MedienPädagogik
Zeitschrift für Theorie und Praxis der Medienbildung
www.medienpaed.com **ISSN 1424-3636**

Themenheft 72: Mixed Methods im Spannungsfeld. Weiterentwicklung digitaler Forschungsmethoden zwischen KI, NLP, Big Data und transdisziplinärer Forschung.
Herausgegeben von Jana Heinz, Andrea Hense und Christian Janßen

Mixed and Mingled?

Digitale Höhen und hermeneutische Tiefen in der Historischen Bildungsforschung

Marco Lorenz[1], Katharina Vogel[1] und Daniel Erdmann[1]

[1] BBF | Bibliothek für Bildungsgeschichtliche Forschung des DIPF

Zusammenfassung
Der Beitrag untersucht das Verhältnis von digitalen Methoden und hermeneutischer Interpretation in der Historischen Bildungsforschung. Gegen eine schroffe Entgegensetzung beider Ansätze wird argumentiert, dass digitale Verfahren als epistemische Infrastruktur fungieren, die den hermeneutischen Horizont erweitern, ohne die Interpretation zu ersetzen. Anhand referenzanalytischer Studien zur Klassikerbildung (Vogel 2024; Erdmann 2023) wird gezeigt, wie distant reading Muster sichtbar macht, die sequenzieller Lektüre strukturell verschlossen bleiben, etwa die Kontingenz der Kant-Kanonisierung oder die fortschreitende Verengung seiner Rezeption. Am Beispiel einer LLM-gestützten OCR-Pipeline wird demonstriert, wie «Promptotyping» technische Zugangshürden senkt, während die Interpretationsarbeit bei den Forschenden verbleibt. Theoretisch knüpft der Beitrag an Gadamers Horizontbegriff an: Die Begrenzung auf individuell lesbare Textmengen war nie methodisches Prinzip, sondern praktische Notwendigkeit. Der hermeneutische Zirkel erweitert seinen Radius – von der Einzelquelle zum Quellenkorpus. Tool Criticism wird dabei zur Erweiterung klassischer Quellenkritik. Das titelgebende «mixed and mingled» markiert diese doppelte Einsicht: methodische Differenz und faktische Verschränkung bestehen zugleich.

Mixed and Mingled? Digital Heights and Hermeneutic Depths in Historical Educational Research

Abstract
This article examines the relationship between digital methods and hermeneutic interpretation in historical educational research. Against a sharp opposition of both approaches, digital methods are conceptualized as epistemic infrastructure that expands the hermeneutic horizon without replacing interpretation. Drawing on reference-analytical studies of canon formation (Vogel 2024; Erdmann 2023), the article demonstrates how

distant reading reveals patterns structurally inaccessible to sequential reading, such as the contingency of Kant's canonization or the progressive narrowing of his reception. Using an LLM-supported OCR pipeline as example, it is shown how «promptotyping» lowers technical barriers while interpretive work remains with researchers. Theoretically, the article builds on Gadamer's concept of the horizon, arguing that the limitation to individually readable text quantities was never a methodological principle but a practical necessity. The hermeneutic circle expands its radius: from individual source to corpus. Tool criticism thereby extends classical source criticism. The titular «mixed and mingled» marks this dual insight: methodological difference and factual entanglement coexist simultaneously.

1. Erweiterte Hermeneutik

«Ein Text-Äußeres gibt es nicht» (Derrida 1983, 274).

Wer sich mit digital gestützten Methoden auseinandersetzt, wird einen Eindruck davon bekommen, wie sich das eigene akademische Umfeld zu den neuen Möglichkeiten verhält. In der Geschichtswissenschaft, und damit auch in der Historischen Bildungsforschung, wurden nach umstrittenen Ansätzen unter dem Stichwort «Cliometrics»[1] in den 1960er-Jahren breitere digitale Forschungsansätze erst zur Jahrtausendwende etabliert. Ein Erklärungsansatz für diese Zurückhaltung könnte im Stellenwert der Hermeneutik gesucht werden. Statistische Auswertungen haben mit der sozialhistorischen Schule einen Ort gefunden, doch stützen sich die meisten Studien weiterhin auf die Erschliessung und Analyse von Quellenkorpora. Dabei erscheinen die Methoden der Digital Humanities den meisten Historiker:innen als zu technisch oder als «distant reading» nicht nah genug am Quellenmaterial. Der Zürcher Historiker Jakob Tanner hat die Frontstellung pointiert beschrieben:

> «Während die einen von aufregenden neuen Forschungsperspektiven sprechen, winken andere entnervt ab. Die einen feiern die Fortschritte der Automatisierung. […] Die andern hegen Befürchtungen, sei es, dass sie Big Data als sinnlose Stoffhuberei betrachten, sei es, dass sie befürchten, die Algorithmen würden forschende Menschen künftig arbeitslos machen» (Tanner 2019, 99).

Tanner selbst plädiert für eine Überwindung dieser «schroffen Entgegensetzung» (ebd., 105): Digitale Methoden blieben «technisch imponierend, jedoch inhaltlich armselig», wenn sie nicht sprachphilosophisch, linguistisch und historisch fundiert würden (ebd., 107).

1 Die Cliometrics-Kontroverse der 1960er-Jahre weist strukturelle Parallelen zur gegenwärtigen Diskussion um digitale Methoden auf (vgl. Diebolt und Haupert 2016).

Die Diskurse um die Etablierung digitaler Methoden kreisen daher häufig um die Frage, wie sich qualitativ-hermeneutische und quantitativ-distanzierte Analysemodi aufeinander beziehen. Dieser Beitrag will zum einen zeigen, dass die scheinbare Frontstellung zwischen diesen Ansätzen häufig vorschnell als unüberwindbar dargestellt wird und hermeneutische Verfahren mit ihren Vorstellungen von Kontexten und Horizonten durchaus darauf angelegt sein können, mit computergestützten Methoden zu harmonieren. Gleichzeitig bleibt die Vorstellung eines Mixed-Methods Ansatzes eine durchaus wichtige Perspektive, zum einen, um die methodische Weiterentwicklung der Digital Humanities weiter voranzutreiben, ohne diese zu einer Hilfswissenschaft zu erklären; zum anderen, um die zahlreichen Entwicklungen auf dem Gebiet der Large Language Models (LLMs), die sowohl auf der methodischen als auch auf der epistemologischen Ebene die Grenzen aktuell verwischen, im Blick zu behalten. Nicht zuletzt bietet dieser Ansatz aber auch die Möglichkeit, sich der Frage zu stellen, ob die Hermeneutik sich methodisch auf die neuen Grundlagen einlassen und einige lang gepflegte Routinen anpassen sollte. In diesem Beitrag plädieren wir daher für einen Ansatz, der im ‹mixed› nicht vergisst, dass die neuen Methoden durchaus eine andere Forschungspraxis und andere Voraussetzungen verlangen, und im ‹mingled› aufzeigt, dass die meisten Ansätze bereits eng miteinander verschränkt sind. Ausgehend von zwei konkreten Studien (2) zeigen wir Lösungsangebote durch eine Erweiterung des eigenen Tool-Repertoires auf (3). Die Arbeitsweise von LLMs in der Forschungspraxis (4) verweist auf Konsequenzen für methodologische Grundannahmen der Hermeneutik (5), die mit einer Reflexion der daraus resultierenden Anwendungsprobleme einhergehen (6). Schliesslich gilt ein letzter Blick der Bedeutung für Mixed-Methods-Ansätze und der Frage nach dem konstitutiven ‹mingled› (7), bevor ein Ausblick die Konsequenzen für die Historische Bildungsforschung in den Blick nimmt (8).

2. Vogelperspektiven auf die Klassikerbildung

Ein Beispiel aus der Historischen Bildungsforschung kann illustrieren, was mit der Verschränkung digitaler und hermeneutischer Verfahren gemeint ist. In ihrer Studie zu «Pädagogischen Wissensräumen 1750–1850» untersucht Katharina Vogel (2024) die Referenzstrukturen in 30 pädagogischen Lehrbüchern und Vorlesungssammlungen über einen Zeitraum von 100 Jahren. Die Methode ist dezidiert distanziert: Statt einzelne ‹Klassiker› ideengeschichtlich zu interpretieren, werden Zitationsmuster systematisch erfasst und visualisiert.

