Curriculum-based Question Generation for Mathematics and Science

Abhishek Singh[1], **Krishnendu Ghosh**[2]
[1] SCSET, Bennett University, Greater Noida, Uttar Pradesh, India
[2] Indian Institute of Information Technology Dharwad, Dharwad, Karnataka, India
abksingh2004@gmail.com, krishnendu@iiitdwd.ac.in

Abstract

The purpose of this study is to leverage the use of Natural Language Processing by using Large Language Models (LLMs) so that we can generate questions with respect to the curriculum for Mathematics and Science subjects. To achieve this, we implemented two principal methodologies from the field of Generative-AI: Fine-tuning and Retrieval Augmented Generation (RAG) are two of these approaches. The Fine-tuning approach which we are using is Transfer-Learning using unsloth(), which offers two types of Hugging Face's trainers for fine-tuning in its GitHub repository: Direct Performance Optimizer (DPO) and Supervised QA Fine-tuning using SFT Trainer. We have used the SFT Trainer, which requires reward modelling or reinforcement learning and works based on labelled data to fine-tune the pre-trained LLM for specific tasks (Li et al., 2024). To generate curriculum-based questions, we have used a supervised and labelled dataset of Mathematics questions (Hendrycks et al., 2021). In contrast, for Science, we have used NCERT (National Council of Educational Research and Training) books issued by CBSE (Central Board of Secondary Education), a fine-tuned model for structured and supervised mathematical questions data, while RAG model for books' PDF for generating contextually appropriate science questions due to containing a large number of factual details in a book.

This research highlights how these approaches can be used to design highly effective instructional tools compatible with generating curriculum-based questions on required subjects. Thus, using these approaches, there is a possibility of developing educational content which is beneficial for students and teachers.

1 Introduction

In the rapidly evolving landscape of education, integrating artificial intelligence (AI) presents promising avenues for enhancing curriculum development and instructional methodologies (Zhang and Aslan, 2021). This research focuses on using the domains of Natural Language Processing (NLP) like Large Language Models (LLMs) and Generative Artificial Intelligence to generate curriculum-based questions in mathematics and science, aiming to support educators in creating diverse and challenging question sets by utilizing Generative AI techniques such as Fine-Tuning and Retrieval Augmented Generation (RAG), this study explores the efficacy and practicality of these methods in automating question generation, ultimately contributing to be the part of those system participating in developing a dynamic and tailored educational experiences.

The questions generated through these AI techniques can be utilized in various educational contexts, such as formative assessments, homework assignments, or practice exams. This versatility provides educators with a valuable tool to enhance teaching and learning by ensuring students are exposed to various question types and difficulty levels. The integration of AI in question generation also allows for the rapid updating of question sets to align with current curricular standards and emerging educational needs.

Furthermore, the use of Fine-Tuning and RAG in this research underscores AI's potential to address specific educational challenges. Fine-tuning, with its ability to generate precise and relevant mathematical problems, demonstrates the utility of structured data and supervised learning in creating specialized content. In contrast, RAG leverages external knowledge sources to produce science questions that are contextually appropriate and diverse, reflecting the dynamic nature of scientific inquiry.

This research will highlight the importance of adopting AI-driven methodologies in education to foster more engaging and effective learning environments. By automating the creation of curriculum-based questions, educators can allocate more time to personalized instruction and interactive learning activities. Additionally, the continuous improvement of AI models through Fine-Tuning and RAG ensures that the generated questions remain relevant and challenging, thereby promoting a deeper understanding of mathematical and scientific concepts among students.

By enhancing the variety and quality of questions available to educators, these technologies contribute to a more robust and adaptable educational framework. This research showcases the potential of Artificial Intelligence in supporting educational innovation and emphasizes the ongoing need for collaboration between educators and AI/ML developers to fully realize the benefits of these advanced technologies in the classroom.

