# KI-Technologien in Bibliotheken zwischen Wunsch und Realität

**Stephan Holländer**

Vor anderthalb Jahren war in der New York Times[1] zu lesen, dass die Bibliothek am Hauptsitz von OpenAi Inc. in San Francisco ein beliebter Aufenthaltsort der Mitarbeitenden sei, der auf Anweisung von Sam Altman ganz ohne Hi-Tech-Technologie auskomme. Der Raum sei in warmen Holztönen gehalten und dem Rose-Leseraum der New York Public Library nachempfunden. Das geneigte Publikum fragt sich, wie es möglich ist, dass in der Bibliothek im Unternehmen, das mit ChatGPT im November 2022 eine Sensation in der Welt der Internet-Argonauten auslöste, keine Künstliche Intelligenz (KI) anzutreffen ist?

Wer sich in den Bibliotheken Deutschlands, Österreichs und der Schweiz (DACH) umschaut, stellt keine großen Unterschiede fest. Zwar verfügt die Mehrheit der Bibliotheken über ein Bibliothekssystem, Zugang zum Internet, aber KI bleibt vom bibliothekarischen Arbeitsumfeld noch weitgehend unsichtbar. Es gibt Ausnahmen. So haben die Nationalbibliotheken Deutschlands und der Schweiz sowie Hochschulbibliotheken erste Projekte mit KI unternommen, von denen noch zu sprechen sein wird, aber in der großen Mehrheit der Bibliotheken gibt es noch keine Bibliothekssysteme mit KI.

Wer sich hingegen in der gesamteuropäischen Bibliothekslandschaft umschaut, vermeint ein Nord-Südgefälle zu erkennen. Die Bibliotheken in den skandinavischen und baltischen Ländern scheinen sich dieser neuen Technologie früher und tatkräftiger angenommen zu haben als Bibliotheken in anderen Ländern. So hat die Finnische Nationalbibliothek in ihrem Labor die Software Annif[2] für das halbautomatische Indexieren auf der Basis von KI entwickelt und diese schlägt Themenüberschriften für Texte aus ihrem Vokabular vor.

Die Schwedische Königliche Nationalbibliothek hat diese Software übernommen und trainiert in ihrem Entwicklungslabor große Sprachmodelle wie BERT[3] und Spracherkennungsmodelle anhand schwedischer Daten aus den Sammlungen der Bibliothek. Die Norwegische Nationalbibliothek hat das gleiche für die norwegische Sprache entwickelt. Die Königliche Nationalbibliothek Dänemarks hat eine innovative KI-Suche[4] eingeführt, die dabei hilft, akademische und forschungsbezogene Fragen mithilfe der Ressourcen der Bibliothek zu beantworten. Ähnliches gibt es auch von den Nationalbibliotheken der baltischen Staaten zu berichten.

Auch die Königliche Bibliothek der Niederlande verwendet die KI zur Optimierung ihrer Suchtechnologie und zur Verwaltung von Millionen von digitalen Büchern, Zeitungen, Zeitschriften und mehr. Die Königliche Bibliothek Belgiens (KBR) hat ein Werkzeug entwickelt, das KI zur Verarbeitung und Validierung bibliografischer Daten nutzt[5]. Die Bibliothek begann zu diesem Zweck mit der Microsoft Power Plattform zu arbeiten.

Die Nationalbibliothek Frankreichs hat über 40 Millionen Dokumente unterschiedlicher Art (Bücher, Manuskripte, Karten, Partituren, Objekte, Tondokumente und Videospiele) in ihren Magazinen und fast 9 Millionen digitalisierte und in Gallica zugängliche Dokumente, sowie 40 Milliarden im Internetarchiv gesammelte URLs, die bei der Suche durch KI besser auffindbar gemacht werden[6].

