

MILHOUSE: LLM-basierte Sacherschließung in der Praxis

Ein Werkstattbericht der Universitätsbibliothek Magdeburg

Uli Niemann, Universitätsbibliothek Magdeburg

Sascha Bosse, Universitätsbibliothek Magdeburg

Zusammenfassung

Wissenschaftliche Bibliotheken verzeichnen einen starken Medienzuwachs, deren inhaltliche Erschließung personell kaum stemmbar ist. Dieser Beitrag stellt MILHOUSE vor, ein an der Universitätsbibliothek Magdeburg entwickeltes, auf großen Sprachmodellen basierendes Werkzeug zur (semi-)automatisierten Zuordnung von Notationen einer Fachsystematik. Die Anwendung reichert Titel mit geeigneten Metadaten an, nutzt strukturierte Prompts mit festen Ausgabeschemata und integriert eine Plausibilitätsprüfung zur Reduktion modellbedingter Halluzinationen. Die Bewertung der Qualität der Ergebnisse wird in einem *Human-in-the-Loop*-Ansatz sichergestellt. MILHOUSE wurde in einem produktionsnahen Einsatz (Q3 2025) von den Fachreferentinnen und Fachreferenten der Universitätsbibliothek Magdeburg praktisch erprobt. Der Beitrag stellt die Potenziale dieses niedrigschwelligen Ansatzes dar und skizziert Gestaltungsprinzipien für eine schnelle, verantwortungsvolle Integration Künstlicher Intelligenz in bibliothekarische Workflows.

Summary

Academic libraries are experiencing a rapid increase in media holdings, making comprehensive subject indexing increasingly difficult to manage with existing staff resources. This paper presents MILHOUSE, a tool developed at the University Library of Magdeburg that leverages large language models for the (semi-)automated assignment of subject classification notations. The tool enriches bibliographic records with suitable metadata, employs structured prompts with fixed output schemes, and integrates a plausibility check to reduce model-induced hallucinations. The quality of the results is ensured through a human-in-the-loop approach. MILHOUSE was practically tested by subject librarians of the University Library of Magdeburg in a near-production setting (Q3 2025). This paper highlights the advantages of this lightweight approach over traditional training pipelines and derives design principles for the rapid and responsible integration of artificial intelligence into library workflows.

Schlagwörter: Automatisierte Sacherschließung; Große Sprachmodelle (LLM); Künstliche Intelligenz in Bibliotheken; Human-in-the-Loop

Zitierfähiger Link (DOI): <https://doi.org/10.5282/o-bib/6220>

Autorenidentifikation: Uli Niemann, ORCID: [0000-0001-9634-2248](https://orcid.org/0000-0001-9634-2248),
Sascha Bosse, ORCID: [0000-0002-2490-363X](https://orcid.org/0000-0002-2490-363X)

Dieses Werk steht unter der Lizenz [Creative Commons Namensnennung 4.0 International](https://creativecommons.org/licenses/by/4.0/).

1. Einleitung

Die digitale Transformation verändert Arbeitsprozesse, Nutzungsweisen und Sammlungsprofile wissenschaftlicher Bibliotheken grundlegend. Mit ihr einher geht ein kontinuierlicher Anstieg des jährlichen Medienzuwachses, der vor allem durch die steigende Bedeutung elektronischer Medien (E-Medien) getragen wird. So verzeichneten die im Deutschen Bibliotheksverband (dbv) organisierten wissenschaftlichen Bibliotheken einen Anstieg des E-Medien-Bestands von rund 77 Mio. im Jahr 2022 auf etwa 84 Mio. im Jahr 2024.¹ Auch die Deutsche Bibliotheksstatistik (DBS) weist darauf hin, dass neu erworbene Printmedien in den letzten Jahren tendenziell rückläufig sind, während der Ausbau digitaler Angebote den Gesamtzuwachs deutlich erhöht.² In Summe verzeichneten wissenschaftliche Bibliotheken damit einen erheblich höheren jährlichen Medienzuwachs als noch vor zehn Jahren (vgl. Abb. 1).

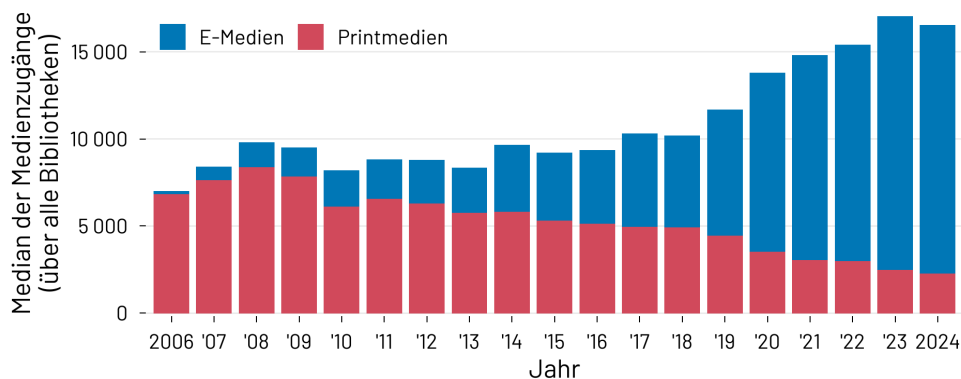


Abb. 1: Median der jährlichen Zugänge von Print- und E-Medien an 280 wissenschaftlichen Bibliotheken in Deutschland (2006–2024). Die Darstellung basiert auf den Kennzahlen der Deutschen Bibliotheksstatistik (DBS Nr. 19 für Print, Nr. 111 für E-Medien).

Ein analoger Trend lässt sich auch international beobachten. So verzeichnet die Association of Research Libraries (ARL) für Forschungsbibliotheken in den USA und Kanada zwischen 1998 und 2018 einen Anstieg der Gesamtausgaben um rund 68 %.³ Dabei wuchsen die Aufwendungen für fortlaufende Ressourcen (primär elektronische Zeitschriften) mit 166 % überproportional stark, während die Ausgaben für einmalige Anschaffungen lediglich um 23 % zunahmen. In US-amerikanischen wissenschaftlichen Bibliotheken überstiegen die Ausgaben für E-Medien erstmals im Zeitraum 2013 bis 2016 erstmals jene für Printmedien.⁴

1 Deutscher Bibliotheksverband e.V.: Bibliotheken 2024, 2024, S. 3, https://www.bibliotheksverband.de/sites/default/files/2024-10/Bibliotheken%202024_web.pdf, Stand: 05.01.2026.

2 hzb – Hochschulbibliothekszentrum des Landes Nordrhein-Westfalen: Deutsche Bibliotheksstatistik (DBS), 2025, <https://www.bibliotheksstatistik.de/>, Stand: 05.01.2026.

3 Association of Research Libraries: ARL Statistics Survey Statistical Trends. Expenditure Trends in ARL Libraries, 1986–2018, Association of Research Libraries, 2019, <https://www.arl.org/arl-statistics-survey-statistical-trends/>, Stand: 05.01.2026.

4 Kortick, Yoel: Trends in Physical and Electronic Resource Usage in U.S. Academic Libraries, in: Ex Libris Blog, 2021, <https://exlibrisgroup.com/blog/changing-priorities-trends-in-physical-and-electronic-resource-usage-in-us-academic-libraries/>, Stand: 05.01.2026.

Die Universitätsbibliothek der Otto-von-Guericke-Universität Magdeburg (UB Magdeburg) versorgt rund 13000 Studierende sowie Forschende und Mitarbeitende der Universität mit gedruckten und elektronischen Medien. Im Jahr 2024 umfasste der Bestand 435.904 gedruckte Bände sowie 426.143 E-Books. Der Wandel hin zu E-Medien zeigt sich in den Erwerbungszahlen der Jahre 2020 bis 2024: Einem jährlichen Zuwachs von durchschnittlich 6.249 Printmedien standen 50.427 digitalen Einheiten gegenüber. Auch die Ausgaben spiegeln dies wider: Durchschnittlich flossen 692.212 € in E-Books, verglichen mit 363.432 € für Printmedien.⁵ Die inhaltliche Erschließung des Bestands wird derzeit von neun Fachreferentinnen und Fachreferenten (FR) durchgeführt.

Abb. 2 zeigt den Anteil von E- und Print-Medien im Bestand der UB Magdeburg nach Erscheinungsjahr. Während bei den Titeln von 1995 noch 78 % gedruckt und nur 17 % elektronisch sind, überholen ab etwa 2005 E-Medien die Printmedien und dominieren seitdem deutlich. Für das Erscheinungsjahr 2024 entfallen nahezu alle Zugänge auf elektronische Titel.