1.1 The Role of Natural Language Processing

Natural Language Processing (NLP) is a field of study in artificial intelligence that enables machines to understand, interpret, and generate human language, and the domain of large language models comes into play. A variety of educational tools, from augmentation (Ghosh, 2022b,a) to problem-solving (Ghosh and Das, 2023), began to emerge as a result of advancements in NLP and its applications. In the context of question generation, Large Language Models with slight improvements and alterations can analyze curricular texts and extract key concepts, terms, and relationships. These insights are then used to formulate questions that test students' understanding of the material. For example, a model can parse a biology textbook, identify important topics such as cellular respiration, be trained on various datasets containing different topics and levels of mathematics questions, and generate relevant multiple-choice or short-answer questions. For example, a natural language processing model might analyze a chapter or dataset on photosynthesis or calculus and generate Questions that cover different aspects of the subject and concepts, such as 'the role of chlorophyll and the stages of the light-dependent reactions' or 'integrate the following function' respectively (Nadkarni et al., 2011).

1.2 Benefits of AI-Driven Question Generation

Consistency and Standardization: AI-driven systems may ensure that the questions generated are consistent in terms of difficulty level and format. This standardization is crucial for maintaining fairness in assessments.

Scalability: AI can generate a huge amount of content in a short period of time, resulting in the allocation of more time to other activities and focusing on quality teaching, which can be beneficial for large educational institutions or online learning platforms that serve a platform for thousands of students (Gravina et al., 2019).

1.3 Practical Applications and Case Studies

Several educational technology companies and research institutions are pioneering AI-driven question generation. For example, platforms like Quizlet and Khan Academy's integrated AI to create practice questions that align with their extensive library of educational content. Research studies have demonstrated that AI-generated questions can be as effective as human-created ones in assessing student knowledge and promoting learning (Lu et al., 2021).

1.4 Challenges and Future Directions

(Pedro et al., 2019)

Quality Control: Ensuring that AI-generated questions are accurate, relevant, and free from bias requires rigorous validation, which is lacking while using the prompting technique due to no use of syllabus information in the Large Language Model.

Complexity of Higher-Order Thinking: Generating questions that test higher-order thinking skills, such as analysis and synthesis, is more challenging than creating factual recall questions and requires proper research both quantitatively and qualitatively.

Integration with Educational Systems: Seamlessly integrating AI question generation with existing Learning Management Systems (LMS) and educational workflows still remains a technical challenge for schools and colleges (Owoc et al., 2019).

2 Background

The automation of the generation of questions with the help of the mentioned techniques of artificial intelligence has the potential to revolutionize educational practices. This makes it possible for educators to transition from delivering content to teaching and addressing individual students. In addition, the conscious control over the frequency of new AI transformers like Meta's LLaMa, Google's Gemini and OpenAI's ChatGPT, where now you can create custom versions of ChatGPTs that combine instructions, extra knowledge, and any combination of skills. This update allows for preserving the relevance of the generated responses toward the up-to-date educational and technological standards and learning outcomes. One of the significant issues was to point out the fact that teachers, together with the developers of AI technologies, act as the creators of these technologies since their efficiency is crucial for the attainment of the goal that is stated – providing the learners with high-quality, flexible, and comprehensive educational experiences. Based on these considerations, it is evident that the role of AI is going to grow as the year's progress within educational settings, and consequently, the positive effects of creating educational content and delivery will improve the learning outcome of students within various fields of study.

The improvement in NLP techniques led to the development of Transformers in 2017, an innovative architecture in this domain which has become the foundation for developing new natural language processing applications and Generative AI tools.

Transformers were first proposed in the paper by Vaswani *et al.* (2017) titled "Attention is All You Need" (Vaswani et al., 2017) and have since become a game changer in the creation of powerful LLMs like GPT and BERT. These models are most useful in the generation of curriculum-based questions in subjects such as mathematics and sciences when done automatically. Transformers perform well in processing big data and identifying the interactions between data points, which enables them to produce a variety of questions relevant to the context. There are methods like fine-tuning that allow transformers to be trained to generate specific and pertinent questions for the curriculum, such as intricate mathematical problems. Moreover, Retrieval Augmented Generation (RAG) uses transformers to improve the question generation process, as the questions are not only diverse but also up-to-date with the latest scientific advancements.

