

Ein Zugang zu Editionen mithilfe neuer Technologien:

Entwicklung eines Chatbots mit Wissensgraphen, LLMs und
Retrieval-Augmented Generation (RAG)
als Prototyp für die Editionswissenschaft

Till Bäumker

19. Februar 2025

Inhaltsverzeichnis

1	Einleitung und Kontext der Arbeit	1
1.1	Was sind Editionen?	1
1.2	Digitale Editionen: Chancen und Herausforderungen	2
1.3	Innovationen durch Chatbots, Wissensgraphen, RAG und LLMs	3
1.4	Ziel der Arbeit: Entwicklung eines Chatbots	4
2	Deutsche Losbücher im Allgemeinen und das <i>Mainzer Kartenlosbuch</i> als Grundlage für den Wissensgraphen	5
2.1	Allgemeine Merkmale deutscher Losbücher	5
2.2	Grundlage für den Wissensgraphen: Das <i>Mainzer Kartenlosbuch</i>	6
3	Technologische Grundlagen und Konzeption des Chatbots	7
3.1	Technologische Grundlagen	7
3.1.1	Wissensgraphen	7
3.1.2	Retrieval-Augmented Generation (RAG): Erweiterung von Sprachmodellen durch externe Wissensquellen	8
3.2	Konzeption des Chatbots	8
3.2.1	Funktionsweise des Zusammenspiels von Wissensgraphen und LLM im Chatbot	8
3.2.2	Benutzeroberfläche	10
3.2.3	Erlebbarkeit des <i>Mainzer Kartenlosbuchs</i> : Der Modus <i>Losbuch spielen</i>	10
4	Technische Implementierung des Chatbots	10
4.1	Preprocessing: Datenaufbereitung für den Wissensgraphen	10
4.2	Integration von Wissensgraph und LLM	12
5	Evaluation des Chatbots	14
5.1	Korrektheit der Antworten	14
5.2	Natürliche Sprachinteraktion	16
5.3	Funktionstest Modus <i>Losbuch spielen</i>	16
5.4	Zusammenfassung der Ergebnisse	16
6	Herausforderungen bei Tests und Umsetzung des Chatbots	17
6.1	Einschränkung der Testergebnisse	17
6.2	Einschränkung der Interpretierbarkeit und Erklärbarkeit	18
6.3	Einschränkung der Übertragbarkeit	19
7	Potenziale und Grenzen des entwickelten Chatbots als Prototyp für die Editionswissenschaft	19

1 Einleitung und Kontext der Arbeit

Diese Arbeit untersucht die Entwicklung eines Chatbots, der exemplarisch einen digitalen Zugang zu einer Edition des *Mainzer Kartenlosbuches* bereitstellt. Es handelt sich um ein humoristisches Werk der Weissagungsliteratur, das um 1510 in Mainz von Johann Schöffer gedruckt wurde. In dieser Arbeit wird die Edition von Matthias Däumer verwendet (Däumer 2021). Der Chatbot wird technisch mithilfe von Wissensgraphen, Retrieval-Augmented Generation (RAG) und LLMs (Large Language Models) realisiert und ermöglicht so einen interaktiven Zugang zu historischen Texten über natürliche Sprache. Ziel der Arbeit ist es, durch die Konzeption, Umsetzung und Evaluation die Potenziale und Grenzen dieser Technologie für die Editionswissenschaft aufzuzeigen.

Anmerkung: Es wird von LLMs im Plural gesprochen, da bei der Erstellung des Chatbots an mehreren Stellen verschiedene LLMs zum Einsatz kamen.

Um den Kontext dieser Arbeit zu verstehen, werden zunächst grundlegende Konzepte und Entwicklungen in der Editionswissenschaft beschrieben.

1.1 Was sind Editionen?

In dieser Arbeit wird der Editions begriff auf Grundlage einer allgemeinen Definition von Sahle verwendet. Sahle beschreibt eine Edition als die „erschließende Wiedergabe historischer Dokumente“. „Dabei meint »historisch« eine zu überbrückende Distanz zwischen unserem heutigen Verständnis und dem Erklärungsbedarf zur historischen Wirklichkeit.“ Der Begriff »Dokumente« bezieht sich auf beschreibbare physische Gegenstände und die dazugehörigen Texte. Wichtig für eine Edition ist die »Wiedergabe«, welche die „vollständige Repräsentation des Editionsgegenstandes [...] auf bildlicher und/oder (meistens) textlicher Ebene“ umfasst. Eine bloße Beschreibung oder ein Katalog erfüllen diese Anforderung nicht.

Darüber hinaus ist die »Erschließung« ein zentraler Bestandteil der Edition und umfasst „alle kritischen Operationen“, wie die Auswahl der Objekte, deren Beschreibung, Entscheidungen bei der Transkription, die Identifikation von referenzierten Objekten im Text, historische Sachanmerkungen sowie die Textkritik. Eine Reproduktion ohne eine kritische Auseinandersetzung wird nach Sahle nicht als Edition betrachtet (Sahle 2017, S. 239).

Ausgehend davon gibt es verschiedene Ansätze in der Editions wissenschaft, die sich je nach Zielstellung unterscheiden. Während einige Editionen den Fokus auf die Rekonstruktion eines idealisierten Urtextes legen, betonen andere die Materialität der Dokumente oder den Entstehungsprozess eines Textes. Beispielsweise wird in der genetischen Edition der Schaffensprozess durch die Darstellung verschiedener Textstufen sichtbar gemacht, während in der diplomatischen Edition die genaue Form und Gestaltung der überlieferten Dokumente

detailgetreu wiedergegeben wird (Sahle 2017, S. 237 f.).

1.2 Digitale Editionen: Chancen und Herausforderungen

Wie Jan Hess feststellt, ist „das Digitale in der Editionswissenschaft inzwischen allgegenwärtig. Digitale bzw. Hybrid-Editionen sind nicht die Ausnahme, sondern eher die Regel geworden“ (Hess 2024, S. 147). Diese Aussage unterstreicht die Bedeutung digitaler Technologien in der Editionswissenschaft.

Für Sahle bieten digitale Editionen der Editionswissenschaft neue Perspektiven, da sie moderne Technologien integrieren und über die bloße Digitalisierung von Druckausgaben hinausgehen und damit einem digitalen Paradigma folgen (Sahle 2017, S. 240 ff.). Im Gegensatz zu gedruckten Editionen, die häufig nur eine eingeschränkte Sicht auf die Überlieferung bieten, ermöglichen digitale Editionen eine flexiblere und umfassendere Darstellung. In der digitalen Umgebung wird die Druckausgabe zu einer von vielen möglichen Präsentationsformen (Sahle 2013, S. 149). Dies zeigt den Prozess der Transmedialisierung: Statt einer festen medialen Form, wie sie die Druckkultur vorgibt, wird der Text als flexible, digital nutzbare Datenstruktur erfasst (Sahle 2017, S. 241).

Während also gedruckte Werke „freeze in time“, bleiben digitale Editionen dynamisch und offen für zum Beispiel Ergänzungen wie Annotationen, Kommentare und Verknüpfungen zu weiteren Materialien (van Zundert und Boot 2011, S. 143). Sie sind keine statischen Produkte, sondern fortlaufende Projekte, die kontinuierlich aktualisiert und erweitert werden können. Darstellungsformen wie Hyperlinks sowie multimediale, multimodale und interaktive Inhalte, die in der Druckkultur nur angedeutet werden können, lassen sich in digitalen Medien umsetzen (Sahle 2013, S. 149 f., 161).

Auch Robinson betont, dass digitale Editionen keine endgültigen, unveränderlichen Texte mehr vorgeben. Stattdessen ermöglichen sie eine offene, interaktive Auseinandersetzung, bei der Nutzer:innen aktiv mit den Inhalten arbeiten können:

„To apply the critical arts and sciences to editing in the electronic domain is not to return to the days when an editor established and printed a text and said, effectively: take it or leave it (Robinson 2002, S. 145).“

Elena Pierazzo beschreibt das Potenzial der Darstellung von Texten in digitalen Medien, indem sie auf multimodale Formate wie die App von Touch Press verweist, die literarische Werke durch interaktive Inhalte neu erschließt. Laut Pierazzo ermöglichen solche Formate eine tiefere Auseinandersetzung mit Texten. Dies zeigt, wie digitale Editionen das Potenzial bieten, vielfältige Zugänge zu Texten zu ermöglichen und dadurch eine breitere Leserschaft anzusprechen (Pierazzo 2016, S. 54 f.).

Dabei ist es wichtig, zwischen digitalisierten und digitalen Editionen zu unterscheiden. An

dieser Stelle beziehe ich mich auf die Unterscheidung von Sahle. Während digitalisierte Editionen lediglich bestehende Druckausgaben in digitale Form übertragen, schöpfen digitale Editionen die Potenziale digitaler Technologien voll aus (Sahle 2013, S. 152). Sie bieten dynamisch generierte Inhalte, interaktive Elemente und multimediale Verknüpfungen, die in der Druckkultur nicht darstellbar sind. Eine digitale Edition ist „nicht ohne wesentliche Informations- und Funktionsverluste in eine typografische Form“ übertragbar (Sahle 2013, S. 149). Sahle wirft in diesem Zusammenhang die Frage auf, ob eine digitalisierte Edition durch erweiterte Funktionen und Verknüpfungen den Status einer digitalen Edition erreichen kann (Sahle 2013, S. 152).