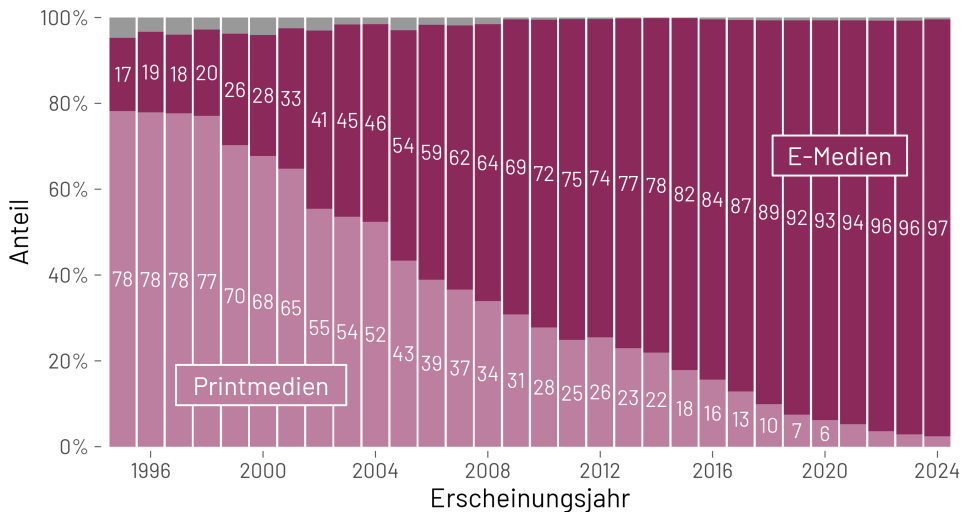


Abb. 2: Anteil elektronischer Medien und Printmedien nach Erscheinungsjahr in der UB Magdeburg für den Bestand in 2024. Grauer Bereich repräsentiert sonstige Medien wie Mikroformen, Ton- und Videodatenträger.

Mit dem starken Medienzuwachs wächst zugleich die Herausforderung, deren Inhalte verlässlich, konsistent und in der Breite zu erschließen. Eine internationale Befragung von 107 Universitätsbibliotheken in Australien, Brasilien, England, Finnland, Norwegen, Portugal, Schweden, Spanien und den USA ergab, dass 58 % der Bibliotheken für E-Books keine eigenen Schlagwörter vergeben, sondern Metadaten von Verlagen bzw. Anbietern übernehmen; zugleich wird die Eignung dieser Angaben teils skeptisch beurteilt.⁶ Im deutschsprachigen Raum verweist eine Umfrage der AG K10plus Sacherschließung⁷ unter 128 Teilnehmerbibliotheken auf einen Rückgang der intellektuellen Sacherschließung und

5 Quelle: Deutsche Bibliotheksstatistik (DBS): Printbestand 2024 (DBS Nr. 18), E-Book-Bestand 2024 (DBS Nr. 113.1), durchschnittlicher jährlicher Zugang Printbestand 2020–2024 (DBS Nr. 19), durchschnittlicher jährlicher Zugang digitaler Bestand 2020–2024 (DBS Nr. 111), durchschnittliche Ausgaben Printmedien 2020–2024 (DBS Nr. 21), durchschnittliche Ausgaben E-Books 2020–2024 (DBS Nr. 113.4).

nennt E-Medien ausdrücklich als Materialgattung, die häufig nicht oder weniger konsistent erschlossen wird.⁸ Wiesenmüller konstatiert, Sacherschließung werde vielerorts "primär als Kostenfaktor" wahrgenommen, Ressourcen würden tendenziell abgezogen und Qualität wie Umfang litten, weshalb computerunterstützte Verfahren notwendig seien, aber mit menschlicher Kontrolle kombiniert werden müssten.⁹ Automatisierte Sacherschließung verspreche Bibliotheken große Effizienzgewinne, zugleich bleibe ein Human-in-the-loop-Ansatz unverzichtbar, um Qualität, Verantwortung und Kontextbewusstsein sicherzustellen.¹⁰

Die UB Magdeburg nutzt eine eigene Lokalsystematik für die physische Aufstellung und fachliche Feinerschließung des Printbestands. Sie wurde im Zuge der Gründung der Otto-von-Guericke-Universität 1993 und der damit verbundenen Zusammenlegung der Vorgängerbibliotheken eingeführt und seither kontinuierlich gepflegt. Diese universale Systematik deckt alle an der Universität vertretenen Fächer ab und bildet lokale Forschungsschwerpunkte besonders differenziert ab. Sie ist hierarchisch in 21 Hauptklassen und bis zu 6 Ebenen gegliedert und umfasst rund 3.500 belegungsfähige Endnotationen.¹¹ Etwa 60 % des Printbestands sind freihandaufgestellt und damit unmittelbar an die Vergabe von Lokalnotationen gebunden.

Die Beibehaltung der Lokalsystematik ist dabei nicht allein historisch begründet, sondern folgt strategischen und nutzerorientierten Erwägungen. Zum einen dient sie den Nutzenden als vertrautes Orientierungssystem. Zum anderen ermöglicht sie es, Printmedien und den wachsenden E-Book-Bestand inhaltlich konsistent als Einheit zu erschließen. Eine Umstellung des umfangreichen Bestands auf eine allgemeine Klassifikation wäre hingegen mit einem unverhältnismäßig hohen organisatorischen und finanziellen Aufwand verbunden.

Da E-Medien keine physische Aufstellung erfordern und ihre Erwerbung anfangs teils losgelöst von den etablierten Print-Abläufen erfolgte, wurde über Jahre hinweg auf eine Vergabe von Lokalnotationen verzichtet. Erst seit 2023 werden einzeln erworbene E-Books sukzessive klassifiziert. Bei umfangreichen Titelkollektionen (z. B. EBS-Modelle, Paketkäufe) stößt die manuelle Erschließung jedoch schnell an Grenzen, da häufig Hunderte bis Tausende Titel gleichzeitig in den Bestand gelangen. Erschwerend kommt hinzu, dass sich das Aufgabenprofil der FR in den letzten Jahren stark diversifiziert hat.¹² Neben der klassischen Bestandsentwicklung umfassen die Tätigkeiten zunehmend Auf-

6 Gil-Leiva, Isidoro; Fujita, Mariângela Spotti Lopes; Ortuño, Pedro Díaz u. a.: The Abandonment of the Assignment of Subject Headings and Classification Codes in University Libraries Due to the Massive Emergence of Electronic Books, in: Knowledge Organization 47 (8), 2021, S. 652–654, <https://doi.org/10.5771/0943-7444-2020-8-646>.

7 Die AG K10plus Sacherschließung ist ein Fachgremium im Verbundsystem K10plus, einer Kooperation des Bibliotheksservice-Zentrums Baden-Württemberg (BSZ) und der Verbundzentrale des Gemeinsamen Bibliotheksverbunds (VZG). Die AG koordiniert Standards, Verfahren sowie die Weiterentwicklung der inhaltlichen Erschließung.

8 Balakrishnan, Uma; Kühn, Armin: Ergebnisse der Umfrage zur sachlichen Erschließung in GBV- und SWB-Teilnehmerbibliotheken, 2019, S. 16–19, <https://opus.k10plus.de/frontdoor/index/index/docId/464>, Stand: 05.01.2026.

9 Wiesenmüller, Heidrun; Hinrichs, Imma: Computerunterstützte Inhaltsererschließung. Bericht über einen Workshop an der UB Stuttgart – mit einem Exkurs zum neuen Inhaltsererschließungskonzept der DNB, in: o-bib 4 (3), 2017, S. 94–105, Zitat S. 94, <https://doi.org/10.5282/o-bib/2017H3S94-105>.

10 Vgl. Kasprzik, Anna; Seeliger, Frank: KI in wissenschaftlichen Bibliotheken, Teil 1: Handlungsfelder, große Player und die Automatisierung der Erschließung, 2022, <https://www.zbw-mediatalk.eu/de/2022/08/ki-in-wissenschaftlichen-bibliotheken-teil-1-handlungsfelder-grosse-player-und-die-automatisierung-der-erschliessung/>, Stand: 05.01.2026.