2.1 Evolution of Machine Learning in Education

A brief review of machine learning in education is provided by starting with the periods grounded on the Turing Test of the 1950s (Turing, 1950), B.F. Skinner's programmed instruction of 1960s (McCarthy et al., 1959), machine learning and NLP 1970s-1990s Seymour Papert's AI as a tool for learner interactivity (Goldstein and Papert, 1977). This breakthrough became possible in the 2000s with the help of LLMs, which can be illustrated by IBM's Watson, an AI for question-answering computer system that uses natural language processing (NLP) and machine learning (ML) to answer questions posed to it. In 2011, IBM's Watson AI system was developed to compete in the TV game show "Jeopardy!" and won the competition. Since then, IBM has continued to develop and refine Watson AI (Apte et al., 2000). Launched in the years 2011 and 2015, respectively, GPT-2 by Penn and common and OpenAI's first Generative Pre-trained Transformer (GPT) models paved the way for this study (Mhlanga, 2023).

2.2 Theoretical Foundation

The theoretical framework of this study is based on the machine learning approach, which in turn is based on the use of algorithms that, when trained on big data, are capable of recognizing intricate patterns and making predictions with a high degree of accuracy. This capability is essential in the process of shifting question generation to fit well within the curricula of education. With the help of big data, these algorithms can be further refined to generate questions that are in line with the curriculum with high relevance and accuracy (Cope and Kalantzis, 2016).

NLP, a core component of AI, is the branch of computer science that deals with the interaction between computers and human languages. NLP, when integrated with machine learning, has given rise to the Large Language Model (LLM), which is capable of solving complex linguistic problems with a high level of accuracy. These models, like GPT and BERT, employ techniques like transformers that enable them to work with large amounts of data and identify patterns within the text. This makes them very useful in areas such as education, especially for the generation of questions that are both contextually relevant and varied (Liddy, 2001).

From this perspective, the use of LLMs builds on their capacity to create clear and curriculum-relevant questions that can be developed for various educational stages and topics, thereby improving the general efficiency of the teaching-learning procedures.

2.3 Research Focus

This study extends the examination of two approaches of Generative-AI, i.e., Fine-Tuning and RAG's capacities for generating curriculum-based questions in mathematics and science. In this context, the Research will identify the advantages and methods of the mentioned approaches, and its purpose is to contribute to the advancement of educators globally through the use of Machine Learning in education.

2.4 Technological Advancements

Large Language Models that can train on recovered datasets in the 2000s introduced the fundamental concepts for both fine-tuning and Retrieval-Augmented Generation (RAG) (Hardesty, 2019). There is a general approach which is used known as fine-tuning; it involves training LLMs for other specific areas, such as education. RAG, on the other hand, makes use of information retrieval research from the early 1970s so that

3 Methodologies

This research investigates two primary methodologies for addressing challenges in automated question generation for educational purposes: Fine-tuning and Retrieval-Augmented Generation (RAG). Both methods utilize large language models (LLMs) to create curriculum-aligned questions in Mathematics and Science.

3.1 Fine-Tuning

Fine-tuning involves further training a pre-trained LLM on a dataset specific to the target task. The fine-tuning process starts with a pre-trained base model.

Fine-tuning, as described in the paper "Growing a Brain: Fine-Tuning by Increasing Model Capacity," (Wang et al., 2017) is a process where a pre-trained neural network is adapted to a new task by gradually increasing its model capacity. Instead of maintaining a fixed architecture, the model starts with a smaller, less complex structure and progressively adds new layers or neurons during the fine-tuning process. This incremental growth allows the model to learn simpler patterns and adapt to more complex representations as needed. The approach is designed to optimize learning efficiency, reduce the risk of overfitting, and enhance the model's ability to generalize to the new task. By carefully controlling when and how the model's capacity is expanded, fine-tuning enables the model to achieve better performance compared to traditional fine-tuning methods.

There are some hugging face trainers that are provided for working with fine-tuned models; one of them is a supervised fine-tuning trainer.

3.1.1 Supervised Q&A Fine-Tuning (SFT Trainer)

The SFT Trainer provides a convenient and efficient way to fine-tune pre-trained language models for specific NLP tasks while utilising the strengths of the Hugging Face Transformers library. Its flexibility and customizability make it a valuable tool for researchers and practitioners alike.

Objective: The SFT Trainer is built on top of the transformers, so we are Fine-tune the model on a specific labelled dataset for generating curriculum-based questions.

