



# Conversational Interfaces in Bioinformatics: Bridging Computational Power and User Accessibility

Angsuman Das

Rajiv Gandhi University, India.

## Abstract

This extensive examination investigates the transformative function of conversational interfaces in the field of bioinformatics, offering an inventive resolution to bridge the gap between intricate computational tools and users from biological, clinical, and medical background who possess limited or no coding proficiency. By capitalizing on the capabilities of natural language processing, these interfaces facilitate an intuitive means of interacting with software, thereby democratizing the accessibility to data storage, analysis, manipulation, genomic sequencing, Proteomic, metabolomics and personalized medicine. By highlighting prominent projects such as GeCoAgent and BioLLMBench, we demonstrate the potential of chatbots and virtual assistants in simplifying the retrieval, analysis, and visualization of data. This review delves into the historical progression of conversational artificial intelligence, its present applications in the field of bioinformatics, and the obstacles encountered in the integration of sophisticated computational tasks. Moreover, it outlines future directions, placing emphasis on the necessity for improved comprehension of natural language and integration with external tools. This study underscores the significance of conversational interfaces in enhancing the accessibility of bioinformatics tools, thereby expediting scientific discovery and innovation in the realm of life sciences. Additionally, this examination explores the ethical considerations surrounding the use of conversational interfaces in bioinformatics, including issues of data privacy, bias mitigation, and ensuring equitable access across diverse user demographics. By addressing these concerns, the integration of conversational AI in bioinformatics can be approached responsibly, fostering trust and acceptance within the scientific community and beyond. Furthermore, this study advocates for ongoing collaboration between computational experts, biologists, clinicians, and ethicists to optimize the development and deployment of conversational interfaces, ensuring their seamless integration into the research and healthcare ecosystems.

**Keywords:** Conversational Interfaces, Bioinformatics, Natural Language Processing, Democratization, Ethical Considerations

**Full length article** \*Corresponding Author, e-mail: [angsudas62@gmail.com](mailto:angsudas62@gmail.com)

## 1. Introduction

The field of bioinformatics has brought about a significant transformation in the realm of biological research and clinical diagnostics, providing unprecedented capabilities in the analysis of data, genomic sequencing, and personalized medicine. This interdisciplinary domain merges the disciplines of biology, computer science, and mathematics to analyze and interpret biological data, particularly the extensive datasets generated by modern techniques such as next-generation sequencing (NGS). However, despite its immense potential, the complexity of bioinformatics tools and the need for computational expertise present substantial obstacles to a wide range of users, including biologists, clinicians, and researchers with limited coding skills [1]. In recent times, conversational interfaces, also known as chatbots or virtual assistants, have emerged as a promising solution to bridge the gap between sophisticated computational tools and end-users. These interfaces utilize

natural language processing (NLP) to enable users to interact with software through human-like conversations, significantly reducing the barriers to accessing complex computational resources [2]. In the context of bioinformatics, conversational interfaces can simplify tasks such as data retrieval, analysis, and visualization, thereby making these processes more accessible to individuals without specialized expertise. The potential of conversational interfaces to democratize access to bioinformatics tools is emphasized by an increasing body of literature. For example, the GeCoAgent project by Politecnico di Milano illustrates the use of a dialogic interface to facilitate the extraction and analysis of genomic data, offering a user-friendly approach to managing vast repositories of genomic information (Crovari et al., 2021). Similarly, the BioLLMBench framework provides a comprehensive evaluation of large language models (LLMs) such as GPT-4, Bard, and LLaMA in executing bioinformatics tasks, highlighting the strengths and limitations of these models in various bioinformatics domains

[3]. The objective of this review is to examine the current state of conversational interfaces in the field of bioinformatics, with a focus on their design, capabilities, and applications. By analyzing recent advancements and case studies, we aim to provide insights into how these technologies can further evolve to meet the needs of the bioinformatics community. Through this analysis, we intend to underscore the significance of conversational interfaces in enhancing the accessibility of bioinformatics tools, thereby empowering a broader range of users to leverage these potent resources in their research and clinical practices.

## **2. Background**

The historical development and progression of conversational interfaces, which facilitate the interaction between humans and computers using natural language, can be traced back to the early stages of computing. One notable example is ELIZA, a program created in the mid-1960s at MIT that emulated a psychotherapist by rephrasing much of the user's input as questions [4]. The inception of ELIZA marked the initial exploration of natural language processing (NLP) capabilities, establishing the basis for future advancements in conversational interfaces. Throughout the years, the domain of NLP and conversational artificial intelligence (AI) has undergone significant evolution, driven by progress in machine learning, deep learning, and the availability of extensive datasets. The emergence of chatbots and virtual assistants such as Siri, Alexa, and Google Assistant and chatGPT in propelled conversational interfaces into the mainstream [5] demonstrating their potential for a diverse range of applications, ranging from customer service to personal productivity.

