

# A Platform for the Biomedical Application of Large Language Models

This manuscript ([permalink](#)) was automatically generated from [biocypher/biochatter-paper@a745e0c](#) 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

---

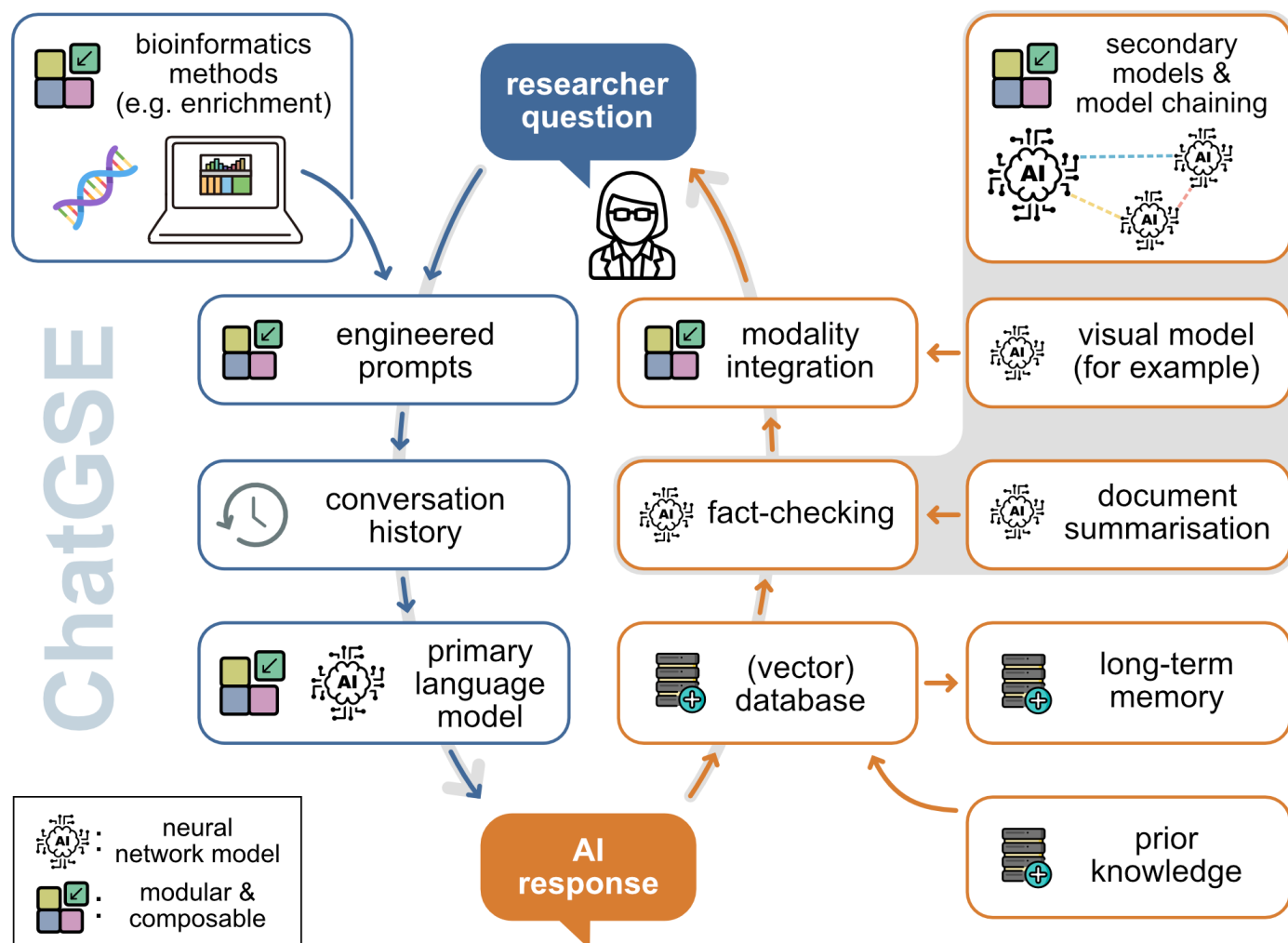
Despite our technological advances, biology and biomedicine continue to pose incredible challenges [1]. We measure more and more data points with ever-increasing resolution to such a degree that their analysis and interpretation have become the bottleneck for their exploitation. One reason for this challenge may be the inherent limitation of human knowledge [2]. 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.