Method: SFT Trainer employs supervised learning with a structured dataset of input-output pairs. For instance, the mathematics dataset might include columns labelled "Problem," "Level," "Type," and "Solution." Fine-tuning utilizes the first three columns to formulate new problems based on specified criteria. Prompts can specify problem type (e.g., Algebra), difficulty level (e.g., Level 5), and a brief question to guide generation.

Implementation: The model was programmed to create new and solvable mathematics problems based on structured prompts and adapt its responses to the desired format. While some topics like geometry may have limitations in generating questions with diagrams, the model could produce various mathematical questions.

3.2 Retrieval-Augmented Generation (RAG)

Objective: In the context of this research, Retrieval-Augmented Generation (RAG) was implemented to enhance the generative capabilities of a trained LLM (Large Language Model) by integrating external knowledge sources through retrieval mechanisms.

Method: The method involved several steps:

1. Initial Setup with a Trained LLM: A pre-trained LLM was used as the baseline model.

2. Dataset Preparation:

    - The syllabus book's chapter data were loaded, converted into a structured format suitable for training and testing and saved as JavaScript Object Notation(JSON).
    - Text chunks were segmented using a text splitter for manageable processing.

- Embedding was generated for text chunks using a pre-trained model.
- A vector database (e.g., ChromaDB) was initialized to efficiently store and retrieve text and embeddings.

3. Integration of Retrieval Mechanisms: Retrieval mechanisms were implemented to connect the LLM with external knowledge sources. This involved retrieving relevant information from the knowledge base to augment the generative process.

4. Testing and Confirmation: The effectiveness of the RAG approach was tested and confirmed to ensure accurate retrieval and incorporation of external knowledge using different chat templates.

5. Usage in Question Generation: RAG was applied to generate science-based questions, for example, by providing specific queries related to topics such as Acids, Bases, and Salts; Metals and Non-metals; and Carbon and Compounds.

6. Output Interpretation: The RAG system successfully generated question-answer pairs based on input queries, demonstrating its capability to leverage retrieved knowledge effectively in generating domain-specific content.

This approach enhanced the RAG model's ability to produce relevant and accurate content, particularly in domains where external knowledge integration is critical, such as science education.

4 Results and Discussion

Mathematics: Fine-tuning proves advantageous for mathematics because supervised and structured data can be mapped (input with output). The formatting prompt function effectively grows the model's capacity on specific tasks like mathematics problem generation, ensuring the creation of precise and relevant mathematical problems with the required input (Wang et al., 2017).

Science: Retrieval Augmented Generation (RAG) outperforms fine-tuning for science questions by leveraging external knowledge sources. The scope of science questions often requires comprehensive factual awareness to generate new questions, as they cannot be mapped to any specific type of question. Fine-tuning may also lead to a decrease in the diversity of questions. Therefore, using a text RAG enables the generation of accurate and contextually appropriate science questions. This dynamic integration of information maintains basic facts and fundamental theoretical information as embeddings, helping bridge gaps in the pre-trained model's knowledge base (Lewis et al., 2020).

The results highlight the different approaches required for optimizing AI models in the fields of mathematics and science. In mathematics, structured data and explicit input-output mappings make fine-tuning a suitable method. This approach enhances the model's ability to generate precise mathematical problems by effectively utilizing supervised data and prompt formatting functions. Conversely, the nature of science questions, which often require a broad and diverse knowledge base, makes Retrieval Augmented Generation (RAG) a more effective strategy. Fine-tuning in this context could reduce the variety of questions due to the inability to map science questions to specific types. Conversely, RAG leverages external knowledge and dynamically integrates information, ensuring that generated questions are accurate and contextually relevant. This method bridges gaps in the pre-trained model's knowledge, maintaining high diversity and factual accuracy.

These findings suggest that while fine-tuning is beneficial for tasks with well-defined structures and supervised data, such as mathematics, RAG is more suitable for areas requiring extensive and varied knowledge, such as science. This differentiation in approach ensures that AI models are optimized for the specific requirements of each field.

4.1 Quantitative Analysis

4.1.1 Retrieval Metrics

Metric	Score
Precision@5	0.2
Recall@5	1.0
MRR	1.0
NDCG@5	1.0

Precision@5: 0.2 Only 20% of the top 5 retrieved documents are relevant. This indicates that there is room for improvement in retrieving more relevant documents among the top results.