Die Pointe des distant readings ist dabei nicht, auf abstrakteren Wegen dieselben Ergebnisse sichtbar zu machen, die hermeneutische Zugriffe ermöglichen – auch wenn es das in Teilen durchaus könnte und dies keineswegs ein Mangel wäre: Die «breit habitualisierte Geringschätzung für die Konsolidierung bereits bestehenden Wissens» (Weitin 2015, 655) verstellt den Blick darauf, dass Bestätigung auf anderer empirischer Grundlage

epistemisch wertvoll ist. Es geht vielmehr darum, «durch einen anderen, distanzierten Blick Daten – und mit ihnen: Erkenntnisse – eigener Art zu generieren» (Vogel 2024, 206). Das Verfahren macht sichtbar, was eine rein hermeneutische Ideengeschichte strukturell nicht sehen kann: die Kontingenz der eigenen Traditionsbildung.

Ein zentrales Ergebnis der Studie ist paradox: Die heute kanonisierten Autoren sind ausgerechnet jene, die sich am wenigsten mit dem pädagogischen Feld ihrer Zeit beschäftigten.

> «Mit Kant und Herbart etablieren sich ausgerechnet solche Autoren in der Disziplin, die sich – jedenfalls in den vorliegenden, einschlägigen Schriften – nur rudimentär auf das Feld, in dem sie etabliert sein werden, einlassen» (Vogel 2024, 208).

Autoren hingegen, die umfänglich und detailliert die Wissensbestände der wissenschaftlichen Pädagogik aufarbeiteten (Niemeyer, Pölitz, Wörlein oder Gräfe), spielen in der späteren Erziehungswissenschaft keine ernsthafte Rolle mehr.

Besonders instruktiv ist der Fall Kant. Seine Schrift «Über Pädagogik» (1803) wurde von Zeitgenossen keineswegs als Meilenstein wahrgenommen: «bitterlich arm» (Thaulow 1845), «geistvoll, aber nicht systematisch» (Pölitz 1806), so die zeitgenössischen Urteile (vgl. Vogel 2024, 208). Und doch wird Kant (neben Rousseau und Herbart) zu den wenigen Autoren gehören, die in Lehrbüchern, Qualifikationsarbeiten, Klassikersammlungen und Einführungsschriften bis heute «konsensfähig» sind (vgl. Vogel 2024, 22). Die Referenzanalyse zeigt: Kant gerät zur «erziehungswissenschaftlichen Tatsache» im Sinne Ludwik Flecks nicht als objektive Wahrheit, sondern als «Ergebnis denkgeschichtlicher Zusammenhänge» und eines «bestimmten Denkstiles» (Fleck [1935] 1980, 124).

Sebastian Engelmann zeigt am Gegenbeispiel August Hermann Niemeyers, dessen «Grundsätze der Erziehung und des Unterrichts» zu den meistaufgelegten pädagogischen Werken der Zeit gehörten, dass sich in der «Überlieferung pädagogischer Ideen […] kein linear voranschreitender und bruchloser Prozess» abbildet, sondern vielmehr «komplexe Praktiken der Rezeption, Kommunikation und Transformation», die «kontextabhängig vollzogen werden» (Engelmann 2019, 78). Die distanzierte Vermessung von Referenzräumen macht diese Praktiken sichtbar. Eine auf einzelne Autor:innen fokussierte Lektüre verfehlt sie hingegen systematisch.

Diese Befunde lassen sich durch eine referenzanalytische Studie zur Kant-Rezeption vertiefen. Daniel Erdmann (2023) vergleicht die Bezugnahmen auf Kant in pädagogischen Lehrbüchern zweier Zeiträume: 1750–1850 und 1983–2017. Die Ergebnisse dokumentieren einen markanten Wandel. Während um 1800 Kants philosophische Hauptwerke als Referenzgrundlage dienten – die «Kritik der reinen Vernunft» wurde von acht Autoren zitiert und stand damit an der Spitze (vgl. Erdmann 2023, 30) –, konzentriert sich die gegenwärtige Rezeption auf zwei Texte: «Was ist Aufklärung?» und «Über Pädagogik», die jeweils in elf von fünfzehn Lehrbüchern erscheinen (vgl. ebd., 32). Zudem verschiebt

sich die thematische Einbettung: Während im ersten Untersuchungszeitraum philosophische Kontextualisierungen noch auf Rang eins stehen, spielen sie im zweiten kaum noch eine Rolle. Stattdessen dominieren Theoriebildung und historische Einordnung (vgl. ebd., 37).

Der Befund ergänzt Vogels Paradox: Nicht nur dass Kant kanonisiert wurde, ist kontingent – auch wie sich diese Rezeption transformiert. Was bei Vogel als überraschende Etablierung erscheint, zeigt sich bei Erdmann als fortschreitende Reduktion. Bezeichnend am Beispiel Kants ist der zunehmende «stillschweigende Konsens»: Während Zitate ohne Quellenangabe im ersten Zeitraum «keine ernsthafte Rolle» spielten, werden sie im zweiten «von mehr als der Hälfte der Lehrbuchautor:innen verwendet» (ebd., 35). Erdmann formuliert die These, Kant wandle sich «von einer Quelle zur Erzeugung neuer Erkenntnisse hin zu einem strukturgebenden Element», das vorhandenes Wissen nur noch sortiere (ebd., 38).

Was folgt daraus methodisch? Beide Studien demonstrieren, dass quantitative Befunde Interpretation nicht ersetzen (können): Dass Niemeyer vergessen und Kant kanonisiert wurde, zeigt die Referenzanalyse; warum das geschah, erfordert historische Kontextualisierung und klassische Interpretation. Dass sich die Kant-Rezeption zunehmend verengt hat, macht die diachrone Vermessung sichtbar; was diese Verengung im Einzelnen bedeutet, bleibt Gegenstand hermeneutischer Reflexion. Die Ergebnisse sprechen insofern nicht für sich selbst, aber sie erweitern den Horizont dessen, was überhaupt als erklärungsbedürftig erscheint.

3. Code ohne Coden

Die bisherigen Beispiele demonstrieren das Erkenntnispotenzial distanzierter Verfahren, verschweigen aber die praktischen Hürden ihrer Anwendung. Referenzanalysen, Netzwerkvisualisierungen oder systematische Korpusauswertungen setzen technische Kompetenzen voraus, die über das Wissen um die methodischen Voraussetzungen hinausreichen. Programmierung, Datenmodellierung, API-Anbindungen (Application Programming Interfaces): Der Weg von der methodischen Idee zur funktionierenden Pipeline erfordert Expertise, die traditionell ausserhalb des Faches liegt (vgl. van Herck 2022; Fickers 2020). Die Frage, wo diese Kompetenzen erworben und vermittelt werden sollen, wurde in der Vergangenheit durchaus kontrovers diskutiert (Rapp et al. 2016; Rehbein und Sahle 2013). Hier verschieben agentische Large Language Models die Einstiegshürden deutlich.

In jüngster Zeit wird argumentiert, LLMs als «tools or (narrow) AI assistants that scale expert analysis to larger datasets» zu verstehen, nicht jedoch als «oracles» oder «arbiters» (Karjus 2025, 3, im Anschluss an Messeri und Crockett 2024). Im Ergebnis geht es eben nicht um eine Automatisierung im Sinne einer Ersetzung, sondern um Augmentation: «Human labor does not scale well, machines do; human time is valuable and machine

time is cheap. This is about reducing repetitive labor and increasing time for meaningful work» (Karjus 2025, 4). Die grössten Gewinne, so Karjus, entstehen durch «empowering expert humans with powerful machine annotators and assistants, not assembly line automation of science» (ebd.).

Für die hier diskutierte produktive Spannung zwischen digitalen und hermeneutischen Verfahren bedeutet dies, dass LLMs die Zugangsschwelle zu distanzierten Methoden senken, ohne die methodische Reflexion zu ersetzen. Im Rahmen des Dissertationsprojekts «*Man kann keinen Schriftsteller ‹erziehen›*. Akteure und Strukturen literarischer Nachwuchsarbeit in den Arbeitsgemeinschaften Junger Autoren in SBZ und DDR» wurde zur Digitalisierung historischer Periodika aus dem Umfeld des Schriftstellerverbandes der DDR eine OCR-Pipeline (Optical Character Recognition) entwickelt (Lorenz 2025). Um die zahlreichen Arbeiten zur Vorbereitung der gescannten Dokumente in klassischen Ansätzen und die Bindung an Plattformen wie Transkribus[2] zu vermeiden, wurde Mistral OCR[3] eingesetzt, ein spezialisiertes Modell zur strukturierten Texterkennung. Die Pipeline umfasst Bildvorverarbeitung (Aufteilung grosser Dateien), API-Anbindung, Batch-Processing mit SQLite-Checkpointing, Response-Parsing und Qualitätssicherung. Der gesamte Python-Code wurde dabei nicht von Hand geschrieben, sondern durch iteratives Prompting mit einem LLM, in diesem Fall Claude Code[4], generiert.