Digitale Editionen eröffnen zahlreiche Möglichkeiten, bringen jedoch auch Herausforderungen mit sich. Da digitale Medien durch ihre „complex, multimedia, networked content“ interaktiv und anpassungsfähig sind, sind „more sophisticated browse and search functions to access all the material and information of an edition“ wichtig, um den Zugang und die Nutzungsmöglichkeiten zu optimieren (Sahle 2016, S. 30). So kann sichergestellt werden, dass Nutzer:innen trotz der wachsenden Komplexität effizient auf Inhalte zugreifen und diese gezielt erschließen können.

Kalmer et al. bestätigen dies und weisen darauf hin, dass fehlende Verlinkungen sowie das Fehlen standardisierter Funktionen wie Suchleisten die Nutzung digitaler Editionen erheblich erschweren können (Pontes Trabula 2024, S. 241).

Eine weitere Herausforderung digitaler Editionen ist ihre disziplinübergreifende Nutzbarkeit. Da sie in verschiedenen Fachbereichen Anwendung finden, sollten sie breit anschlussfähig sein. Sahle fordert daher, dass digitale Editionen „as widely accessible and usable as possible for all interested disciplines“ sein sollen. Dies ist besonders wichtig, da ihre Erstellung oft mit erheblichem zeitlichen und finanziellen Aufwand verbunden ist und grundlegende Arbeiten wie digitale Repräsentationen und Transkriptionen meist nur einmal erfolgen. Um eine doppelte Bearbeitung derselben Materialien zu vermeiden und die nachhaltige Nutzung zu sichern, sollten digitale Editionen so konzipiert sein, dass sie über disziplinäre Grenzen hinweg genutzt werden können (Sahle 2016, S. 21 ff.).

1.3 Innovationen durch Chatbots, Wissensgraphen, RAG und LLMs

Digitale Editionen basieren auf Extensible Markup Language (XML), einer Auszeichnungssprache zur strukturierten Erfassung von Texten. Es gibt verschiedene XML-Dialekte, wobei die Empfehlungen der *Text Encoding Initiative* (TEI) als Leitstandard für die Erstellung digitaler Editionen gelten. Mittlerweile werden digitale Editionen in der Regel als Webanwendungen veröffentlicht, die HTML (Hypertext Markup Language) zur Darstellung, CSS (Cascading Style Sheets) für das Layout und JavaScript für interaktive Funktionen

nutzen (Sahle 2017, S. 245) (Kuczera 2016, S. 1). Neuere, graphbasierte Ansätze bieten ebenfalls vielversprechende Möglichkeiten, komplexe Verknüpfungen zwischen Entitäten und deren Beziehungen darzustellen und stellen eine Alternative zu XML dar (Kuczera 2016, S. 7).

Um die im vorherigen Kapitel genannten Herausforderungen (erweiterte Such- und Browsing-Funktionen, Zugänglichkeit) der digitalen Editionswissenschaft zu adressieren, könnten interaktive Lösungen entwickelt werden, die sowohl die Zugänglichkeit als auch die Funktionalität erweitern. Eine mögliche Lösung wären Chatbots, die Wissensgraphen mit Retrieval-Augmented Generation (RAG) und LLMs kombinieren. Diese Technologien ermöglichen das Durchsuchen strukturierter Daten mithilfe natürlicher Sprache und die Generierung verständlicher Antworten (Wilcock 2024, S. 3). Gleichzeitig könnten Chatbots die Zielgruppe digitaler Editionen erweitern, da sie auch für Nutzer:innen ohne technische oder editionswissenschaftliche Vorkenntnisse zugänglich sind.

Das Potenzial solcher Technologien zeigt sich bereits in anderen Anwendungsbereichen. Im Paper von Bhat et al. wird ein Chatbot vorgestellt, der personalisierte und kontextbezogene Informationen bereitstellt, etwa Öffnungszeiten oder Menüempfehlungen im Restaurantbereich. Dabei nutzt der Chatbot *Neo4j*-Wissensgraphen, um relevante Daten effizient zu organisieren, während der RAG-Ansatz präzise Antworten in Echtzeit ermöglicht (Bhat u. a. 2024, S. 8 ff.). Solche RAG-basierten Chatbots haben sich insbesondere in domänenspezifischen Anwendungen als leistungstark erwiesen (Kulkarni u. a. 2024).

1.4 Ziel der Arbeit: Entwicklung eines Chatbots

Diese Arbeit untersucht die Entwicklung eines prototypischen Chatbots, der Wissensgraphen mit RAG und LLMs kombiniert. Ziel der Arbeit ist es, durch die Konzeption, Implementierung und Evaluierung eines Prototypen die technische Umsetzbarkeit sowie die Potenziale solcher Systeme für die Editionswissenschaft zu analysieren. Dabei wird geprüft, ob der Chatbot die beschriebenen Herausforderungen digitaler Editionen – wie disziplinübergreifender Zugang und innovative Suchfunktionen – erfolgreich adressieren kann. Als Grundlage dient die Edition des *Mainzer Kartenlosbuchs*, die 1510 von Johann Schöffer gedruckt und von Matthias Däumer editiert wurde. Diese Edition bietet durch ihre umfassenden Metadaten und die digitale Verfügbarkeit eine gute Basis für die Untersuchung (Däumer 2021, S. 87).

Dabei ist jedoch zu betonen, dass die Aussagekraft dieses Projekts begrenzt ist, da es sich lediglich um einen Prototypen handelt und keine Nutzerstudien oder umfassenden Usability-Tests durchgeführt wurden.

Um dieses Vorhaben zu adressieren, werden im Verlauf der Arbeit folgende Schritte unternommen:

1. **Konzeption des Chatbots:** Entwicklung eines systematischen Designs, das auf einem Wissensgraphen und Retrieval-Augmented Generation basiert.
2. **Implementierung:** Technische Realisierung des Prototyps zur Demonstration der Machbarkeit eines solchen Systems.
3. **Evaluierung:** Funktionalitätstestung des Chatbots, um die technische Leistungsfähigkeit zu überprüfen. Dabei wird die Effektivität des Systems in Bezug auf die Adressierung der Herausforderungen digitaler Editionen analysiert.
4. **Einschränkungen:** Reflexion über die Grenzen der dargestellten Umsetzung und der Testergebnisse.

Es stellt sich auch die Frage, welchen Status ein solcher Chatbot innerhalb der Editions wissenschaft einnimmt: Handelt es sich um eine eigene digitale Edition oder lediglich um eine digitalisierte Edition? Diese Frage kann leider im Rahmen dieser Arbeit nicht adressiert werden.

2 Deutsche Losbücher im Allgemeinen und das *Mainzer Kartenlosbuch* als Grundlage für den Wissensgraphen

Als inhaltliche Grundlage für den zu entwickelnden Wissensgraphen und den Chatbot-Prototyp dient die Edition des *Mainzer Kartenlosbuchs*, eines humoristischen Werks der Weissagungsliteratur (Däumer 2021, S. 90). Um das Werk angemessen in seinen historischen und literarischen Kontext einzuordnen werden zunächst Losbücher im Allgemeinen betrachtet.

2.1 Allgemeine Merkmale deutscher Losbücher

Der Begriff »Losbuch« ist im Deutschen seit dem 13. Jahrhundert belegt und bezeichnet eine Textsorte, die fest im Alltagsleben einer breiten Bevölkerungsschicht verankert war (Heiles 2018, S. 23, S. 26).

Ausgehend von den von Heiles zusammengetragenen Definitionen (u.a. in Bolte (1903), Bolte (1925) und Sotzmann (1851)) lassen sich Losbücher in der germanistischen Forschung verkürzt als Werke beschreiben, die Orakelsprüche enthalten. Diese werden durch Zufallsmechanismen wie Würfeln oder Kartenziehen zugänglich gemacht. Sie bieten Antworten auf Fragen zur Zukunft und vereinen Elemente der Prophezeiung, Unterhaltung und moralischen Reflexion (Heiles 2018, S. 30 ff.).

Trotz ihrer vielseitigen Anwendung waren Losbücher oft umstritten, da sie mit Aberglauben

assoziiert wurden. Bereits im 13. Jahrhundert wurde vor ihrer Nutzung gewarnt (Heiles 2018, S. 24).

Die Struktur von Losbüchern erfordert diskontinuierliches Lesen (Heiles 2018, S. 40). Grundsätzlich lassen sich zwei Typen laut Heiles unterscheiden: Losbücher mit vorgegebenen Fragen und solche ohne, bei denen die Nutzer frei entscheiden können, welche Frage sie stellen (Heiles 2018, S. 39 ff.).

2.2 Grundlage für den Wissensgraphen: Das *Mainzer Kartenlosbuch*

Das *Mainzer Kartenlosbuch* ist ein humoristisches Werk der Weissagungsliteratur und wurde um 1510 von Johann Schöffer in Mainz gedruckt (Däumer 2021, S. 90). Die Grundlage für den Wissensgraphen bildet die moderne Edition von Matthias Däumer, die 2021 im Hirzel Verlag veröffentlicht wurde. Diese Edition bietet eine umfassende und strukturierte Aufbereitung des Werks und enthält nicht nur den vollständigen Text, sondern auch eine Einleitung mit Informationen zur Druck- und Überlieferungsgeschichte sowie zu den sprachlichen und kulturellen Verbindungen zu älteren Losbüchern (Däumer 2021, S. 88 ff.). Diese detaillierte Aufbereitung macht die Edition zu einer guten Basis für die Modellierung eines Wissensgraphen und die interaktive Präsentation im Rahmen eines Chatbot-Prototyps.