Recall@5: 1.0 All relevant documents were retrieved within the top 5 results. This is excellent and suggests that the retrieval system successfully finds all relevant documents, though they might not be ranked optimally.

Mean Reciprocal Rank (MRR): 1.0 The first relevant document is consistently ranked first among the retrieved documents. This is very good and shows that when relevant documents are present, they are found at the top rank.

Normalized Discounted Cumulative Gain (NDCG@5): 1.0 All relevant documents are perfectly ranked within the top 5 results. This is excellent and indicates that the ranking of relevant documents is ideal.

4.1.2 Generation Metrics

Metric	Score
BLEU	0.043
ROUGE-1 (Precision)	0.507
ROUGE-1 (Recall)	0.497
ROUGE-1 (F-measure)	0.502
ROUGE-L (Precision)	0.264
ROUGE-L (Recall)	0.259
ROUGE-L (F-measure)	0.261
Perplexity	12.92
BERT-Score	0.243

BLEU (Bilingual Evaluation Understudy): 0.043 BLEU score measures the overlap of n-grams between the generated text and reference text. A score of 0.043 is relatively low, indicating low overlap with the reference text.

ROUGE (Recall-Oriented Understudy for Gisting Evaluation)

- ROUGE-1: Precision=0.507, Recall=0.497, F-measure=0.502
- ROUGE-L: Precision=0.264, Recall=0.259, F-measure=0.261

ROUGE scores evaluate the overlap of n-grams and other linguistic units between the generated text and reference text. The scores indicate moderate overlap, with room for improvement in both precision and recall.

Perplexity: 12.92 Perplexity measures how well the language model predicts a sample, with lower values indicating better performance. A perplexity of 12.92 is reasonably good, indicating that the language model performs adequately in predicting the generated text.

BERT-Score: 0.243 BERT-Score uses pre-trained BERT embeddings to compare the similarity between the generated text and reference text. A score of 0.243 indicates moderate similarity between the generated and reference texts.

4.1.3 Summary

Retrieval Metrics show good recall but could benefit from improved precision to enhance the relevance of top-ranked documents.
Generation Metrics suggest that while the language model performs adequately in terms of perplexity, there is room for improvement in BLEU, ROUGE, and BERT Score metrics to increase the quality and similarity of the generated text to the reference text.

4.2 Qualitative Analysis

To evaluate the generated questions qualitatively, we recruited 12 teachers who have taught the respective subjects for at least five years in different schools. The survey was conducted using a 5-point Likert scale-based satisfaction questionnaire (where 1 refers to the lowest rating and 5 to the highest), consisting of 9 questions, as shown in Table 1.

The responses from the subjects have been depicted in the following figures: Figure 1-8.

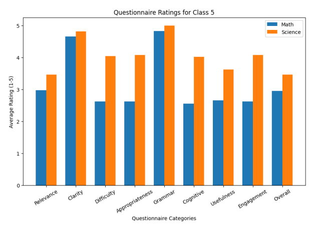

Figure 1: Qualitative evaluation for Grade Level 5

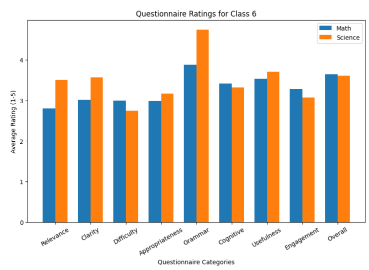

Figure 2: Qualitative evaluation for Grade Level 6

From these figures, it is evident that it is possible to generate relevant curriculum-based questions of different types and grade levels.

Table 1: Qualitative questionnaire to evaluate the overall quality of the generated questions

Q1	How relevant is the generated question to the specified curriculum?
Q2	Is the question clearly worded and easy to understand?
Q3	How would you rate the difficulty level of the question?
Q4	Is the difficulty level appropriate for the target grade?
Q5	Is the grammar and syntax of the question correct?
Q6	Does the question appropriately challenge the cognitive abilities of the students?
Q7	How useful is the question in assessing students' knowledge and understanding?
Q8	How engaging do you find the question for students?
Q9	How would you rate the overall quality of the generated question?
Q10	Please provide any suggestions you have for improving the generated questions.