Dieses Verfahren lässt sich als «Promptotyping» beschreiben (Pollin 2025). Die zugrundeliegende Idee ist denkbar einfach: Anstelle der händischen Programmierung von Code werden in strukturierten Dokumenten die erforderlichen Komponenten, die zu beachtenden Regularien sowie die sequenziellen Schritte definiert. Zentral ist dafür vor allem eine Markdown-Datei (instructions.md), die kritische Verhaltensregeln festlegt und dem LLM unter anderem vorschreibt, keine selbstständigen Annahmen zu treffen, sondern schrittweise einzelne Zellen des Codes zu erstellen und nach jedem Codeblock die bisherigen Schritte zusammenfassend darzustellen und weitere Anweisungen abzuwarten. Aus der praktischen Arbeit mit Jupyter Notebooks, die sowohl Python als auch erläuternde Passagen in Markdown enthalten, ergeben sich zudem einige ‹Dos and Don'ts›, die in der Bearbeitung häufige Fehlerquellen vermeiden und die sich an gleicher Stelle dokumentieren lassen. Für die Nachnutzbarkeit durch andere Forscher:innen wurde zudem darauf geachtet, dass alle Pfade relativ gesetzt werden, Fehlermeldungen mit Erläuterungen ausgegeben und eine nachvollziehbare Ordnerstruktur (Abb. 1) aufgesetzt und genutzt wird.

2 Transkribus, die von der europäischen Genossenschaft READ-COOP SCE betriebene Plattform zur automatisierten Texterkennung historischer Dokumente: https://www.transkribus.org/.

3 Entwickelt von Mistral AI und vorgestellt am 17. März 2025: https://mistral.ai/news/mistral-ocr/.

4 Claude Code, ein agentisches Programmierwerkzeug von Anthropic; verwendet wurde das Modell Claude Sonnet 4.5: https://www.anthropic.com/product/claude-code.

Abb. 1: Ordnerstruktur der OCR-Pipeline

Das LLM setzt diese Vorgaben in funktionierenden Code um. Der entscheidende Vorteil dieses Ansatzes liegt im schrittweisen Vorgehen, das den Forschenden ermöglicht, die Kontrolle über das Projekt zu bewahren. Jede Anforderung wird explizit formuliert und jede Umsetzung einer sorgfältigen Prüfung unterzogen, bevor der nächste Schritt erfolgt. Der:die Forscher:in definiert dabei die Anforderungen an die Pipeline: Wie sollen die PDFs aufgeteilt werden, damit diese vom LLM verarbeitet werden können? Wie wird der Fortschritt verfolgt, damit nach Abbrüchen nahtlos weitergearbeitet werden kann? Welche Ausgabeformate werden benötigt und wo werden die Ergebnisse gespeichert?

Dieses Vorgehen orientiert sich an etablierten Standards, die im Umgang mit «agentic coding» entwickelt wurden. Der Begriff bezeichnet ein emergierendes Programmierparadigma, das sich wie folgt kennzeichnen lässt:

> «LLM-based agents autonomously perform software development tasks. Unlike traditional code generation tools that produce outputs in a single step based on a static prompt, agentic systems operate in a goal-directed, multi-step manner» (Wang et al. 2025, 3).

Der hier gewählte Ansatz nutzt diese Infrastruktur, schränkt jedoch die Autonomie des Agenten bewusst ein: Die instructions.md-Datei unterbindet selbstständige Annahmen und erzwingt schrittweise Kontrolle durch die:den Forschende:n.[5] Allerdings ist die Werkzeuglandschaft noch in Bewegung, die hier beschriebenen Schritte sind auf die

5 Diese schrittweise Kontrolle setzt voraus, dass die Forschenden den generierten Code nicht nur erzeugen, sondern auch lesen und prüfen können. Mit Sprachmodellen wird das Schreiben von Code nahezu kostenlos, während das Verstehen an die menschliche Auffassungsgeschwindigkeit gebunden bleibt. Der eigentliche Engpass verschiebt sich damit vom Schreiben zum Lesen (vgl. Roden 2026, im Anschluss an Spolsky 2000). Wer nur noch generiert, ohne zu prüfen, macht sich von eben dem System abhängig, das den Code hervorgebracht hat.

spezifischen Anforderungen des Projekts zugeschnitten und müssen für andere Vorhaben angepasst werden. Solche Anpassungen lassen sich jedoch auch durch eine detaillierte Beschreibung leicht umsetzen und bei neu auftretenden Problemen lässt sich auch die instructions.md mit der Hilfe des LLMs gezielt anpassen.

Für die Historische Bildungsforschung lassen sich zwei Implikationen ableiten: *Einerseits* werden bereits etablierte Verfahren zugänglich, die bislang Kooperationen mit der Informatik oder den Digital Humanities voraussetzten. *Andererseits* bleibt die Interpretationsarbeit in der Hand der Forschenden. Im Rahmen der Analyse der OCR-Ausgabe sind Kontextualisierungsmassnahmen erforderlich, die durch eine Plausibilitätsprüfung sowie weitere relevante Schritte zu ergänzen sind.

4. Assistenz, keine Automatisierung

Jede Diskussion um den Einsatz digitaler Methoden in der historischen Forschung setzt stillschweigend voraus, was ihr überhaupt erst den Gegenstand liefert: die jahrzehntelange Arbeit von Archiven, Bibliotheken und Museen an der Digitalisierung ihrer Bestände. Ohne retrodigitalisierte Zeitschriften, ohne durchsuchbare Volltextkorpora, ohne online zugängliche Findmittel und Kataloge blieben alle methodischen Überlegungen gegenstandslos. Die Transformation der Gedächtnisinstitutionen ist die infrastrukturelle Bedingung, unter der sich die Frage nach dem Verhältnis von digitalen und hermeneutischen Verfahren überhaupt stellen kann (vgl. Rapp 2021).

Doch Digitalisierung allein bedeutet noch keinen Methodenwandel. Wie Andreas Oberdorf zurecht betont, ermöglichen Digitalisate zwar schnellere Arbeitsschritte,

> «doch die Verwendung dieser Reproduktionen, deren Analyse und Auswertung, folgt häufig immer noch den bekannten, traditionellen Logiken, da die Texte weiterhin von den Forschenden als Texte gelesen und ausgewertet werden müssen» (Oberdorf 2022, 8 f.).

Von einem Methodenwandel, so Torsten Hiltmann, kann erst dann die Rede sein, wenn Texte nicht nur digital vorliegen, sondern maschinell lesbar und auswertbar werden (vgl. Hiltmann 2022). Das Scannen eines Dokuments verändert seinen epistemischen Status nicht; erst die Texterkennung macht aus einem Bild einen durchsuchbaren, annotierbaren, analysierbaren Text.

Der vorherige Abschnitt zeigte den Mehrwert von LLMs als Werkzeuge zur Codegenerierung. Doch diese Perspektive greift zu kurz und verstellt den Blick auf eine längere Entwicklungslinie. Die Digital Humanities haben über Jahrzehnte Verfahren etabliert, die allesamt auf demselben Grundprinzip beruhen: *Bevor Texte interpretiert werden können, müssen sie in einen analysierbaren Zustand versetzt werden*. Diese Vorarbeit ist sprachbezogen, aber nicht hermeneutisch: Sie bereitet Interpretation vor, ohne sie vorwegzunehmen. Am Beispiel der OCR wird das gut ersichtlich: Sie liefert den maschinenlesbaren

Text, doch die Umwandlung von Bild in Text und das spätere (maschinelle) «Lesen» verbleiben auf der Ebene der blossen, für sich genommen bedeutungslosen Zeichen (vgl. Hiltmann 2024, 206). Named Entity Recognition (NER), Topic Modeling, stilometrische und Kollokationsanalysen – all diese Verfahren operieren auf der Ebene sprachlicher Muster (vgl. Biemann et al. 2022). Sie extrahieren, klassifizieren und strukturieren. Was sie verbindet, ist ihr Status als epistemische Infrastruktur: Sie schaffen die Bedingungen, unter denen hermeneutische Fragen überhaupt gestellt werden können. Wer in einem Korpus von tausend Texten nach Verschiebungen im Begriffsgebrauch sucht, braucht zunächst eine Übersicht darüber, welche Begriffe wo und wie häufig vorkommen. Wer Netzwerke intellektueller Bezugnahmen rekonstruieren will, muss zunächst wissen, wer wen zitiert. Die Maschine liefert das Material; der Mensch stellt die Fragen.