11 Universitätsbibliothek Magdeburg: Systematik der Universitätsbibliothek Magdeburg, 2025, https://www.ub.ovgu.de/ub_media/Literatursuche/Systematik.pdf, Stand: 05.01.2026.

gabenbereiche wie Open Science und Forschungsunterstützung (Open-Access-Transformation, Forschungsdatenmanagement), die Mitgestaltung der Digitalisierung von Bibliotheksprozessen,¹³ sowie Förderung von Medien- und Informationskompetenz.

Daraus ergibt sich ein Spannungsfeld: Einerseits ist eine flächendeckende manuelle Erschließung der wachsenden Medienbestände unter diesen Bedingungen unrealistisch, andererseits bleibt eine konsistente inhaltliche Erschließung aus strategischer Sicht unverzichtbar.¹⁴

- *Nutzerorientierung*: Nutzende erwarten formatunabhängige, thematisch integrierte Sichten. Wer ein Buch im Regal findet, möchte wissen, ob es ergänzende oder aktuellere E-Books zum selben Thema gibt.
- *Auffindbarkeit und Sichtbarkeit*: Eine klassifikatorische Erschließung verbessert die thematische Suche von E-Medien insbesondere in lokalen Katalog- und Discovery-Systemen (z. B. über fachliche Facetten oder Filter). Gleichzeitig liefert sie strukturierte, maschinenlesbare Daten, die von spezialisierten Suchdiensten und KI-basierten Systemen zur Kontextualisierung genutzt werden können.
- *Bestandsanalyse und -entwicklung*: Nur eine systematische Erschließung ermöglicht präzise Lückenanalysen und eine profilorientierte Erwerbung.
- *Nutzungsanalyse*: Die fachliche Zuordnung bildet die Basis für differenzierte Auswertungen und datengestützte Steuerungsentscheidungen.
- *Bestandspflege*: Eine fachliche Zuordnung ist Voraussetzung dafür, E-Medien eindeutig einem Fachreferat zuzuweisen und damit die Überprüfung auf Relevanz, Aktualität, Nutzbarkeit und Mehrfachlizenzierung systematisch vorzunehmen.
- *Prozessharmonisierung*: Einheitliche Verfahren erhöhen die Konsistenz in der Medienbearbeitung und reduzieren Doppelstrukturen.

Auf maschinellem Lernen basierende Verfahren zur (semi-)automatischen Sacherschließung gelten seit längerem als vielversprechender Ansatz, um die genannten Probleme zu adressieren, Fachreferate zu entlasten und Prozesse zu beschleunigen.¹⁵ Neuere Entwicklungen im Bereich generativer KI und großer Sprachmodelle (Large Language Models, LLMs) knüpfen an diese Arbeiten an und eröffnen zusätzliche Möglichkeiten für eine stärker kontextualisierte und skalierbare automatische Erschließung.

12 Sühl-Strohmenger, Wilfried; Tappenbeck, Inka (Hg.): Praxishandbuch Wissenschaftliche Bibliothekar:innen, Berlin, Boston 2024, <https://doi.org/10.1515/9783110790375>.

13 Vgl. Bosse, Sascha; Lücke, Christine; Thomas, Linda: Library Inventory Statistics Application. Eine Fallstudie über die agile Projektarbeit im Rahmen des Fachreferats, in: o-bib 9 (2), 2022, S. 1–15, <https://doi.org/10.5282/o-bib/5774>.

14 Vgl. Franke-Maier, Michael; Kasprzik, Anna; Ledl, Andreas u. a. (Hg.): Editorial, in: Qualität in der Inhaltsererschließung, Berlin, Boston 2021, S. 1–18, <https://doi.org/10.1515/9783110691597-001>.

15 Golub, Koraljka: Automated Subject Indexing. An Overview, in: Cataloging & Classification Quarterly 59 (8), 2021, S. 702–719, <https://doi.org/10.1080/01639374.2021.2012311>.

Ziel dieses Beitrags ist die Vorstellung und Einordnung des in der UB Magdeburg entwickelten Tools MILHOUSE als niedrigschwellige, KI-basierte Lösung. Zunächst werden dazu in Abschnitt 2 etablierte Ansätze zur automatisierten Sacherschließung mittels ressourcenintensiver, trainingsbasierter Frameworks diskutiert. Abschnitt 3 stellt die technische Architektur, die Funktionsweise und die Workflow-Integration von MILHOUSE im Detail vor. Darauf aufbauend wird in Abschnitt 4 über die konkreten Praxiserfahrungen, quantitative Nutzungsdaten und das qualitative Feedback der FR aus dem produktionsnahen Einsatz berichtet. Diese Ergebnisse und Limitationen werden in Abschnitt 5 diskutiert. Der Beitrag schließt mit einem Fazit und einem Ausblick auf künftige Entwicklungsstufen.

2. Verwandte Arbeiten zur automatisierten Sacherschließung

Aktuelle Ansätze zur automatischen Sacherschließung von Bibliotheksbeständen basieren überwiegend auf klassischen Verfahren der Textklassifikation und des Information Retrieval.¹⁶ In der Regel kommen statistische oder maschinelle Lernmethoden zum Einsatz, bei denen Modelle auf vorab annotierten Korpora trainiert werden, um anschließend neue Dokumente automatisch mit Schlagwörtern, Notationen einer Fachsystematik oder allgemein Klassen einer Taxonomie zu versehen.

Annif, ein von der Nationalbibliothek Finnlands entwickeltes Open-Source-Toolkit für die automatische Sacherschließung, hat sich in den letzten Jahren als weit verbreitetes Referenzwerkzeug etabliert.¹⁷ Annif erzielt in Evaluationen überzeugende Ergebnisse¹⁸ und wird inzwischen in einer Reihe produktiver Anwendungen eingesetzt, etwa in der Sacherschließungsmaschine (EMa) der Deutschen Nationalbibliothek (DNB)¹⁹ sowie im System AutoSE des Leibniz-Informationszentrums Wirtschaft (ZBW).²⁰ Flankierend dazu fungiert im deutschsprachigen Verbundkontext der Digitale Assistent DA-3 als Instrument zur webbasierten, maschinellen Unterstützung der intellektuelle Inhaltserschließung.²¹

Die bisherigen Erfahrungen der DNB und des ZBW zeigen zugleich, dass der Aufbau eigener Systeme zur automatisierten Sacherschließung mit erheblichem organisatorischem, technischem und finanziellem Aufwand verbunden ist. Auch an der Staatsbibliothek zu Berlin wird im Rahmen des mehrjährigen Projekts Mensch.Maschine.Kultur (2022–2025) eine eigene organisatorische und personelle Infrastruktur für KI-gestützte Inhaltsanalyse und Sacherschließung aufgebaut.²² Der langfristige Charakter dieser Vorhaben und die Notwendigkeit der Einwerbung externer Fördermittel, etwa im ZBW-Programm MetaFor (11,7 Mio. € Fördervolumen 2027–2030),²³ lassen darauf schließen, dass es sich um

16 Golub: Automated Subject Indexing. An Overview, in: *Cataloging & Classification Quarterly* 59 (8), 2021, S. 702–719, <https://doi.org/10.1080/01639374.2021.2012311>

17 Suominen, Osmo: Annif: DIY automated subject indexing using multiple algorithms, in: *LIBER Quarterly* 29 (1), 2019, S. 1–25, <https://doi.org/10.18352/lq.10285>.

18 Golub, Koraljka; Suominen, Osmo; Mohammed, Ahmed Taiye u. a.: Automated Dewey decimal classification of Swedish library metadata using Annif software, in: *Journal of Documentation* 80 (5), 2024, S. 1057–1079, <https://doi.org/10.1108/JD-01-2022-0026>.

19 Deutsche Nationalbibliothek: Start der Erschließungsmaschine EMa, 2022, https://jahresbericht.dnb.de/Webs/jahresbericht/DE/2022/Hoehepunkte/Erschliessungsmaschine/erschliessungsmaschine_node.html, Stand: 05.01.2026.

20 Kasprzik, Anna: Automating Subject Indexing at ZBW. Making Research Results Stick in Practice, in: *LIBER Quarterly* 33 (1), 2023, S. 1–18, <https://doi.org/10.53377/lq.13579>.

