Annals of Biomedical Engineering (2023) 51:2647–2651
https://doi.org/10.1007/s10439-023-03284-0

BMES BIOMEDICAL ENGINEERING SOCIETY

LETTER TO THE EDITOR

ChatGPT, Bard, and Large Language Models for Biomedical Research: Opportunities and Pitfalls

Surendrabikram Thapa[1] · **Surabhi Adhikari**[2]

Received: 10 June 2023 / Accepted: 13 June 2023 / Published online: 16 June 2023

Abstract

Large Language Models (LLMs) such as ChatGPT and Bard have emerged as groundbreaking interactive chatbots, capturing significant attention and transforming the biomedical research landscape. These powerful tools offer immense potential for advancing scientific inquiry, but they also present challenges and pitfalls. Leveraging large language models, researchers can streamline literature reviews, summarize complex findings, and even generate novel hypotheses, enabling the exploration of uncharted scientific territories. However, the inherent risk of misinformation and misleading interpretations underscores the critical importance of rigorous validation and verification processes. This article provides a comprehensive overview of the current landscape and delves into the opportunities and pitfalls associated with employing LLMs in biomedical research. Furthermore, it sheds light on strategies to enhance the utility of LLMs in biomedical research, offering recommendations to ensure their responsible and effective implementation in this domain. The findings presented in this article contribute to the advancement of biomedical engineering by harnessing the potential of LLMs while addressing their limitations.

Keywords Large language models · Biomedical research · Natural language processing · ChatGPT · Bard

Introduction

In recent years, language models (LMs) have emerged as powerful tools in natural language processing, revolutionizing various fields such as information retrieval, language generation, and machine translation [1]. Transformer-based models like Bidirectional encoder representations from transformers (BERT) have demonstrated remarkable capabilities in understanding and generating human language. However, the recent emergence of large language models (LLMs) represents a transformative leap in the capabilities of language processing systems [2].

Large language models, such as ChatGPT and Bard, have taken the field by storm, propelling the boundaries of what language models can achieve [3]. These models are built upon the foundations laid by previous language models like BERT, capitalizing on their successes and expanding their capabilities. Large language models stand out for their massive scale, utilizing significant computational resources and extensive training data [4]. By training on vast corpora of text, often comprising millions or billions of sentences, LLMs acquire an unparalleled understanding of human language.

Unlike their predecessors, LLMs possess extensive knowledge and understanding of diverse topics, gained through pre-training on large text corpora [5]. This breakthrough lies in their ability to generate coherent and contextually relevant responses. Researchers and developers have recognized the potential of LLMs and embraced their application in various domains, including biomedical research.

Large language models have promising prospects in the biomedical research landscape. With their enhanced language comprehension, they can efficiently navigate through the scientific literature, assisting researchers in literature reviews and information synthesis. These models enable faster literature reviews, key finding extraction, and even the generation of novel hypotheses. Large language models offer researchers efficient access to relevant knowledge, accelerating scientific discovery and facilitating interdisciplinary collaborations.

Associate Editor Stefan M. Duma oversaw review of this article.

✉ Surendrabikram Thapa
surendrabikram@vt.edu

[1] Virginia Tech, Blacksburg, VA 24060, USA

[2] Columbia University, New York, NY 10027, USA

However, along with the immense opportunities, the utilization of LLMs in biomedical research also presents challenges and pitfalls. The vastness of knowledge encoded within these models comes with the inherent risk of misinformation, biased interpretations, and a lack of transparency in decision-making processes. Rigorous validation and verification processes become paramount to ensure the reliability and accuracy of the insights generated by LLMs.

This article aims to provide a comprehensive overview of the current landscape of LLMs in biomedical research. It delves into the opportunities presented by LLMs in biomedical research while addressing the associated pitfalls, emphasizing the need for rigorous validation and verification. Furthermore, the article sheds light on strategies to enhance the utility of LLMs in biomedical research and offers recommendations to ensure their responsible and effective implementation in this domain. By harnessing the potential of LLMs while addressing their limitations, this research contributes to the advancement of biomedical engineering, fostering a deeper understanding of their role in scientific discovery and promoting responsible use in the pursuit of transformative healthcare innovations.