### **2.1. An Overview of the Challenges in Bioinformatics that Require User-Friendly Interfaces**

Bioinformatics confronts distinctive challenges arising from the characteristics of biological data and the complexity of analysis techniques. The exponential increase in data volume, particularly with the introduction of high-throughput sequencing technologies, presents significant obstacles in terms of data storage, processing, and interpretation. Additionally, bioinformatics encompasses a wide array of analysis methods, encompassing tasks such as sequence alignment, gene expression analysis, and intricate predictive modeling, as well as machine learning applications. These tasks often necessitate specialized knowledge and programming skills, thereby limiting access to a subset of researchers possessing computational expertise [1]. The demand for user-friendly interfaces in the field of bioinformatics is driven by the aspiration to make these powerful tools and methodologies accessible to a broader audience, including biologists, clinicians, and researchers who may not possess extensive programming skills. By simplifying the interaction with bioinformatics tools, not only does it democratize access to these resources, but it also accelerates scientific discoveries and enhances clinical decision-making by enabling more researchers to directly analyze and interpret data.

## **3. Review of Conversational Interfaces in Bioinformatics**

### **3.1. Code Interpreter and Bioinformatics Applications**

The incorporation of conversational interfaces in the field of bioinformatics presents a promising opportunity to streamline the interaction with complex computational tasks. However, there are limitations to the current capabilities of these interfaces, particularly in handling specialized bioinformatics tasks that require the execution of third-party packages, access to extensive databases, and processing of large datasets. Notable obstacles that hinder the adoption of conversational interfaces in more advanced bioinformatics applications include the absence of support for installing additional packages and limited storage capacity. Recent studies have shown efforts to enhance conversational interfaces for bioinformatics. An evaluation of ChatGPT, for example, demonstrates its potential in automating bioinformatics programming tasks and assisting life scientists with coding. The evaluation involved 184 programming exercises from an introductory bioinformatics course and found that ChatGPT could successfully complete 75.5% of the tasks on its first attempt. With additional prompts, this success rate increased to 97.3%. This showcases the ability of conversational AI to significantly streamline the coding process for bioinformatics applications [6].

### **3.2. Innovative Conversational Agents for Data Extraction and Analysis**

GeCoAgent represents a notable advancement in the utilization of dialogic interfaces for genomic data extraction and analysis. By providing a conversational interface, GeCoAgent facilitates the navigation and analysis of complex genomic datasets for clinicians and biologists, offering a user-friendly approach to accessing powerful computational tools. Key features of GeCoAgent include multimodal support, which enhances user interaction by combining text-based dialogue with visual data representations, and the simplification of data analysis results into formats such as spreadsheets and graphics [7].

### **3.3. Web Platforms and Mobile Applications**

The BioIMAX platform exemplifies the application of Web 2.0 approaches in facilitating collaborative data analysis in bioinformatics. By combining the collaborative nature of the web with the computational power of desktop applications, BioIMAX enables scientists to visually explore multivariate image data, share findings, and engage in discussions with collaborators from any location. This approach underscores the importance of accessibility and collaboration in bioinformatics research [8].

### **3.4. Domain-Specific Conversational Assistants**

Specialized conversational assistants, such as BioGraphBot [9] and RiboChat [10] are designed to address specific bioinformatics databases and analytical tasks, showcasing the versatility of conversational interfaces in tackling diverse challenges within the field. BioGraphBot simplifies the interaction with bioinformatics graph databases by translating natural language queries into Gremlin, while RiboChat offers a chat-style interface for the analysis and

annotation of ribosome profiling data. These tools highlight how conversational AI can enhance user engagement and improve the efficiency of data analysis in bioinformatics.

### 3.5. Benchmarking Conversational Interfaces in Bioinformatics

The BioLLMBench framework provides a comprehensive evaluation of the performance of large language models, such as GPT-4, Bard, and LLaMA, in bioinformatics tasks. This benchmarking analysis reveals the strengths and weaknesses of each model across various bioinformatics domains, providing valuable insights into the suitability of conversational AI for specific applications. Notably, GPT-4 demonstrated proficiency in domain knowledge tasks and machine learning model development, highlighting the potential of conversational interfaces to contribute significantly to bioinformatics research [3].

## 4. Challenges and Future Directions

The integration of conversational interfaces into bioinformatics has demonstrated considerable potential to make complex computational tools more accessible to a wider audience. However, several challenges remain that need to be addressed to fully realize this potential. Additionally, the future directions of this field are promising, with advancements in artificial intelligence (AI) and machine learning poised to further enhance the capabilities of these interfaces.

### 4.1. Current Limitations

One of the primary limitations of current conversational interfaces in bioinformatics is the need for more sophisticated natural language understanding (NLU). While significant progress has been made, these systems often struggle with understanding complex, domain-specific queries and instructions. This limitation can hinder their ability to accurately interpret user intents, especially in bioinformatics, where the terminology and data are highly specialized [11]. Another challenge is the integration of third-party bioinformatics tools and databases. Bioinformatics analysis frequently requires the use of specialized software and access to vast databases. Current conversational interfaces may not fully support the integration with these external resources, limiting their usefulness for more complex or specialized tasks [12].