Das *Mainzer Kartenlosbuch* steht in der Tradition älterer Texte und geht auf das sogenannte Tierlosbuch zurück, bekannt als *Basler Druck* von 1485. Alle 48 Lossprüche des *Mainzer Kartenlosbuchs* sind auf diese Vorlage zurückzuführen. Auch Teile des Kolophons zeigen eine Anlehnung an das Tierlosbuch. Zudem zeigt das Werk einen starken Einfluss eines Nürnberger Drucks, der selbst eine Umarbeitung des Tierlosbuchs darstellt und als direkte Vorlage für den Mainzer Druck diente. Sprachliche Analysen zeigen, dass der Text seinen Ursprung in Nürnberg hat und von Schöffer in Mainz mit geringfügigen dialektalen Modifikationen neu aufgelegt wurde.

Das in der Edition von Däumer dargestellte Kartenlosbuch wurde 1890 von Adolf Hofmeister als Faksimile herausgegeben und befindet sich heute als einziges noch existierendes Exemplar in der British Library, London (Däumer 2021, S. 88). In der Edition wurden die Lossprüche aus dem historischen Druck in eine modernisierte Fassung überführt.

Das Kartenlosbuch bietet zwei Nutzungsarten: Einerseits kann es mit einem klassischen deutschen Kartendeck verwendet werden, andererseits ermöglicht eine im Buch enthaltene drehbare Losscheibe die Weissagung auch ohne ein Kartenspiel. Der Schwerpunkt dieser Arbeit liegt auf der ersten Nutzungsart. Hierbei wählt der Nutzer eine Karte aus einem deutschen Kartenblatt, die dann mit einem zugehörigen Vers aus dem Losbuch verknüpft

wird.

Oft stehen die Weissagungen der Karten im Gegensatz zu deren eigentlicher Bedeutung, wodurch die Weissagung humorvoll hinterfragt wird. Stilistisch ist das Werk geprägt von performativen Sprachstrukturen, die den dialogischen Charakter der Texte unterstreichen (Däumer 2021, S. 90 f.).

3 Technologische Grundlagen und Konzeption des Chatbots

Der prototypische Chatbot zur Edition des *Mainzer Kartenlosbuchs* basiert auf der Kombination von Wissensgraphen, LLMs und RAG. Diese Technologien ermöglichen es strukturierte Daten mithilfe natürlicher Sprache zu durchsuchen und zu verarbeiten. Im Folgenden werden die technologischen Grundlagen erklärt und mit der Umsetzung des Chatbots verknüpft.

3.1 Technologische Grundlagen

3.1.1 Wissensgraphen

Wissensgraphen sind bereits Werkzeuge in der Editionswissenschaft. Sie ermöglichen die strukturierte Modellierung und Verknüpfung von Daten (Hodel 2021, S. 30). Außerdem repräsentieren sie Informationen als Netzwerke, in denen Knoten Entitäten wie Textstellen oder Begriffe darstellen und Kanten deren Beziehungen abbilden. Diese Beziehungen können Intertextualitäten oder thematische Verknüpfungen sein. (Fernandes und Bernardino 2018, S. 373).

Diese Netzwerke erleichtern die Analyse und Nutzung historischer Dokumente, da sie Verknüpfungen mit externen Ressourcen ermöglichen und komplexe Datenbeziehungen sichtbar machen (Hodel 2021, S. 30). Fernandes et al. empfehlen *Neo4j* zur Modellierung und Abfrage von Graphen, da es benutzerfreundlich ist, eine leistungsstarke Abfragesprache *Cypher* bietet und schnelle sowie effiziente Datenabfragen ermöglicht (Fernandes und Bernardino 2018, S. 380).

In der vorliegenden Arbeit wird *Neo4j* genutzt, um die Inhalte der Edition des *Mainzer Kartenlosbuchs* in einem Wissensgraphen zu modellieren. Dieser bildet die Grundlage für hybride Suchverfahren, die sowohl vektorbasierte als auch graphbasierte Ansätze kombinieren.

3.1.2 Retrieval-Augmented Generation (RAG): Erweiterung von Sprachmodellen durch externe Wissensquellen

RAG erweitert die Funktionalität großer Sprachmodelle (LLMs), indem sie diese mit externen Wissensquellen verknüpft. Dabei kann ein Wissensgraph als Datenbasis dienen. Relevante Informationen werden aus dem Graphen abgerufen und anschließend durch die sprachlichen Fähigkeiten des LLMs kontextualisiert und für die Nutzer:innen in verständlicher Form aufbereitet (Gao u. a. 2024, S. 1).

RAG kombiniert die strukturierte Darstellung von Entitäten und deren Beziehungen mit der Fähigkeit von LLMs, präzise und verständliche Antworten in natürlicher Sprache zu generieren. Diese Synergie ermöglicht es gezielte Wissensabfragen zu formulieren, die weit über klassische Suchfunktionen hinausgehen (Wilcock 2024, S. 8). Dong et al. heben hervor, dass graphbasierte RAG-Ansätze insbesondere in den Bereichen Generierungsqualität, Wissenskonsistenz und Argumentationsfähigkeit klassischen Modellen überlegen sind (Dong u. a. 2024, S. 4).

Für den in dieser Arbeit entwickelten Prototyp wird RAG eingesetzt, um die aus dem Wissensgraphen extrahierten Informationen in den Kontext der Nutzeranfrage zu integrieren. Dadurch werden natürliche Sprachabfragen ermöglicht, die präzise und kontextspezifische Antworten aus dem Wissensgraphen generieren.

3.2 Konzeption des Chatbots

Der Chatbot wird als Schnittstelle zwischen den Nutzer:innen und dem Wissensgraphen entwickelt, um den Zugang zu den im Wissensgraph gespeicherten Informationen der Edition des *Mainzer Kartenlosbuchs* zu erleichtern. Durch den Einsatz von RAG wird die Suche über natürliche Spracheingaben ermöglicht. Auf diese Weise wird unter anderem die in der Einleitung erwähnte Herausforderung adressiert, die fortgeschrittenere Such- und Navigationsfunktionen erfordert, um auf alle Materialien und Informationen einer Edition zuzugreifen (Sahle 2016, S. 30).

Im Folgenden wird die wesentliche Architektur und Funktionsweise des Chatbots beschrieben, um ein grundlegendes Verständnis für dessen Aufbau zu vermitteln.

3.2.1 Funktionsweise des Zusammenspiels von Wissensgraphen und LLM im Chatbot

Das Zusammenspiel von Wissensgraphen und LLM bildet das Fundament des Chatbots. Der Prozess lässt sich in vier Schritte unterteilen. Die Konzeption orientiert sich dabei an den Ansätzen von Wilcock (Wilcock 2024).

- **Erstellung des Wissensgraphen:** Der Wissensgraph wird mit den relevanten

Inhalten der Edition des *Mainzer Kartenlosbuchs* gefüllt. Diese Inhalte umfassen Textstellen, Metadaten und Verweise, die in einer dynamischen Netzwerkstruktur modelliert werden.

- **Datenabruf aus dem Wissensgraphen:** Eine Nutzeranfrage in natürlicher Sprache wird von einem LLM interpretiert und in eine maschinenlesbare Abfrage umgewandelt. Diese Abfrage wird an den Wissensgraphen gesendet, der sowohl vektorbasierte als auch graphbasierte Suchverfahren nutzt, um relevante Datenpunkte zu identifizieren.
- **Erweiterte Informationsbereitstellung durch RAG:** Die extrahierten Daten werden als Kontext in einen erweiterten Prompt integriert, den das LLM nutzt, um eine präzise Antwort zu generieren.
- **Generierung der Antwort:** Das LLM verarbeitet die ursprüngliche Anfrage zusammen mit den Kontextdaten und liefert eine klare, verständliche Antwort. Dabei werden sowohl die semantischen als auch die relationalen Aspekte der Daten berücksichtigt (Wilcock 2024, S. 3 ff.).

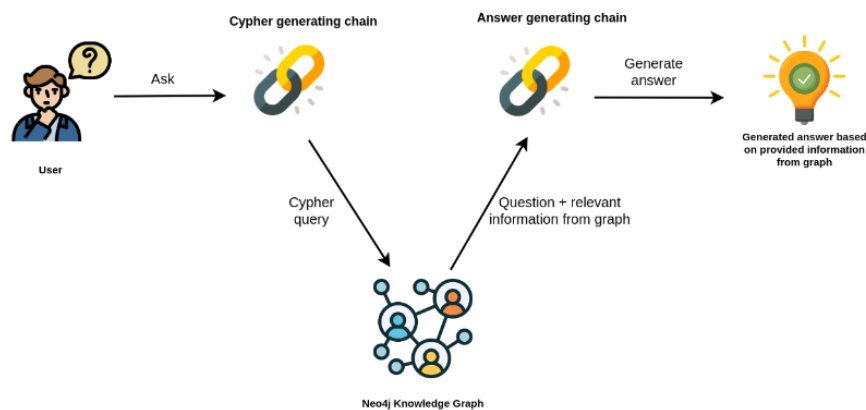


Abbildung 1: Illustration der Funktionsweise des Zusammenspiels von Wissensgraphen und LLM im Chatbot (Wilcock 2024, S. 4)

Das Bild zeigt die zentralen Schritte der Interaktion zwischen Nutzer:innen, Wissensgraph und LLM im Chatbot. Eine Nutzeranfrage wird interpretiert, in eine maschinenlesbare Cypher-Abfrage für den Wissensgraphen übersetzt, und die relevanten Datenpunkte werden extrahiert. Anschließend generiert das LLM eine präzise Antwort, indem es die ursprüngliche Anfrage mit den Graphdaten kombiniert.

Im Bild fehlt jedoch eine explizite Darstellung des Schritts „Erweiterte Informationsbereitstellung durch RAG“, bei dem die extrahierten Daten als Kontext in den Prompt des LLM integriert werden.

3.2.2 Benutzeroberfläche

Die Benutzeroberfläche des Chatbots bildet die Brücke zwischen den Nutzer:innen und dem technischen Fundament. Nutzer:innen können Fragen eingeben, zwischen verschiedenen Modi wechseln und Antworten in natürlicher Sprache erhalten.