Research Questions

This section introduces the research questions that will guide our exploration of the utilization of large language models (LLMs) in biomedical research. These questions aim to uncover the opportunities, challenges, and strategies associated with employing LLMs in this field. The research questions we will address are as follows:

(a) What are the opportunities presented by LLMs in biomedical research, such as streamlining literature reviews, summarizing complex findings, and generating novel hypotheses?
(b) What are the potential pitfalls and challenges associated with employing LLMs in biomedical research, including the risk of misinformation and misleading interpretations?
(c) What validation and verification processes can be implemented to ensure the reliability and accuracy of insights generated by LLMs in the context of biomedical research?

By exploring these research questions, we aim to gain a deeper understanding of the impact and potential of LLMs in biomedical research.

Opportunities with Large Language Models

Large language models (LLMs) have emerged as a powerful tool with immense potential for advancing biomedical research. In this section, we will explore the specific opportunities that LLMs offer in the biomedical field.

Streamlining Literature Reviews

One of the notable opportunities presented by LLMs is their ability to streamline literature reviews. Traditional literature reviews can be time-consuming and labor-intensive, requiring researchers to sift through numerous scientific articles to extract relevant information. Large language models, with their contextual understanding and knowledge representation, can aid in automating this process by rapidly extracting key information from vast amounts of text data. By identifying relevant studies, extracting pertinent findings, and summarizing them, LLMs significantly reduce the time and effort required for literature review tasks.

Summarizing Complex Findings

Large language models have the capacity to summarize complex biomedical findings, facilitating researchers' grasp and communication of key insights. Scientific articles often contain intricate information, making it challenging for researchers to quickly distill and convey the essential points. Large language models can analyze and synthesize the content, distilling it into concise and coherent summaries that effectively convey crucial scientific knowledge. This capability not only saves time but also enhances the accessibility and dissemination of critical scientific findings.

Generating Novel Hypotheses

Another significant opportunity lies in the potential of LLMs to generate novel hypotheses. Leveraging their pre-trained knowledge on vast text corpora, including scientific literature, LLMs can suggest new research directions and generate innovative hypotheses based on existing scientific understanding. By exploring connections between various concepts and identifying patterns, LLMs provide researchers with fresh insights and avenues to explore. This capability can spark new lines of inquiry, foster interdisciplinary collaborations, and drive innovation in the biomedical field.

Advancing Biomedical Natural Language Processing (BioNLP)

In the field of Biomedical Natural Language Processing (BioNLP), LLMs offer tremendous opportunities for efficient language processing in the biomedical domain [6]. Large language models can contribute to tasks such as named entity recognition, relation extraction, semantic parsing, and biomedical question-answering systems. Their comprehensive understanding of language and biomedical concepts enables researchers to extract and interpret biomedical information from unstructured text, facilitating information retrieval, knowledge discovery, and decision support in the biomedical field.

By capitalizing on these opportunities, researchers can enhance their productivity, access critical knowledge, and explore new frontiers in biomedical research. However, careful consideration must be given to responsible implementation, rigorous validation processes, and ethical considerations to ensure the reliability and integrity of LLM outputs in biomedical research and BioNLP.

In the subsequent sections, we will discuss the potential pitfalls associated with LLMs in biomedical research ("Pitfalls of Large Language Models" section), explore future directions and provide recommendations for their optimal utilization ("Future Directions and Recommendations" section), and conclude our comprehensive examination of LLMs in the context of biomedical engineering, highlighting their opportunities and addressing their limitations ("Conclusion" section).

Pitfalls of Large Language Models

While large language models (LLMs) offer promising opportunities in biomedical research, it is important to acknowledge and address the potential pitfalls associated with their use. In this section, we will delve into the key challenges and limitations that researchers should consider when utilizing LLMs in the biomedical field.

Misinformation and Biased Outputs

One of the primary concerns with LLMs is the risk of generating misinformation or biased outputs. Large language models learn from vast amounts of text data, including both reliable and unreliable sources. Consequently, they may inadvertently produce inaccurate or misleading information. Additionally, if the training data contains biases, such as gender or racial biases present in scientific literature, LLMs can perpetuate and amplify these biases in their generated outputs. Researchers must exercise caution and employ rigorous validation and verification processes to ensure the reliability and accuracy of the information derived from LLMs.

