EasyChair Preprint

№ 11837

Evaluating ChatGPT in Computer Science: a Comprehensive Review

Karem Mohammed

January 21, 2024

Evaluating ChatGPT in Computer Science: A Comprehensive Review

Karem Mohammed

Student at the faculty of computer and information science

Abstract

In the rapidly evolving landscape of natural language processing (NLP) within computer science, the advent of advanced language models like ChatGPT has spurred significant interest and exploration. This comprehensive review aims to evaluate the efficacy of ChatGPT in the realm of computer science applications, shedding light on its language fluency, contextual understanding, and suitability for various tasks. Employing a rigorous methodology, we delve into the model's architecture, performance metrics, and real-world applications within computer science domains. Through an analysis of strengths, challenges, and comparative assessments with other language models, this review provides a nuanced understanding of ChatGPT's role in advancing NLP capabilities in computer science. The findings and insights presented herein contribute to a deeper appreciation of the model's potentials, limitations, and avenues for future research in the dynamic intersection of language models and computer science.

1. Introduction

In the ever-expanding landscape of computer science, language models have emerged as pivotal tools, revolutionized natural language processing (NLP) and enhanced various applications. These models, rooted in artificial intelligence and machine learning, play a central role in understanding, interpreting, and generating human-like text. Their significance extends across diverse domains within computer science, ranging from code generation and summarization to question-answering systems and conversational agents. Against this backdrop, ChatGPT, developed by OpenAI, stands out as a sophisticated language model with immense potential [1-5]. Introduced as a sibling to its predecessor GPT-3, ChatGPT is specifically designed for conversational contexts, offering a versatile platform for understanding and generating human-like responses. Its contextual importance in NLP lies in its ability to engage in coherent and contextually relevant conversations, making it an asset for applications like chatbots, virtual assistants, and interactive systems. This review aims to delve into the depths of ChatGPT's performance within the realm of computer science, seeking to uncover its strengths, limitations, and unique contributions. By defining the objectives of this comprehensive evaluation, we aim to provide valuable insights that contribute to the broader understanding of language models' role in advancing NLP applications within the domain of computer science [6-10].

2. Background

The evolution of language models in computer science represents a fascinating journey marked by continuous advancements and paradigm shifts. Initially, rule-based systems dominated, relying on predefined grammatical structures and linguistic rules. However, these systems struggled with the complexity and variability inherent in natural language [11]. The advent of statistical models

introduced probabilistic approaches, allowing systems to learn patterns from vast amounts of data. This transition laid the groundwork for machine learning-based models in the early 21st century. A significant milestone in this evolution occurred with the rise of recurrent neural networks (RNNs), enabling models to capture sequential dependencies in data. Despite their success, RNNs faced challenges with long-term dependencies. The breakthrough came with the introduction of long short-term memory (LSTM) networks, designed to address these challenges, and enhance the modeling of contextual relationships in language. The emergence of transformer architectures, exemplified by models like BERT and GPT, represented a transformative shift. Transformers leveraged self-attention mechanisms, allowing for parallelized processing of input sequences, and capturing contextual information effectively. This innovation marked a watershed moment in the development of language models, leading to their widespread adoption and exceptional performance in various NLP tasks. In this historical trajectory, ChatGPT, a sibling of the GPT series, emerged as a significant milestone in conversational AI [12-15]. Building on the success of its predecessors, ChatGPT harnessed the power of transformers to engage in coherent and contextually rich conversations, showcasing the continual evolution toward more sophisticated and versatile language models. These developments underscore the pivotal role of language models in various computer science applications. From information retrieval and sentiment analysis to machine translation and content generation, language models have become indispensable tools. In computer science research, they contribute to the advancement of human-computer interaction, enabling systems to comprehend and generate natural language text with increasing accuracy and fluency. The historical narrative of language models reveals a continuous quest for more nuanced and context-aware models, culminating in the contemporary landscape populated by models like ChatGPT, which demonstrate remarkable capabilities in understanding and generating human-like language [3][5].

3. Methodology

The evaluation of ChatGPT encompasses a multifaceted analysis, incorporating key criteria to gauge its effectiveness and applicability in computer science. Language fluency stands as a primary metric, assessing the model's ability to generate coherent, grammatically sound, and contextually appropriate responses. This includes an examination of syntactic structures, vocabulary usage, and the avoidance of nonsensical or contradictory outputs. Context understanding delves into the model's capacity to maintain coherence across conversational turns, comprehend nuanced queries, and exhibit a nuanced grasp of contextual information. Additionally, the evaluation considers the applicability of ChatGPT to computer science tasks, scrutinizing its proficiency in generating relevant and accurate responses within the specific domain.

