Key Technology Analysis and Educational Application Research of Question Answering System under the Background of ChatGPT

Xinhan Pan [1], *

[1] The School of Information Management and Artificial Intelligence, Zhejiang University of Finance & Economics, Hangzhou, 310018, China

*200110900212@zufe.edu.cn

Abstract. The advent of advanced artificial intelligence(AI) models such as ChatGPT has brought a transformative impact on various sectors, with the field of education being a notable one. Given the system's impressive efficiency and accuracy in answering questions, ChatGPT has garnered significant attention as a prospective tool for knowledge acquisition and learning tutoring. Traditional educational methodologies, while effective in many respects, exhibit certain limitations, particularly in personalized learning and scalability. These limitations underline the necessity for innovative solutions, a role that AI-powered question answering systems like ChatGPT are well-positioned to fulfill. This review aims to delve into an in-depth evaluation of the potential and applications of question answering systems in the educational landscape. The objective is not merely to identify the possibilities, but also to thoroughly investigate the key technologies and methodologies underpinning these systems. From machine learning(ML) algorithms and natural language processing to user interaction design and data privacy safeguards, various facets of these question answering systems will be examined.

Keywords: AI, ChatGPT, Question-Answering system, LLM, Education

1 Introduction

AI is increasingly reshaping societal norms and practices, with advancements like ChatGPT, a groundbreaking large-scale language model (LLM), exemplifying this trend. ChatGPT's potential for addressing complex queries and replacing highly repetitive tasks is indicative of its revolutionary impact [1]. This manuscript intends to unpack the pivotal technologies underpinning question answering systems, contextualized within ChatGPT, and delve into their prospective applications in the educational sector.Initially, the paper elucidates the basic principles and architectural components of ChatGPT, then delves into its strengths and constraints with respect to question answering tasks. Subsequently, a comprehensive review of vital technologies

P. Kar et al. (eds.), *Proceedings of the 2023 International Conference on Image, Algorithms and Artificial Intelligence (ICIAAI 2023)*, Advances in Computer Science Research 108,
https://doi.org/10.2991/978-94-6463-300-9_42

integral to question answering systems, such as Automatic Speech Recognition(ASR) and Natural Language Processing(NLP), is undertaken.

Through the lens of these key techniques, the manuscript uncovers salient challenges and proffers solutions for constructing efficient and accurate question answering systems specifically tailored for educational purposes, all within the context of ChatGPT. Concluding the paper is an in-depth exploration of the potential applications of question answering systems within the education domain. In particular, the focus is placed on their possibilities for enhancing learning assistance, facilitating personalized instruction, and promoting effective knowledge acquisition [2]. This examination underscores the paper's commitment to fostering an enriched understanding of AI's transformative role in education..

2 Principle Overview

2.1 Basic Principles and Classification

Question-answering systems can be bifurcated into two categories: voice-based and text-based systems. Text-based Question Answering Systems extract answers from a predefined knowledge base, document, or corpus by analyzing and processing text-based input questions. NLP and information retrieval techniques are commonly used to comprehend the questions, retrieve pertinent information, and generate responses [3].

Rule-based question answering systems adhere to predefined rules and pattern matching to recognize and answer particular question types. The answers in these systems are static and predefined. Statistical-based question answering systems leverage statistical methods and ML techniques to generate responses by analyzing large corpuses to learn associations between questions and answers. These systems can learn and enhance their performance based on training data [4]. Knowledge graph-based question-answer systems utilize structured knowledge representations such as knowledge graphs to facilitate question understanding and answer generation. These systems can extract relevant information from the knowledge graph and apply the relevant information to question answering.

Voice-based Question Answering Systems principally use ASR technology to convert voice input into text, which is then fed into a text-based question-answering system to generate answers. These systems also require voice synthesis technology to convert the generated text responses into voice output. Text-to-Speech (TTS) question answering systems convert text responses to voice output, allowing users to receive answers in verbal form. Speech-to-text (STT) question answering systems convert user voice input into text through voice recognition technology, and then apply a text-based question answering system to process and answer the questions. The transformative question answering system, ChatGPT, is based on the Transformer architecture. This deep neural network architecture, proficient in handling natural language tasks, uses self-attention mechanisms [5].

