

Transparent Application of Text Generation Tools in Scientific Research

Johann-Mattis List

Chair for Multilingual Computational Linguistics

University of Passau

In this opinion piece, I share my view on the application of language models and text generation services in scientific research. In my opinion, scientific research that lives up to the promises of open science must provide full documentation of all prompts and exchanges that were used to create a given study. A mere mention that AI tools have been used in study design, writing, or coding is not enough.

Introduction

There is a divide in scholars between those who embrace language models and specifically chatbots in their own research, arguing that they help them in wording their texts and the like, and those who are strictly against their use in scientific work. I would count myself to the latter camp, as I also do not use language models to write any of my texts, and I switch off all AI features in search engines, where I can. But independently of my own opinion on this matter, I was asking myself if there was any form of a transparent way to use language models.

In the end of last year, I have argued that I see no way to justify the use of chatbots and language models as long as they are not constructed by open principles (see Liesenfeld, Lopez, and Dingemans 2023) and as long as they do not allow for the direct retrieval of existing sources that serve as anchor points to ground their contents (List 2025). Given the practice that I observe from many colleagues, not many scientists share my opinion. Similar to the increased use of generated texts in editing Wikipedia articles (Brooks, Eggert, and Peskoff 2024), I observe more and more colleagues reporting of intensive chat discussions that they have with ChatGPT or other chatbots, or using chatbots to code or create figures for their studies.

Since I have always been enjoying to produce my scientific work without the help of others, I never felt the need to turn to chatbots to ask for help, neither with wordings in

a text, nor with code or figures. Given that I also may spent hours in the internet in the search for particular kinds of information that a seemingly all-knowing person might answer within seconds, I can, however, see why people feel the desire to speed up certain aspects of their individual investigation workflows, even if I do not see a concrete use to turn to chatbots in my own work so far.

When recently even the German Research Foundation officially confirmed that its main committee had approved the use of chatbots, provided certain major rules are followed (DFG 2025), I thought it might be time to review these rules more closely, contrasting them with my own requirements with respect to transparency and openness in scientific research.

2 Lax Guidelines on Generated Text Use

The German Research Foundation emphasizes that the use of text generation machines in review writing must be in concordance with four major principles, confidentiality, transparency, critical quality assessment, and responsibility. Confidentiality refers to the content of the proposals that reviewers are asked to review. Here, the rules do not allow to share these proposals with chatbot services that store the information permanently, ruling out naive usage of chatbots like Grok or ChatGPT via their web interfaces. Transparency refers to the obligation to disclose chatbot usage by scientists acting as reviewers. Critical quality assessment means that reviewers are required to thoroughly check all generated contents in their reviews. Responsibility emphasizes that the reviewers remain responsible for their reviews and that they cannot blame the bots they used to write their assessments.

As far as I understand these guidelines, it is clear that they basically rule out the most popular text generation services offered by big tech companies such as Google or OpenAI, since they store the questions that users have been asked and use them also as feedback in the further development of their services. Given the push of companies like Microsoft to include chatbots in all their services, one may also ask oneself to which degree the creation of word documents with the help of Microsoft's Office program can be done in a way conform with the confidentiality requirement of the German Research Foundation. In any case, the guidelines – if honestly followed by scientists – drastically restrict the application of chatbots in review generation, probably close enough to make their application impossible for those who follow the rules. If scientists follow those rules is a different question, of course, but that people stick to good scientific practice is anyway largely outside of the control of research agencies.

3 Conflicts with Good Scientific Practice and Open Science

Despite the fact that the guidelines are – in principle – much stricter than their wording might suggest, they bear two major shortcomings that may have an unwanted effect on good scientific practice. On one hand, the apparently lax wording of the guidelines invites misunderstandings and abuse. On the other hand, the guidelines fall short of providing concrete and transparent proposals for the use of language generation services that would not contradict or undermine good scientific practice.

That the guidelines are ill-worded, forcing many people to believe that the DFG now allows its reviewers to delegate reviews to ChatGPT or other irresponsible language generation services, can be easily seen from the huge negative reaction that the publication of the guidelines received among scientists (see Bahr 2026, as a representative example). It seems obvious that clear examples for use cases that contradict or are in line with the guidelines are missing.

For me, however, it is even more annoying that no attempts are undertaken to clarify how text generation machines could be employed in a transparent way that would not contradict the current notion of good scientific practice as they are published by the German Research Foundation itself (DFG Team Wissenschaftliche Integrität 2024). From the perspective of open science principles, it seems clear to me that good scientific practice in the use of text generation machines requires very clear guidelines with very high bars regarding transparent documentation. More concretely, in my opinion, the only transparent way of using text generation machines that are trained on public data would consist in submitting all prompts along with the texts produced by scientists. This would mean that scientists who have used chatbots excessively to create research papers would have to disclose in a transparent and clear manner, what labor has been delegated to text generation services, and how the output was later included in the results. I do not have fixed ideas of how I would expect prompts to be shared right now. With students, I typically ask for screenshots of the prompt discussions, but an even better format might consist in screenshots along with versions in plain text, where question and answer are being transparently shared.