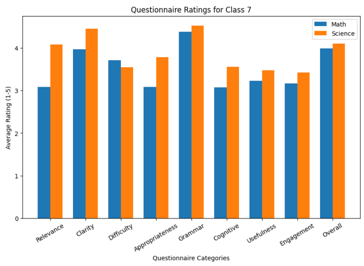

Figure 3: Qualitative evaluation for Grade Level 7

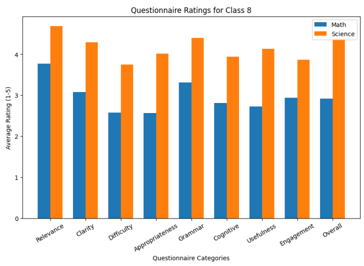

Figure 4: Qualitative evaluation for Grade Level 8

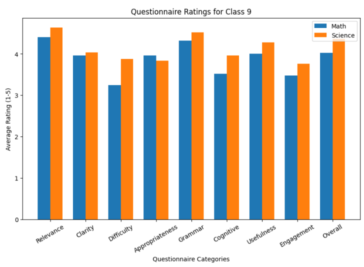

Figure 5: Qualitative evaluation for Grade Level 9

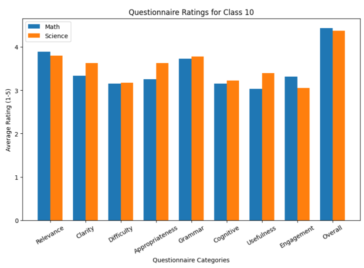

Figure 6: Qualitative evaluation for Grade Level 10

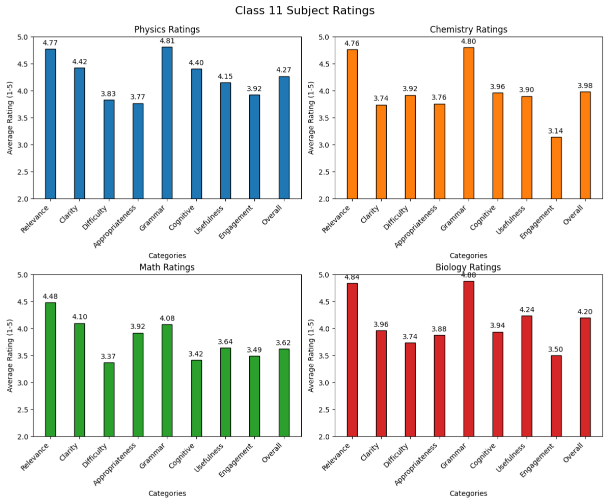

Figure 7: Qualitative evaluation for Grade Level 11

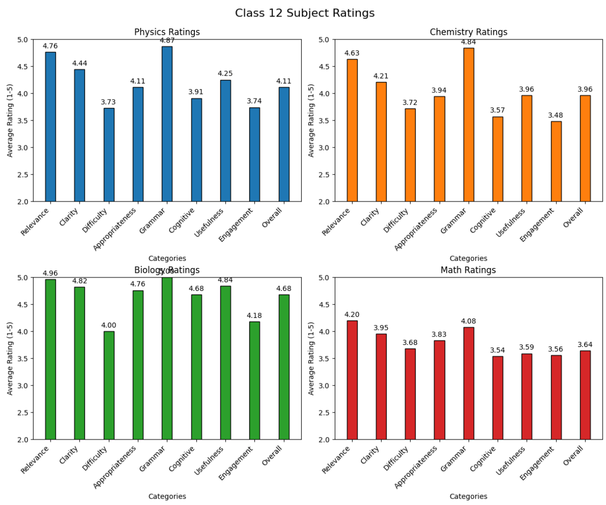

Figure 8: Qualitative evaluation for Grade Level 12

5 Conclusion

Artificial intelligence can tailor questions to individual student needs, helping to address specific learning gaps. Adaptive learning systems can use Generative-AI and Natural Language Processing/LLMs to provide personalized question sets that challenge students appropriately based on their performance.

This study also contributes to the growing body of research exploring the intersection of Artificial Intelligence and Machine Learning in education, specifically focusing on enhancing curriculum development and assessment practices.