21 DA-3 – Digitaler Assistent zur intellektuellen Sacherschließung, 2024, <https://www.da-3.de/>, Stand: 05.01.2026.

22 Staatsbibliothek zu Berlin: Mensch.Maschine.Kultur. Künstliche Intelligenz für das digitale Kulturerbe, 2022, <https://mmk.sbb.berlin/>, Stand: 05.01.2026.

substanzielle, dauerhafte Ressourcenbindungen handelt und nicht um kurzfristige Pilotprojekte. Solche Beispiele verdeutlichen, dass eigenentwickelte Lösungen für die bibliothekarische Erschließung in der Regel mit hohen Investitions-, Betriebs- und Wartungskosten einhergehen.

Dagegen bieten großskalige KI-Sprachmodelle etablierter Anbieter wie Anthropic, Google oder OpenAI das Potenzial, generische Aufgaben über standardisierte Programmierschnittstellen (Application Programming Interfaces, APIs) und modulare Integrationsansätze effizient zu lösen, auch in der bibliothekarischen Erschließung. Beispielsweise lassen sich innerhalb kurzer Zeit spezialisierte Varianten von ChatGPT (sogenannte "(Custom) GPTs") rasch konfigurieren: Ausgestattet mit spezifischen Anweisungen, Beispielen, Regeln und angebundenen Datenquellen fungieren sie als gezielte Klassifikationsassistenten für Fachsystematiken wie die Dewey Decimal Classification (DDC).²⁴ Da hierbei weder eigene Trainingspipelines noch lokale Serverinfrastrukturen erforderlich sind, sinken die technischen Einstiegshürden erheblich. Dadurch können Bibliotheken KI-gestützte Erschließungsinstrumente vergleichsweise zügig in bestehende Workflows integrieren und produktiv erproben.

Diese Entwicklung erklärt das wachsende Interesse an Ansätzen auf der Basis großer Sprachmodelle. Diese nutzen kontextreiche semantische Repräsentationen und bleiben durch Transferlernen auch bei begrenzten Mengen kuratierter Trainingsdaten leistungsfähig. Das Potenzial solcher Modelle für die automatische Sacherschließung verdeutlichte der internationale Wettbewerb SemEval-2025 Task 5, der gezielt untersuchte, wie effektiv LLMs inhaltliche Schlagwörter für wissenschaftliche Publikationen generieren.²⁵ Als besonders erfolgreich erwiesen sich kombinierte Verfahren, bei denen LLMs für die semantische Vorverarbeitung (z. B. Schlüsselwortgenerierung, multilinguale Verarbeitung und Erzeugung zusätzlicher Trainingsdaten) genutzt und die so gewonnenen Informationen anschließend von bewährten Klassifikationsverfahren weiterverarbeitet werden.

Ein zentrales Risiko beim Einsatz generativer KI in der Sacherschließung bleiben jedoch Halluzinationen – das Vorschlagen von Notationen, die in der verwendeten Systematik nicht existieren. Die Ursache liegt darin, dass generative Modelle Wortfolgewahrscheinlichkeiten berechnen, ohne inhärent gegen geschlossene Fachsystematiken zu validieren.²⁶ Vor diesem Hintergrund berichten Dobreski und Hastings, dass frei verfügbare KI-Chatbots (u. a. ChatGPT 3.5, Gemini 1.0) bei einfachen Aufgaben zur DDC häufig ungenaue oder halluzinierte Notationen erzeugten.²⁷ Diese Ergebnisse sind jedoch im Kontext des Untersuchungszeitraums (Mai 2024) zu betrachten: Die Modelle mussten Regelwerke und Vokabulare implizit aus ihren Trainingsdaten rekonstruieren. Neuere Modellgenerationen mit ver-

23 Leibniz-Gemeinschaft, Senat: Stellungnahme zur ZBW – Leibniz-Informationszentrum Wirtschaft, Kiel/Hamburg, Senatssternungnahme 18. März 2025, Leibniz-Gemeinschaft, 2025, S. A–9, https://www.leibniz-gemeinschaft.de/fileadmin/documents/3131-2964/ZBW_-_Senatssternungnahme_18_03_2025_mit_Anlagen.pdf, Stand: 05.01.2026.

24 LibAuto: AI-Powered Dewey Decimal Classifier Tool, 2024, <https://www.libauto.in/2024/12/ai-powered-dewey-decimal-classifier-tool.html>, Stand: 05.01.2026.

25 D'Souza, Jennifer; Sadruddin, Sameer; Israel, Holger u. a.: SemEval-2025 Task 5. LLMs4Subjects - LLM-based Automated Subject Tagging for a National Technical Library's Open-Access Catalog, in: Rosenthal, Sara; Rosá, Aiala; Ghosh, Debanjan u. a. (Hg.): Proceedings of the 19th International Workshop on Semantic Evaluation (SemEval-2025), Vienna, Austria 2025, S. 2570–2583, <https://aclanthology.org/2025.semeval-1.328/>.

26 Vgl. Hübner, Conrad: ALIMA – Sacherschließung unterstützt durch große Sprachmodelle. Ein Werkstattbericht, 2025, https://blog.dini.de/EPub_FIS/2025/05/19/alima-sacherschliessung/, Stand: 05.01.2026.

27 Dobreski, Brian; Hastings, Christopher: AI Chatbots and Subject Cataloging. A Performance Test, in: Library Resources & Technical Services 69 (2), 2025, S. 1–14, <https://doi.org/10.5860/lrts.69n2.8440>.

besserten Reasoning-Fähigkeiten²⁸ sowie Ansätze, die Fachsystematiken explizit bereitstellen (z. B. Retrieval Augmented Generation²⁹), blieben unberücksichtigt.

Zudem verdeutlichte SemEval-2025 Task 5, dass die Ergebnisqualität von umfangreichen Trainingsdaten abhängt und der Ressourcenbedarf vieler Lösungen deren operative Nutzung einschränkt. Für kleine und mittelgroße Bibliotheken ist der Betrieb oder das fortlaufende Training komplexer Modell-Ensembles kaum realistisch. Erschwerend hinzu kommen häufige Qualitätsmängel in lokalen Trainingsdaten wie Inkonsistenzen, historische Verzerrungen in der Notationsvergabe (u. a. durch Personalwechsel) oder eine schiefe Klassenverteilung ("Long Tail"). Daraus ergibt sich ein Bedarf an leicht integrierbaren, rein LLM-basierten Werkzeugen, die ohne aufwändige Anpassungs- oder Trainingsphasen direkt in bestehende Arbeitsabläufe eingebunden werden können. Genau diesen Bedarf adressiert die Anwendung MILHOUSE, die als praktikabler Ansatz auf der Basis kommerzieller LLMs entwickelt wurde.

3. MILHOUSE: KI-basierte Unterstützung der inhaltlichen Erschließung

MILHOUSE (Maschinelle Indexierung mittels Lernender Heuristiken zur Optimierung der Unterstützung von SacherschließungsEntscheidungen) ist eine Webanwendung zur KI-gestützten inhaltlichen Erschließung von Büchern (vgl. Abb. 3). Das System generiert für eine Titelliste auf Basis einer gewählten Fachsystematik Vorschläge für Notationen und liefert zu jedem Vorschlag eine kurze Begründung.

28 Ein Reasoning-Modell ist ein Sprachmodell, das darauf ausgelegt ist, komplexe Probleme zu lösen, indem es logische Denkprozesse simuliert: Es zerlegt schwierige Aufgaben in kleine Schritte und prüft verschiedene Lösungswege, um das beste Ergebnis zu finden. (Vgl. VDE ITG: Große Sprachmodelle und Reasoning, 2025, <https://www.vde.com/de/itg/publikationen/itg-kurzinformationen/grosse-sprachmodelle-und-reasoning>, Stand: 15.01.2026)

29 Retrieval Augmented Generation (RAG) erweitert Sprachmodelle, indem es ihnen erlaubt, vor der Antwortgenerierung gezielt Informationen aus externen Datenquellen abzurufen. Auf diese Weise können auch spezialisierte Fachsystematiken berücksichtigt werden, auf die das Modell ursprünglich nicht trainiert wurde und gezielt für eine präzise klassifikatorische Sacherschließung genutzt werden. (Vgl. Fraunhofer IESE: Retrieval Augmented Generation (RAG): Chat mit eigenen Daten, 2024, <https://www.iese.fraunhofer.de/blog/retrieval-augmented-generation-rag/>, Stand: 15.01.2026)

Abb. 3: Screenshot der Webanwendung MILHOUSE

Die Eingabe der Titel-Identifikatoren erfolgt entweder direkt als Liste von ISBNs bzw. Pica-Produktionsnummern (PPNs) (vgl. Abb. 3 (1)) oder durch Hochladen einer Excel-Datei mit bestehenden bibliografischen Metadaten (2). Nach der Auswahl der gewünschten Klassifikation (3) kann der Prozess gestartet werden. Zur Verfügung stehen derzeit die Lokalsystematik der UB Magdeburg³⁰ sowie die Basisklassifikation (BKL). Über die erweiterten Einstellungen (4) lassen sich Parameter wie die Anzahl der Notationsvorschläge pro Titel, die maximale Hierarchietiefe der Systematik oder die Einschränkung auf spezifische Fachsegmente definieren.