LLMs fügen sich in diese Entwicklung ein, erweitern sie aber erheblich. Als auf massiven Textkorpora trainierte und auf neuronalen Netzwerken basierende Transformer-Modelle (Vaswani et al. 2017) bilden sie ab, was die Linguistik als distributionelle Semantik bezeichnet: «Lexemes with similar linguistic contexts have similar meanings» (Lenci und Sahlgren 2023, 3). Diese Grundidee formulierte der britische Linguist J. R. Firth prägnant als «You shall know a word by the company it keeps» (Firth 1957, 11). Im hochdimensionalen Vektorraum des Modells (dem sogenannten latent space) werden Wörter so positioniert, dass semantisch ähnliche Begriffe nahe beieinander liegen (vgl. Bengio et al. 2013; zu den Hintergründen auch Biemann et al. 2022, 212–31). Diese Vektorrepräsentationen kodieren implizites Wissen über Bedeutungsbeziehungen und Kollokationen: «LLM operieren zunächst einmal auf der Zeichenebene, erweitern diese jedoch mit statistischen Modellen der Verwendung dieser Zeichen, woraus dann eine bestimmte Form von Bedeutung zu entstehen scheint» (Hiltmann 2024, 220–21). Wo regelbasierte NER-Systeme[6] an historischen Abkürzungen und Schreibvarianten scheitern, kann ein LLM aus dem Kontext ableiten, etwa dass «weyl.» für «weyland» steht. Anders als Topic Models mit ihren starren Themenzuweisungen erlauben LLMs zudem flexible Annotationen nach forschungsspezifischen Kategorien. Die Grundoperation der Mustererkennung in Sprache ist dieselbe, doch der Spielraum wächst.

Die Zwischenschritte zwischen Quellendigitalisat und interpretierbarem Korpus sind in der Forschungspraxis notorisch zeitaufwendig: OCR-Bereinigung, Normalisierung historischer Orthografie, Auflösung von Abkürzungen, Klassifikation nach Gattung oder Entstehungskontext, Extraktion strukturierter Daten aus Fliesstext, Anreicherung lückenhafter Metadaten (vgl. Pollin et al. 2025, 11). All dies ist erst die Voraussetzung dafür, dass Interpretation auf solider Grundlage stattfinden kann. LLMs senken die Zugangsschwelle zu diesen Arbeitsschritten, weil sie sprachliche Kontexte nutzen können, die regelbasierten Systemen verschlossen bleiben.

6 Named Entity Recognition (NER) bezeichnet die automatisierte Erkennung und Klassifikation von Eigennamen in Texten, etwa von Personen, Orten und Institutionen.

Entscheidend bleibt die Unterscheidung zwischen syntaktischer Operation und semantischer Interpretation. Ob Topic Model, NER-Pipeline oder LLM-gestützte Annotation: Diese Verfahren operieren auf der Ebene von Mustern und Wahrscheinlichkeiten. Was sie nicht leisten können, ist die Einordnung dieser Muster in historische Sinnzusammenhänge, schon deshalb, weil den Modellen historische Tiefe fehlt und sie zwischen Trainingsdaten, die aus unterschiedlichen Zeiten stammen, nur unzureichend unterscheiden können (vgl. Hiltmann 2024, 230–31; Meding und Daugs 2026, 77–79).[7] Dass ein Name in einem Text auftaucht, können sie extrahieren; warum dieser Name dort auftaucht und was seine Nennung bedeutet, bleibt Sache der Forschenden.

Gabriele Gramelsberger hat diese Entwicklung wissenschaftsphilosophisch als «Automatisierung kognitiver Funktionen des Wahrnehmens, Erkennens, Zuordnens oder Schließens» beschrieben (Gramelsberger 2019, 191). Anders als die Automatisierung menschlicher Arbeitskraft im 20. Jahrhundert ziele die gegenwärtige Transformation auf «epistemische Akteurialität»: Algorithmen werden selbst zu Akteuren der Erkenntnisproduktion. Diese Diagnose trifft einen realen Befund und nährt zugleich die Befürchtung, maschinelle Verfahren könnten hermeneutische Arbeit ersetzen.

Sybille Krämer hat demgegenüber die epistemische Differenz zwischen menschlicher und maschineller Sprachverarbeitung jüngst medienphilosophisch präzisiert: LLMs prozessieren «tokenstatistisch und das heißt ohne Interpretation»; ihre Ausgaben sind «inferenziell aus den impliziten Bezügen innerhalb von Datenkorpora abgeleitet» (Krämer 2025, 183). Krämers Analyse bezieht sich auf Chatbots – also auf die dialogische Interaktion mit sprachgenerativen Systemen (ein ähnliches Beispiel auch bei Hiltmann 2024). Der hier beschriebene Einsatz ist jedoch ein anderer: LLMs fungieren nicht als responsive Gesprächspartner, sondern als Werkzeuge zur Strukturierung, Annotation und Transformation von Daten. Die zugrundeliegende Operationslogik bleibt jedoch dieselbe. Gerade deshalb trifft Krämers Analogie: Der Umgang mit LLMs lasse sich «mit unserem Verhältnis zu Bibliotheken vergleichen: Gesammelte Ressourcen sozialer Gedächtnisse, die von keinem Individuum mehr überblickt, gleichwohl jedoch individuell genutzt werden können» (Krämer 2025, 183). Für den hier vertretenen Ansatz folgt daraus, dass die Alterität zwischen menschlicher Bedeutungssensitivität und maschineller Bedeutungsneutralität nicht Defizit ist, sondern Bedingung produktiver Arbeitsteilung.

Diese Arbeitsteilung ist damit auch keine Notlösung. Sie macht Zeit frei für das, was Annotation ermöglichen soll: die hermeneutische Durchdringung des Materials. Die Verschränkung von digitalen und interpretativen Verfahren ist dabei kein Desiderat, das erst noch eingelöst werden müsste. Sie ist längst gängige Praxis geworden: in jeder Studie, die auf aufbereiteten Korpora basiert, in jeder Visualisierung, die gedeutet wird,

7 Zukünftige Entwicklungen bleiben abzuwarten. Ein Beispiel ist das «vintage language model» Talkie (entwickelt von Levine, Duvenaud und Radford), das ausschliesslich auf vor 1931 erschienenem Textmaterial trainiert wurde und dessen Wissensstand bewusst auf diesen historischen Zeitpunkt begrenzt ist; Live-Demo unter https://talkie-lm.com/chat.

in jeder quantitativen Auffälligkeit, die zum Ausgangspunkt qualitativer Analyse wird. LLMs intensivieren diese Verschränkung, indem sie die infrastrukturellen Arbeitsschritte flexibler und zugänglicher machen, ohne den hermeneutischen Kern zu berühren.

5. Horizonte in Bewegung

Wie aber verhält sich diese in der Forschungspraxis beobachtete Verschränkung zu den Grundannahmen hermeneutischen Denkens? Eine Revision der Hermeneutik verlangt sie nicht, denn deren vermeintliche Grenzen waren nie Prinzipien, sondern praktische Notwendigkeiten.

Hans-Georg Gadamers Begriff des Horizonts kann hier als Ausgangspunkt dienen. Der Horizont bezeichnet bei Gadamer den Umkreis dessen, was von einem Standpunkt aus überhaupt in den Blick kommen kann. Verstehen vollzieht sich immer innerhalb eines Horizonts, aber dieser Horizont ist keine feste Grenze. Er ist, wie Gadamer formuliert, «etwas, in das wir hineinwandern und das mit uns mitwandert. Dem Beweglichen verschieben sich die Horizonte» (Gadamer [1960] 2010, 309). Was als Kontext verfügbar ist, bestimmt, was als erklärungsbedürftig erscheint und welche Fragen überhaupt gestellt werden können. Der hermeneutische Zirkel – die Bewegung zwischen Teil und Ganzem, zwischen Vorverständnis und Text – operiert innerhalb dieses Horizonts.