References

Chidanand Apte, Leora Morgenstern, and Se June Hong. 2000. Ai at ibm research. *IEEE Intelligent Systems and their applications*, 15(6):51–57.

Bill Cope and Mary Kalantzis. 2016. Big data comes to school: Implications for learning, assessment, and research. *Aera Open*, 2(2):2332858416641907.

Krishnendu Ghosh. 2022a. *Augmenting learning materials to support integrated and multimodal learning*. Ph.D. thesis, IIT Kharagpur.

Krishnendu Ghosh. 2022b. Remediating textbook deficiencies by leveraging community question answers. *Education and Information Technologies*, 27(7):10065–10105.

Krishnendu Ghosh and Sayantan Das. 2023. Automatic generation of algebraic representation for physics problems. *Authorea Preprints*.

Ira Goldstein and Seymour Papert. 1977. Artificial intelligence, language, and the study of knowledge. *Cognitive science*, 1(1):84–123.

Daniele Gravina, Ahmed Khalifa, Antonios Liapis, Julian Togelius, and Georgios N. Yannakakis. 2019. Procedural content generation through quality diversity. In *2019 IEEE Conference on Games (CoG)*, pages 1–8.

Larry Hardesty. 2019. The history of amazon's recommendation algorithm. *Amazon Science*, 22.

Dan Hendrycks, Collin Burns, Saurav Kadavath, Akul Arora, Steven Basart, Eric Tang, Dawn Song, and Jacob Steinhardt. 2021. Measuring mathematical problem solving with the math dataset. *NeurIPS*.

Patrick Lewis, Ethan Perez, Aleksandra Piktus, Fabio Petroni, Vladimir Karpukhin, Naman Goyal, Heinrich Küttler, Mike Lewis, Wen-tau Yih, Tim Rocktäschel, et al. 2020. Retrieval-augmented generation for knowledge-intensive nlp tasks. *Advances in Neural Information Processing Systems*, 33:9459–9474.

Jiaxiang Li, Siliang Zeng, Hoi-To Wai, Chenliang Li, Alfredo Garcia, and Mingyi Hong. 2024. Getting more juice out of the sft data: Reward learning from human demonstration improves sft for llm alignment. *arXiv preprint arXiv:2405.17888*.

Elizabeth D Liddy. 2001. Natural language processing.

Owen HT Lu, Anna YQ Huang, Danny CL Tsai, and Stephen JH Yang. 2021. Expert-authored and machine-generated short-answer questions for assessing students learning performance. *Educational Technology & Society*, 24(3):159–173.

J McCarthy, DG Bobrow, DC Luckham, ML Minsky, RK Brayton, K Maling, N Rochestert, L Hodes, DMR Park, CE Shannon, et al. 1959. Xiii. artificial intelligence. *Quarterly Progress Report*, (53):122.

David Mhlanga. 2023. Open ai in education, the responsible and ethical use of chatgpt towards lifelong learning. In *FinTech and artificial intelligence for sustainable development: The role of smart technologies in achieving development goals*, pages 387–409. Springer.

Prakash M Nadkarni, Lucila Ohno-Machado, and Wendy W Chapman. 2011. Natural language processing: an introduction. *Journal of the American Medical Informatics Association*, 18(5):544–551.

Mieczysław L Owoc, Agnieszka Sawicka, and Paweł Weichbroth. 2019. Artificial intelligence technologies in education: benefits, challenges and strategies of implementation. In *IFIP International Workshop on Artificial Intelligence for Knowledge Management*, pages 37–58. Springer.

Francesc Pedro, Miguel Subosa, Axel Rivas, and Paula Valverde. 2019. Artificial intelligence in education: Challenges and opportunities for sustainable development.

Alan Mathison Turing. 1950. Mind. *Mind*, 59(236):433–460.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. *Advances in neural information processing systems*, 30.

Yu-Xiong Wang, Deva Ramanan, and Martial Hebert. 2017. Growing a brain: Fine-tuning by increasing model capacity. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2471–2480.

Ke Zhang and Ayse Begum Aslan. 2021. Ai technologies for education: Recent research & future directions. *Computers and Education: Artificial Intelligence*, 2:100025.