3.2.3 Erlebbarkeit des *Mainzer Kartenlosbuchs*: Der Modus *Losbuch spielen*

Neben der Beantwortung allgemeiner Fragen bietet der Chatbot einen speziellen Modus namens *Losbuch spielen*. Dieser Modus greift die historische Funktion des *Mainzer Kartenlosbuchs* auf und macht sie interaktiv erlebbar. Im Gegensatz zur Beantwortung allgemeiner Fragen wird dieser Modus nicht mithilfe von Wissensgraphen und Retrieval-Augmented Generation (RAG) umgesetzt. Aus diesem Grund liegt der Schwerpunkt der Arbeit nicht auf diesem Modus. Die Umsetzung umfasst:

- **Zufällige Losauswahl:** Der Chatbot wählt ein Los zufällig aus und zeigt dessen Inhalt an, einschließlich des zugehörigen Textes und der Bilder.
- **Übersetzung ins Neuhochdeutsche:** Um die Zugänglichkeit zu erhöhen, übersetzt ein LLM die historische Weissagung ins Neuhochdeutsche.
- **Visuelle Darstellung:** Ergänzend zur textlichen Präsentation wird ein Bild des gezogenen Loses angezeigt, um die Weissagung visuell zu unterstützen.

In der Beschreibung wird stets von einem LLM gesprochen, jedoch kommen an anderen Stellen auch weitere LLMs zum Einsatz, etwa bei der Erstellung des Wissensgraphen. Das folgende Kapitel zeigt unter anderem, an welchen Stellen diese eine Rolle spielen, wobei der Fokus auf der konkreten technischen Umsetzung des Chatbots liegt.

4 Technische Implementierung des Chatbots

Die technische Umsetzung des Chatbots folgt der zuvor beschriebenen Konzeption. Sie umfasst die Integration eines Chatbots, der ein LLM nutzt. Dieses wird durch einen Wissensgraphen mit RAG erweitert. Die Implementierung erfolgte in *Python*.

4.1 Preprocessing: Datenaufbereitung für den Wissensgraphen

Das Preprocessing stellt sicher, dass die Rohdaten in einem geeigneten Format für den Wissensgraphen und die semantische Suche aufbereitet werden. Zunächst wird das PDF-Dokument des *Mainzer Kartenlosbuchs* (Däumer 2021) mit `pdfplumber` verarbeitet, um den Text seitenweise zu extrahieren. Zusätzlich werden enthaltene Bilder erkannt, gruppiert und ausgeschnitten, um visuelle Informationen zu sichern. Diese Bildgruppen werden mithilfe

von *Pillow* als Bilddateien gespeichert, um ihre spätere Verwendung im *Losbuch spielen*-Modus zu ermöglichen. Darüber hinaus wird eine JSON-Datei erstellt, die die einzelnen Losprüche, Bildpfade sowie einen Platzhalter für die Übersetzung der Weissagungen ins Neuhochdeutsche enthält. Diese Datei bildet die Datengrundlage für den *Losbuch spielen*-Modus.

Der extrahierte Text wird anschließend mithilfe von *TokenTextSplitter* aus der *LangChain*-Bibliothek in kleinere, handhabbare Abschnitte (Chunks) unterteilt. Diese Textabschnitte werden dann durch *LLMGraphTransformer*, ebenfalls aus der *LangChain*-Bibliothek, in eine für den Graphen kompatible Struktur umgewandelt. Diese Struktur ermöglicht eine einfache Integration in den *Neo4j*-Wissensgraphen, wo sowohl Textdaten als auch Metadaten gespeichert und abgerufen werden können (Bratanič 2024).

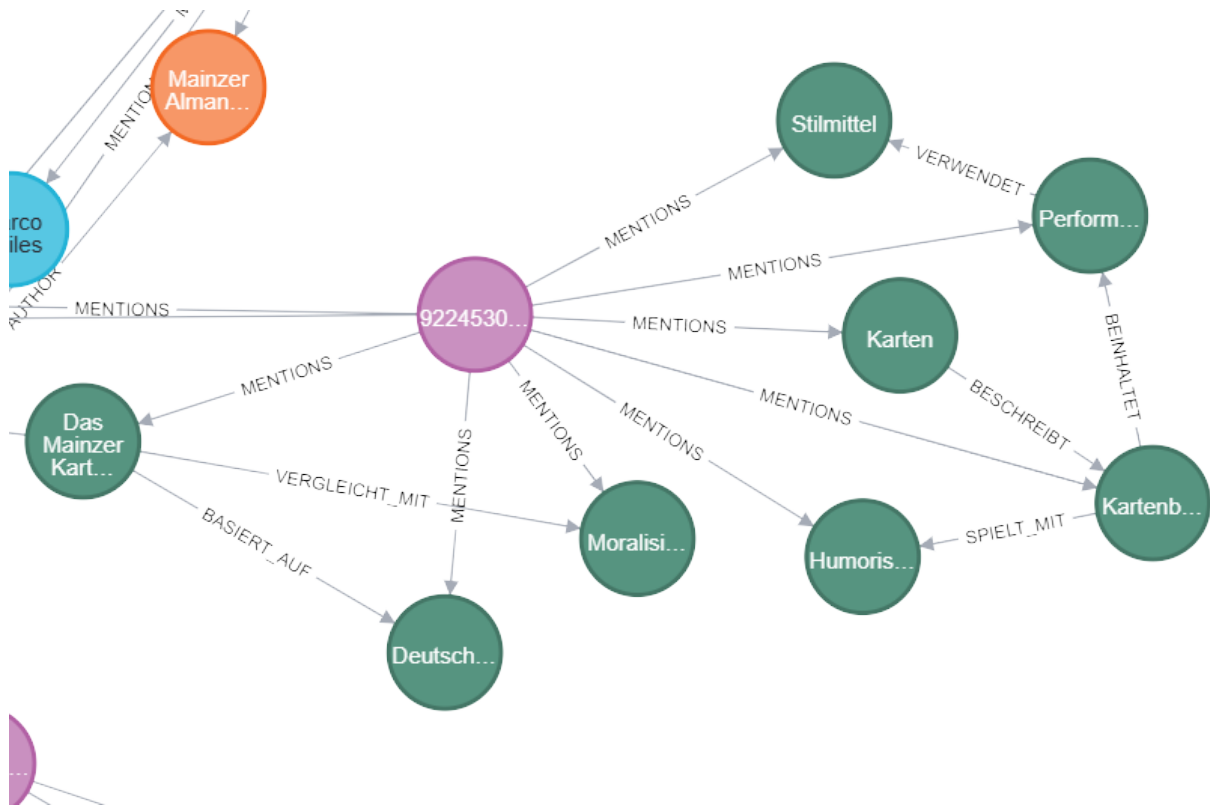


Abbildung 2: Ausschnitt der visuellen Darstellung des Neo4j-Wissensgraphen

Abbildung 2 zeigt einen Ausschnitt der grafischen Darstellung des Wissensgraphen, in dem Knoten (als Kreise dargestellt) Entitäten repräsentieren und durch Kanten miteinander verbunden sind, die die Beziehungen zwischen diesen Entitäten verdeutlichen. Die Zahl in der Mitte des Graphen stellt die Vektorrepräsentation eines Textteils dar, die durch ein Embedding-Modell erzeugt wurde.

4.2 Integration von Wissensgraph und LLM

In der Implementierung des Chatbots wurde das zuvor beschriebene Konzept in eine funktionale Anwendung überführt. Die Grundstruktur besteht aus zwei Hauptmodi: einem allgemeinen Fragenmodus und einem *Losbuch*-Modus.

Neo4j dient in diesem System als Plattform zur Speicherung und Abfrage des Wissensgraphen. Die Implementierung der Datenbankbindung und die Durchführung von Abfragen erfolgt mithilfe des offiziellen Python-Treibers für *Neo4j*, der über die Bibliothek `neo4j` bereitgestellt wird. Die Suche im Wissensgraphen wird mit `Neo4jVector` aus der *LangChain*-Bibliothek durchgeführt. Dabei wird ein vektorbasierter Index verwendet, um relevante Informationen aus dem Graphen zu finden. Dieser Index wird mit `OpenAIEmbeddings` erstellt, das die Texte des Wissensgraphen in numerische Vektoren umwandelt. Durch die Kombination von vektorbasierter und graphbasierter Suche kommt ein hybrides Suchverfahren zum Einsatz, das sowohl die semantische Ähnlichkeit der Inhalte als auch die Beziehungen zwischen den Entitäten im Graphen berücksichtigt.

Der extrahierte Kontext wird anschließend mithilfe verschiedener Komponenten der *LangChain*-Bibliothek an ein Large Language Model (LLM) übergeben, welches Antworten generiert. Die Kommunikation mit dem LLM erfolgt über `ChatOpenAI`. Der Kontext und die Nutzerfrage werden durch `ChatPromptTemplate` in einer strukturierten Eingabevorlage organisiert, die sicherstellt, dass das LLM die Aufgabe präzise versteht. Schließlich verbindet `LLMChain` diese Komponenten in einem durchgängigen Workflow, indem es die Eingabevorlage an das LLM weiterleitet und die generierten Antworten verarbeitet.

Wenn der Kontext aus dem Wissensgraphen nicht ausreicht, um eine präzise Antwort zu liefern, erhält der Nutzer eine standardisierte Fehlermeldung, die darauf hinweist, dass Anfragen ausschließlich auf den Informationen der eingepflegten Edition des *Mainzer Kartenlosbuchs* basieren müssen. Dadurch wird sichergestellt, dass das Modell nicht auf eigene Daten zugreift und keine fehlerhaften Angaben liefert, die über den Wissensgraphen hinausgehen.