Gadamer bestimmt den Horizont dabei nicht medial. Es geht nicht um Bücher, nicht um Handschriften, nicht um eine bestimmte Materialität der Überlieferung. Der Horizont umfasst, was dem:der Interpretierenden zugänglich ist, und diese Zugänglichkeit ist historisch variabel. Ein Gelehrter des 18. Jahrhunderts, der auf die Bestände einer Universitätsbibliothek angewiesen war, operierte in einem anderen Horizont als ein:e Forscher:in des 21. Jahrhunderts mit Zugang zu digitalisierten Archiven. Beide betreiben Hermeneutik, nur in unterschiedlichen Horizonten.

Diese Variabilität liegt im Begriff selbst. Der Horizont ist gerade nicht das, was die:der Interpretierende aus eigener Kraft überblicken kann, sondern das, was die Bedingungen der jeweiligen Situation erschliessen. Zu diesen Bedingungen gehören: welche Texte überliefert sind, welche davon katalogisiert und auffindbar sind, welche physisch oder digital zugänglich sind, und welche davon in der verfügbaren Zeit tatsächlich gelesen werden können. Die Begrenzung auf das, was ein einzelner Mensch lesen kann, war nie ein methodisches Prinzip der Hermeneutik. Sie war eine Bedingung, unter der hermeneutische Arbeit faktisch stattfand. Mit digitalen Verfahren verschiebt sich diese Bedingung. Die Menge des tatsächlich Gelesenen wächst zwar auch mit LLMs nicht, wohl aber die Breite der Auswahl: Wie riesige Sachregister verweisen sie auf relevante Stellen, statt stellvertretend zu lesen.

Was aber bedeutet es für die hermeneutische Arbeit, wenn sich der Horizont verschiebt? Die Antwort liegt weniger in der Menge des Zugänglichen als in dessen Struktur. Der Horizont bestimmt bei Gadamer, was in den Blick kommt, und zugleich, wie es

erscheint: als fragwürdig oder selbstverständlich, als erklärungsbedürftig oder gegeben. Ein erweiterter Horizont verändert deshalb nicht primär die Antworten, sondern die Fragen.

Das lässt sich an den diskutierten Beispielen zeigen. Dass Kant kanonisiert wurde, während Niemeyer vergessen wurde, ist keine Frage, die bei der Lektüre einzelner Texte entstehen kann. Sie setzt voraus, dass beide Rezeptionslinien zugleich sichtbar sind – als Muster in einem Korpus, nicht als Eigenschaften einzelner Werke. Erst der erweiterte Horizont macht die Kontingenz sichtbar; erst die Sichtbarkeit macht sie erklärungsbedürftig. Die hermeneutische Frage – warum diese Kanonisierung? – folgt der methodischen Vorarbeit, die sie ermöglicht hat.

Ähnlich verhält es sich mit Erdmanns Befund zur Rezeptionsverarmung. Dass sich die Kant-Rezeption thematisch verengt hat, ist nichts, was die Lektüre eines einzelnen Lehrbuchs zeigen könnte. Es erfordert den diachronen Vergleich über zwei Jahrhunderte – einen Horizont, der über das hinausgeht, was individuelle Lektüre erschliessen kann. Der Befund selbst ist noch keine Interpretation, aber er konfiguriert den Raum, in dem Interpretation stattfinden kann.

Die Pointe ist nicht, dass digitale Methoden mehr Material liefern, sondern dass sie den Horizont so rekonfigurieren, dass neue Konstellationen sichtbar werden – Muster, Lücken, Verschiebungen, die bei sequenzieller Lektüre strukturell unsichtbar bleiben. Das ist kein quantitativer Gewinn, der sich in «mehr Texten» messen liesse. Es ist ein qualitativer Unterschied: Andere Fragen werden möglich.

Die hermeneutische Grundoperation bleibt davon unberührt. Verstehen vollzieht sich weiterhin als Bewegung zwischen Teil und Ganzem, als Revision des Vorverständnisses im Licht des Gelesenen. Aber das ‹Ganze›, auf das sich diese Bewegung bezieht, ist nicht mehr nur der einzelne Text oder das Œuvre einer Autorin oder eines Autors. Es kann nun das Korpus sein, das Referenznetzwerk, die diachrone Entwicklung eines Diskurses. Der hermeneutische Zirkel, so liesse sich formulieren, erweitert seinen Radius: von der Einzelquelle zum Quellenkorpus. Dadurch gewinnt er Fragen, die sich an der Einzelquelle allein nicht hätten stellen lassen.

Allerdings gerät dabei eine medientheoretische Annahme unter Spannung. Krämers Diagrammatologie hatte die «Kulturtechnik der Verflachung» als epistemischen Gewinn beschrieben: Die Zweidimensionalität der Fläche schaffe einen «überschaubaren und vollständig kontrollierbaren Raum» (Krämer 2009, 94). Bei LLMs ist diese Übersicht prinzipiell verloren. Der Embedding-Raum, in dem semantische Nähe als geometrische Distanz berechnet wird (vgl. Widdows 2004; Wiedemann und Fedtke 2021), umfasst hunderte bis tausende Dimensionen. Übrig bleibt die Fläche der Schnittstelle: Text als In- und Output. Gerade darin liegt jedoch eine Kontinuität zur hermeneutischen Praxis: Auch den Sinnhorizont eines Textes überblicken wir nie vollständig. Der Horizont ist, wie Gadamer betont, gerade das, worüber wir nicht verfügen. Die Fläche der Schnittstelle bleibt der Ort, an dem philologische Urteilskraft ansetzt.

Daraus folgt aber auch, dass digitale Methoden mitnichten, wie von manchen befürchtet, zwingend für jede Forschungsfrage herangezogen werden müssen. Vielmehr gilt es, weiterhin darüber nachzudenken, welche Quellen erkenntnisleitend und zweckdienlich zur Beantwortung der gestellten Frage herangezogen werden müssen und welche nicht.

6. Interpretation vor der Interpretation

Als epistemische Infrastruktur sind digitale Verfahren allerdings nicht neutral. Jeder Arbeitsschritt zwischen Quellendigitalisat und analysierbarem Korpus trifft Vorentscheidungen, die selbst der Reflexion bedürfen. Das gilt für LLMs ebenso wie für die längere Entwicklung, in der sie stehen, von regelbasierten NLP-Systemen über statistische Verfahren bis zu neuronalen Sprachmodellen. Die Probleme, die im Folgenden diskutiert werden, betreffen digitale Methoden in den Geisteswissenschaften generell.

Für diese Reflexionsarbeit hat sich in den Digital Humanities der Begriff der «Tool Criticism» etabliert. Traub und van Ossenbruggen haben ihn 2015 eingeführt:

> «With tool criticism we mean the evaluation of the suitability of a given digital tool for a specific task. Our goal is to better understand the impact of any bias of the tool on the specific task, not to improve the tools performance» (Traub und van Ossenbruggen 2015, 1).

Anders als bei der klassischen Quellenkritik, die den Überlieferungskontext und die Entstehungsbedingungen eines Dokuments reflektiert, richtet sich Tool Criticism auf die Werkzeuge, mit denen Quellen erschlossen werden. Zur Frage, was eine Quelle aussagt, tritt die Frage, was das Werkzeug aus ihr macht. Koolen, van Gorp und van Ossenbruggen haben diesen Ansatz weiterentwickelt und Reflexion als «central concept in digital tool criticism» bestimmt (Koolen et al. 2019, 380). Ihr Modell verbindet Forschungsfragen, Methoden, Daten und Werkzeuge in einem interdependenten Netzwerk, das durch reflexive Praxis integriert wird. Tool Criticism fügt der klassischen Quellenkritik damit eine weitere Frage hinzu: «How was a (version of a) source made?» (ebd., 373).

Ein erstes Problemfeld betrifft die Tokenisierung – die Zerlegung von Text in verarbeitbare Einheiten (vgl. Andresen 2024, 25–27). Wo wird geschnitten? Historische Texte operieren mit Kontraktionen, Abkürzungen und Worttrennungen am Zeilenende, die modernen Tokenizern fremd sind. Jede Entscheidung, ob «weyl.» als ein Token oder als Abkürzung für «weyland» behandelt wird, ist bereits interpretativ. Ähnlich verhält es sich mit Named Entity Recognition: Die an das Material angelegten Kategorien sind moderne Klassifikationen, die historische Ordnungen nicht abbilden müssen. Wenn ein NER-System auf gegenwärtigen Textkorpora trainiert wurde, erkennt es, was in diesen Korpora häufig vorkommt und übersieht, was randständig ist oder anderen Kategorien folgt.

