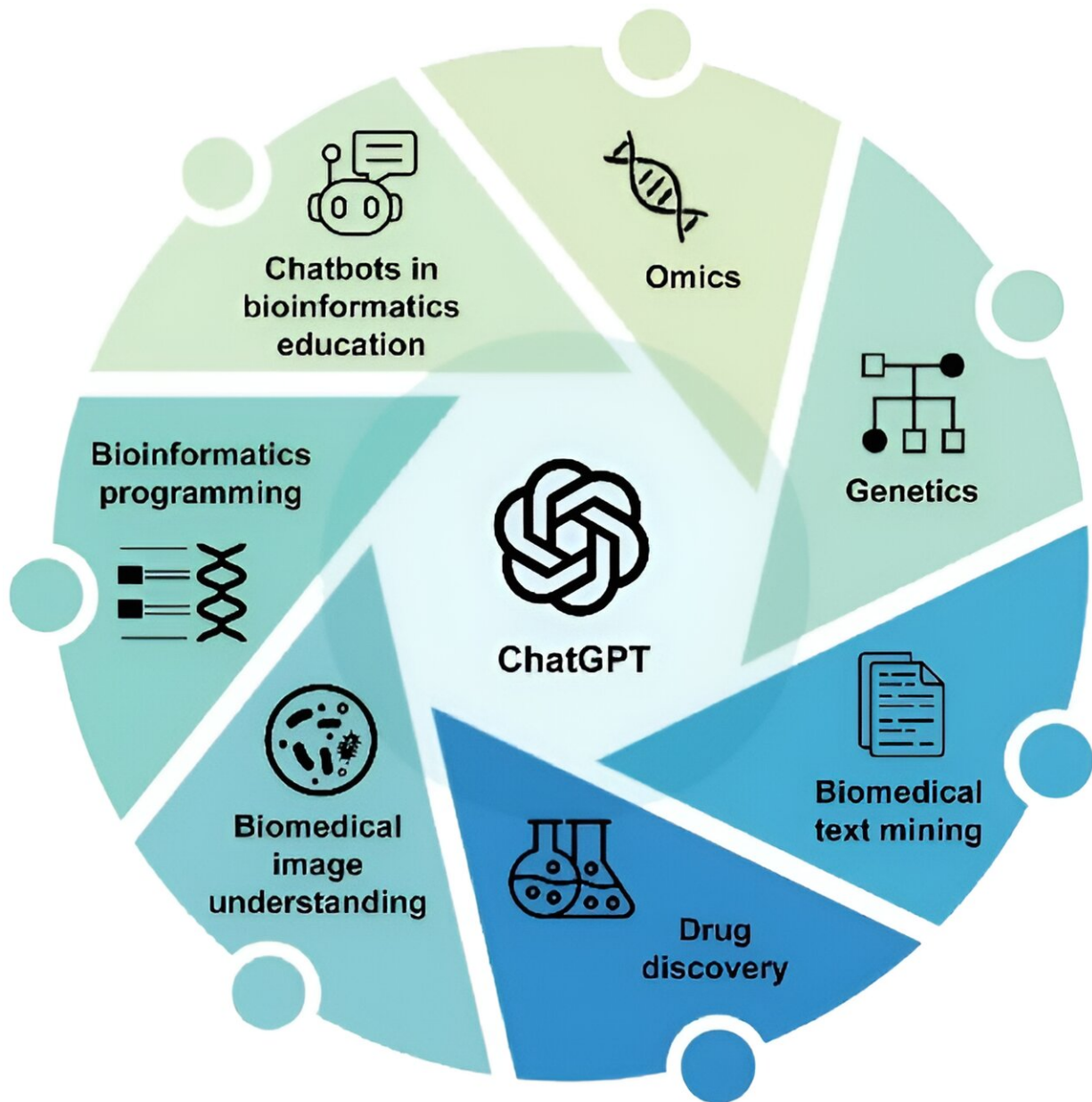# Researchers highlight ChatGPT's role in bioinformatics and biomedical informatics

July 25 2024

Topics of ChatGPT use in bioinformatics and biomedical informatics. Credit: Wang J, Cheng Z, Yao Q, Liu L, Xu D, Hu G

In a systematic review published in *Quantitative Biology*, researchers from West Virginia University (WVU) and their collaborators critically examined the transformative impact of ChatGPT in the fields of bioinformatics and biomedical informatics. The study offers a comprehensive overview of ChatGPT's applications in key areas such as omics, biomedical text mining, drug discovery, image analysis, coding, and bioinformatics education.

"2023 marks the first year for the bioinformatics and biomedical informatics community to explore the potential of ChatGPT in advancing the field," said Gangqing "Michael" Hu, an assistant professor at WVU. "Significant efforts have been made to evaluate ChatGPT's competency in completing common bioinformatics tasks."

ChatGPT has shown impressive capabilities in tasks related to knowledge-mining and text-generation, such as cell type annotation, question answering, and caption generation. However, it exhibited limitations in tasks requiring quantitative analysis and reasoning.

"While ChatGPT has shown some promise, there are areas that require significant improvement," noted Jinge Wang, the first author of the review. "We need more efficient prompt engineering and model fine-tuning to minimize hallucination and improve response accuracy."

Hallucination refers to instances where the model generates responses that sound plausible but are factually incorrect. It happens when chatbots produce responses not grounded in their training data.

Hu highlighted the use of coding for questions that can be addressed through coding to reduce hallucinations. "That tool, formerly known as Code Interpreter, is now a default feature of ChatGPT," he said.

The researchers also emphasized the importance of human expert augmentation in enhancing chatbots' performance. Earlier last year, Hu proposed a framework called OPTIMAL—Optimization of Prompts Through Iterative Mentoring and Assessment—to improve communication with ChatGPT for bioinformatics coding, also published in *Quantitative Biology*.

"A well-curated knowledge base also helps," added Wang. "Large language models are good learners. A few good examples from reputable sources can substantially boost accuracy, known as retrieval-augmented generation."

Working with Hu and Wang were Zien Cheng, a student volunteer in the WVU Department of Microbiology, Immunology and Cell Biology; Qiuming Yao, of the University of Nebraska-Lincoln; Li Liu, of Arizona State University; and Dong Xu, of the University of Missouri.

"We aimed to provide a comprehensive overview of the current use of ChatGPT in bioinformatics and identify its strengths and weaknesses with evidence from multiple independent sources," said Wang. "It is clear that ChatGPT and similar models can be effectively integrated into bioinformatics workflows, but not in all cases."

The researchers remain optimistic about the future of ChatGPT in bioinformatics and biomedical informatics. "AI is a fast-moving field, and we anticipate significant advancements in the coming years," said Hu. "We foresee substantial progress in the coming years in detecting and addressing issues like chatbot hallucinations, achieving accuracy comparable to human experts and the speed of machines."

More information: Jinge Wang et al, Bioinformatics and biomedical informatics with ChatGPT: Year one review, *Quantitative Biology* (2024). DOI: 10.1002/qub2.67

Provided by Higher Education Press