During the training of ChatGPT, OpenAI prepared extensive dialogue datasets as input, which were sourced from online chat logs, social media platforms, or domain-specific dialogue data. OpenAI subsequently constructed a dialogue generation model based on the Transformer architecture.The Transformer model revolves around self-attention mechanisms and feed-forward neural networks. Self-attention mechanisms calculate the relevance between different positions in the input sequence and aggregate information through weighted sums, enabling the model to capture global semantic relationships. The feed-forward neural networks further process and enrich the input through non-linear transformations. A voice-based question-answering system fosters a more natural and interactive learning experience, thereby enhancing student engagement and interest. OpenAI uses a self-supervised learning approach during ChatGPT's training to maximize the predicted probabilities of the next response. The model predicts the next dialogue response based on the previous dialogue history. Residual connections and layer normalization assist the model in learning and optimization. Additionally, position encodings represent positional information in the input and output sequences. The training process updates the model's parameters using gradient descent optimization and cross-entropy loss functions, gradually approximating the optimal solution.

2.2 Related Technologies

NLP, ML, and Deep Learning(DL) technologies are instrumental to question answering systems. These technologies merge to equip question answering systems with remarkable capabilities for understanding natural language, reasoning, and generating accurate responses.

NLP, a field focused on enabling computers to comprehend and process human language, incorporates techniques such as text analysis, lexical analysis, syntactic analysis, and semantic analysis. These techniques are designed to assist computers in understanding and generating human language. ML, a method enabling computers to learn and enhance performance through data, has wide applications in question answering systems. Here, ML algorithms can train models capable of understanding questions and generating accurate answers. Common ML techniques include classification, clustering, and sequence models. DL, a subset of ML, leverages neural network models for advanced data representation and feature learning. In question answering systems, DL is widely used, especially pre-trained models based on large-scale corpora, such as the Transformer model. These models facilitate more accurate and contextually relevant response generation. The attention mechanism, a pivotal technique in DL, enables models to handle long text sequences better. It allows the model to focus more on the relevant information in the input when generating responses, thereby improving the accuracy and coherence of the responses [6].

Word embedding is another technique mapping words into a low-dimensional vector space. Word embedding can map semantically similar words to similar vectors, providing a superior semantic representation. In question answering systems, word embedding aids in understanding questions and constructing semantically relevant responses.

ASR is another crucial technology used to convert speech input into textual representations, thereby enabling question answering systems to process and answer voice questions, as seen in Siri and Cortana.ASR technology employs components like acoustic models, language models, and pronunciation dictionaries to convert speech signals into corresponding text representations. Acoustic models analyze and model the speech signal, language models model and evaluate possible word sequences, while pronunciation dictionaries provide the correspondence between words and their pronunciation. Utilizing ASR technology, the question answering system can receive a user's speech input and convert the speech into a textual form for questions. This ability allows users to interact verbally with the question answering system, thus enhancing the system's usability and user experience.

The amalgamation of these techniques equips question answering systems with the robust capability to handle complex natural language questions and generate accurate and coherent responses. In practice, other related techniques and methods, such as named entity recognition, sentiment analysis, and logical reasoning, can further augment the system's performance and intelligence [7].

2.3 Introduction to ChatGPT and Other Intelligent Question and Answer Systems

There are a variety of question-answering systems available, and whether they are text-based or voice-based, these systems have been able to satisfy the needs of a sizable percentage of people.

A language model powered by AI, ChatGPT is built on the Transformer architecture and is intended to have lively, coherent discussions. In order to gain a thorough knowledge of linguistic patterns, semantic linkages, and contextual signals, ChatGPT makes use of a massive pretraining procedure on a variety of text corpora. ChatGPT excels at interactive question answering scenarios because it generates responses that fit the context, giving the impression of a natural conversation [8].

There are a number of additional intelligent question answering systems that have significantly advanced the area in addition to ChatGPT. Watson, Bard, Siri, and Cortana are a few noteworthy examples. To comprehend user inquiries and produce pertinent solutions, these systems make use of a range of strategies and tactics.

Microsoft's intelligent personal assistant Cortana uses ML and NLP to offer specialized support across a variety of devices. Cortana interfaces with numerous Microsoft services and products and provides voice-based interaction [9].

Siri is an intelligent virtual assistant created by Apple that uses a combination of speech recognition, natural language understanding, and ML techniques to reply to user orders and enquiries. Siri offers a variety of services, including creating reminders, sending messages, and giving advice and information.

A chatbot named Bard is built using a substantial language model. The approach employed by Bard is based on LaMDA (Language Models for Dialog Applications), which is created by Google.

3 Application in the Education Industry

3.1 Requirements and Current Situation

The existing method of traditional classroom instruction frequently lacks involvement. The typical teacher-led instructional approach places the teacher in the forefront, dispensing knowledge while the pupils act as passive recipients. The engagement and communication between students, teachers, and other students are restricted by this one-way delivery.