Für die Verarbeitung werden die angegebenen Identifikatoren zunächst um Metadaten angereichert. Über die SRU-Schnittstelle³¹ des K10Plus-Katalogs ruft MILHOUSE bibliografische Informationen (Titel, beteiligte Personen, ISBN) ab. Optional kann zur besseren Kontextualisierung eine Kurzzusammenfassung (Synopsis) über die Google Books API³² ergänzt werden (5). Auf dieser Grundlage erzeugt die Anwendung einen strukturierten Prompt, der sowohl die Titeldaten als auch Codes und Beschreibungen der ausgewählten Fachsystematik enthält. Abb. 4 zeigt einen exemplarischen Aufbau des Prompts. Um einen konsistenten und maschinenlesbaren Output sicherzustellen, wird dem Modell ein erwartetes Ausgabeschema im JSON-Format übergeben:³³ Angefordert werden dabei für jedes Werk die PPN und ISBN, die vorgeschlagene Notation (code), deren Bezeichnung (meaning) sowie eine kurze, textuelle Begründung (explanation).

30 Universitätsbibliothek Magdeburg: Systematik der Universitätsbibliothek Magdeburg (s. Anm. 11).

31 SRU – Search/Retrieve via URL, <https://wiki.k10plus.de/display/K10PLUS/SRU>, Stand: 05.01.2026.

32 Google LLC: Google Books APIs v1 Reference, <https://developers.google.com/books>, Stand: 05.01.2026.

33 OpenAI: Structured model outputs, <https://platform.openai.com/docs/guides/structured-outputs>, Stand: 05.01.2026.

Gegeben ist die folgende Fachsystematik:

```

```{.csv}
"code","meaning"
"B.020","Naturwissenschaften: Allgemeines, Biographien, Geschichte"
"B.030","Naturwissenschaften: Allgemeines, Zeitschriften, Reihen, Serien"
"B.040","Naturwissenschaften: Allgemeines, Nachschlagewerke, Tabellenwerke"

... 2 541 weitere Notationseinträge (aus Platzgründen nicht aufgeführt) ...

"Y. 53","Künstliche Intelligenz. Intelligente Datenanalyse. Erkennende Systeme"
"Y. 54","Künstliche Intelligenz. Intelligente Datenanalyse. Sprachverarbeitung"
"Y. 55","Künstliche Intelligenz. Intelligente Datenanalyse. Neuronale Datenverarbeitung"
```

Finde für die folgenden Titel die drei passendsten Notationen und gib die Ergebnisse als
JSON-Objekt zurück:

```{.csv}
"ppn","isbn","title","author","synopsis"
"1923379771","9783658461560","Regionale Vielfalt? Neue Trends subnationaler
Parteiensysteme in Deutschland","Anna-Sophie Heinze | Uwe Ju...","Die Länderebene gehört
in der deutschen Parteienforschung nach wie vor zu den wenig beachteten
Beschäftigungsfeldern. Wenngleich sie für die Parteiorganisationen, Wahlen,
Parteiensysteme und nicht zuletzt die Pandemiebekämpfung von zentraler Bedeutung ist,
liegt der Fokus wissenschaftlicher und ö..."
"188481087X","9783031542633","Hadamard Products of Projective Varieties | Frontiers in
Mathematics","Cristiano Bocchi; Enrico Carlini","This monograph deals with the Hadamard
products of algebraic varieties..."
```

Gib für jedes Buch die PPN, ISBN, die Zuordnung (notation), die Beschreibung (meaning)
und eine Kurzbegründung als JSON zurück.

```

Abb. 4: Beispiel eines strukturierten Prompt-Setups in MILHOUSE

Die Ergebnisse samt Notationsvorschlägen und Begründungen werden tabellarisch dargestellt (6). Darüber hinaus führt das System eine Plausibilitätsprüfung durch: Nicht in der Systematik existente Einträge werden in der Spalte "Halluzination?" explizit markiert. Dies ermöglicht FR, invalide Zuordnung schnell zu erkennen. Auf eine vollständige Filterung dieser Halluzinationen wird bewusst verzichtet, um Transparenz zu gewährleisten und Daten für die kontinuierliche Weiterentwicklung des Systems zu sammeln. Die Ergebnisse können als Excel-Datei exportiert (7) und in bestehende Arbeitsabläufe integriert werden. Über ein Feedback-Formular (8) lassen sich die Vorschläge mittels einer fünfstufigen Likert-Skala (1 = "Trifft überhaupt nicht zu" bis 5 = "Trifft voll zu") sowie eines Freitextfeldes bewerten. Diese Rückmeldungen dienen der Qualitätssicherung und Feinjustierung des Systems.

Die Auswahl des Sprachmodells folgte praktischen Rahmenbedingungen (universitärer OpenAI-Zugang) und empirischen Tests: Reasoning-fähige Modelle zeigten die stabilsten Ergebnisse; o4-mini erwies sich als optimaler Kompromiss aus Qualität, Latenz und Kosten³⁴ und wurde zum Zeitpunkt

34 OpenAI: OpenAI API Pricing Documentation, <https://platform.openai.com/docs/pricing>, Stand: 05.01.2026.

der Entwicklung (Q2 2025) nur von o3 und Gemini 2.5 Pro übertroffen.³⁵ Ein kritischer Faktor ist die Batch-Größe: Praxistests zeigten, dass die Qualität bis ca. 20 Titel pro Prompt stabil bleibt, während größere Batches das Halluzinationsrisiko erhöhen.³⁶

Aus Gründen der Kostentransparenz weist MILHOUSE nach jedem Durchlauf die Nutzungsdaten aus (9), darunter Input-/Output-Token, Rechenzeit, geschätzte API-Kosten sowie das verwendete Modell. Eine beispielhafte Anfrage mit der vollständigen Lokalsystematik der UB Magdeburg und einem Batch von 20 Titeln erzeugt ca. 76 200 Input-Token und 7 100 Output-Token, was bei o4-mini (1,10 US-Dollar je 1 Mio. Input-Token bzw. 4,40 US-Dollar je 1 Mio. Output-Token) Kosten von ca. 0,08 US-Dollar verursacht. Durch Einschränkung der Fachbereiche oder Hierarchiestufen lässt sich der Tokenverbrauch weiter senken.

MILHOUSE wurde in der Programmiersprache R³⁷ entwickelt und mittels des Shiny-Pakets³⁸ als interaktive Webanwendung realisiert. Der Betrieb erfolgt intern an der UB Magdeburg auf einem Shiny Server.³⁹ Das R-Paket ellmer⁴⁰ ermöglicht eine modellagnostische Architektur, sodass die angebundenen LLMs flexibel ausgetauscht werden können, was insbesondere vor dem Hintergrund des rasanten Entwicklungstempos und der kurzlebigen Benchmarks im LLM-Bereich als Vorteil erscheint. Außerdem lässt sich die Anwendung somit unabhängig von den institutionellen Rahmenbedingungen auf andere Einrichtungen übertragen. Zudem ist die einfache Integration weiterer Systematiken möglich. Datenschutzrechtlich ist der Betrieb von MILHOUSE unbedenklich, da ausschließlich öffentlich zugängliche Katalogdaten und Fachsystematiken ohne Personenbezug verarbeitet werden.

4. Praxiserfahrungen

MILHOUSE wurde in einem sehr kurzen Entwicklungszyklus agil realisiert: Konzeption und Implementierung erfolgten durch einen einzelnen Fachreferenten parallel zum laufenden Dienstbetrieb. Bereits nach rund zwei Monaten stand eine erste funktionsfähige Version zur Verfügung, die am 26. Juni 2025 den FR der UB Magdeburg vorgestellt wurde. Da die technische Infrastruktur (Shiny-Server, Authentifizierung, Grundkonfiguration) bereits vorhanden war, beschränkte sich der initiale Einrichtungsaufwand auf wenige Stunden. Die Entwicklung umfasste im Wesentlichen die Anwendungslogik, die Modellanbindung sowie iterative Tests mit realen Titeldaten. Seit dem 14. August 2025 werden sämtliche Systemanfragen protokolliert und das Feedback der FR strukturiert erfasst, wodurch erstmals belastbare Nutzungs- und Qualitätsdaten vorliegen.