Large Language Models (LLMs) of the current generation, on the other hand, can access enormous amounts of knowledge, encoded (incomprehensibly) in their billions of parameters [5]. Trained correctly, they can recall and combine virtually limitless knowledge from their training set. ChatGPT has taken the world by storm, and many biomedical researchers already use LLMs in their daily work, for general as well as bioinformatics-specific tasks [7]. However, the current, predominantly manual, way of interacting with LLMs is virtually non-reproducible, and their behaviour can be erratic. For instance, they are known to hallucinate: they make up facts as they go along, and, to make matters worse, are convinced - and convincing - regarding the truth of their hallucinations [7]. While current efforts towards AGI (Artificial General Intelligence) manage to ameliorate some of the shortcomings by ensembling multiple models [9] with long-term memory stores [10], the current generation of AI does not inspire adequate trust to be applied to biomedical problems without supervision [8]. Additionally, biomedicine demands greater care in data privacy, licensing, and transparency than most other real-world issues.

A major aim of computational biology is to distil high-dimensional molecular measurements into a humanly digestible form by projecting the measurements into a lower-dimensional space composed of gene programs, pathways, or other functional groupings of biological entities, for example via gene set enrichment analyses. However, even this distilled knowledge requires advanced expertise and thorough literature research to effectively interpret and exploit, and benchmarking the methods' performance is non-trivial [11].

To improve and accelerate this interpretation and exploration, we have developed [biochatter](#), a platform for communicating with LLMs specifically tuned to biomedical research, the use of which we demonstrate in a conversational web interface, [ChatGSE](#) (Figure 1). The platform guides the human researcher intuitively through the interaction with the model, while counteracting the problematic behaviours of the LLM. Since the interaction is mainly based on plain text, it can be used by virtually any researcher. We engineer prompts around the queries of the user to improve model performance


with regard to biomedicine, and automate the integration of popular bioinformatics methods, such as differential expression and gene set enrichment (Supplementary Note [[sec:sup-prompts?](#)])).



**Figure 1: The ChatGSE composable platform architecture (simplified).** The user submits a question about a topic of interest (e.g., an experiment) along with the low-dimensional results of a bioinformatics analysis (top left). The platform's main response circuit (blue) composes a number of specifically engineered prompts and passes them (and a conversation history) to the primary LLM, which generates a response for the user based on all inputs. This response is simultaneously used to prompt the secondary circuit (orange), which fulfils auxiliary tasks to complement the primary response. In particular, using search, the secondary circuit queries a database as prior knowledge repository and compares annotations to the primary response. The knowledge graph can also serve as long-term memory extension of the model. Further, an independent LLM receives the primary response for fact-checking, which can be supplemented with context-specific information by a document summarisation model. If this "second opinion" differs from the primary response, a warning is issued. The platform is composable in all aspects, in principle allowing arbitrary extensions to other, specialised models for additional tasks orchestrated by the primary LLM.