The following issues may result from inadequate interactivity:

Low interest among students Students may lose interest in studying and lack initiative and drive if there aren't enough possibilities for contact and engagement.

The needs of each individual are not met: Different students have various learning preferences, rates of learning, and levels of comprehension. Some students may struggle or advance more slowly as a result of ineffective instructional methods that fail to take into account the needs and unique variations of each student.

The depth of learning outcomes is insufficient. Interactive teaching encourages students to think critically, ask questions, and explore. Students' capacity to gain a deeper understanding and apply knowledge may be constrained by a lack of interactive training.

Lack of timely input: Interactive training can let students and instructors communicate in real-time for feedback and direction. In a classroom with minimal interaction, students might not get prompt responses to their inquiries or evaluations of how well they are learning.

Considering the effects of the COVID-19 outbreak. Due to the quarantine policy, a sizable number of students were required to attend online classes. According to this study's investigation, educational disparities were further exacerbated in online classes, as suggested by Dr. Orla Doyle [1] and COVID-19. Children with lower socioeconomic status may be most impacted by school closings and the ensuing loss of instructional time, according to one study.Since there aren't enough resources for education in general, this study demonstrated how variations in educational resources at home might exacerbate inequality. The gap in home education resources can often lead to a gap in a student's overall literacy, and the gap in home education during COVID-19 determines the gap in the level of education a child receives. Teachers are present at school to offer guidance, but they cannot be with every student at all times.As long as home education levels differ among students in a school with a similar level of education, the gap between them will widen. To further emphasize this point, consider that in countries with limited educational resources, such as India, where there are frequently one to two or three hundred students per class, the availability of home education is even more limited, the country's rate of literacy growth is inevitably lower than in places with abundant educational resources. Such an educational system is bound to make it impossible to foster skills due to the lack of customized assistance for pupils, the inability to address their questions, and other issues [10].

But in recent years, the question-answering technology could be able to find a solution to this issue.

3.2 Application Analysis

There is research that shows ChatGPT outperforms other comparable models in all language tests, particularly in low-resource languages, where its performance advantage is more significant, demonstrating that ChatGPT can be used for basic instruction in the majority of countries around the world. The majority of the current question-answering systems on the market have been shown to be capable.

Most question-answering systems currently available have been shown to be effective, and there is research on them. For example, the ChatGPT performs better than other comparison models in all language tests, particularly in low-resource languages, where its performance advantage is more pronounced, demonstrating that ChatGPT can be used for basic instruction in the majority of the world.

Since code work was once thought to be the most creative work, the addition of code training has also significantly improved ChatGPT's capacity for logical reasoning. In terms of performance comparisons, ChatGPT outperforms other models in most problem types, with the exception of cause and effect (why) and date (date), where its performance advantage is more pronounced. What is known is that OpenAI is continually enhancing ChatGPT's performance, so the precision of question-answering tools like ChatGPT is certain to increase going forward. This will make ChatGPT the ideal teaching tool—a question-answering tool that knows everything, is cautious and patient, and has strong logical reasoning.The finest instructor will unquestionably be a system like this one for answering questions.

Of course, this question-answering system's capacity for error-correcting coding was examined. The author of this post then uses his own tests to show the effectiveness of ChatGPT for code mentoring.Programmers can uncover problems by using ChatGPT to help with debugging, foresee bugs, and explain difficulties. Due to its understanding of and ability to analyze code fragments, as well as its knowledge representation and natural language generating abilities, it is perfectly suited for these tasks.

Teachers can use ChatGPT in a number of ways to bolster and enhance their pedagogical and assessment methods, according to Herft. He provides us with some good options for use. For instance, teachers can use ChatGPT's capabilities to develop open-ended question prompts that are consistent with the learning objectives and success standards of the unit of instruction. Furthermore, ChatGPT can be used to create effective rubrics that succinctly and clearly outline what students must do to succeed at the various required levels of proficiency. Once more, educators can use ChatGPT to develop "prompts for formative assessment activities that provide ongoing feedback to inform teaching and learning".

3.3 Future Application Prospects

According to the analysis presented above, it is evident that intelligent question-answering systems will undoubtedly have a significant impact on the education sector. However, it is also undeniable that existing intelligent question-answering systems suffer from numerous drawbacks, and the authors have noted that while ChatGPT can be a useful tool for resolving programming errors, it is not a perfect solution.