35 AI, Epoch: GPQA Diamond – AI performance on a set of Ph.D.-level science questions, <https://epoch.ai/benchmarks/gpqa-diamond>, Stand: 05.01.2026.

36 Kalai, Adam Tauman; Nachum, Ofir; Vempala, Santosh S. u. a.: Why Language Models Hallucinate, in: arXiv preprint arXiv:2509.04664, 2025, <https://arxiv.org/abs/2509.04664>; Tschochohei, Maximilian; Adams, Lisa Christine; Bressen, Keno Kyrill u. a.: KI-gestützte klinische Entscheidungsunterstützungssysteme. Herausforderungen und Potenziale, in: Bundesgesundheitsblatt – Gesundheitsforschung – Gesundheitsschutz 68, 2025, S. 872–879, <https://doi.org/10.1007/s00103-025-04092-8>.

37 R Core Team: R: A Language and Environment for Statistical Computing, Vienna, Austria 2025, <https://www.R-project.org/>, Stand: 05.01.2026.

38 Chang, Winston; Cheng, Joe; Allaire, JJ u. a.: shiny. Web Application Framework for R, 2025, <https://cran.r-project.org/web/packages/shiny/index.html>, Stand: 05.01.2026.

39 Posit Software, PBC: Shiny Server, Boston, MA 2025, <https://posit.co/products/open-source/shiny-server/>, Stand: 05.01.2026.

40 Wickham, Hadley; Vaughan, Davis; Posit Software, PBC: ellmer. A Unified Interface to Large Language Models in R, Version 0.3.0, 2025, <https://ellmer.tidyverse.org>, Stand: 05.01.2026.

Zwischen dem 14. August und dem 20. Oktober 2025 generierte MILHOUSE für insgesamt 1592 Titel Notationsvorschläge zur Lokalsystematik. Alle neun FR, die an der UB Magdeburg Sacherschließungsaufgaben wahrnehmen, setzten das Tool in diesem Zeitraum produktiv ein. Der Nutzungsumfang fiel dabei unterschiedlich aus: Während ein Fachreferat Vorschläge für 490 Titel erstellte, waren es im Referat mit der geringsten Nutzung lediglich 7 Titel. Durchschnittlich wurden pro FR 176,9 Titel bearbeitet (Standardabweichung 197,0). Diese Werte deuten sowohl auf einen klaren Bedarf, andererseits aber auch auf unterschiedliche Integrationsgrade in die jeweiligen Arbeitsabläufe hin.

In der praktischen Erprobung lag die Halluzinationsquote bei etwa 2,6 % aller Vorschläge. Positiv fiel dabei auf, dass über die Hälfte dieser fehlerhaften Vorschläge durch eine erneute Generierung korrigiert werden konnte. Solche Fehler sind somit im FR-Workflow schnell identifizierbar, was den gewählten Ansatz eines kuratierten Einsatzes (*Human in the Loop*) gegenüber einer Vollautomatisierung bestätigt.

Die wenigen weiteren beobachteten Unstimmigkeiten ließen sich überwiegend auf wiederkehrende Muster zurückführen. Zwar wurde das Fachgebiet oft korrekt identifiziert, jedoch häufig eine zu allgemeine Hierarchieebene gewählt; so wurden beispielsweise chemischen Titeln eher allgemeine Überblicksnotationen als spezifische Fachklassen zugewiesen. Echte Fehlklassifikationen traten vor allem bei Titeln auf, für die keine inhaltliche Anreicherung (z. B. über Google-Books-Zusammenfassungen) verfügbar war, was in Einzelfällen zu einer Zuordnung zu fachfremden Disziplinen führte. Zudem zeigte sich, dass fachlich korrekte Notationen zwar vorgeschlagen wurden, jedoch nicht immer die vorderen Rangplätze belegten.

Zur qualitativen Beurteilung der Vorschläge bewerteten zwei FR die Systemvorschläge anhand der in Abschnitt 3 vorgestellten fünfstufigen Likert-Skala. Eine Person bewertete 18 Titel mit einem Durchschnittswert von 4,8, eine weitere Person bewertete 87 Titel mit einem Durchschnittswert von 4,5. Auch wenn die Stichprobenzahlen begrenzt sind, deuten die Rückmeldungen sowie das ausschließlich positive qualitative Feedback darauf hin, dass die inhaltliche Qualität der Zuordnungsvorschläge von den FR mehrheitlich als sehr hoch eingeschätzt wird.

Neben der laufenden Sacherschließung kam MILHOUSE auch in einem retrospektiven Projekt zum Einsatz. Die UB Magdeburg hat aus Vorgängereinrichtungen umfangreiche Magazinbestände übernommen, die bislang keiner Notation der Lokalsystematik zugeordnet waren. Solche unsystematisierten Titel gelten als problematisch: Ohne fachliche Zuordnung fehlt die Verantwortlichkeit im Fachreferat, sie werden daher nicht in der turnusmäßigen Bestandsrevision berücksichtigt und sind für die Nutzenden schwerer auffindbar. Mithilfe von MILHOUSE konnten rund 35.000 dieser Alttitle mit einer Lokalnotation versehen werden. So ließ sich in kurzer Zeit eine Aufgabe bewältigen, die manuell kaum leistbar gewesen wäre.

Ein weiterer Aspekt der Praxiseinführung betrifft die kontinuierliche Verbesserung des Tools auf der Basis von Anwenderrückmeldungen. Neben der quantitativen Bewertung werden auch Freitextkommentare erfasst, wie etwa folgende Rückmeldung:

“Die Reihenfolge der Ergebnisse stimmt nicht mit der Reihenfolge der Eingabe überein. Das hat ein wenig für Verwirrung gesorgt.”

Solches Feedback fließt direkt in die Weiterentwicklung ein, etwa durch Anpassungen der Ergebnisdarstellung oder der Interaktion im Interface. Gleichzeitig wird die Ergebnisqualität kontinuierlich beobachtet, um bei Bedarf nachzusteuern oder auch einen Wechsel des Sprachmodells zu prüfen. Potenzial für weitere Verbesserungen liegt in der Anreicherung der Prompts um zusätzliche Metadaten wie Schlagwörter, Abstracts oder strukturierte Inhaltsverzeichnisse. Dem stehen jedoch unvollständige Daten sowie potenziell steigende Kosten durch längere Prompts gegenüber. MILHOUSE wird somit nicht als statisches Produkt verstanden, sondern als agil weiterentwickelte Arbeitsumgebung für die FR.

5. Diskussion

Die Implementierung von MILHOUSE an der UB Magdeburg zeigt, dass eine LLM-gestützte Sacherschließung in einem Human-in-the-Loop-Rahmen erhebliche Effizienzgewinne ermöglicht.⁴¹ Insbesondere für die Erschließung umfangreicher retrospektiver Bestände, deren manuelle Bearbeitung ökonomisch nicht darstellbar wäre, erweist sich der Ansatz als tragfähig. Dem stehen vernachlässigbare und transparent kalkulierbare Grenzkosten auf Modellseite gegenüber. Die hohe Qualität der Vorschläge und eine beherrschbare Halluzinationsrate (2,6 %) bestätigen die Praxistauglichkeit. Damit erfüllt MILHOUSE die Forderung, computergestützte Verfahren nicht als Autopiloten, sondern als skalierbares Assistenzsystem unter fachlicher Kontrolle zu etablieren.⁴²

Strategisch bedeutsamer als die reine Modellperformanz ist der gewählte Integrationsansatz. Während die Implementierung eigenentwickelter Klassifikations-Pipelines (z. B. auf Annif-Basis) in vielen Bibliotheken erfolgreich etabliert ist und Datenhoheit gewährleistet, ist sie zugleich mit erheblichen organisatorischen und technischen Aufwänden verbunden.⁴³ Der mit MILHOUSE verfolgte API-basierte Ansatz bietet hierzu eine ergänzende Alternative, die insbesondere für Einrichtungen mit begrenzten Entwicklungs- und Betriebsressourcen einen schnellen Einstieg erlaubt. Ein weiterer Vorteil besteht in der Modellagnostik: Kommt ein leistungsfähigeres LLM auf den Markt, kann es mit geringem Aufwand und ohne erneutes domänenspezifisches Modelltraining (*Zero-Shot*) in die bestehende Architektur integriert werden.