### 4.2. Future Directions

Advancements in AI and machine learning offer promising avenues to overcome these limitations. Improved NLU algorithms, trained on domain-specific datasets, can enhance the ability of conversational interfaces to understand and process complex bioinformatics queries. Furthermore, the development of more flexible and modular conversational systems could facilitate easier integration with third-party tools and databases, expanding the range of tasks these interfaces can support. The potential of AI to improve the performance of conversational interfaces extends beyond just better understanding and integration. Machine learning techniques can also be used to personalize the interface for

individual users, adapting to their preferences and improving the user experience. Additionally, AI can aid in the development of predictive models that anticipate user needs, suggesting relevant analyses or highlighting important data insights.

## 5. Conclusions

The reviewed papers collectively underscore the significant potential of conversational interfaces in transforming the landscape of bioinformatics. These studies highlight the ability of conversational AI to automate programming tasks, interpret complex bioinformatics illustrations, and offer practical tips for harnessing the power of such technologies in computational biology. From the development of innovative conversational agents like GeCoAgent to the comprehensive benchmarking efforts of BioLLMBench, the research demonstrates progress in making bioinformatics tools more accessible and user-friendly. Conversational interfaces represent a pivotal step towards democratizing access to bioinformatics, allowing researchers, clinicians, and students without extensive computational expertise to leverage powerful analytical tools and databases. This accessibility is crucial for fostering innovation and accelerating discoveries in fields ranging from genomics to personalized medicine. However, the journey is far from complete. The call to action for the scientific community is clear: continued research and development in this area are imperative. Advancements in natural language understanding, integration capabilities with third-party tools, and the adoption of user-centered design principles are necessary to overcome existing limitations and unlock the full potential of conversational interfaces in bioinformatics. As we move forward, the collaboration between bioinformaticians, computer scientists, and end-users will be key to refining these technologies. By focusing on user needs and leveraging cutting-edge AI research, conversational interfaces can become even more powerful allies in the quest to make bioinformatics accessible to all, ultimately accelerating the pace of scientific discovery and innovation in the life sciences.

## References

- [1] R. Pereira, J. Oliveira, M. Sousa. (2020). Bioinformatics and computational tools for next-generation sequencing analysis in clinical genetics. *Journal of clinical medicine*. 9 (1) 132.
- [2] L.C. Klopfenstein, S. Delpriori, S. Malatini, A. Bogliolo. (2017). The rise of bots: A survey of conversational interfaces, patterns, and paradigms. In *Proceedings of the 2017 conference on designing interactive systems*. 555-565.
- [3] S. Mangul, V. Sarwal, V. Munteanu, T. Suhodolschi, D. Ciorba, E. Eskin, W. Wang. (2024). BioLLMBench: A Comprehensive Benchmarking of Large Language Models in Bioinformatics.
- [4] C. Bassett. (2019). The computational therapeutic: exploring Weizenbaum's ELIZA as a history of the present. *AI & SOCIETY*. 34 803-812.
- [5] V. Goar, N.S. Yadav, P.S. Yadav. (2023). Conversational AI for natural language processing: A review of ChatGPT. *International Journal on*

- Recent and Innovation Trends in Computing and Communication. 11 109-117.
- [6] L. Wang, X. Ge, L. Liu, G. Hu. (2023). Code interpreter for bioinformatics: are we there yet?. *Annals of Biomedical Engineering*. 1-3.
- [7] P. Crovari, S. Pidò, P. Pinoli, A. Bernasconi, A. Canakoglu, F. Garzotto, S. Ceri. (2021). GeCoAgent: a conversational agent for empowering genomic data extraction and analysis. *ACM Transactions on Computing for Healthcare (HEALTH)*. 3 (1) 1-29.
- [8] C. Loyek, N.M. Rajpoot, M. Khan, T.W. Nattkemper. (2011). BioIMAX: A Web 2.0 approach for easy exploratory and collaborative access to multivariate bioimage data. *BMC bioinformatics*. 12 1-11.
- [9] A. Messina, A. Augello, G. Pilato, R. Rizzo. (2018). Biographbot: a conversational assistant for bioinformatics graph databases. In *Innovative Mobile and Internet Services in Ubiquitous Computing: Proceedings of the 11th International Conference on Innovative Mobile and Internet Services in Ubiquitous Computing (IMIS-2017)*. 135-146.
- [10] M. Xie, L. Yang, G. Chen, Y. Wang, Z. Xie, H. Wang. (2022). RiboChat: a chat-style web interface for analysis and annotation of ribosome profiling data. *Briefings in Bioinformatics*. 23 (2) bbab559.
- [11] Z. Wang, Z. Xu, Q. Ai, V. Srikumar. (2023). An In-depth Investigation of User Response Simulation for Conversational Search. *arXiv preprint arXiv:2304.07944*.
- [12] A. Mannekote, M. Celepkolu, J.B. Wiggins, K.E. Boyer. (2023). Exploring Usability Issues in Instruction-Based and Schema-Based Authoring of Task-Oriented Dialogue Agents. In *Proceedings of the 5th International Conference on Conversational User Interfaces*. 1-6.