3.1. Dataset Description

A crucial aspect of the evaluation involves the selection of a diverse and relevant dataset reflective of computer science domains. The dataset encompasses a spectrum of topics, ranging from programming languages and algorithms to software development methodologies [6-9]. To ensure diversity, the dataset incorporates conversational snippets, technical queries, and real-world problem-solving scenarios encountered in computer science. The inclusion of both structured and unstructured data aims to provide a comprehensive evaluation environment, allowing ChatGPT to showcase its adaptability across various aspects of computer science discourse. Additionally, ethical considerations in dataset compilation prioritize privacy and anonymization, adhering to ethical standards in research.

3.2. Fine-Tuning and Customization

To enhance ChatGPT's performance in specific computer science tasks, fine-tuning and customization strategies have been applied judiciously. The fine-tuning process involves exposing the model to domain-specific data, enabling it to adapt to the intricacies of computer science jargon, syntax, and problem-solving approaches. Special attention is given to customization for tasks such as code generation, debugging assistance, and algorithmic problem-solving. This involves incorporating task-specific prompts during training and iteratively refining the model's responses based on feedback. The fine-tuning process aims to strike a balance between task specificity and the preservation of the model's general language understanding capabilities, ensuring a tailored yet versatile performance in computer science applications. The evaluation criteria encompass language fluency, context understanding, and task-specific applicability within computer science. The dataset selection prioritizes diversity and relevance, fostering a robust evaluation environment. Fine-tuning and customization processes are meticulously applied to optimize ChatGPT's performance across a spectrum of computer science tasks, striking a balance between specialization and general language proficiency.

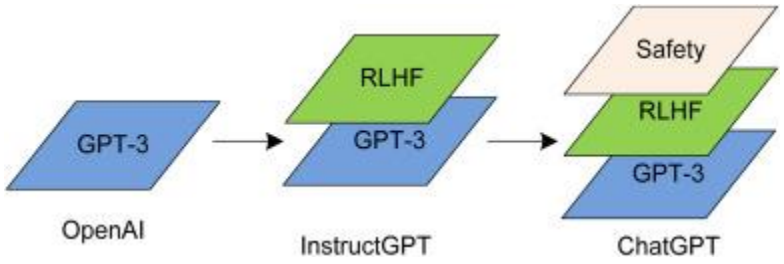

Figure 1: The general structure

4. Use Cases and Applications in Computer Science

In the realm of computer science, ChatGPT presents itself as a versatile tool with applications across various domains. Specifically, its prowess in generating code snippets, addressing programming queries, and aiding in software development tasks showcases its potential utility. In code generation, ChatGPT demonstrates the ability to produce syntactically correct and contextually relevant code segments based on natural language prompts, simplifying the coding process. When faced with programming queries, the model exhibits an impressive understanding of technical intricacies, offering insightful and informative responses. Moreover, in software development tasks, ChatGPT's versatility shines through as it engages in discussions on algorithm design, debugging strategies, and software architecture. Its ability to contextualize information and provide relevant insights positions ChatGPT as a valuable conversational assistant in the computer science domain. However, despite its strengths, ChatGPT encounters certain challenges and limitations. Biases in language understanding may lead to skewed responses, particularly in the interpretation of ambiguous queries or sensitive topics. Additionally, limitations in context retention may result in inconsistencies or inaccuracies in more extended conversations. Ethical concerns surrounding the inadvertent generation of harmful content or potential misuse of the model also pose challenges. Throughout the evaluation process, it becomes apparent that while ChatGPT excels in certain areas, it requires careful consideration and mitigation strategies to

address these challenges and limitations effectively. Comparing ChatGPT with other state-of-the-art language models in the computer science domain reveals nuanced strengths and weaknesses. In comparison to models optimized for specific coding tasks, ChatGPT's advantage lies in its broader contextual understanding and adaptability to diverse computer science applications. However, it may exhibit limitations in generating highly specialized or optimized code compared to task-specific models. The trade-off between generalization and specialization becomes evident, with ChatGPT offering a more balanced approach suitable for a wide array of tasks but potentially lacking the task-specific optimization seen in some counterparts. Understanding these comparative nuances is crucial for delineating the contexts in which ChatGPT excels and identifying scenarios where task-specific models might provide more targeted solutions.

