

A Platform for the Biomedical Application of Large Language Models

This manuscript ([permalink](#)) was automatically generated from [biocypher/biochatter-paper@fa69d7a](#) on October 27, 2023.

Authors

- **Sebastian Lobentanzer**

 [0000-0003-3399-6695](#) ·  [slobentanzer](#) ·  [slobentanzer](#)

Heidelberg University, Faculty of Medicine and Heidelberg University Hospital, Institute for Computational Biomedicine, Heidelberg, Germany

- **Julio Saez-Rodriguez**

 [0000-0002-8552-8976](#) ·  [saezrodriguez](#) ·  [saezlab](#)

Heidelberg University, Faculty of Medicine and Heidelberg University Hospital, Institute for Computational Biomedicine, Heidelberg, Germany

✉ — Correspondence possible via [GitHub Issues](#)

Abstract

The wealth of knowledge we have amassed in the context of biomedical science has grown exponentially in the last decades. Consequently, understanding and contextualising scientific results has become increasingly difficult for any single individual. In contrast, current Large Language Models (LLMs) can remember an enormous amount of information, but have notable shortcomings, such as a lack of generalised awareness, logical deficits, and a propensity to hallucinate. To improve biomedical analyses, we propose to combine human ingenuity and machine memory by means of an open and modular conversational platform, [biochatter](#), exemplified in the web application [ChatGSE](#). We safeguard against common LLM shortcomings using general and biomedicine-specific measures and allow automated integration of popular bioinformatics methods. Ultimately, we aim to improve the AI-readiness of biomedicine and make LLMs more useful and trustworthy in research applications.

Main

Despite our technological advances, biology and biomedicine continue to pose incredible challenges (Gallagher, 2023). We measure more and more data points with ever-increasing resolution to such a degree that their analysis and interpretation have become the bottleneck for their exploitation. One reason for this challenge may be the inherent limitation of human knowledge (Marois and Ivanoff, 2005). Even seasoned domain experts cannot know the implications of every molecule, be it metabolite, DNA, RNA, or protein, even in their own domain. In addition, biological events are context-dependent, for instance with respect to a cell type or specific disease.

Citation by URL [\[1\]](#).

Citation by alias [\[2\]](#).

References

1. **Study reveals cancer's 'infinite' ability to evolve**
BBC News
(2023-04-12) <https://www.bbc.com/news/health-65252510>
2. **Study reveals cancer's 'infinite' ability to evolve**
James Gallagher
BBC News (2023-04-12) <https://www.bbc.com/news/health-65252510>