4 Transparent Use of Generated Texts

While I have not been able to come up with a perfect solution in this regard, it seems very clear to me that it is not enough, if scientists add a statement in a footnote that says that they used ChatGPT to clean up their references or fixed some wordings.

There are two main reasons why I think that it is important to share the actual prompts in order to make text generation usage transparent. On one hand, text generation is more intrusive than information retrieval via classical search engines. Since the sources of generated texts are typically entirely unclear, more rigor is required to allow scholars to

understand what parts were written by the authors of a study and what parts were generated. On the other hand, transparency in science has an educational aspect. Scientists learn from their colleagues how to carry out their own research. Transparency on prompts here helps scholars to learn from the prompts of their colleagues and from the tools they use to support them in writing, to enhance their own work.

As researchers who subscribe to publicly funded open science, we are cooks who prepare their meals in the open, sharing all ingredients and recipes to make sure that all people out there in the world who want to prepare their own meals can prepare them at home without having to buy them in our store. Since we are publicly funded and large parts of the world still seem to believe that scientific progress is an enterprise that should be pursued in a collaborative, open competition for the best ideas, we should not be selective about the ingredients we use to create our dishes. In hiding parts of our ingredients, we contribute to the growing mistrust in science that can be observed more and more frequently these days. If the promise in the openness of science that many scientists subscribe to is taken seriously, everybody who uses chatbots to write their texts should live up to it.

How chat conversations should be stored concretely, however, is a question I cannot answer at this point. For the time being, I think any way would be okay, be it screenshots or questions and answers pasted in text files. Since generated texts constitute a completely new problem of transparency that has no precedence that I would know of in the past, it seems that practitioners of science must come up with their own solutions in the future.

With more people seeking transparency in the use of text generation services, I would hope that new conventions can emerge quickly over the next years, leading to new conventions and standards that complete basic rules of citation, data sharing, and replicable code in the future.

5 Conclusion

It is absolutely clear to me that my perspective on the usage of text generation tools in scientific research may be different from that of many colleagues. As somebody who can be seen as the opposite of an early adopter with respect to text generation services, my opinion may seem much stricter than it should be to those who enjoy their daily chats with chatbots. However, as somebody who has been fighting hard for transparency in science, at least as far as my own discipline is involved, I see no other way than making strict demands. If we want preserve the major rules of good research practice that the scientific community has been trying hard to establish in the past decades, we have to share the sources of our insights in a transparent manner. These sources include the prompts that we used to trigger text generation by large language models.

References

- Bahr, Amrei. 2026. "Begutachtung mittels KI: Zur fragwürdigen Entscheidung der DFG." Arbeit in Der Wissenschaft, January. <https://steady.page/de/arbeit-in-der-wissenschaft/posts/9ce6550f-973e-4d16-9d09-c5556fa10373>.
- Brooks, Creston, Samuel Eggert, and Denis Peskoff. 2024. "The rise of AI-generated content in Wikipedia." In Proceedings of the First Workshop on Advancing Natural Language Processing for Wikipedia, edited by Lucie Lucie-Aimée, Angela Fan, Tajuddeen Gwadabe, Isaac Johnson, Fabio Petroni, and Daniel van Strien, 67–79. Miami, Florida, USA: Association for Computational Linguistics. <https://doi.org/10.18653/v1/2024.wikinlp-1.12>.
- DFG, Deutsche Forschungsgemeinschaft. 2025. "Künstliche Intelligenz in der Begutachtung." Informationen Für Die Wissenschaft 2025 (102). <https://www.dfg.de/de/aktuelles/neuigkeiten-themen/info-wissenschaft/2025/ifw-25-102>.
- DFG Team Wissenschaftliche Integrität. 2024. "Leitlinien Zur Guten Wissenschaftlichen Praxis [Report, Version 2]." Leitlinien. Bonn: Deutsche Forschungsgemeinschaft. <https://doi.org/https://doi.org/10.5281/zenodo.14281892>.
- Liesenfeld, Andreas, Alianda Lopez, and Mark Dingemans. 2023. "Opening up ChatGPT: Tracking openness, transparency, and accountability in instruction-tuned text generators." In Proceedings of the 5th International Conference on Conversational User Interfaces, 1–6. CUI '23. Eindhoven: Association for Computing Machinery. <https://doi.org/10.1145/3571884.3604316>.
- List, Johann-Mattis. 2025. "Kann man verantwortungsvolle Wissenschaft mit KI betreiben?" Von Wörtern Und Bäumen 9 (12). <https://doi.org/https://doi.org/10.58079/15cdq>.

Funding Information
This project has received funding from the European Research Council (ERC) under the European Union's Horizon Europe research and innovation programme (Grant agreement No. 101044282). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.