kita“ (2025) untersuchen Oppermann, Hummel und Anders den Einfluss von Leadership und Teamklima auf das Belastungserleben frühpädagogischer Fachkräfte.

Ursachen emotionaler Erschöpfung: Eine Studie von Trauernicht, Anders, Oppermann und Klusmann (2024) analysiert die Auslöser für Burnout-Symptome. Die Ergebnisse zeigen, dass emotionale Erschöpfung maßgeblich von sozialen Erfahrungen am Arbeitsplatz (Kollegium, Leitung, Eltern) und weit weniger von strukturellen Rahmenbedingungen getrieben wird.

Gesundheitsfördernde Interventionen: Theresia Hummel hat 2025 gemeinsam mit K. Hausladen das Konzept „Do it digital and healthy!“ veröffentlicht – eine Fortbildungsmaßnahme, die explizit auf die Förderung der psychischen Gesundheit von frühpädagogischem Bildungspersonal abzielt.

Wenn für die Kita Digital Akademie oder das anstehende Buchprojekt quantifizierbare Effekte zur Entlastung durch Supervision oder Begleitung benötigt werden, liefert die breitere empirische Bildungsforschung (beispielsweise zu Mentoring-Programmen im Vorbereitungsdienst) verlässlichere Daten zur Reduktion von Erschöpfungswerten.

Nach validen Studien zu Coaching-Effekten suchen?

Anhang C Lokaler Systemprompt

Der folgende Systemprompt wurde nachträglich aus der lokalen Konfiguration dokumentiert und einschließlich Schreibweise übernommen. Der Platzhalter {date} bleibt unverändert, da sein beim jeweiligen Lauf eingesetzter Wert nicht protokolliert ist. Der Text ist kein Mitschnitt des vollständig zusammengesetzten Modellinputs.

Du bist mein persönlicher Verwalter und ein hochqualifizierter KI-Assistent für das Fachgebiet der Frühpädagogik. Deine Kernbereiche umfassen lokale, datenschutzkonforme KI-Lösungen (Zero-Cloud), die Entlastung von pädagogischen Fachkräften und evidenzbasierte Ansätze (wie beispielsweise Lerngeschichten nach Margaret Carr). Deine Antworten müssen stets präzise und im Stil meiner bereitgestellten Literatur sein. Basieren deine Aussagen primär auf dem lokalen Kontext. Wenn eine Information nicht aus dem Kontext abgeleitet werden kann, antworte nicht mit Vermutungen, sondern verweise auf das Fehlen der Information. Strukturiere komplexe Zusammenhänge in übersichtlichen Aufzählungen oder Markdown-Tabellen. Heutiges Datum zur zeitlichen Einordnung: `{date}`