Based on my own testing, ChatGPT struggles to answer questions of a subjective nature, just like other question-answering systems may struggle to give accurate and pertinent answers to questions that are subjective, emotional, or value-based. The quality of ChatGPT output will depend on the quality of the training data and the system's design. These inquiries entail individual viewpoints, emotional states, and cultural variations and call for more flexible and subjective reasoning. For instance, if I ask ChatGPT what language is the best in the world, it will directly respond with " AI language model without offering any subjective opinion. "

Firstly, technical biases and defects may occur within question-answering systems. These systems are designed using training data, so any flaws or biases in the data can significantly impact the results. Reviewing and improving the training data is an essential part of risk management strategies to reduce the potential for biases and errors. A functional test of ChatGPT revealed a tendency to generate biased statements about China. It does not refrain from or resist making political statements about the country, even though the model does not provide analysis on certain political events. This indicates the training data was not screened sufficiently for such statements [5]. Thus, risk management is not yet fully implemented in some areas, presenting serious potential threats to various nations.

Question-answering systems might also handle private or sensitive information. In responding to queries, the system could collect, store, or transmit personal data about the user. Therefore, risk management should incorporate privacy protection measures to ensure the security and confidentiality of user data. Moral and legal concerns may also arise with question-answering systems. They could provide information on legal situations, medical conditions, or other specialized topics that often require expert training and knowledge. These limitations should be taken into account in risk management to ensure the system delivers information with appropriate caveats and disclaimers. It's crucial to develop standards for utilizing models like ChatGPT in scientific research, considering accountability, integrity, transparency, and honesty, as highlighted thoroughly by van Dis et al. Future advancements in question-answering technologies will likely enable effective risk management, privacy safeguarding for users, and objective answer generation, while also enhancing multimodal interactivity..

4 Conclusion

This paper commences with an analysis of ChatGPT's key technologies,related technologies and classifications of question-answering system, shedding light on the issue of unequal educational resources in the modern era, a problem further exacerbated during the COVID-19 pandemic.This paper stresses that the current state of education fails to address the broad needs of the populace. The review proceeds to scrutinize the reasons behind the aptitude of existing question-answering systems in satisfying the requirements of the educational sector, alongside an exploration of potential risk factors. Justifications for future regulations governing question-answering systems are also provided, in conjunction with a gaze into their future. Further research is warranted in this promising field to ascertain the future direction of question-answering systems in education and to enhance their effectiveness.In the future, the potential value of question answering systems in education will provide valuable insights and references for the development and downstream research of the LLM represented by ChatGPT.

Reference

1. Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., ... & Amodei, D. (2020). Language models are few-shot learners. arXiv preprint arXiv:2005.14165.
2. Lewis, P., Goyal, N., Ghazvininejad, M., Mohamed, A., Levy, O., Stoyanov, V., & Zettlemoyer, L. (2020). Retrieval-augmented generation for knowledge-intensive nlp tasks. arXiv preprint arXiv:2005.11401.
3. Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., ... & Stoyanov, V. (2019). RoBERTa: A robustly optimized BERT pretraining approach. arXiv preprint arXiv:1907.11692.
4. Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., & Sutskever, I. (2019). Language models are unsupervised multitask learners. OpenAI blog, 1(8), 9.
5. Reddy, S., Täckström, O., Collins, M., & Petrov, S. (2019). On the compatibility of constituent and dependency parsers. Transactions of the Association for Computational Linguistics, 7, 59-70.
6. Roberts, A., Reeves, R., Gardner, M., & others. (2020). How Much Knowledge Can You Pack Into the Parameters of a Language Model? arXiv preprint arXiv:2002.08910.
7. Talmor, A., & Berant, J. (2020). MultiQA: An Empirical Investigation of Generalization and Transfer in Reading Comprehension. arXiv preprint arXiv:2005.00247.
8. Wang, A., Singh, A., Michael, J., Hill, F., Levy, O., & Bowman, S. R. (2019). GLUE: A multi-task benchmark and analysis platform for natural language understanding. arXiv preprint arXiv:1804.07461.
9. Wellner, B., Talmor, A., & Berant, J. (2020). Comparing prize-collecting and profit-maximization for training reading comprehension models. arXiv preprint arXiv:2005.10219.
10. Yang, Z., Dai, Z., Yang, Y., Carbonell, J. G., Salakhutdinov, R., & Le, Q. V. (2019). Xlnet: Generalized autoregressive pretraining for language understanding. arXiv preprint arXiv:1906.08237.