Damit verlagert sich der Aufwand weg vom Modelltraining und Serverbetrieb hin zu Prompt-Engineering und Prozessvalidierung. MILHOUSE demonstriert einen pragmatischen Mittelweg, der Skalierungsengpässe beseitigt, ohne eine komplexe eigene Infrastruktur vorauszusetzen oder die kuratorische Verantwortung der Fachreferate aufzugeben.

41 Vgl. Brador, Isabel: Could Artificial Intelligence Help Catalog Thousands of Digital Library Books? An Interview with Abigail Potter and Caroline Sacucci, 2024, <https://blogs.loc.gov/thesignal/2024/11/could-artificial-intelligence-help-catalog-thousands-of-digital-library-books-an-interview-with-abigail-potter-and-caroline-sacucci/>, Stand: 05.01.2026.

42 Wiesenmüller; Hinrichs: Computerunterstützte Inhaltserschließung. Bericht über einen Workshop an der UB Stuttgart – mit einem Exkurs zum neuen Inhaltserschließungskonzept der DNB; Gil-Leiva u. a.: The Abandonment of the Assignment of Subject Headings and Classification Codes in University Libraries Due to the Massive Emergence of Electronic Books; Mosqueira-Rey, Eduardo; Hernández-Pereira, Elena; Alonso-Ríos, David u. a.: Human-in-the-loop machine learning. A state of the art, in: *Artificial Intelligence Review* 56 (4), 2023, S. 3005–3054, <https://doi.org/10.1007/s10462-022-10246-w>.

43 Deutsche Nationalbibliothek: Start der Erschließungsmaschine EMa; Kasprzik: Automating Subject Indexing at ZBW: Making Research Results Stick in Practice.

Trotz der positiven Befunde bestehen Grenzen. Eine umfassende empirische Evaluation wurde bislang nicht durchgeführt: Die qualitative Bewertung stützt sich auf lediglich zwei Experten, umfasst nur begrenzte Stichproben und kann fachgebietsspezifischem Bias unterliegen. Künftige Arbeiten sollten daher systematische Evaluationsmaße sowie Messungen der Bearbeitungszeit im Realbetrieb berücksichtigen.

Jenseits technischer Aspekte erfordert der KI-Einsatz eine normative Einordnung. Kasprzik mahnt an, neben direkten finanziellen Aufwänden auch ökologische und soziale Folgekosten sowie institutionelle Risiken in die Kalkulation einzubeziehen.⁴⁴ Suominen ergänzt dies durch Leitlinien für einen "zivilisierten" Einsatz von KI, die das Gemeinwohlprinzip, die Präferenz für minimalistische Systeme ("small AI") und die kritische Reflexion von Abhängigkeiten gegenüber kommerziellen Anbietern betonen.⁴⁵

Die aktuelle Nutzung von LLMs kommerzieller Anbieter in MILHOUSE ist ein pragmatischer Startpunkt, kein Dogma. Der modulare Aufbau erlaubt einen Wechsel auf Open-Weight-Modelle bzw. lokales Hosting, sobald diese konkurrenzfähig sind. Dies verringert Risiken wie Vendor-Lock-in oder Kostenvolatilität und sichert die langfristige Balance zwischen technologischer Leistungsfähigkeit und digitaler Souveränität.

6. Fazit und Ausblick

Die Praxiserfahrungen mit MILHOUSE an der UB Magdeburg belegen, dass der strategische Hauptvorteil von LLM-APIs nicht in der reinen Klassifikationsleistung liegt, sondern vielmehr in ihrer schnellen, kostentransparenten und niedrigschwelligen Integration. Dieser Ansatz schließt die kritische Lücke zwischen wachsendem Erschließungsbedarf (insbesondere bei E-Medien und Altbeständen) und begrenzten personellen Ressourcen.

LLM-gestützte Sacherschließung liefert, unter Einbeziehung der FR zur Qualitätskontrolle, verlässlich tragfähige Ergebnisse. Halluzinationen bleiben ein Randphänomen, werden durch die integrierte Plausibilitätsprüfung früh sichtbar und lassen sich im Workflow der Fachreferate gut beherrschen. Die Vorteile werden unmittelbar sichtbar: FR werden spürbar entlastet und bislang unsystematisierte Rückstände können in vertretbarer Zeit abgearbeitet werden, beispielhaft die nachträgliche Klassifikation von rund 35000 Titeln. Dies steigert die Sichtbarkeit und Auffindbarkeit der Medien in Katalog- und Discovery-Systemen, was wiederum Bestandssteuerung und Nutzungsanalysen verbessert. Allerdings erfordert die Abhängigkeit von externen Modellen und Anbietern klare Governance-Strukturen und Kostenkontrolle.

Perspektivisch bieten sich schrittweise Ausbaustufen an, wie die Anbindung weiterer Klassifikationen (z. B. obere Hierarchiestufen der DDC) sowie Active-Learning-Verfahren, die das Feedback der FR systematisch zur Qualitätssteigerung nutzen. Ergänzend wäre eine systematische Vergleichsstudie zwi-

44 Kasprzik, Argie: Die KI(rche) im Dorf lassen, in: o-bib 12 (4), 2025, <https://doi.org/10.5282/O-BIB/6201>.

45 Suominen, Osmo: Building Civilized AI. Five Principles for Working with AI in Cultural Heritage and Libraries, 2024, https://docs.google.com/presentation/d/e/2PACX-1vRA1o11pODOJ0FmFc8dRj-xNZRU57lsxzDACiYt6d-Bdfq1ujw3gGp-SedTQnXDG0MrRg3_WA11GQS/pub, Stand: 05.01.2026.

schen einem rein LLM-basierten Ansatz und hybriden Verfahren (z. B. LLM + Annif) sinnvoll, um deren jeweilige Vor- und Nachteile in Bezug auf Qualität der Klassifikationsvorschläge, Ressourcenbedarf und Wartungsaufwand zu bewerten.

7. Literaturverzeichnis

- DA-3 – Digitaler Assistent zur intellektuellen Sacherschließung, 2024, <https://www.da-3.de/>, Stand: 05.01.2026.
- SRU – Search/Retrieve via URL, <https://wiki.k10plus.de/display/K10PLUS/SRU>, Stand: 05.01.2026.
- AI, Epoch: GPQA Diamond – AI performance on a set of Ph.D.-level science questions, <https://epoch.ai/benchmarks/gpqa-diamond>, Stand: 05.01.2026.
- Association of Research Libraries: ARL Statistics Survey Statistical Trends. Expenditure Trends in ARL Libraries, 1986–2018, Association of Research Libraries, 2019, <https://www.arl.org/arl-statistics-survey-statistical-trends/>, Stand: 05.01.2026.
- Balakrishnan, Uma; Kühn, Armin: Ergebnisse der Umfrage zur sachlichen Erschließung in GBV- und SWB-Teilnehmerbibliotheken, 2019, <https://opus.k10plus.de/frontdoor/index/index/docId/464>, Stand: 05.01.2026.
- Bosse, Sascha; Lücke, Christine; Thomas, Linda: Library Inventory Statistics Application. Eine Fallstudie über die agile Projektarbeit im Rahmen des Fachreferats, in: o-bib 9 (2), 2022, S. 1–15, <https://doi.org/10.5282/o-bib/5774>.
- Brador, Isabel: Could Artificial Intelligence Help Catalog Thousands of Digital Library Books? An Interview with Abigail Potter and Caroline Saccucci, 2024, <https://blogs.loc.gov/thesignal/2024/11/could-artificial-intelligence-help-catalog-thousands-of-digital-library-books-an-interview-with-a-bigail-potter-and-caroline-saccucci/>, Stand: 05.01.2026.
- Chang, Winston; Cheng, Joe; Allaire, JJ u. a.: shiny: Web Application Framework for R, 2025, <https://cran.r-project.org/web/packages/shiny/index.html>, Stand: 05.01.2026.
- D'Souza, Jennifer; Sadruddin, Sameer; Israel, Holger u. a.: SemEval-2025 Task 5. LLMs4Subjects - LLM-based Automated Subject Tagging for a National Technical Library's Open-Access Catalog, in: Rosenthal, Sara; Rosá, Aiala; Ghosh, Debanjan u. a. (Hg.): Proceedings of the 19th International Workshop on Semantic Evaluation (SemEval-2025), Vienna, Austria 2025, S. 2570–2583, <https://aclanthology.org/2025.semeval-1.328/>.
- Deutsche Nationalbibliothek: Start der Erschließungsmaschine EMA, 2022, https://jahresbericht.dnb.de/Webs/jahresbericht/DE/2022/Hoehepunkte/Erschliessungsmaschine/erschliessungsmaschine_node.html, Stand: 05.01.2026.
- Deutscher Bibliotheksverband e.V.: Bibliotheken 2024, 2024, https://www.bibliotheksverband.de/sites/default/files/2024-10/Bibliotheken%202024_web.pdf, Stand: 05.01.2026.
- Dobreski, Brian; Hastings, Christopher: AI Chatbots and Subject Cataloging. A Performance Test, in: Library Resources & Technical Services 69 (2), 2025, S. 1–14, <https://doi.org/10.5860/lrts.69n2.8440>.
- Franke-Maier, Michael; Kasprzik, Anna; Ledl, Andreas u. a. (Hg.): Editorial, in: Qualität in der Inhaltserschließung, Berlin, Boston 2021, S. 1–18, <https://doi.org/10.1515/9783110691597-001>.
- Fraunhofer IESE: Retrieval Augmented Generation (RAG). Chat mit eigenen Daten, 2024, <https://www.iese.fraunhofer.de/blog/retrieval-augmented-generation-rag/>, Stand: 15.01.2026.