In dieser Anwendung wird das *OpenAI GPT-4-o-mini Modell* genutzt, das sowohl zur Beantwortung von Nutzerfragen als auch zur Übersetzung der Weissagungen ins Neuhochdeutsche im Modus *Losbuch spielen* dient (OpenAI 2024).

Für den *Losbuch spielen*-Modus wird eine JSON-Datei verwendet, die die Weissagungen, Symbole und Bildpfade enthält. Der Chatbot wählt zufällig ein Los aus der Datei aus, übersetzt die Weissagung ins Neuhochdeutsche mithilfe eines LLMs (ebenfalls *GPT-4-o-mini*) und gibt zusätzlich das entsprechende Bild und Symbol aus. Diese Funktionalität sorgt für eine interaktive und historische Benutzererfahrung.

Die Benutzeroberfläche wurde mithilfe der Python-Bibliothek `streamlit` entwickelt, welche

die Erstellung einer interaktiven Webanwendung ermöglicht. Die Anwendung wird in der *Streamlit Community Cloud* bereitgestellt, sodass keine lokale Installation erforderlich ist, um den Chatbot zu nutzen.

Für die Erstellung des Codes habe ich mich an einem Blogeintrag aus dem *LangChain*-Blog orientiert (LangChain 2024).

Der Code kann unter folgendem Link eingesehen werden: <https://github.com/TillBaerumker/Orakulus>. Die Anwendung selbst ist online unter <https://orakulusmainz.streamlit.app> erreichbar. Der Chatbot trägt den Namen *Orakulus*.

Alle verwendeten Module sind im Literaturverzeichnis noch einmal aufgeführt.

5 Evaluation des Chatbots

Die Evaluation des Chatbots zielt darauf ab, dessen Funktionalität zu bewerten.

Zum einen wird die »Korrektheit der Antworten« untersucht, um sicherzustellen, dass der Chatbot relevante Inhalte präzise bereitstellt und bei fehlenden Informationen transparent reagiert. Wie in der Einleitung erwähnt, betont Sahle, dass digitale Editionen „more sophisticated browse and search functions to access all the material and information of an edition“ erfordern (Sahle 2016, S. 30). Gleichzeitig weisen Kalmer et al. darauf hin, dass fehlende Verlinkungen und standardisierte Funktionen, wie Suchleisten, für Nutzer:innen oft hinderlich sind (Pontes Trabula 2024, S. 241).

Zum anderen wird die »natürliche Sprachinteraktion« des Chatbots geprüft, um zu bewerten, ob der Chatbot eine intuitive und verständliche Kommunikation ermöglicht. Sahle fordert, digitale Editionen „as widely accessible and usable as possible for all interested disciplines“ zu gestalten (Sahle 2016, S. 22). Durch eine natürlichsprachige Interaktion können m.E. Hürden abgebaut und eine Kommunikation mit Angehörigen verschiedener Disziplinen ermöglicht werden.

Im Folgenden wird detailliert beschrieben, wie die Tests durchgeführt wurden und welche Einschränkungen dabei zu beachten sind.

5.1 Korrektheit der Antworten

Die Korrektheit der Antworten wurde durch Tests mit insgesamt **40 Fragen** überprüft, die dem Chatbot gestellt wurden. Jede Antwort wurde anschließend auf ihre inhaltliche Korrektheit hin überprüft und mit dem Originaltext der Edition abgeglichen. Die Fragen wurden in zwei Gruppen unterteilt:

- **20 Fragen zu enthaltenen Informationen:** Diese Fragen bezogen sich auf Informationen, die im Wissensgraphen enthalten sein sollten. Diese dienten dazu, die Korrektheit der Antworten des Chatbots zu überprüfen.
- **20 Fragen zu nicht enthaltenen Informationen:** Diese Fragen behandeln Inhalte, die nicht im Wissensgraphen gespeichert sein sollten. Diese sollen überprüfen, ob der Chatbot keine falschen Aussagen trifft.

Definition der Metriken: Zur Analyse wurden die Metriken **Precision**, **Recall**, **Accuracy** und **F1-Score** verwendet:

- **Precision:** Der Anteil der korrekt als relevant klassifizierten Antworten (True Positives) unter allen als positiv klassifizierten Antworten (True Positives + False Positives) (Goutte und Gaussier 2005, S. 347).

- **Recall:** Der Anteil der korrekt als relevant klassifizierten Antworten (True Positives) unter allen tatsächlich relevanten Antworten (True Positives + False Negatives) (Goutte und Gaussier 2005, S. 347).
- **Accuracy:** Der Anteil der korrekt klassifizierten Anfragen (True Positives und True Negatives) unter allen Anfragen (Ren und M. Jose 2005, S. 443).
- **F1-Score:** Der harmonische Mittelwert von Precision und Recall, der eine Balance zwischen beiden Metriken darstellt (Tharwat 2021, S. 174).

Ergebnisse der Testfragen: Die Ergebnisse der Testfragen sind wie folgt:

- **19 True Positives:** Fragen, die im Wissensgraphen enthalten sein sollten und korrekt beantwortet wurden.
- **1 False Negative:** Fragen, die im Wissensgraphen enthalten sein sollten und nicht beantwortet wurden.
- **20 True Negatives:** Fragen, die nicht im Wissensgraphen enthalten sein sollten und korrekt nicht beantwortet wurden. In allen Fällen wurde die standardisierte Fehlermeldung ausgegeben, wodurch sichergestellt wird, dass der Chatbot ausschließlich auf Inhalte des Wissensgraphen zurückgreift.
- **0 False Positives:** Fragen, die nicht im Wissensgraphen enthalten sein sollten und fälschlicherweise beantwortet wurden.

Berechnung der Metriken:

- **Accuracy:**

$$\text{Accuracy} = \frac{\text{True Positives} + \text{True Negatives}}{\text{Gesamtzahl der Fragen}} = \frac{19 + 20}{40} = 0.975 \text{ (97.5\%)}$$

- **Precision:**

$$\text{Precision} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Positives}} = \frac{19}{19 + 0} = 1.0 \text{ (100\%)}$$

- **Recall:**

$$\text{Recall} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Negatives}} = \frac{19}{19 + 1} = 0.95 \text{ (95\%)}$$

- **F1-Score:**

$$\text{F1} = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} = 2 \cdot \frac{1.0 \cdot 0.95}{1.0 + 0.95} = 0.974 \text{ (97.4\%)}$$

Die Ergebnisse zeigen, dass der Chatbot in **97.5 % der Fälle korrekte Antworten** liefert. Besonders hervorzuheben ist die **Precision von 100 %**, die verdeutlicht, dass keine falschen positiven Antworten generiert wurden. Der **Recall von 95 %** zeigt, dass fast alle relevanten Informationen korrekt abgerufen wurden. Der **F1-Score von 97.4 %** bestätigt ein ausgewogenes Verhältnis zwischen Präzision und Recall.

5.2 Natürliche Sprachinteraktion

Die Qualität der vom Chatbot generierten Antworten wurde anhand eines subjektiven Bewertungsschemas von mir beurteilt. Der Maßstab umfasst drei Kategorien:

- **0:** Die Antwort ist unverständlich oder maschinell/unmenschlich formuliert.
- **1:** Die Antwort ist grammatikalisch korrekt, klingt jedoch holprig oder unnatürlich.
- **2:** Die Antwort ist klar formuliert, natürlich klingend und sprachlich flüssig.

Im Rahmen der Tests wurden 20 Antworten des Chatbots bewertet. Alle Antworten erhielten die höchste Bewertung von **2**, was zeigt, dass die Antworten des Chatbots im Rahmen der Tests durchweg als verständlich, grammatikalisch korrekt und natürlich klingend bewertet wurden.

Es wurde auf eine automatische Evaluation durch große Sprachmodelle (LLMs) verzichtet, da diese derzeit nur eingeschränkt als Messinstrumente geeignet sind (Bavaresco u. a. 2024, S. 4). Allerdings birgt die alleinige Bewertung durch eine Einzelperson potenzielle Subjektivität. Zukünftige Tests könnten durch die Einbindung mehrerer unabhängiger Evaluatoren:innen oder durch eine hybride Form der Bewertung, bei der ein LLM in Kombination mit einem menschlichen Gutachter eingesetzt wird, ergänzt werden. Dies war im Rahmen dieser Arbeit jedoch nicht möglich.

5.3 Funktionstest Modus *Losbuch spielen*

Neben den Tests zur sprachlichen Interaktion und der Präzision der Antworten wurde auch der Modus *Losbuch spielen* des Chatbots getestet. Der Chatbot hat alle Lose korrekt dargestellt. Die Korrektheit der Übersetzung der Weissagung ins Neuhochdeutsche konnte an dieser Stelle nicht getestet werden, da eine solche Bewertung den Rahmen dieser Arbeit überschreiten würde.

5.4 Zusammenfassung der Ergebnisse

Die Ergebnisse verdeutlichen die hohe Leistungsfähigkeit des Chatbots. Mit einer Accuracy von 97,5 %, einer Precision von 100 %, einem Recall von 95 % und einem F1-Score von 97,4 % zeigt das System, dass es im Rahmen der Tests relevante Inhalte zuverlässig

bereitstellt und gleichzeitig irrelevante Antworten vermeidet. Der hohe Recall verdeutlicht, dass fast alle relevanten Informationen korrekt abgerufen wurden, während die Precision von 100 % zeigt, dass keine falschen positiven Antworten generiert wurden. Zudem zeigt der getestete *Losbuch spielen* Modus die interaktive Funktionalität des Systems, auch wenn die Übersetzung der Weissagungen ins Neuhochdeutsche im Rahmen der Arbeit nicht getestet werden konnte.

