OCR lehren, obwohl der Chatbot es kann? Offene Lernmaterialien für Datenkompetenz in den Geisteswissenschaften

Evgenia Samoilova[1], Jonas Michael Rohe[2], Frank Fischer[2], Sonja Schimmler[3], Ulrike Lucke[1]

[1] Universität Potsdam · [2] Freie Universität Berlin · [3] Fraunhofer FOKUS / TU Berlin
Kontakt: evgenia.samoilova@uni-potsdam.de

Der Wissenschaftsrat hat im Juli 2026 „intellektuelle Souveränität“ als Leitidee für die Hochschulbildung im Zeitalter generativer KI hervorgehoben: dazu gehören die Stärkung fachlicher Expertise, der Aufbau von KI-Kompetenzen sowie die hochschulübergreifende Entwicklung geeigneter KI-Infrastrukturen [WR26]. In den Geisteswissenschaften wächst die Verfügbarkeit digitaler Quellen schneller als die Kompetenzen zu ihrer kritischen Nutzung. In vielen geisteswissenschaftlichen Studiengängen, auch an Brandenburger Hochschulen, sind Daten- und KI-Kompetenzen jedoch bislang nicht systematisch im Curriculum verankert. Verschiedene Initiativen, etwa an der Universität Potsdam oder der FU Berlin, haben bereits fachspezifische Konzepte entwickelt und erprobt, um digitale Methoden stärker in die geisteswissenschaftliche Lehre zu integrieren [Lu26]. Solche Ansätze bleiben jedoch häufig an einzelne Lehrveranstaltungen, Module oder engagierte Personen gebunden und lassen sich nur begrenzt auf andere Kontexte übertragen. Offene Bildungsmaterialien (OER) bieten hier das Potenzial für die nachhaltigere Verankerung entsprechender Kompetenzen.

Das Berlin-Brandenburgische Datenkompetenzzentrum QUADRIGA entwickelt solche OER als interaktive Online-Lehrbücher, die den Erwerb von Datenkompetenzen anhand fachspezifischer Forschungsbeispiele unterstützen. Grundlage ist das QUADRIGA-Datenkompetenzframework [Pe26], das fächerübergreifend beschreibt, welche Kompetenzen für den Umgang mit Forschungsdaten erforderlich sind. Die Vermittlung erfolgt fachspezifisch. In Weiterbildungen für Forschende lassen sich unsere OER gut einsetzen, wenn Zielgruppe und Material aufeinander abgestimmt sind. Eine größere Herausforderung stellt die Nachnutzung in heterogenen Gruppen dar. Wir untersuchen sie an zwei Kontexten, in denen dieselbe OER [Sk25] eingesetzt wurde: dem DH Early Career Fellowship an der Universität Potsdam und dem Seminar „Daten erzählen" an der FU Berlin. Inhaltlich führt die OER durch einen Arbeitsschritt, der inzwischen auch mit generativer KI möglich ist: Historische Zeitungen werden mit einer eigenen OCR-Pipeline in Python digitalisiert, abschließend geht die OER darauf ein, was Sprachmodelle hier leisten können. Im Seminar wurde dieser Vergleich vertieft. Aktuelle Studien zeigen, dass Sprachmodelle historische Drucke oft genauer erkennen, zugleich aber weniger offensichtliche Fehler machen und nicht immer reproduzierbare Ergebnisse liefern [Lv25]. Die Gegenüberstellung macht so praktisch erfahrbar, was intellektuelle Souveränität im Umgang mit KI bedeutet: Stärken und Grenzen verschiedener Verfahren einzuschätzen und begründet zu entscheiden, welches für die jeweilige Aufgabe geeignet ist. Fellowship und Seminar nutzten Cloud-Umgebungen, sodass keine Installation nötig war.

Ablauf: 2–3 Min. Anmoderation; 15 Min. Impuls; 12 Min. Diskussion entlang der drei Fragen: Welche Kompetenzen sollten Studierende selbst entwickeln, wenn KI Teile einer Aufgabe übernehmen kann, und wie werden solche Lernziele fachlich bestimmt? Wie lassen sich OER so gestalten, dass Lehrende sie in heterogenen Gruppen leicht nachnutzen und um KI-Aspekte erweitern können? Welche technischen Voraussetzungen braucht die nachhaltige Nutzung von OER mit Daten und KI, und welche davon sollten einzelne Hochschulen selbst bereitstellen bzw. hochschulübergreifend entwickelt werden?

Literatur

[Lv25] Levchenko, M.: Evaluating LLMs for Historical Document OCR: A Methodological Framework for Digital Humanities. arXiv:2510.06743, 2025.
[Lu26] Lucke, U.; Hafer, J.; Möring, S.; Trilcke, P. (Hrsg.): Digitale Fachkonzepte. Forschendes Lernen und digitale Methoden in den Geisteswissenschaften. Potsdamer Beiträge zur Hochschulforschung 7, Universitätsverlag Potsdam, 2026. ISBN 978-3-86956-569-9.
[Pe26] Petras, V.; Neuroth, H.; Seltmann, M.; Schnaitter, H.; Walter, P.: QUADRIGA Datenkompetenzframework (Version 4). Zenodo, 2026. https://doi.org/10.5281/zenodo.19470557
[Sk25] Skorinkin, D.; Sluyter-Gäthje, H.; Trilcke, P.: Quantitative Analyse der Medienwellen der Spanischen Grippe (1918/19). Eine Fallstudie (Version v1.0.0-beta.3). Zenodo, 2025. https://doi.org/10.5281/zenodo.15703503
[WR26] Wissenschaftsrat: Intellektuelle Souveränität: Empfehlungen für die Hochschulbildung in Zeiten von generativer KI (Drs. 3319-26). Köln, 2026. https://doi.org/10.57674/1evx-t906