

Einsatz von Generativer Künstlicher Intelligenz zur automatisierten Erstellung mathematischer Übungsaufgaben

Exposé zur Anmeldung der Bachelorarbeit

INES ROHRBACH

HTWK Leipzig

24. April 2024

1. ANGABEN ZUR PERSON

Ines Rohrbach

Matrikel: 20INB

Matrikelnummer: XXXXX

ines.rohrbach@st.ud.htwk-leipzig.de

2. ANGABEN ZU DEM ZWEITGUTACHTER

M. Sc. Alexander Pögelt

Wissenschaftlicher Mitarbeiter der HTWK Leipzig

alexander.poegelt@htwk-leipzig.de

3. THEMATISCHE VORSTELLUNGEN ZUR BACHELORARBEIT

3.1 Motivation

Generative Künstliche Intelligenz (GKI) spielt eine immer größer werdende Relevanz und bietet ein breites Spektrum an Anwendungsmöglichkeiten. Dazu zählt auch die Lehre, in der sie Chancen für individuelle Unterstützung beim Lernen mit sofortigem Feedback, Möglichkeit für Fragen und der Bereitstellung von angepassten Aufgaben für die jeweiligen Bedürfnisse bietet. Auch für Lehrende können arbeits- und zeitaufwendige Prozesse unterstützt werden. Darunter fällt auch die Erstellung von Aufgaben wie zum Beispiel Online-Übungsaufgaben. Die automatisierte Aufgabenerzeugung durch GKI ist nicht trivial, da die Sicherstellung der Korrektheit der Aufgaben und ihre Schwierigkeitsstufe nicht gegeben ist. Diese Arbeit soll sich auf die automatisierte Generierung der Aufgaben konzentrieren und untersuchen, wie diese maschinell überprüft werden können. Darüber hinaus soll auch das Bereitstellen von Aufgaben erleichtert werden, indem die Aufgaben bereits korrekt formatiert sind.

3.2 Zu erreichende Ziele

Die Arbeit soll sich damit befassen, mittels GKI mathematische Aufgaben im Wissensbereich *Faktenwissen* zu generieren. Dazu soll erforscht werden, wie die Qualität und Einordnung der Aufgaben maschinell überprüft werden können und mit menschlichen Wahrnehmungen übereinstimmen. Außerdem soll betrachtet werden, wie die erstellten Aufgaben als QTI-Format (*Question and Test Interoperability*) erstellt oder umgewandelt werden können, um das Bereitstellen der Aufgaben weiter zu vereinfachen. Dabei sollen für die Generierung *Large Language Modelle* (LLM), wie GPT-4, verwendet werden. Schlussendlich soll die Arbeit in das *tech4comp* Framework eingebunden werden.

3.3 Zu untersuchende Teilgebiete

Für die Generierung der Aufgaben ist es notwendig, dass sich vor allem mit *Prompt Engineering* beschäftigt wird, damit die Erzeugung möglichst gute Ergebnisse erzielt und *Halluzinationen* meidet.

Dafür muss die Klassifikation der Aufgaben betrachtet werden. Ebenso ist eine möglichst maschinelle Überprüfung deren Einordnung und Qualität der Aufgaben zu betrachten. Entsprechend sind Metriken zur Bewertung der Ergebnisse zu bestimmen.

3.4 Stand der Wissenschaft

Mitte 2023 wurde im Rahmen des *tech4comp*-Verbunds ein Paper geschrieben, welches sich mit der allgemeinen Aufgabengenerierung beschäftigt und überprüft, welche Qualität die Aufgaben, bewertet von Experten, haben. Auf diesem Paper soll diese Arbeit basieren. Aber auch außerhalb des Verbunds wird die Aufgabenerzeugung mittels Generativer KI steigend untersucht. Gerade Multiple-Choice-Aufgaben werden viel betrachtet, da sie eine Methode sind, mit der möglichst einfach das Verständnis überprüft werden kann. Dafür ist eine bestimmte Qualität der *Distractors* notwendig. Diese sollen möglichst genau die falschen Antworten bieten, die bei häufigen Fehlern oder falschem Verständnis aufkommen. So kann neben der Rückmeldung der richtigen Antwort ebenfalls zusätzlich adequates Feedback gegeben werden. Insgesamt ist dabei die Frage, inwiefern sich maschinell überprüfen lässt, ob die Aufgabe (i) ordentlich das gewünschte Wissen abdeckt, (ii) die Lösung richtig ist und (iii) die Begründung bzw. *Chain-of-Thought* richtig und hilfreich ist.

4. METHODIK

Die Arbeit beinhaltet eine Literaturrecherche, welche GKI sowie LLM, Qualitätsbewertung und die Klassifikation der Aufgaben umfasst. Basierend darauf sollen Aufgaben durch GKI, dabei insbesondere von GPT-4, erzeugt werden. Zusätzlich soll betrachtet werden, ob Aufgaben durch das Hinzunehmen von spezialisierten LLMs, wie Llemma, bereichert werden. Die Aufgaben sollen schlussendlich im QTI-Format vorliegen, wobei untersucht werden soll, ob dies ebenfalls von der GKI erreicht werden kann. Ansonsten sollen die Aufgaben als JSON oder Markdown erzeugt werden und anschließend in QTI umgewandelt werden. Desweiteren soll maschinell die Qualität und Einordnung überprüft werden. Dafür wird ebenfalls die Bewertung durch einen Experten in dem Gebiet benötigt.

5. DATEN

Die benötigten Daten sind existierende Aufgaben, als Referenz für die GKI, und Information über den Lehrinhalt für die zu generierenden Aufgaben. Beides sollte im Rahmen des *tech4comp*-Verbunds vorhanden sein und zur Verfügung stehen.

6. LITERATUR

- Meissner et al. LLM-Generated Competence-Based E-Assessment Items for Higher Education Mathematics: Methodology and Evaluation. (submitted for publication)
- Kevin Hwang, Sai Challagundla, Maryam M. Alomair, Lujie Karen Chen, Fow-Sen Choa. Towards AI-Assisted Multiple Choice Question Generation and Quality Evaluation at Scale: Aligning with Bloom's Taxonomy. *NeurIPS'23 Workshop on Generative AI for Education (GAIED)*, [Link](#)
- McNichols et al. Automated Distractor and Feedback Generation for Math Multiple-choice Questions via In-context Learning. *NeurIPS'23 Workshop on Generative AI for Education (GAIED)*, [Link](#)

- Davis Ernest (2024): Mathematics, Word Problems, Common Sense, and Artificial Intelligence. In *BULLETIN (New Series) OF THE AMERICAN MATHEMATICAL SOCIETY*, Volume 61, Number 2, 287—303, doi:10.1090/bull/1828

7. SPERRVERMERK

Ein Sperrvermerk ist nicht notwendig.

8. ZEITPLAN

Vorarbeiten:

04.2024 : Themenfindung; Erstellung/Abstimmung Exposé

Bearbeitung des Bachelorarbeitsthemas und Schreiben der Bachelorarbeit:

04. 2024 : Anmeldung der Arbeit an der HTWK Leipzig, Literaturarbeit, erste Implementierungen

05. 2024 : Literaturarbeit, Implementierungen, Gliederung der Arbeit, Erste Kapitel schreiben, Aufgaben überprüfen lassen

06. 2024 : Implementierungen, Arbeit weiter schreiben

07. 2024 : Ergebnisse dokumentieren, Arbeit zu Ende schreiben und FRISTGERECHTE ABGABE

nach Abgabe der Arbeit:

07. 2024 : Zeit für Begutachtung, Vorbereitung des Kolloquiums

08. 2024: Kolloquium