Die getesteten Fragen und die entsprechenden Ergebnisse können im Repository unter *Testing Orakulus* eingesehen werden: <https://github.com/TillBaeumker/Orakulus>.

6 Herausforderungen bei Tests und Umsetzung des Chatbots

In diesem Kapitel werden die verschiedenen Einschränkungen der durchgeführten Tests sowie der Implementierung des Chatbots beschrieben.

6.1 Einschränkung der Testergebnisse

Die Testergebnisse zeigten, dass der Chatbot in 97,5 % der Fälle korrekt reagierte, mit einer Precision von 100 % und einem Recall von 95 %. Diese Werte unterstreichen die Präzision und Zuverlässigkeit des Systems bei der Erfüllung der gestellten Aufgaben. Dennoch ist die Aussagekraft der Ergebnisse durch die begrenzte Abdeckung der Tests eingeschränkt. Der zugrunde liegende Wissensgraph umfasst insgesamt 171 Knoten und 323 Beziehungen. Nicht getestete Bereiche könnten daher potenzielle Schwächen des Systems enthalten, die bislang unberücksichtigt blieben. Wie Tharwat (2021, S. 1) betont, können Metriken wie Precision und Recall insbesondere bei unausgeglichene Datensätzen oder unzureichender Testabdeckung problematisch sein, da sie Schwächen des Systems nicht vollständig abbilden.

Ein weiterer Aspekt, der berücksichtigt werden muss, ist die Bewertung durch nur eine Einzelperson. Dies birgt das Risiko subjektiver Verzerrungen. Wie Pichler und Reiter (2022, S. 15) hervorheben, wird in der Forschung häufig das Inter-Annotator-Agreement genutzt, um die Konsistenz der Ergebnisse zu sichern. In dieser Arbeit wurde jedoch auf ein solches Verfahren verzichtet, da die Durchführung mit mehreren Evaluators:innen einen erheblichen zeitlichen und organisatorischen Mehraufwand bedeutet hätte.

Durch diese Einschränkungen bleiben potenzielle Unschärfen in den Testergebnissen bestehen, die in zukünftigen Arbeiten durch eine breitere Testabdeckung und den Einbezug mehrerer Evaluators:innen adressiert werden könnten.

6.2 Einschränkung der Interpretierbarkeit und Erklärbarkeit

Şahin beschreibt Interpretierbarkeit und Erklärbarkeit als zwei verwandte Konzepte. Interpretierbarkeit meint, wie gut Menschen die inneren Mechanismen und Entscheidungen eines Modells verstehen. Erklärbarkeit (in Englisch »explainability«) geht darüber hinaus. Sie beschreibt, wie klar und verständlich diese Entscheidungen dargestellt werden können. Während es bei der Interpretierbarkeit um das Verstehen der Logik eines Modells geht, bedeutet Erklärbarkeit, dass die Entscheidung des Modells für den Nutzer:innen nachvollziehbar ist. (Şahin, Arslan und Özdemir 2024, S. 4).

Im Bereich des Deep Learnings ist die Interpretierbarkeit häufig eingeschränkt, was vor allem dem Black-Box-Charakter vieler Modelle zuzuschreiben ist. Bei vielen tiefen neuronalen Netzwerken ist es schwierig, die genauen Entscheidungsprozesse nachzuvollziehen, da diese Modelle eine hohe Anzahl an Parametern und Schichten enthalten, die die Entscheidungen beeinflussen (Li u. a. 2022, S. 3198).

Darüber hinaus ist die Erklärbarkeit von Modellen wie *GPT-4-o-mini* aufgrund ihrer Closed-Source-Natur eingeschränkt. Da der Quellcode sowie die Modellgewichte nicht öffentlich zugänglich sind, wird es für Nutzer:innen und Entwickler:innen erschwert, die zugrunde liegenden Entscheidungsprozesse nachzuvollziehen. Wie im Artikel Nowak und Sprinkart (2024, S. 781) betont wird:

„[...]weder der Programmcode noch die Modellgewichte unter offenen Lizenzen zur freien Nutzung verfügbar sind.“

Durch die Verwendung von *GPT-4-o-mini* und eines Embedding-Modells von *OpenAI* weist die Implementierung des Chatbots Einschränkungen in Bezug auf Interpretierbarkeit und Erklärbarkeit auf. Aufgrund des Closed-Source-Charakters und der Komplexität dieser Modelle sind die Entscheidungsprozesse nicht vollständig nachvollziehbar. Das betrifft folgende Punkte:

- **Konstruktion des Wissensgraphen:** *GPT-4-o-mini* wird zur Erstellung des Wissensgraphen eingesetzt.
- **Modus „Losbuch spielen“:** In diesem Modus wird *GPT-4-o-mini* zur Übersetzung der Weissagungen ins Neuhochdeutsche genutzt.
- **Allgemeiner Fragenmodus:** *GPT-4-o-mini* übernimmt die semantische Verarbeitung und Generierung der Antworten.
- **Embedding-Modell von *OpenAI*:** Das verwendete Embedding-Modell wandelt Texte des Wissensgraphen in numerische Vektoren um.

Nutzer:innen sind daher auf die Ergebnisse angewiesen, können jedoch die zugrunde liegenden Entscheidungsgrundlagen nur begrenzt überprüfen.

6.3 Einschränkung der Übertragbarkeit

Die aktuelle Implementierung ist speziell auf das *Mainzer Kartenlosbuch* ausgerichtet. Besonders der Modus *Losbuch spielen* wurde so entwickelt, dass er die historische Funktion des Kartenlosbuchs nachbildet. Dieser Modus verwendet eine statische JSON-Datenstruktur, die die Weissagungen, Symbole und Bilder des Kartenlosbuchs enthält und eng mit den spezifischen Inhalten des Buches verknüpft ist. Auch die Benutzeroberfläche wurde gezielt auf diese Edition des Kartenlosbuchs angepasst, um eine benutzerfreundliche Interaktion mit den Inhalten zu ermöglichen. Daher ist die Übertragbarkeit dieses Modus auf andere historische Quellen begrenzt, da eine Anpassung der Datenstruktur erforderlich wäre.

Beim Wissensgraphen zeigt sich jedoch eine andere Situation: Obwohl der derzeitige Graph auf einer speziellen Datenstruktur für das Kartenlosbuch basiert, ermöglicht der verwendete Workflow eine vergleichsweise einfache Integration alternativer Wissensgraphen. Diese Möglichkeit zur Integration alternativer Graphen ist auch der *read.me*-Datei zu entnehmen. Allerdings wurde die Skalierbarkeit des Systems nicht getestet, sodass unklar bleibt, wie gut der Ansatz mit einer stark wachsenden Datenbasis oder einer höheren Nutzerlast umgehen könnte.

Diese Spezifikationen schränken die unmittelbare Anwendbarkeit des Ansatzes auf andere Kontexte ein.

7 Potenziale und Grenzen des entwickelten Chatbots als Prototyp für die Editionswissenschaft

In der vorliegenden Arbeit wurde ein Chatbot entwickelt, der mithilfe von Wissensgraphen, Retrieval-Augmented Generation (RAG) und LLMs die Inhalte der Edition des *Mainzer Kartenlosbuchs* durch natürliche Sprachinteraktion zugänglich macht. Ziel war es, die technische Umsetzbarkeit sowie die Potenziale solcher Systeme für die Editionswissenschaft zu untersuchen. Dies geschah durch die Konzeption, Implementierung und Evaluierung eines Prototypen. Im Folgenden skizziere ich die Potenziale, Einschränkungen und möglichen Ableitungen für zukünftige Arbeiten des entwickelten Chatbots als Prototyp.

Die Umsetzbarkeit des Prototypen wurde durch positive Testergebnisse belegt. Die Korrektheit der Antworten wurde durch Tests überprüft, die unter anderem eine hohe Genauigkeit und Präzision zeigten und somit die Funktionalität des Systems im Rahmen der durchgeführten Tests bestätigten.

Trotz dieser positiven Ergebnisse zeigt die begrenzte Testbasis methodische Schwächen auf. Wie Goutte et al. und Tharwat betonen, sind Metriken wie Precision, Recall und F1-Score bei kleinen Datensätzen nicht ausreichend, um die Systemleistung umfassend zu

bewerten (Goutte und Gaussier 2005, S. 345)(Tharwat 2021). Größere Datensätze und unabhängige Bewertungen könnten die Qualität der Testergebnisse zukünftiger Studien erheblich verbessern und damit die Aussagekraft über die Funktionalität des Systems erhöhen.

Ein weiterer Punkt ist die Bewertung der natürlichen Sprachinteraktion, die in dieser Arbeit nur durch eine Einzelperson durchgeführt wurde, was potenziell subjektiv beeinflusst sein kann. Arstein hebt hervor, dass das Inter-Annotator-Agreement eine sinnvolle Maßnahme darstellt, um die Konsistenz der Ergebnisse zu überprüfen und die Objektivität zu erhöhen (Arstein 2017, S. 298). Für künftige Arbeiten wäre es daher ratsam, mehrere unabhängige Evaluator:innen einzubeziehen, um Verzerrungen zu minimieren und fundiertere Ergebnisse zu erzielen.

Ein zentrales Potenzial des Chatbots für die Editionswissenschaft liegt in seiner Fähigkeit, Inhalte aus einem Wissensgraphen durch natürliche Sprachinteraktion zugänglich zu machen. Zwar existieren bereits graphbasierte Ansätze in der digitalen Editionswissenschaft (Kuczera 2017), doch der Chatbot erweitert diese um eine neue Präsentationsform. Diese neue Form der Zugänglichkeit könnte digitale Editionen über benutzerfreundliche Schnittstellen einem breiteren, interdisziplinären und weniger technisch versierten Publikum zugänglich machen. Darüber hinaus erfüllt die natürliche Sprachabfrage im Wissensgraphen m.E. Sahles Forderung nach „sophisticated browse and search functions“ (Sahle 2016, S. 30). Die Funktionalität der „search function“ konnte zumindest durch das Testing belegt werden.