Damit ist das grundsätzliche Problem der Trainingskorpora angesprochen (vgl. Schubbach 2024, 218–20). Die distributionelle Semantik, auf der LLMs beruhen, lernt Bedeutung aus Kookkurrenz: Was häufig in ähnlichen Kontexten auftritt, wird als semantisch ähnlich repräsentiert. Für historische Forschung bedeutet das zweierlei. Zum einen spiegeln die massiven Trainingskorpora gegenwärtige Wissensordnungen und nicht die des Untersuchungsgegenstandes. Zum anderen verschwindet Randständiges: Autor:innen, Begriffe, Diskurse, die in der Überlieferung marginalisiert wurden, sind in den Modellen unterrepräsentiert. Eine mögliche Reaktion auf diesen Umstand liegt im Nachtraining auf historischen Korpora oder im gezielten Einsatz von Few-Shot- und Zero-Shot-Verfahren, bei denen dem Modell wenige oder keine Trainingsbeispiele aus dem Zielbereich gegeben werden. Wie weit solche Ansätze Randständiges sichtbar machen, wird von der Qualität und Breite der verfügbaren historischen Korpora abhängen und bleibt empirisch zu prüfen.

Ein zweites Problemfeld betrifft die Texterkennung selbst. Die im vorherigen Abschnitt beschriebene OCR-Pipeline arbeitet bislang primär mit Antiqua-Drucken. Frakturschriften lassen sich zwar verarbeiten, aber mit erhöhter Fehlerquote. Durch spezialisierte Modelle und visuelles Encoding liessen sich hier Verbesserungen erzielen. Handschriften hingegen stellen eine qualitativ andere Herausforderung dar: Die Variabilität der Schriftzüge erfordert eigene Modelle und angepasste Architekturen. Hier wird es voraussichtlich verschiedene Anwendungsfälle geben – von gut lesbaren Kanzleischriften bis zu schwer entzifferbaren Notizen –, die jeweils eigene Lösungen verlangen.

Schliesslich ein LLM-spezifisches Problem: die Reproduzierbarkeit (vgl. Hiltmann 2024, 231). Anders als deterministische Algorithmen produzieren Sprachmodelle bei wiederholter Anwendung auf denselben Input nicht notwendig denselben Output. Für wissenschaftliche Nachvollziehbarkeit ist das ein ernstes Hindernis. Lösungsansätze existieren: Die Temperatur (ein Parameter, der die Zufälligkeit der Ausgabe steuert) kann nahe Null gesetzt werden, um deterministischere Ergebnisse zu erzielen. Bei Annotationsaufgaben können mehrfache Durchläufe mit anschliessender Aggregation die Stabilität erhöhen. Entscheidend ist in jedem Fall die Dokumentation: Modellversion, Parameter, Prompts müssen festgehalten werden, damit Ergebnisse überprüfbar bleiben. Das unterscheidet den hier beschriebenen Einsatz – LLMs als Werkzeuge zur Datenaufbereitung – grundlegend von der dialogischen Interaktion mit Chatbots, bei der Reproduzierbarkeit keine vergleichbare Rolle spielt.

Die Reflexion dieser Vorentscheidungen ist Teil der hermeneutischen Verantwortung. Tool Criticism erweitert die klassische Quellenkritik um eine weitere Dimension: Nicht nur die Quelle, auch das Werkzeug muss kritisch befragt werden. Die Infrastruktur strukturiert vor, was überhaupt als Befund erscheinen kann. Diese Vorstrukturierung transparent zu machen, gehört zur methodischen Sorgfalt.

Ein letzter Blick gilt der Arbeit des Interpretierens generell, also der Zuschreibung von Bedeutung zu Entitäten (vgl. dazu Jacke 2024). Für die Digital Humanities steht die ungeklärte Frage im Raum, inwiefern geisteswissenschaftliches Interpretieren eine Spezifik aufweist, die es vom Interpretieren von Daten (basierend auf computationellen Methoden als naturwissenschaftlich verstanden) unterscheidet. Ob die Operation der Interpretation einer quantifizierten Repräsentation eines Textes (und deren Visualisierung) sich unterscheidet von der Interpretation des gelesenen Textes, muss nicht zwingend mit ja beantwortet werden. In jedem Fall spielt sowohl für die Datenzusammenstellung, -aufbereitung wie auch die -auswertung die Interpretation eine zentrale Rolle, gleichwohl «Interpretation in wissenschaftlichen Kontexten […] sich grundsätzlich dadurch aus[zeichnet], dass ihre Evaluation nicht unproblematisch ist» (Jacke 2024, Abs. 20).

7. Mixed and Mingled

Das Eingangszitat dieses Beitrags – Derridas «Ein Text-Äußeres gibt es nicht» – ist kein schmückendes Motto, sondern methodologisch produktiv. Was Derrida zeichentheoretisch formuliert, lässt sich auf die hier diskutierten Verfahren beziehen: Bedeutung entsteht nicht durch Referenz auf eine textexterne Realität, sondern durch Differenzen innerhalb eines Systems von Zeichen. Sprache vermittelt, strukturiert und formt Erfahrung. Jeder Versuch, auf einen Ursprung oder ein Fundament zu verweisen, ist selbst wieder textuell strukturiert und diskursiv, symbolisch, kulturell codiert. Für die Interpretation folgt daraus, dass jede Lektüre immer schon in einem Netz von Bedeutungspraktiken situiert ist – ein voraussetzungsloses Lesen gibt es nicht.

Die distributionelle Semantik, auf der Large Language Models beruhen, operationalisiert diese Einsicht technisch. Was bei Derrida *différance* heisst – die permanente Verschiebung von Bedeutung, das Verweisen von Zeichen auf Zeichen ohne letzte Stabilität –, findet im Embedding-Raum seine technische Entsprechung: ein Raum reiner Relationalität, in dem es kein ‹Aussen› gibt, auf das Bedeutung gegründet werden könnte. Daraus folgt keine Gleichsetzung von Dekonstruktion und maschinellem Lernen, wohl aber eine methodologische Konsequenz: Wenn Texte immer schon in Referenzräumen, Diskursformationen und Rezeptionslinien situiert sind, dann machen digitale Verfahren nicht einen externen ‹Kontext› zugänglich, sondern erweitern den Horizont dessen, was als Text überhaupt lesbar wird.

In einem Sonderband zu Mixed Methods stellt sich die Frage, was der hier skizzierte Ansatz zu diesem Paradigma beiträgt und worin er sich von gängigen Designs unterscheidet. Klassische Mixed-Methods-Modelle operieren häufig sequenziell: Quantitative Verfahren identifizieren Muster, qualitative Verfahren vertiefen ausgewählte Fälle; oder umgekehrt: explorative Interviews generieren Hypothesen, die anschliessend statistisch geprüft werden.

Der hier vertretene Ansatz beschreibt etwas anderes: nicht Kombination, sondern konstitutive Interdependenz. Digitale Verfahren liefern nicht Befunde, die anschliessend hermeneutisch «vertieft» werden. Sie konfigurieren den Horizont, innerhalb dessen hermeneutische Fragen überhaupt entstehen können. Die Frage nach dem «Warum» setzt die Sichtbarkeit des «Dass» voraus; die hermeneutische Arbeit beginnt nicht nach der digitalen, sondern durch sie.

Das Fragezeichen im Titel dieses Beitrags markiert eine methodologische Pointe. «Mixed» erkennt an, dass die Differenz bestehen bleibt. Statistische Muster sprechen nicht für sich selbst; Interpretation ist nicht automatisierbar. Die Alterität zwischen maschineller Bedeutungsneutralität und menschlicher Bedeutungssensitivität erweist sich damit als Bedingung produktiver Arbeitsteilung. Wer glaubt, LLMs könnten hermeneutische Arbeit ersetzen, verwechselt Mustererkennung mit Verstehen (vgl. dazu auch Hiltmann 2024).