Lack of Explainability and Interpretability

Another challenge associated with LLMs is the lack of explainability and interpretability. These models operate as complex black boxes, making it difficult to understand the reasoning behind their outputs. This lack of transparency can be problematic, especially in critical biomedical decision-making processes. Researchers must be mindful of the limitations of LLMs and seek ways to enhance explainability, interpretability, and transparency to ensure confidence and trust in the generated results.

Ethical Considerations and Data Privacy

Ethical considerations and data privacy are crucial aspects when utilizing LLMs in biomedical research. Researchers must handle sensitive patient information and ensure compliance with privacy regulations. The use of LLMs raises concerns regarding data protection, the potential for re-identification, and the responsible use of patient data. Appropriate data anonymization, secure data storage, and adherence to ethical guidelines are paramount to safeguarding patient privacy and maintaining the trust of research participants.

Bias Amplification and Inequality

Large language models have the potential to amplify existing biases present in the data they are trained on. If the training data contains biases related to demographics, disease prevalence, or treatment outcomes, the generated outputs may reflect and perpetuate these biases. This poses challenges in ensuring equitable and unbiased healthcare outcomes. Researchers must be vigilant in recognizing and mitigating bias in the training data and the generated outputs to promote fairness and inclusivity in biomedical research and healthcare applications.

Limitations in Domain-Specific Knowledge

While LLMs possess extensive pre-trained knowledge, they may have limitations in domain-specific knowledge, especially in rapidly evolving fields of biomedical research. The models' training data may not adequately capture the latest advancements or specialized knowledge in specific biomedical disciplines. Researchers must be mindful of the need for domain expertise and carefully evaluate and validate the outputs of LLMs in the context of specific biomedical domains to ensure accuracy and reliability.

Lack of Access to Up-to-Date Information and Subscription-Based Content

Large language models, such as ChatGPT, have a knowledge cutoff, meaning they lack access to the most recent medical literature beyond their training data, which typically ends at a specific date. Medical knowledge is constantly evolving, and new research studies, treatments, or guidelines may have been published since then. Researchers should verify the information provided by LLMs with current and reputable medical sources. Additionally, LLMs may not have access to subscription-based content, limiting their ability to provide detailed information from specific articles.

Limited Ability to Interpret Complex Medical Studies and Navigate Search Strategies

Medical literature can be highly technical and complex, requiring specialized knowledge to interpret and understand. While LLMs can generate text, they may struggle with comprehending the intricacies of certain studies and accurately interpreting their findings. Moreover, LLMs might not have the ability to guide users through effective search strategies to navigate medical literature.

Future Directions and Recommendations

As large language models (LLMs) continue to advance and be integrated into biomedical research, several future directions and recommendations can enhance their utility and address the challenges they present. In this section, we discuss potential avenues for improvement and provide recommendations for researchers and developers working with LLMs in the biomedical field.

Advancing Model Training and Fine-Tuning

Future research should focus on advancing the training and fine-tuning of LLMs specifically for biomedical applications. Fine-tuning LLMs on domain-specific biomedical data, such as curated biomedical literature or electronic health records, can improve their performance and domain-specific understanding. Additionally, efforts should be made to incorporate the most up-to-date medical literature into LLM training pipelines, allowing them to provide more accurate and current information.

Enhancing Explainability and Transparency

Addressing the lack of explainability and interpretability in LLMs is critical. Researchers should explore methods to make LLMs more transparent and understandable, enabling users to gain insights into the decision-making processes of the models. Techniques such as attention mechanisms, interpretability modules, or model-agnostic explainability methods can shed light on how LLMs arrive at their outputs. Promoting transparency in the development and deployment of LLMs can foster trust and enable researchers to better evaluate and interpret the generated results.

Mitigating Bias and Ensuring Fairness

Efforts to mitigate biases in LLMs should be a priority. Researchers must carefully curate and preprocess training data to minimize biases present in the data and address potential sources of inequality. Regular audits and evaluations should be conducted to detect and rectify biases that may arise during model training and deployment. Collaboration between domain experts, ethicists, and data scientists can help establish guidelines and best practices for developing unbiased LLMs that promote fairness and inclusivity in biomedical research and healthcare.