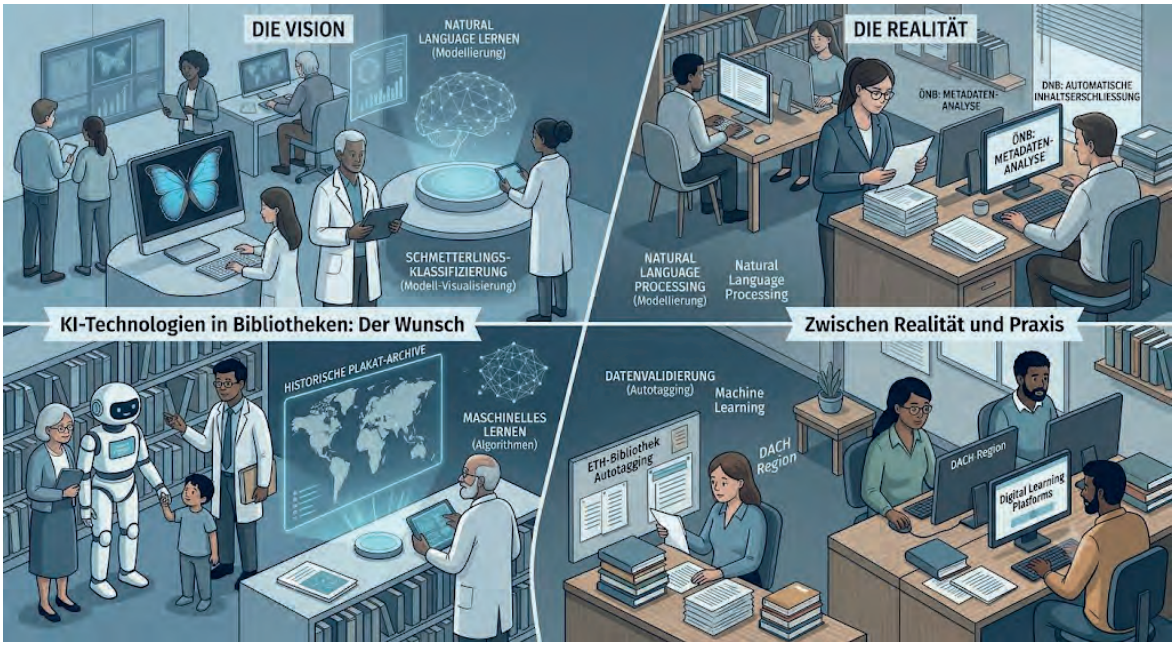
Eine Umfrage in italienischen Bibliotheken[7] zeigt, dass die Einführung künstlicher Intelligenz durch eine Reihe interner und externer Faktoren beeinflusst wird. Das persönliche Interesse der Bibliothekare und ihre selbstständige Weiterbildung im Bereich Technologie lässt keine große Offenheit für die Einführung von KI vermuten. Negative Erfahrungen mit unzulänglichen automatischen Antworten bei Chatbots in der Vergangenheit haben möglicherweise zur Skepsis beigetragen.

Beim Blick auf die Bibliotheken Deutschlands, Österreichs und der Schweiz ergibt sich wie eingangs erwähnt ein unterschiedliches Bild:

## Deutschland

Die Deutsche Nationalbibliothek hat in einem vierjährigen Projekt Technologien des Maschinellen Lernens und

1 New York Times Inside OpenAI's library https://www.nytimes.com/2024/05/15/technology/openai-library-office.html.
2 Dies ist Software zur automatischen Schlagwort-Indexierung und Klassifikation, siehe https://annif.org/.
3 Das Sprachmodell BERT ist beschrieben unter https://crl.acrl.org/index.php/crl/article/view/25748.
4 https://bib.kglakademi.dk/discovery/search?vid=45KBDK_KADK:KGLAKADEMI&lang=da.
5 https://www.vaia.be/en/blog/ai-in-de-bibliotheek-boeken-catalogiseren-met-behulp-van-ai.
6 https://actualitte.com/article/125182/technologie/culture-et-intelligence-artificielle-la-france-s-en-remet-a-microsoft.
7 https://www.bibliotecheoggi.it/it/articolo/16125/ia-in-biblioteca-quale-impatto-i-risultati.




*Copyright: Illustration: Erstellt mit Google Imagen / Gemini*

des Natural Language Processing für die inhaltliche Erschließung von Medien erforscht. Die Grundlage des maschinellen Analyseprozesses sind dabei unstrukturierte Texte, wie Volltexte oder Inhaltsverzeichnisse der im Bestand befindlichen Medien sowie auch Metadaten, sofern diese vorhanden sind. Tagungen in Leipzig im Oktober letzten Jahres und in Frankfurt im Januar dieses Jahres dienen dem Erfahrungsaustausch auch zu Erschließung, Computervision und Einsatz im GLAM-Bereich, aber auch zu Daten, Prozessen, Modellen und Hardware für die automatische Inhaltserschließung.