«Mingled» erkennt zugleich, dass die Verschränkung längst gängige Praxis ist. Jede Studie, die auf aufbereiteten Korpora basiert, ist bereits «mingled»; jede Visualisierung, die gedeutet wird, jede quantitative Auffälligkeit, die zum Ausgangspunkt qualitativer Analyse wird. Die Frage ist nicht, ob digitale und hermeneutische Verfahren verbunden werden sollen, sondern wie reflektiert diese Verbindung geschieht. Das «and» im Titel insistiert darauf, dass beides zugleich gilt: Die Methoden bleiben unterschieden und sind doch schon ineinander verwoben.

Klassische Forschungspraxis kannte einen Endpunkt der Quellenarbeit: Das Korpus wurde zusammengestellt, die Texte wurden gelesen, Ergebnisse schriftlich festgehalten. Diese Grenze war oft praktisch erzwungen – durch die Kapazität individueller Lektüre, durch das, was Archive und Bibliotheken zugänglich machten. Die hier beschriebene Arbeitsweise verschiebt diese Grenze, hebt sie aber nicht auf. Jede Interpretation kann neue Fragen generieren, die neue Suchanfragen verlangen, neue Annotationen, neue Muster. Die Entscheidung, wann es reicht, wird damit zu einer expliziten methodischen Wahl – nicht mehr durch materielle Grenzen erzwungen, sondern durch das Urteil der Forschenden.

Das ist kein Verlust an Verbindlichkeit, sondern Gewinn an Reflexivität. Tool Criticism, die Befragung der eigenen Werkzeuge, wird zum integralen Bestandteil des hermeneutischen Prozesses. Die Frage, was das Werkzeug aus der Quelle macht, gehört zur Frage, was die Quelle aussagt. Am Ende entscheidet der:die Forscher:in, welches Argument trägt.

8. Ausblick

Für die Historische Bildungsforschung folgt daraus eine doppelte Aufgabe (vgl. ähnlich Simons et al. 2026). Einerseits gilt es, die technischen Möglichkeiten zu nutzen, die LLMs und andere digitale Werkzeuge bieten. Die oben diskutierten Probleme – Trainingskorpora, Reproduzierbarkeit, die Vorstrukturierung durch Werkzeuge – verlangen dabei methodische Sorgfalt und kritische Reflexion. Andererseits bleibt die Ausbildungsfrage

virulent: Wo werden die Kompetenzen vermittelt, die ein reflektierter Umgang mit diesen Werkzeugen voraussetzt? Die Antwort kann nicht lauten, dass alle Historiker:innen programmieren lernen müssen; sie kann aber auch nicht lauten, dass technische Fragen an Spezialist:innen delegiert werden, während die ‹eigentliche› Interpretation unberührt bleibt.

Die digitalen Höhen und hermeneutischen Tiefen, von denen der Untertitel spricht, sind keine Gegensätze, zwischen denen gewählt werden müsste. Sie bezeichnen die Bewegungsrichtungen einer Forschungspraxis, die zwischen Übersicht und Nahsicht wechselt, zwischen Muster und Bedeutung, zwischen dem, was die Maschine sichtbar macht, und dem, was nur der Mensch verstehen kann. Die produktive Spannung zwischen beiden aufrechtzuerhalten – nicht aufzulösen – ist die methodische Herausforderung, vor der die Historische Bildungsforschung steht.[8]

Literaturverzeichnis

Andresen, Melanie. 2024. *Computerlinguistische Methoden für die Digital Humanities: Eine Ein-führung für Geisteswissenschaftler:innen*. Tübingen: Gunter Narr. https://doi.org/10.24053/9783823395799.

Bengio, Yoshua, Aaron Courville, und Pascal Vincent. 2013. «Representation Learning: A Review and New Perspectives». *IEEE Transactions on Pattern Analysis and Machine Intelligence* 35 (8): 1798–1828. https://doi.org/10.1109/TPAMI.2013.50.

Biemann, Chris, Gerhard Heyer, und Uwe Quasthoff. 2022. *Wissensrohstoff Text: Eine Einführung in das Text Mining*. Wiesbaden: Springer Vieweg. https://doi.org/10.1007/978-3-658-35969-0.

Derrida, Jacques. 1983. *Grammatologie*. Übersetzt von Hans-Jörg Rheinberger und Hanns Zischler. Frankfurt a. M.: Suhrkamp.

Diebolt, Claude, und Michael Haupert. 2016. «Clio's Contributions to Economics and History». *Revue d'économie politique* 126 (5): 971–89. https://doi.org/10.3917/redp.265.0971.

Engelmann, Sebastian. 2019. «Alles wie gehabt? Zur Konstruktion von Klassikern und Geschichte(n) der Pädagogik». In *Erinnern, Umschreiben, Vergessen: Die Stiftung des disziplinären Gedächtnisses als soziale Praxis*, herausgegeben von Markus Rieger-Ladich, Anne Rohstock, und Karin Amos, 65–93. Weilerswist: Velbrück. https://doi.org/10.5771/9783748901662.

Erdmann, Daniel. 2023. «Von Referenz-Prominenz und Rezeptions-Differenz: Immanuel Kant in (quasi-)erziehungswissenschaftlichen Lehrbüchern». In *Grenzziehungen und Grenzbeziehungen des Disziplinären: Verhältnisbestimmungen (in) der Erziehungswissenschaft*, herausgegeben von Susann Hofbauer, Felix Schreiber, und Katharina Vogel, 26–41. Beiträge zur Theorie und Geschichte der Erziehungswissenschaft 49. Bad Heilbrunn: Klinkhardt. https://doi.org/10.35468/6042.

8 Die sprachliche Überarbeitung und die Prüfung der Einhaltung der redaktionellen Vorgaben erfolgten mit Unterstützung von Claude (Anthropic; Modelle Claude Opus 4.5 bis 4.8). Der Einsatz von Claude Code bei der Entwicklung der OCR-Pipeline ist in Fussnote 4 dokumentiert. Die Verantwortung für Inhalt, Text und die Software liegt bei den Autor:innen.

Fickers, Andreas. 2020. «Update für die Hermeneutik. Geschichtswissenschaft auf dem Weg zur digitalen Forensik?» *Zeithistorische Forschungen/Studies in Contemporary History* 17 (1): 157–68. https://doi.org/10.14765/zzf.dok-1765.

Firth, John R. 1957. «A Synopsis of Linguistic Theory 1930–1955». In *Studies in Linguistic Analysis*, 1–32. Publications of the Philological Society, Special Volume. Oxford: Basil Blackwell.

Fleck, Ludwik. (1935) 1980. *Entstehung und Entwicklung einer wissenschaftlichen Tatsache*. Frankfurt a. M.: Suhrkamp.

Gadamer, Hans-Georg. (1960) 2010. *Wahrheit und Methode: Grundzüge einer philosophischen Hermeneutik*. 7. Aufl. Tübingen: Mohr Siebeck.

Gramelsberger, Gabriele. 2019. *Philosophie des Digitalen zur Einführung*. Hamburg: Junius.

Hiltmann, Torsten. 2022. «Vom Medienwandel zum Methodenwandel: Die fortschreitende Digitalisierung und ihre Konsequenzen für die Geschichtswissenschaften in historischer Perspektive». In *Digital History: Konzepte, Methoden und Kritiken Digitaler Geschichtswissenschaft*, herausgegeben von Karoline Dominika Döring, Stefan Haas, Mareike König und Jörg Wettlaufer, 13–44. *Studies in Digital History and Hermeneutics* 6. Berlin: De Gruyter Oldenbourg. https://doi.org/10.1515/9783110757101-002

Hiltmann, Torsten. 2024. «Hermeneutik in Zeiten der KI: Large Language Models als hermeneutische Instrumente in den Geschichtswissenschaften». In *KI:Text: Diskurse über KI-Textgeneratoren*, herausgegeben von Gerhard Schreiber und Lukas Ohly, 201–32. Berlin: De Gruyter. https://doi.org/10.1515/9783111351490-014

Jacke, Janina. 2024. «Interpretation». In *Begriffe der Digital Humanities. Ein diskursives Glossar*, herausgegeben von AG Digital Humanities Theorie des Verbandes Digital Humanities im deutschsprachigen Raum e. V. Zeitschrift für digitale Geisteswissenschaften / Working Papers 2. Wolfenbüttel. https://doi.org/10.17175/wp_2023_006_v2.

Kant, Immanuel. 1803. *Über Pädagogik*. Hrsg. von Friedrich Theodor Rink. Königsberg: Friedrich Nicolovius.