- Gil-Leiva, Isidoro; Fujita, Mariângela Spotti Lopes; Ortuño, Pedro Díaz u. a.: The Abandonment of the Assignment of Subject Headings and Classification Codes in University Libraries Due to the Massive Emergence of Electronic Books, in: Knowledge Organization 47 (8), 2021, S. 646–667, <https://doi.org/10.5771/0943-7444-2020-8-646>.
- Golub, Koraljka: Automated Subject Indexing. An Overview, in: Cataloging & Classification Quarterly 59 (8), 2021, S. 702–719, <https://doi.org/10.1080/01639374.2021.2012311>.
- Golub, Koraljka; Suominen, Osma; Mohammed, Ahmed Taiye u. a.: Automated Dewey decimal classification of Swedish library metadata using Annif software, in: Journal of Documentation 80 (5), 2024, S. 1057–1079, <https://doi.org/10.1108/JD-01-2022-0026>.
- Google LLC: Google Books APIs v1 Reference, <https://developers.google.com/books>, Stand: 05.01.2026.
- hbz – Hochschulbibliothekszentrum des Landes Nordrhein-Westfalen: Deutsche Bibliotheksstatistik (DBS), 2025, <https://www.bibliotheksstatistik.de/>, Stand: 05.01.2026.
- Hübler, Conrad: ALIMA – Sacherschließung unterstützt durch große Sprachmodelle. Ein Werkstattbericht, 2025, https://blog.dini.de/EPub_FIS/2025/05/19/alima-sacherschliessung/, Stand: 05.01.2026.
- Kalai, Adam Tauman; Nachum, Ofir; Vempala, Santosh S. u. a.: Why Language Models Hallucinate, in: arXiv preprint arXiv:2509.04664, 2025, <https://arxiv.org/abs/2509.04664>.
- Kasprzik, Anna: Automating Subject Indexing at ZBW: Making Research Results Stick in Practice, in: LIBER Quarterly 33 (1), 2023, S. 1–18, <https://doi.org/10.53377/lq.13579>.
- Kasprzik, Anna; Seeliger, Frank: KI in wissenschaftlichen Bibliotheken, Teil 1: Handlungsfelder, große Player und die Automatisierung der Erschließung, 2022, <https://www.zbw-mediatalk.eu/de/2022/08/ki-in-wissenschaftlichen-bibliotheken-teil-1-handlungsfelder-grosse-player-und-die-automatisierung-der-erschliessung/>, Stand: 05.01.2026.
- Kasprzik, Argie: Die KI(rche) im Dorf lassen, in: o-bib 12 (4), 2025, <https://doi.org/10.5282/O-BIB/6201>.
- Kortick, Yoel: Trends in Physical and Electronic Resource Usage in U.S. Academic Libraries, in: Ex Libris Blog, 2021, <https://exlibrisgroup.com/blog/changing-priorities-trends-in-physical-and-electronic-resource-usage-in-us-academic-libraries/>, Stand: 05.01.2026.
- Leibniz-Gemeinschaft, Senat: Stellungnahme zur ZBW – Leibniz-Informationszentrum Wirtschaft, Kiel/Hamburg, Senatsstellungnahme 18. März 2025, Leibniz-Gemeinschaft, 2025, https://www.leibniz-gemeinschaft.de/fileadmin/documents/3131-2964/ZBW_-_Senatsstellungnahme_18_03_2025_mit_Anlagen.pdf, Stand: 05.01.2026.
- LibAuto: AI-Powered Dewey Decimal Classifier Tool, 2024, <https://www.libauto.in/2024/12/ai-powered-dewey-decimal-classifier-tool.html>, Stand: 05.01.2026.
- Mosqueira-Rey, Eduardo; Hernández-Pereira, Elena; Alonso-Ríos, David u. a.: Human-in-the-loop machine learning. A state of the art, in: Artificial Intelligence Review 56 (4), 2023, S. 3005–3054, <https://doi.org/10.1007/s10462-022-10246-w>.
- OpenAI: OpenAI API Pricing Documentation, <https://platform.openai.com/docs/pricing>, Stand: 05.01.2026.
- OpenAI: Structured model outputs, <https://platform.openai.com/docs/guides/structured-outputs>, Stand: 05.01.2026.
- Posit Software, PBC: Shiny Server, Boston, MA 2025, <https://posit.co/products/open-source/shiny-server/>, Stand: 05.01.2026.

- R Core Team: R: A Language and Environment for Statistical Computing, Vienna, Austria 2025, <https://www.R-project.org/>, Stand: 05.01.2026.
- Staatsbibliothek zu Berlin: Mensch.Maschine.Kultur. Künstliche Intelligenz für das digitale Kulturerbe, 2022, <https://mmk.sbb.berlin/>, Stand: 05.01.2026.
- Sühl-Strohmenger, Wilfried; Tappenbeck, Inka (Hg.): Praxishandbuch Wissenschaftliche Bibliothekar:innen, Berlin, Boston 2024, <https://doi.org/10.1515/9783110790375>.
- Suominen, Osmo: Building Civilized AI. Five Principles for Working with AI in Cultural Heritage and Libraries, 2024, https://docs.google.com/presentation/d/e/2PACX-1vRA1o11pODoJ0FmFc8dRj-xNZRU57lsxzDACKiYt6d-Bdfq1ujw3GpSedTQnXDG0MrRg3_WAI1GQS/pub, Stand: 05.01.2026.
- Suominen, Osmo: Annif: DIY automated subject indexing using multiple algorithms, in: LIBER Quarterly 29 (1), 2019, S. 1–25, <https://doi.org/10.18352/lq.10285>.
- Tschochohei, Maximilian; Adams, Lisa Christine; Bressemer, Keno Kyrill u. a.: KI-gestützte klinische Entscheidungsunterstützungssysteme. Herausforderungen und Potenziale, in: Bundesgesundheitsblatt – Gesundheitsforschung – Gesundheitsschutz 68, 2025, S. 872–879, <https://doi.org/10.1007/s00103-025-04092-8>.
- Universitätsbibliothek Magdeburg: Systematik der Universitätsbibliothek Magdeburg, 2025, https://www.ub.ovgu.de/ub_media/Literatursuche/Systematik.pdf, Stand: 05.01.2026.
- VDE ITG: Große Sprachmodelle und Reasoning, 2025, <https://www.vde.com/de/itg/publikationen/itg-kurzinformationen/grosse-sprachmodelle-und-reasoning>, Stand: 15.01.2026.
- Wickham, Hadley; Vaughan, Davis; Posit Software, PBC: ellmer. A Unified Interface to Large Language Models in R, Version 0.3.0, 2025, <https://ellmer.tidyverse.org>, Stand: 05.01.2026.
- Wiesenmüller, Heidrun; Hinrichs, Imma: Computerunterstützte Inhaltserschließung. Bericht über einen Workshop an der UB Stuttgart – mit einem Exkurs zum neuen Inhaltserschließungskonzept der DNB, in: o-bib 4 (3), 2017, S. 94–105, <https://doi.org/10.5282/o-bib/2017H3S94-105>.