Ein weiterer Schritt zum Zugang zu Generativer Intelligenz wurde an deutschen Hochschulen und Universitäten gemacht[8]. Die Universitäten und Fachhochschulen sind aber noch weit von einer breiten Grundversorgung ihrer knapp 3,5 Millionen Forschenden und Studierenden in Deutschland mit generativer KI entfernt. Drei einzelne Projekte gibt es in Niedersachsen[9]. Eines des KI-Servicezentrums für sensible und kritische Infrastrukturen (KISSKI) in Göttingen, eines der Kunsthochschule HAWK in Hildesheim, Holzminden und Göttingen und ein privates Modell namens Educa-AI[10].

Die Lehre an Universitäten wird auch durch die Integration von interaktiven digitalen Lernplattformen umgekrempelt. Als Beispiel sei das Angebot von Lernplattformen für Medizin angeführt, wie zum Beispiel AMBOSS[11] oder via medici[12], aber auch Plattformen für andere Studienfächer.

## Österreich

Die Forscher an der Universität Innsbruck haben beispielsweise mithilfe von Supercomputern KI-Modelle auf das Erkennen von Schmetterlingsarten trainiert[13]. Basis war ein Datensatz mit über 540.000 Fotos aus dem Citizen-Science-Projekt „Schmetterlinge Österreichs". Die Forscher nutzten dazu diesen Datensatz, den über 25.000 Freiwillige im Rahmen eines Citizen-Science-Projekts binnen zehn Jahren zusammengetragen hatten. Nach dem Training der KI-Modelle bestimmte die Anwendung 97 Prozent der Schmetterlingsarten richtig. Das Projekt zeigt deutlich, welches Potenzial in der Integration von KI und Citizen Science für die Wissenschaft steckt.

## Schweiz

Die Schweizerische Nationalbibliothek (NB) und die ETH-Bibliothek Zürich haben KI-Projekte im Bereich Computervision umgesetzt. Die NB hat im Rahmen eines Research Fellow-Projekts des Bundes ihre Sammlung historischer Plakate mittels KI-Technologie erschlossen[14]. Auch die ETH-Bibliothek Zürich hat sich mit den Fragen der Erschließung von Bildern auseinandergesetzt. Das Bildarchiv der ETH hat mit seinem E-Pics-Angebot ein

8 Siehe https://www.forschung-und-lehre.de/lehre/die-ki-pioniere-sitzen-in-niedersachsen-6618.

9 https://www.forschung-und-lehre.de/lehre/die-ki-pioniere-sitzen-in-niedersachsen-6618.

10 eine Ausgründung aus den Hochschulen Göttingen und Clausthal-Zellerfeld.

11 AMBOSS ist eine interaktive Lernplattform für das Medizinstudium und fasst in über 1300 Kapiteln das gesamte vorklinische und klinische Wissen zusammen. 2012 in Köln gegründet, heute international aktiv.

12 Ein digitales Angebot des Thieme Verlags mit großem Nutzerzuwachs um klassische Papierprodukte mit einer digitalen Erweiterung anzubieten.

13 Service: Barkmann, F., Lindner, A., Würflinger, R. et al. „Machine learning training data: over 500,000 images of butterflies and moths (Lepidoptera) with species labels". Link zur Originalpublikation: https://www.nature.com/articles/s41597-025-05708-z.

14 b.i.t.online 27, Heft 4/2024.

erstes Projekt unternommen[15]. Das E-Pics-Angebot beinhaltet zwei übernommene Fotoarchive, einerseits das Bildarchiv der ehemaligen Swissair und andererseits der Bestand der Pressefoto-Agentur Comet.

Die ETH-Bibliothek in Zürich nutzt seit 2021 KI-Technologien. Das Bildarchiv der ETH-Bibliothek verwendet KI zur automatisierten Bildklassifikation («Autotagging») als Ergänzung zur intellektuellen Verschlagwortung für seine Plattform der verschiedenen Bildkataloge. Über eine Million Bilder unter anderem aus Beständen der ETH, der Fotoagentur Comet und der ehemaligen Swissair[16] wurden mit dem Computer-Vision-Modell «General» von Clarifai und intellektueller Indexierung mit Tags versehen. Gewisse Metadatenfelder werden zudem mit maschineller Übersetzung ins Englische übersetzt. Bei beiden Anwendungen hat die Qualitätskontrolle gezeigt, dass dies unterschiedlich gut gelingt. Die maschinelle Übersetzung gelingt bis auf einzelne Übersetzungsfehler gut. Beim Autotagging ergaben sich jedoch unterschiedliche Erkennungsraten, unterschiedliche Resultate, die von nahezu hundertprozentiger Genauigkeit bis hin zu Übersetzungsfehlern und falschen Zuordnungen reichten[17]. Hier wurde manuell korrigiert.