Karjus, Andres. 2025. «Machine-Assisted Quantitizing Designs: Augmenting Humanities and Social Sciences with Artificial Intelligence». *Humanities and Social Sciences Communications* 12: Art. 277. https://doi.org/10.1057/s41599-025-04503-w.

Koolen, Marijn, Jasmijn van Gorp, und Jacco van Ossenbruggen. 2019. «Toward a Model for Digital Tool Criticism: Reflection as Integrative Practice». *Digital Scholarship in the Humanities* 34 (2): 368–85. https://doi.org/10.1093/llc/fqy048.

Krämer, Sybille. 2009. «Operative Bildlichkeit: Von der ‹Grammatologie› zu einer ‹Diagrammatologie›? Reflexionen über erkennendes ‹Sehen›». In *Logik des Bildlichen: Zur Kritik der ikonischen Vernunft*, herausgegeben von Martina Hessler und Dieter Mersch, 94–122. Bielefeld: transcript. https://doi.org/10.1515/9783839410516-003.

Krämer, Sybille. 2025. *Der Stachel des Digitalen*. Frankfurt a. M.: Suhrkamp.

Lenci, Alessandro, und Magnus Sahlgren. 2023. *Distributional Semantics*. Studies in Natural Language Processing. Cambridge: Cambridge University Press. https://doi.org/10.1017/9780511783692.

Lorenz, Marco. 2025. «OCR Pipeline for Historical Print Periodicals Using Mistral AI» (v1.0.2). Zenodo. https://doi.org/10.5281/zenodo.17632399.

Meding, Holle, und Aurel Daugs. 2026. «On the Use and Limitations of Large Language Models in Historical Scholarship». In *Understanding Science with Large Language Models? Potentials for the History, Philosophy, and Sociology of Science*, herausgegeben von Arno Simons, Adrian Wüthrich, Michael Zichert und Gerd Graßhoff, 71–99. Bielefeld: transcript.

Messeri, Lisa, und Molly J. Crockett. 2024. «Artificial Intelligence and Illusions of Understanding in Scientific Research». *Nature* 627 (8002): 49–58. https://doi.org/10.1038/s41586-024-07146-0.

Oberdorf, Andreas. 2022. «Der Digital Turn in der historischen Bildungsforschung: Einleitung». In *Digital Turn und Historische Bildungsforschung: Bestandsaufnahme und Forschungsperspektiven*, herausgegeben von Andreas Oberdorf, 7–15. Bad Heilbrunn: Klinkhardt. https://doi.org/10.35468/5952-01.

Pölitz, Karl Heinrich Ludwig. 1806. *Die Erziehungswissenschaft, aus dem Zwecke der Menschheit und des Staates practisch dargestellt*. 2 Bände. Leipzig: J. C. Hinrichs.

Pollin, Christopher. 2025. «Promptotyping: Von der Idee zur Anwendung». DHCraft Blog, 24. April 2025. https://dhcraft.org/excellence/blog/Promptotyping/.

Pollin, Christopher, Franz Fischer, Patrick Sahle, Martina Scholger, und Georg Vogeler. 2025. «When it was 2024 – Generative AI in the Field of Digital Scholarly Editions». *Zeitschrift für digitale Geisteswissenschaften* 10. https://doi.org/10.17175/2025_008.

Rapp, Andrea. 2021. «Digital Humanities und Bibliotheken: Traditionen und Transformationen». 027.7 *Zeitschrift für Bibliothekskultur* 8 (1). https://doi.org/10.21428/1bfadeb6.486c17e5.

Rapp, Andrea, Sabine Bartsch, und Luise Borek. 2016. «Aus der Mitte der Fächer, in die Mitte der Fächer: Studiengänge und Curricula – Digital Humanities in der universitären Lehre». Bibliothek Forschung und Praxis 40 (2): 172–78. https://doi.org/10.1515/bfp-2016-0030.

Rehbein, Malte, und Patrick Sahle. 2013. «Digital Humanities lehren und lernen: Modelle, Strategien, Erwartungen». In *Evolution der Informationsinfrastruktur: Kooperation zwischen Bibliothek und Wissenschaft*, herausgegeben von Heike Neuroth, Norbert Lossau und Andrea Rapp, 209–28. Glückstadt: Verlag Werner Hülsbusch. https://doi.org/10.17875/gup2012-313.

Roden, Golo. 2026. «Code lesen statt Code schreiben: Die unterschätzte Senior-Disziplin». heise online, 16. Mai 2026. https://heise.de/-11288309.

Schubbach, Arno. 2024. «Maschinelles Lernen». In *Digitalität von A bis Z*, herausgegeben von Florian Arnold, Johannes C. Bernhardt, Daniel Martin Feige und Christian Schröter, 213–22. Bielefeld: transcript. https://doi.org/10.1515/9783839467657-022.

Simons, Arno, Adrian Wüthrich, und Michael Zichert. 2026. «Doing Science Studies with Large Language Models: An Introduction». In *Understanding Science with Large Language Models? Potentials for the History, Philosophy, and Sociology of Science*, herausgegeben von Arno Simons, Adrian Wüthrich, Michael Zichert und Gerd Graßhoff, 13–32. Bielefeld: transcript.

Spolsky, Joel. 2000. «Things You Should Never Do, Part I». Joel on Software (Blog), 6. April 2000. https://www.joelonsoftware.com/2000/04/06/things-you-should-never-do-part-i/.

Tanner, Jakob. 2019. «Binäre Codes und komplexes Denken: Digital Humanities und Geschichtswissenschaft». In *Linguistische Kulturanalyse*, herausgegeben von Juliane Schröter, Susanne Tienken, Yvonne Ilg, Joachim Scharloth und Noah Bubenhofer, 91–110. Berlin: De Gruyter. https://doi.org/10.1515/9783110585896-005.

Thaulow, Gustav. 1845. *Erhebung der Pädagogik zur philosophischen Wissenschaft. Oder Einleitung in die Philosophie der Pädagogik. Zum Behuf seiner Vorlesungen*. Berlin: Beit und Comp.

Traub, Myriam C., und Jacco van Ossenbruggen. 2015. «Workshop on Tool Criticism in the Digital Humanities». CWI Technical Report. https://ir.cwi.nl/pub/23500.

Van Herck, Sytze. 2022. «Historians as Computer Users: Organizing Sources with Digital Asset Management for a History of Computing». In *Digital History and Hermeneutics: Between Theory and Practice*, herausgegeben von Andreas Fickers und Juliane Tatarinov, 219–36. Studies in Digital History and Hermeneutics 2. Berlin: De Gruyter Oldenbourg. https://doi.org/10.1515/9783110723991-011.

Vaswani, Ashish, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, und Illia Polosukhin. 2017. «Attention Is All You Need». Version 7. Preprint, arXiv. https://doi.org/10.48550/ARXIV.1706.03762.

Vogel, Katharina. 2024. *Pädagogische Wissensräume 1750–1850: Empirische Studien zum Referenzraum wissenschaftlich-pädagogischer Lehrbücher, Vorlesungssammlungen und Einführungsschriften*. Erziehungswissenschaftliche Studien 15. Göttingen: Universitätsverlag Göttingen. https://doi.org/10.17875/gup2024-253.

Wang, Huanting, Jingzhi Gong, Huawei Zhang, Jie Xu, und Zheng Wang. 2025. «AI Agentic Programming: A Survey of Techniques, Challenges, and Opportunities». Version 2. Preprint, arXiv. https://doi.org/10.48550/arXiv.2508.11126.

Weitin, Thomas. 2015. «Digitale Literaturwissenschaft». *Deutsche Vierteljahrsschrift für Literaturwissenschaft und Geistesgeschichte* 89 (4): 651–56. https://doi.org/10.1007/BF03396502.

Widdows, Dominic. 2004. *Geometry and Meaning*. CSLI Lecture Notes 172. Stanford: CSLI Publications.

Wiedemann, Gregor, und Cornelia Fedtke. 2021. «From Frequency Counts to Contextualized Word Embeddings. The Saussurean Turn in Automatic Content Analysis». In *Handbook of Computational Social Science, Volume 2. Data Science, Statistical Modelling, and Machine Learning Methods*, herausgegeben von Uwe Engel, Anabel Quan-Haase, Sunny Xun Liu, und Lars Lyberg, 366–85. London: Routledge. https://doi.org/10.4324/9781003025245.