Collaboration with Healthcare Professionals

Collaboration between LLM developers and healthcare professionals is crucial. Engaging clinicians, biomedical researchers, and other domain experts throughout the development and evaluation process of LLMs can ensure their outputs align with the needs and requirements of the biomedical field. This collaboration can help identify specific biomedical challenges that LLMs can address effectively, guide the development of specialized biomedical language models, and promote responsible and ethical use of LLMs in healthcare settings.

Continuous Validation and Verification

To mitigate the risks of misinformation and inaccuracies, researchers must establish robust validation and verification processes for LLM outputs. This includes comparing LLM-generated information with reliable and up-to-date medical sources, conducting independent evaluations, and seeking expert opinions when necessary. Establishing validation frameworks and guidelines specific to LLMs in biomedical research can ensure the accuracy and reliability of the information derived from these models.

Ethical Considerations and Data Governance

Ethical considerations and data governance should remain at the forefront when using LLMs in biomedical research. Researchers must adhere to strict privacy regulations, protect sensitive patient information, and ensure secure data handling and storage practices. Establishing clear guidelines

and ethical frameworks for the collection, storage, and usage of biomedical data in LLM development and deployment is essential to maintain trust, confidentiality, and respect for patient privacy.

By actively pursuing these future directions and implementing the recommended practices, researchers can unlock the full potential of LLMs in biomedical research while addressing their limitations and ensuring their responsible and effective implementation. This will contribute to transformative advancements in healthcare and accelerate the progress of biomedical engineering.

Conclusion

In conclusion, the utilization of large language models (LLMs) in biomedical research offers exciting opportunities for streamlining literature reviews, summarizing complex findings, and generating novel hypotheses. However, it is important to address challenges such as limited access to up-to-date information, potential biases, and the need for robust validation processes. Future directions include advancing model training to develop domain-specific LLMs, enhancing explainability, mitigating biases, fostering collaboration with healthcare professionals, and implementing validation processes. Additionally, ethical considerations and data governance play a crucial role in responsible LLM implementation. By effectively addressing these challenges and adopting recommended practices, researchers can harness the potential of LLMs to drive transformative advancements in biomedical engineering. This includes the development of domain-specific LLMs tailored to biomedical research, leading to improved healthcare outcomes and scientific discoveries.

Funding None.

Declarations

Conflict of interest The authors hereby declare that no conflicts of interest exist in relation to the subject matter of this manuscript. No benefits, whether in any form or direct/indirect, have been or will be received from any commercial party.

Ethical Approval As this study does not involve the use of individual-level data, ethical approval is not required for this research.

References

1. Adhikari, S., S. Thapa, U. Naseem, H. Y. Lu, G. Bharathy, and M. Prasad. Explainable hybrid word representations for sentiment analysis of financial news. *Neural Netw.* 164:115–123, 2023.
2. Zhao, W. X., K. Zhou, J. Li, T. Tang, X. Wang, Y. Hou, Y. Min, B. Zhang, J. Zhang, Z. Dong, et al. A survey of large language models. arXiv preprint arXiv:2303.18223, 2023.
3. Dwivedi, Y. K., N. Kshetri, L. Hughes, E. L. Slade, A. Jeyaraj, A. K. Kar, A. M. Baabdullah, A. Koohang, V. Raghavan, M. Ahuja, et al. So what if chatgpt wrote it? multidisciplinary perspectives on opportunities, challenges and implications of generative conversational ai for research, practice and policy. *Int. J. Inf. Manag.* 71:102642, 2023.
4. Mbakwe, A. B., I. Lourentzou, L. A. Celi, O. J. Mechanic, and A. Dagan. ChatGPT Passing USMLE Shines a Spotlight on the Flaws of Medical Education. San Francisco: Public Library of Science, 2023.
5. Thapa, S., U. Naseem, and M. Nasim. From humans to machines: can ChatGPT-like LLMs effectively replace human annotators in NLP tasks? In: Workshop Proceedings of the 17th International AAAI Conference on Web and Social Media, 2023.
6. Chen, Q., H. Sun, H. Liu, Y. Jiang, T. Ran, X. Jin, X. Xiao, Z. Lin, Z. Niu, and H. Chen. A comprehensive benchmark study on biomedical text generation and mining with ChatGPT. bioRxiv, 2023.