# References

---

1. **Study reveals cancer's 'infinite' ability to evolve**  
BBC News  
(2023-04-12) <https://www.bbc.com/news/health-65252510>
2. **Capacity limits of information processing in the brain**  
René Marois, Jason Ivanoff  
*Trends in Cognitive Sciences* (2005-06) <https://doi.org/d5gmqt>  
DOI: [10.1016/j.tics.2005.04.010](https://doi.org/10.1016/j.tics.2005.04.010) · PMID: [15925809](https://pubmed.ncbi.nlm.nih.gov/15925809/)
3. **PaLM: Scaling Language Modeling with Pathways**  
Aakanksha Chowdhery, Sharan Narang, Jacob Devlin, Maarten Bosma, Gaurav Mishra, Adam Roberts, Paul Barham, Hyung Won Chung, Charles Sutton, Sebastian Gehrmann, ... Noah Fiedel  
*arXiv* (2022) <https://doi.org/kxfx>  
DOI: [10.48550/arxiv.2204.02311](https://doi.org/10.48550/arxiv.2204.02311)
4. **LaMDA: Language Models for Dialog Applications**  
Romal Thoppilan, Daniel De Freitas, Jamie Hall, Noam Shazeer, Apoorv Kulshreshtha, Heng-Tze Cheng, Alicia Jin, Taylor Bos, Leslie Baker, Yu Du, ... Quoc Le  
*arXiv* (2022) <https://doi.org/kmfc>  
DOI: [10.48550/arxiv.2201.08239](https://doi.org/10.48550/arxiv.2201.08239)
5. **GPT-4 Technical Report**  
OpenAI  
*arXiv* (2023) <https://doi.org/grx4cb>  
DOI: [10.48550/arxiv.2303.08774](https://doi.org/10.48550/arxiv.2303.08774)
6. **Reference-free and cost-effective automated cell type annotation with GPT-4 in single-cell RNA-seq analysis**  
Wenpin Hou, Zhicheng Ji  
*Cold Spring Harbor Laboratory* (2023-04-21) <https://doi.org/gsznzg>  
DOI: [10.1101/2023.04.16.537094](https://doi.org/10.1101/2023.04.16.537094) · PMID: [37131626](https://pubmed.ncbi.nlm.nih.gov/37131626/) · PMCID: [PMC10153208](https://pubmed.ncbi.nlm.nih.gov/PMC10153208/)
7. **How will generative AI disrupt data science in drug discovery?**  
Jean-Philippe Vert  
*Nature Biotechnology* (2023-05-08) <https://doi.org/gsznzd>  
DOI: [10.1038/s41587-023-01789-6](https://doi.org/10.1038/s41587-023-01789-6) · PMID: [37156917](https://pubmed.ncbi.nlm.nih.gov/37156917/)
8. **Foundation models for generalist medical artificial intelligence**  
Michael Moor, Oishi Banerjee, Zahra Shakeri Hossein Abad, Harlan M Krumholz, Jure Leskovec, Eric J Topol, Pranav Rajpurkar  
*Nature* (2023-04-12) <https://doi.org/gr4td4>  
DOI: [10.1038/s41586-023-05881-4](https://doi.org/10.1038/s41586-023-05881-4) · PMID: [37045921](https://pubmed.ncbi.nlm.nih.gov/37045921/)
9.  **Langchain** <https://python.langchain.com/>
10. **AutoGPT Official**  
AutoGPT Official  
(2023-09-18) <https://autogpt.net/>
11. **Toward a gold standard for benchmarking gene set enrichment analysis**  
Ludwig Geistlinger, Gergely Csaba, Mara Santarelli, Marcel Ramos, Lucas Schiffer, Nitesh Turaga, Charity Law, Sean Davis, Vincent Carey, Martin Morgan, ... Levi Waldron

*Briefings in Bioinformatics* (2020-03-09) <https://doi.org/ggs7tp>  
DOI: [10.1093/bib/bbz158](https://doi.org/10.1093/bib/bbz158) · PMID: [32026945](https://pubmed.ncbi.nlm.nih.gov/32026945/) · PMCID: [PMC7820859](https://pubmed.ncbi.nlm.nih.gov/PMC7820859/)

12. **The TRUST Principles for digital repositories**

Dawei Lin, Jonathan Crabtree, Ingrid Dillo, Robert R Downs, Rorie Edmunds, David Giaretta, Marisa De Giusti, Hervé L'Hours, Wim Hugo, Reyna Jenkyns, ... John Westbrook

*Scientific Data* (2020-05-14) <https://doi.org/ggwrjt>

DOI: [10.1038/s41597-020-0486-7](https://doi.org/10.1038/s41597-020-0486-7) · PMID: [32409645](https://pubmed.ncbi.nlm.nih.gov/32409645/) · PMCID: [PMC7224370](https://pubmed.ncbi.nlm.nih.gov/PMC7224370/)

## Supplementary Materials

---

In our Supplementary Notes, we explain the functions of our platform in more detail. Please note that several of the mentioned features, particularly more advanced ones, are in early developmental stages. For practical reasons, we divide our software into two distinct packages, the user interface (UI) component ChatGSE and the Python backend library biochatter. For an up-to-date overview and preview of current functionality of the platform, please visit the online preview. Both libraries are developed in Python (version 3.10), according to modern standards of software development [12]. We use Streamlit (version 1.21.0, <https://streamlit.io>) for the web UI. We include a code of conduct and contributor guidelines to offer accessibility and inclusivity to all that are interested in contributing to the framework.

{#sec:sup-prompts}

## Prompt Engineering