Dennoch sind Nutzerstudien erforderlich, wie sie beispielsweise von Cameron et al. durchgeführt wurden (Cameron u. a. 2019), um die Akzeptanz und Wirksamkeit des Chatbots in der Praxis zu bewerten. Die Umsetzbarkeit des Systems konnte jedoch im Rahmen der verwendeten Edition und der durchgeführten Tests erfolgreich demonstriert werden.

Eine weitere Einschränkung des Chatbots liegt in der Abhängigkeit von geschlossenen Technologien wie *GPT-4-o-mini*, die aufgrund ihres Black-Box-Charakters die Erklärbarkeit und Interpretierbarkeit der Ergebnisse einschränken. Offene, zugängliche Modelle könnten dazu beitragen, diese Einschränkungen zu verringern und die Anpassungsfähigkeit des Systems zu verbessern.

Ein wichtiger Aspekt für die Editionswissenschaft ist darüber hinaus die Frage der Übertragbarkeit des Systems. Der Chatbot ist derzeit speziell auf die Edition des *Mainzer Kartenlosbuchs* ausgerichtet. Der zugrunde liegende Wissensgraph könnte jedoch auch auf andere Editionen angewendet werden, sofern entsprechende Anpassungen vorgenommen werden. Der Modus *Losbuch spielen* weist jedoch eine eingeschränkte Übertragbarkeit auf, da solche spezifischen Funktionen weitreichende Anpassungen erfordern würden. Falls es sich nicht um ein Losbuch mit einem vergleichbaren Losmechanismus handelt, wäre der

Modus für andere Projekte nicht relevant. Zudem wurde die Skalierbarkeit des Systems nicht untersucht, sodass offen bleibt, inwieweit der Chatbot mit einer größeren Datenbasis oder einer höheren Nutzerlast effizient arbeiten könnte.

8 Fazit und Ausblick

Der entwickelte Chatbot zeigt das Potenzial, digitale Editionen durch den Einsatz von Wissensgraphen, RAG und LLMs interaktiv zugänglich zu machen. Die Testergebnisse bestätigen die Umsetzbarkeit des Systems. Dennoch bestehen methodische Einschränkungen, insbesondere hinsichtlich der Testbasis und der subjektiven Bewertung der Sprachinteraktion. Künftige Arbeiten sollten auf größere Datensätze, Nutzerstudien und die Nutzung offener Modelle setzen, um die Transparenz und Übertragbarkeit des Systems zu erhöhen. Insgesamt bietet der Chatbot einen vielversprechenden Ansatz für die Weiterentwicklung interaktiver digitaler Editionen.

Darüber hinaus müsste noch überlegt werden, welche Rolle ein solcher Chatbot schlussendlich in der Editionswissenschaft spielen soll. Man könnte ihn beispielsweise als zusätzliche Funktion einer bestehenden digitalen Edition integrieren, um bei Unklarheiten Fragen zur Edition zu stellen. Eine Schnittstelle, die über natürliche Sprachinteraktion angesprochen werden kann, könnte den Zugang zu den Inhalten vereinfachen und das Verständnis fördern – ähnlich wie der *Microsoft Copilot*, der als unterstützende Funktion in Anwendungen genutzt wird (Warren 2023).

Ausgehend von der in der Einleitung angesprochenen Unterscheidung zwischen digitalisierten und digitalen Editionen stellt sich im Kontext dieser Arbeit die Frage, ob der entwickelte Chatbot mit seiner Fähigkeit, natürliche Sprachabfragen im Wissensgraphen zu ermöglichen, das Potenzial hat, eine eigenständige digitale Edition zu sein. Während digitalisierte Editionen Druckausgaben lediglich in digitale Form übertragen, schöpfen digitale Editionen die Potenziale digitaler Technologien aus. Laut Sahle kann eine digitale Edition „nicht ohne wesentliche Informations- und Funktionsverluste in eine typografische Form zurückgeführt werden“ (Sahle 2013, S. 149).

Um diese Frage zu adressieren, könnte es lohnenswert sein, das Konzept der Transmedialität zu betrachten, welches bereits eingangs thematisiert wurde. Laut Sahle bezeichnet dies „eine Repräsentationsform von Dokumenten und Texten, die möglichst wenig Vorgaben dazu macht, in welcher Form diese Dokumente und Texte in einer Publikation medialisiert werden“ (Sahle 2013, S. 164). In diesem Sinne könnte der Wissensgraph des Chatbots als eine Edition betrachtet werden, deren Übertragbarkeit auf andere Medienformen jedoch noch getestet werden muss. Ob und inwieweit die dargestellte Weiterverarbeitung der genutzten Edition dieses und weitere notwendige Kriterien erfüllt, um als eigenständige digitale Edition zu gelten, sollte in zukünftigen Arbeiten untersucht werden.

Insgesamt birgt ein solcher Chatbot Potenzial und adressiert auch die Forderungen aus der Wissenschaft, wie sie z.B. von Joris van Zundert formuliert wurden. Er fordert, dass digitale Editionen nicht nur als Repräsentationen des Textes verstanden werden, sondern als performative Elemente, die durch Code erzeugt und verändert werden. Eine »computational edition« erfasst wissenschaftliche Handlungen und Entscheidungen, indem diese in Code eingeschrieben werden, der die wissenschaftliche Praxis nachstellt. Die dargestellte Vorgehensweise geht in diese Richtung, indem Wissensgraphen und interaktive Elemente durch Code erzeugt werden (Van Zundert 2018, S. 20).

Software und Bibliotheken

ChatOpenAI (2025). „ChatOpenAI | LangChain“. Zugriff am 04. Januar 2025. Verfügbar unter: <https://python.langchain.com/docs/integrations/chat/openai/>.

ChatPromptTemplate (2025). „ChatPromptTemplate — LangChain Documentation“. Zugriff am 04. Januar 2025. Verfügbar unter: https://python.langchain.com/api_reference/core/prompts/langchain_core.prompts.chat.ChatPromptTemplate.html.

Clark, Jeffrey A. (2024). „Pillow“. Zugriff am 14. Dezember 2024. Verfügbar unter: <https://pillow.readthedocs.io/en/stable/index.html>.

LangChain (2025). „Constructing Knowledge Graphs | LangChain“. Zugriff am 04. Januar 2025. Verfügbar unter: <https://python.langchain.com/docs/tutorials/graph/>.

LangChain (2025). „LLMChain — LangChain Documentation“. Zugriff am 04. Januar 2025. Verfügbar unter: https://python.langchain.com/api_reference/langchain/chains/langchain.chains.llm.LLMChain.html.

LangChain (2025). „Neo4j Vector Index | LangChain“. Zugriff am 04. Januar 2025. Verfügbar unter: <https://python.langchain.com/docs/integrations/vectorstores/neo4jvector/>.

LangChain (2025). „TokenTextSplitter | LangChain“. Zugriff am 04. Januar 2025. Verfügbar unter: https://js.langchain.com/v0.1/docs/modules/data_connection/document_transformers/token_splitter/.

Neo4j (2025). „Neo4j Python Driver Manual“. Documentation for the Neo4j Python driver, including installation, usage, and advanced features. Zugriff am 04. Januar 2025. Verfügbar unter: <https://neo4j.com/docs/python-manual/current/>.

Singer-Vine, Jeremy und die pdfplumber Contributors (2025). „Pdfplumber“. Zugriff am 04. Januar 2025. Verfügbar unter: <https://github.com/jsvine/pdfplumber>.

Streamlit (2025). „Streamlit: A Faster Way to Build and Share Data Apps“. Zugriff am 04. Januar 2025. Verfügbar unter: <https://streamlit.io/>.

Literatur

- Arstein, Ron (2017). „Inter-Annotator Agreement“. In: *Handbook of Linguistic Annotation*. Hrsg. von Nancy Ide und James Pustejovsky. Springer, S. 297–314. ISBN: 978-94-024-0879-9.
- Bavaresco, Anna u. a. (Juni 2024). *LLMs Instead of Human Judges? A Large Scale Empirical Study across 20 NLP Evaluation Tasks*. DOI: 10.48550/arXiv.2406.18403. arXiv: 2406.18403 [cs]. (Besucht am 04.12.2024).
- Bhat, Vani u. a. (Juli 2024). „Retrieval Augmented Generation (RAG) Based Restaurant Chatbot with AI Testability“. In: *2024 IEEE 10th International Conference on Big Data Computing Service and Machine Learning Applications (BigDataService)*, S. 1–10. DOI: 10.1109/BigDataService62917.2024.00008. (Besucht am 29.11.2024).
- Bolte, Johannes (1903). „Zur Geschichte Der Losbücher“. In: *Zur Geschichte Der Losbücher*. Bd. 4. Thübingen, S. 276–341.
- (1925). „Zur Geschichte Der Punktier Und Losbücher“. In: *Die Volkskunde und ihre Grenzgebiete*.
- Bratanič, Tomaž (2024). *Enhancing the Accuracy of RAG Applications With Knowledge Graphs*. <https://neo4j.com/developer-blog/enhance-rag-knowledge-graph/>. (Besucht am 27.11.2024).
- Cameron, Gillian u. a. (2019). „Assessing the Usability of a Chatbot for Mental Health Care“. In: *Internet Science*. Hrsg. von Svetlana S. Bodrunova u. a. INSCI 2018 International Workshops - INSCI 2018 International Workshops St. Petersburg, Russia, October 24–26, 2018. Cham (Switzerland): Springer, S. 121–132. ISBN: 978-3-030-17704-1.
- Däumer, Matthias (2021). „Mainzer Kartenlosbuch: Eyn losz buch ausz der karten gemacht“. In: *Gedruckte deutsche Losbücher des 15. und 16. Jahrhunderts*. Hrsg. von Marco Heiles, Björn Reich und Matthias Standke. Bd. 1, S. 87–122.
- Dong, Yuxin u. a. (Nov. 2024). *Advanced RAG Models with Graph Structures: Optimizing Complex Knowledge Reasoning and Text Generation*. DOI: 10.48550/arXiv.2411.03572. arXiv: 2411.03572 [cs]. (Besucht am 13.12.2024).
- Fernandes, Diogo und Jorge Bernardino (2018). „Graph Databases Comparison: Allegro-Graph, ArangoDB, InfiniteGraph, Neo4J, and OrientDB:“. In: *Proceedings of the 7th International Conference on Data Science, Technology and Applications*. Porto, Portugal: SCITEPRESS - Science and Technology Publications, S. 373–380. ISBN: 978-989-758-318-6. DOI: 10.5220/0006910203730380. (Besucht am 22.11.2024).
- Gao, Yunfan u. a. (März 2024). *Retrieval-Augmented Generation for Large Language Models: A Survey*. DOI: 10.48550/arXiv.2312.10997. arXiv: 2312.10997 [cs]. (Besucht am 13.12.2024).