Überdies bietet die ETH-Bibliothek KI-Werkzeuge, Workshops und Selbstlernkurse an[18], die bei der Literaturrecherche, beim Abfassen wissenschaftlicher Arbeiten und im Lehrbetrieb hilfreich sind. Lizenzierte Werkzeuge im Angebot sind beispielsweise Scopus AI, Scite und Connected Papers.

Die Swiss Library Service Plattform (SLSP) testet und analysiert KI-Tools in der Produktpalette, die sie in Swisscovery anbietet. Damit werden die Services von SLSP optimal begleitet. Durch gezielte KI-Vermittlung in SLSP sorgt die neue Technologie für eine gesteigerte Effektivität in der täglichen Arbeit. Zudem evaluiert SLSP mit ihren Partnern das Potential für weitere, innovative Implementationen von KI in allen Bereichen der Datenverarbeitung, Datennutzung und Prozesssteuerung.

Einige Universitätsbibliotheken bieten zusätzliche KI-Recherche-Werkzeuge auf Plattformen von externen Anbietern an[19]. Der Schweizer Bibliotheksverband Bibliosuisse hat im vergangenen Jahr verschiedene Weiterbildungsprogramme zu KI für Mitarbeitende aus öffentlichen und wissenschaftlichen Bibliotheken zur Funktionsweise von KI in Bibliotheken durchgeführt, die gut besucht waren und für die auch teilweise Wartelisten geführt werden mussten. Das Angebot umfasste nicht nur technische, sondern beispielsweise auch ethische Themen.

Die KI-Technologie verändert sich weiterhin in schnellem Tempo. Am Horizont zeichnen sich neue Trends und Weiterentwicklungen ab, die auch bereits installierte KI-Anwendungen in Bibliotheken verändern werden. Retrieval Augmented Generation (RAG) lautet das entsprechende Stichwort.

Retrieval Augmented Generation ist eine Technik zur Verarbeitung natürlicher Sprache (Natural Language Processing NLP), die die Stärken von abfragebasierten und generativen Modellen der KI kombiniert. RAG-KI kann genaue Ergebnisse liefern, die bereits vorhandenes Wissen optimal nutzen. Sie kann dieses Wissen aber auch verarbeiten und konsolidieren, um einzigartige, kontextbezogene Antworten, Anweisungen oder Erklärungen in menschenähnlicher Sprache zu erstellen, anstatt nur die abgefragten Daten zusammenzufassen. Die RAG-KI-Technologie erweitert das Potenzial generativer KI und verspricht[20]:

- Zugriff auf Informationen, die möglicherweise aktueller sind als die Daten, die zum Trainieren des LLM verwendet werden.
- Kontinuierliche Aktualisierung der Daten im Wissensspeicher der RAG, ohne dass nennenswerte Kosten entstehen.
- Kontextbezogenere Daten als diejenigen in einem generalisierten LLM im Wissensrepository der RAG.
- Identifizierung der Quelle der Informationen in der Vektordatenbank des RAG und, da die Datenquellen bekannt sind, Korrektur oder Löschung von fehlerhaften Angaben im RAG.

Setzt sich diese Technologie auch in Bibliotheken durch, wird es interessant sein, zu sehen, wie sich dadurch die Arbeits- und Rechercheprozesse in Bibliotheken verändern werden.

**Stephan Holländer**
Basel
stephan@stephan-hollaender.ch

15 b.i.t.online 27, Heft 6/2024 und Nicole Graf, Alles unter Kontrolle? KI im Einsatz im Bildarchiv der ETH-Bibliothek in o-bib https://www.o-bib.de/bib/article/view/5921.

16 https://e-pics3.ethz.ch/de/home/

17 https://www.research-collection.ethz.ch/entities/publication/34182aa4-30a9-41b4-b5b6-7b347f5ca91f

18 https://library.ethz.ch/ki-angebot.html

19 So beispielweise die Universitätsbibliothek Basel auf den-Recherche-Plattformen Consensus, Scite, Elicit, Scopus AI, Web of Science, Research Rabbit und Open Knowledge Maps, sowie Kursen für Nutzerinnen und Nutzer der Bibliothek.

20 What is RAG (Retrieval-Augmented Generation)?, siehe https://aws.amazon.com/what-is/retrieval-augmented-generation/.