



GEBF 2026

13. Kongress der Gesellschaft für Empirische Bildungsforschung



ABSTRACTBAND



Bildungsforschung für technologiebedingte gesellschaftliche Entwicklungen
TU München

16.-18. März 2026 // Hauptkongress
18.-19. März 2026 // Nachwuchskongress

ID: 312 Poster 1 (Hauptkongress): 46

Poster (Hauptkongress)

Thema - 1. Wahl: Digitale Bildung

Thema - 2. Wahl: Hochschulbildung

Datenanalyse: Qualitative Methoden

Stichworte: Retrieval-Augmented Generation, Generative Sprachmodelle, Finanzielle Bildung, Hochschullehre, Antwortqualität

Eigene Lehrmaterialien als RAG-Wissensbasis: Chancen und Grenzen der Hochschullehre im Finanzkontext

Frank Dopatka, Michael Gröschel

Technische Hochschule Mannheim, Deutschland

Die Förderung finanzieller Bildung gilt in Deutschland als zentrales gesellschaftliches Anliegen, da Studien wiederholt Defizite im Finanzwissen der Bevölkerung aufzeigen. Hochschulen stehen vor der Aufgabe, didaktisch hochwertige und zugleich moderne Lernangebote bereitzustellen. Während digitale Lernplattformen und E-Learning-Formate in den letzten Jahren stark an Bedeutung gewonnen haben, eröffnet der Fortschritt im Bereich generativer Sprachmodelle neue Perspektiven für adaptive und dialogische Lernumgebungen. Allerdings sind diese Modelle anfällig für Halluzinationen. Sie erzeugen sprachlich überzeugende, aber faktisch fehlerhafte Inhalte. Gerade im Finanzkontext, in dem die Korrektheit von Informationen von höchster Relevanz ist, stellt dies eine gravierende Einschränkung dar.

Ein vielversprechender Ansatz, um die Antwortqualität generativer Sprachmodelle zu steigern, ist die Verwendung von Retrieval-Augmented Generation (RAG). RAG verbindet die Stärken großer Sprachmodelle mit der Möglichkeit, externe Wissensquellen einzubinden. Anstatt ausschließlich auf im Modell gespeichertes Wissen zurückzugreifen, werden gezielt Dokumente aus einer vordefinierten Datenbank abgerufen, die anschließend in die Antwortgenerierung einfließen. Damit können die Inhalte von Lehrenden selbst bestimmt, aktualisiert und kontrolliert werden, wodurch sich sowohl inhaltliche Präzision als auch didaktische Steuerbarkeit erhöhen.

Dazu wurde ein prototypisches RAG-System entwickelt, das speziell auf den deutschen Finanzbildungsbereich zugeschnitten ist. Als Wissensbasis dienten die Vorlesungsunterlagen, die nach einer Umwandlung in Markdown-Dateien in eine Vektordatenbank überführt und mithilfe eines Sentence Transformers vektorisiert wurden. Die technische Umsetzung erfolgte in einer Python-Umgebung unter Nutzung von LangChain zur Orchestrierung und der OpenAI-API als Schnittstelle zum Sprachmodell.

Die Evaluation des Systems erfolgte anhand von fünfzig Prompts, die typische Fragen aus dem Bereich der finanziellen Bildung abbilden. Dabei wurden drei Modellvarianten miteinander verglichen: GPT-4o ohne zusätzliche Kontextdaten, GPT-4o in Kombination mit RAG sowie GPT-4.5 als leistungsstärkstes Referenzmodell. Die Antworten wurden nach den vier Dimensionen der faktischen Korrektheit, Kontexttreue, semantischen Stringenz und formalen Vollständigkeit bewertet. Diese differenzierte Bewertung ermöglichte es, Stärken und Schwächen der Einbindung eigener Lehrmaterialien sichtbar zu machen.

Die Ergebnisse zeigen, dass die RAG-Variante eine signifikante Verbesserung der faktischen Korrektheit gegenüber der Baseline (GPT-4o ohne Kontextdaten) erzielt. Insbesondere bei komplexen, regulierungsbezogenen Fragen ließen sich Fehler reduzieren, da das Modell auf die bereitgestellten Lehrunterlagen zurückgriff. Auch die Nachvollziehbarkeit der Quellen erhöhte das Vertrauen in die Antworten. Gleichzeitig traten jedoch Grenzen zutage: Da die Wissensbasis ausschließlich aus eigenen Materialien bestand, waren die Antworten inhaltlich auf diesen Rahmen beschränkt. Dies führte teilweise zu oberflächlichen oder unvollständigen Ausführungen, wenn die hinterlegten Unterlagen nicht alle Aspekte einer Frage abdeckten. Die Vollständigkeit der Antworten lag daher unterhalb der Werte des leistungsstärkeren GPT-4.5, das ohne RAG auf ein breiteres internes Wissensspektrum zugreifen konnte.

Diese Ergebnisse verdeutlichen die Chance, aber auch die Herausforderungen der Einbindung eigener Lehrmaterialien in RAG-Systeme. Einerseits erhalten Lehrende die Möglichkeit, Inhalte didaktisch zu steuern und die faktische Korrektheit der Antworten abzusichern. Andererseits hängt die Qualität der Ausgaben unmittelbar von der Qualität und Breite der eingebundenen Materialien ab. Für die Hochschullehre kann RAG einen Beitrag zur Sicherung von Verlässlichkeit und Transparenz leisten, erfordert jedoch eine sorgfältige Auswahl, Aufbereitung und kontinuierliche Pflege der verwendeten Dokumente.

Für die Praxis lässt sich ableiten, dass RAG-Systeme vor allem dann sinnvoll eingesetzt werden können, wenn sie als Ergänzung zu etablierten Lehrmethoden dienen. Sie eignen sich, um Studierenden niedrighschwelligem Zugang zu den Inhalten von Lehrveranstaltungen zu ermöglichen und häufig gestellte Fragen automatisiert zu beantworten. Gleichzeitig dürfen sie nicht als Ersatz für umfassende Lehre verstanden werden, da ihre Qualität eng an die hinterlegte Wissensbasis gekoppelt bleibt.

Aus Forschungsperspektive liefert die vorgestellte Arbeit methodische Grundlagen für die Evaluation von RAG-Systemen im Hochschulkontext. Die Verwendung standardisierter Bewertungskriterien kann als Ausgangspunkt für vergleichende Studien dienen. Künftige Arbeiten sollten prüfen, wie die Einbindung vielfältiger Materialtypen wie Folien, Übungsaufgaben oder annotierte Skripte die Antwortqualität beeinflusst. Zudem sind empirische Studien mit realen Studierenden notwendig, um die tatsächliche Lernwirksamkeit solcher Systeme zu messen.