- Goutte, Cyril und Eric Gaussier (2005). „A Probabilistic Interpretation of Precision, Recall and F-Score, with Implication for Evaluation“. In: *Advances in Information Retrieval (Vol. # 3408): 27th European Conference on IR Research, ECIR 2005, Santiago de Compostela, Spain, March 21-23, 2005, Proceedings*. Hrsg. von David E. Losada und Juan M. Fernandez-Luna. Lecture Notes in Computer Science 3408. Berlin, Heidelberg: Springer Berlin Heidelberg, S. 345–359. ISBN: 978-3-540-25295-5 978-3-540-31865-1. DOI: 10.1007/b107096.
- Heiles, Marco (2018). *Das Losbuch: Manuskriptologie einer Textsorte des 14. bis 16. Jahrhunderts*. Beihefte zum Archiv für Kulturgeschichte Heft 83. Köln Weimar Wien: Böhlau Verlag. ISBN: 978-3-412-51044-2 978-3-412-50904-0.
- Hess, Jan (Nov. 2024). „Von OCR und HTR bis NER und LLM: Zur (Teil-)Automatisierung digitaler Editions- und Transformationsprozesse“. In: *editio* 38.1, S. 147–162. ISSN: 1865-9446. DOI: 10.1515/editio-2024-0009. (Besucht am 20.01.2025).
- Hodel, Tobias (Okt. 2021). „Eine Zunft im digitalen Wandel: die Editionswissenschaften 2.0 in der Schweiz“. In: DOI: 10.5281/ZENODO.5590079. (Besucht am 24.11.2024).
- Kuczera, Andreas (Apr. 2016). „Graphbasierte digitale Editionen“. In: *Mittelalter. Interdisziplinäre Forschung und Rezeptionsgeschichte*.
- (Sep. 2017). „Graphentechnologien in den Digitalen Geisteswissenschaften“. In: *ABI Technik* 37.3, S. 179–196. ISSN: 2191-4664, 0720-6763. DOI: 10.1515/abitech-2017-0042. (Besucht am 11.01.2025).
- Kulkarni, Mandar u. a. (Jan. 2024). *Reinforcement Learning for Optimizing RAG for Domain Chatbots*. DOI: 10.48550/arXiv.2401.06800. arXiv: 2401.06800. (Besucht am 29.11.2024).
- LangChain (März 2024). *Enhancing RAG-based Application Accuracy by Constructing and Leveraging Knowledge Graphs*. <https://blog.langchain.dev/enhancing-rag-based-applications-accuracy-by-constructing-and-leveraging-knowledge-graphs/>. (Besucht am 06.12.2024).
- Li, Xuhong u. a. (Dez. 2022). „Interpretable Deep Learning: Interpretation, Interpretability, Trustworthiness, and Beyond“. In: *Knowledge and Information Systems* 64.12, S. 3197–3234. ISSN: 0219-3116. DOI: 10.1007/s10115-022-01756-8. (Besucht am 13.01.2025).
- Nowak, Sebastian und Alois M. Sprinkart (Okt. 2024). „Große Sprachmodelle von OpenAI, Google, Meta, X und Co.: Die Rolle von „closed“ und „open“ Modellen in der Radiologie“. In: *Die Radiologie* 64.10, S. 779–786. ISSN: 2731-7048, 2731-7056. DOI: 10.1007/s00117-024-01327-8. (Besucht am 06.12.2024).
- OpenAI (2024). *GPT-4o Mini: Advancing Cost-Efficient Intelligence*. <https://openai.com/index/gpt-4o-mini-advancing-cost-efficient-intelligence/>. (Besucht am 12.12.2024).

- Pichler, Axel und Nils Reiter (Dez. 2022). „From Concepts to Texts and Back: Operationalization as a Core Activity of Digital Humanities“. In: *Journal of Cultural Analytics* 7.4. ISSN: 2371-4549. DOI: 10.22148/001c.57195. (Besucht am 25.09.2024).
- Pierazzo, Elena (2016). „Modelling Digital Scholarly Editing: From Plato to Heraclitus“. In: *Digital Scholarly Editing - Theories and Practices*. Hrsg. von Elena Pierazzo und Matthew James Driscoll. Cambridge: Open Book Publishers, S. 41–58.
- Pontes Trabula, Kian (Nov. 2024). „Digitales Edieren gestern, heute und morgen. Tagung an der Herzog August Bibliothek, Wolfenbüttel, 25./26. September 2023“. In: *editio* 38.1, S. 239–244. ISSN: 0931-3079, 1865-9446. DOI: 10.1515/editio-2024-0016. (Besucht am 04.01.2025).
- Ren, Reede und Joemon M. Jose (2005). „Football Video Segmentation Based on Video Production Strategy“. In: *Advances in Information Retrieval (Vol. # 3408): 27th European Conference on IR Research, ECIR 2005, Santiago de Compostela, Spain, March 21-23, 2005, Proceedings*. Hrsg. von David E. Losada und Juan M. Fernandez-Luna. Lecture Notes in Computer Science 3408. Berlin, Heidelberg: Springer Berlin Heidelberg, S. 433–446. ISBN: 978-3-540-25295-5 978-3-540-31865-1. DOI: 10.1007/b107096.
- Robinson, Peter (Jan. 2002). „What Is a Critical Digital Edition?“. In: *Variants: The Journal of the European Society for Textual Scholarship*. (Besucht am 07.01.2025).
- Şahin, Emrullah, Naciye Nur Arslan und Durmuş Özdemir (Nov. 2024). „Unlocking the Black Box: An in-Depth Review on Interpretability, Explainability, and Reliability in Deep Learning“. In: *Neural Computing and Applications*. ISSN: 1433-3058. DOI: 10.1007/s00521-024-10437-2. (Besucht am 16.01.2025).
- Sahle, Patrick (2013). *Digitale Editionsformen. Teil 2: Befunde, Theorie und Methodik*. Schriften des Instituts für Dokumentologie und Editorik 8. Norderstedt: BoD, Books on Demand. ISBN: 978-3-8482-5252-7.
- (2016). „What Is a Scholarly Digital Edition?“. In: *Digital Scholarly Editing - Theories and Practices*. Cambridge: Open Book Publishers, S. 19–40.
 - (2017). „Digitale Editionen“. In: *Digital Humanities: Eine Einführung*. Hrsg. von Fotis Jannidis, Hubertus Kohle und Malte Rehbein. SpringerLink Bücher. Stuttgart: J.B. Metzler, S. 234–249. ISBN: 978-3-476-02622-4 978-3-476-05446-3. DOI: 10.1007/978-3-476-05446-3.
- Sotzmann, Johann Daniel Ferdinand (1851). „Die Loosbücher Des Mittelalters“. In: *Serapeum* 12.1851, S. 305–316, 321–332, 337–342.
- Tharwat, Alaa (Jan. 2021). „Classification Assessment Methods“. In: *Applied Computing and Informatics* 17.1, S. 168–192. ISSN: 2634-1964, 2210-8327. DOI: 10.1016/j.aci.2018.08.003. (Besucht am 17.12.2024).
- van Zundert, J. und P. Boot (2011). „The Digital Edition 2.0 and the Digital Library: Services, Not Resources“. In: *Digitale Edition Und Forschungsbibliothek. (Beiträge Der*

- Fachtagung Im Philosophicum Der Universität Mainz Am 13. Und 14. Januar 2011*). Harrassowitz, S. 141–152.
- Van Zundert, Joris J. (2018). „Why the Compact Disc Was Not a Revolution and Cityfish Will Change Textual Scholarship, or What Is a Computational Edition?“ In: *Ecdotica* 1/2018. ISSN: 1825-5361. DOI: 10.7385/99283.
- Warren, Tom (März 2023). *Microsoft Announces Copilot: The AI-powered Future of Office Documents*. <https://www.theverge.com/2023/3/16/23642833/microsoft-365-ai-copilot-word-outlook-teams>. (Besucht am 16.01.2025).
- Wilcock, Graham (2024). „New Technologies for Spoken Dialogue Systems: LLMs, RAG and the GenAI Stack“. In: *14th International Workshop on Spoken Dialogue Systems (IWSDS 2024)*.