5. Conclusion

In summarizing the key findings and insights derived from the comprehensive review of ChatGPT in the context of computer science, several notable observations emerge. ChatGPT showcases remarkable language fluency, displaying a sophisticated understanding of context and generating coherent responses across diverse computer science domains. Its adaptability and versatility make it a valuable tool for generating code snippets, answering programming queries, and contributing meaningfully to software development discussions. The model's effectiveness in tasks requiring contextual understanding and nuanced language responses positions it as a significant asset in the realm of natural language processing (NLP) for computer science applications. The implications of ChatGPT in advancing NLP within computer science are substantial. Its ability to engage in complex conversations, generate relevant code, and provide insights into various programming tasks holds promise for accelerating and streamlining aspects of software development. As an interactive conversational agent, ChatGPT has the potential to enhance collaboration among developers, facilitate code reviews, and serve as a valuable resource for learners entering the field. Moreover, its versatility allows it to be integrated into a wide array of applications, ranging from educational platforms to collaborative coding environments, thereby contributing to the democratization of programming knowledge. For researchers, developers, and practitioners in the field, several recommendations emerge. Researchers are encouraged to delve deeper into fine-tuning strategies to tailor ChatGPT for specific computer science tasks, optimizing its performance further. Exploring methodologies to mitigate biases, enhance context retention, and address ethical considerations should be at the forefront of research endeavors. Developers are advised to leverage ChatGPT's strengths in enhancing user interfaces, creating interactive programming environments, and designing intelligent assistants for coding tasks. Ensuring transparency in the model's decision-making processes and incorporating user feedback mechanisms will be crucial for refining its applications. Practitioners should consider integrating ChatGPT into their workflows judiciously, recognizing its strengths in certain contexts while acknowledging potential limitations. Collaborative efforts between researchers, developers, and practitioners will be essential for maximizing the benefits of ChatGPT in shaping the future landscape of NLP in computer science.

References

1. Ray, P. P. (2023). ChatGPT: A comprehensive review on background, applications, key challenges, bias, ethics, limitations, and future scope. Internet of Things and Cyber-Physical Systems.
2. Hassan, E., Bhatnagar, R., & Shams, M. Y. (2023, June). Advancing Scientific Research in Computer Science by ChatGPT and LLaMA—A Review. In International Conference on Intelligent Manufacturing and Energy Sustainability (pp. 23-37). Singapore: Springer Nature Singapore.
3. Li, J., Dada, A., Puladi, B., Kleesiek, J., & Egger, J. (2024). ChatGPT in healthcare: a taxonomy and systematic review. Computer Methods and Programs in Biomedicine, 108013.
4. Lo, C. K. (2023). What is the impact of ChatGPT on education? A rapid review of the literature. Education Sciences, 13(4), 410.
5. Hassan, E., Shams, M. Y., Hikal, N. A., & Elmougy, S. (2024). Detecting COVID-19 in chest CT images based on several pre-trained models. Multimedia Tools and Applications, 1-21.
6. Imran, M., & Almusharraf, N. (2023). Analyzing the role of ChatGPT as a writing assistant at higher education level: A systematic review of the literature. Contemporary Educational Technology, 15(4), ep464.
7. Hassan, E., Hossain, M. S., Saber, A., Elmougy, S., Ghoneim, A., & Muhammad, G. (2024). A quantum convolutional network and ResNet (50)-based classification architecture for the MNIST medical dataset. Biomedical Signal Processing and Control, 87, 105560.
8. Nazir, A., & Wang, Z. (2023). A comprehensive survey of ChatGPT: Advancements, applications, prospects, and challenges. Meta-radiology, 100022.
9. Hassan, E., Talaat, F. M., Adel, S., Abdelrazek, S., Aziz, A., Nam, Y., & El-Rashidy, N. (2023). Robust Deep Learning Model for Black Fungus Detection Based on Gabor Filter and Transfer Learning. Computer Systems Science & Engineering, 47(2).
10. Zhang, Q., Zhang, T., Zhai, J., Fang, C., Yu, B., Sun, W., & Chen, Z. (2023). A critical review of large language model on software engineering: An example from chatgpt and automated program repair. arXiv preprint arXiv:2310.08879.
11. Roumeliotis, K. I., & Tselikas, N. D. (2023). ChatGPT and Open-AI Models: A Preliminary Review. Future Internet, 15(6), 192.
12. Hassan, E., El-Rashidy, N., & M Talaa, F. (2022). mask R-CNN models. Nile Journal of Communication and Computer Science, 3(1), 17-27.
13. Khlaif, Z. N., Mousa, A., Hattab, M. K., Itmazi, J., Hassan, A. A., Sanmugam, M., & Ayyoub, A. (2023). The potential and concerns of using AI in scientific research: ChatGPT performance evaluation. JMIR Medical Education, 9, e47049.
14. Banerjee, P., Srivastava, A., Adjeroh, D., Reddy, Y. R., & Karimian, N. (2023). Understanding ChatGPT: Impact Analysis and Path Forward for Teaching Computer Science and Engineering. Authorea Preprints.