

Guidelines for using ChatGPT and other large language models to support scientific writing

Rational

The development of large language models is rapidly progressing and applications in scientific core processes such as the compilation of scientific articles and funding applications become increasingly attractive. The large potential of supporting scientific workflows with AI-generated output is obvious (e.g. fast generation of first drafts, support in literature summary and synthesis, etc.) and it might change all elements of scientific text creation: proposal writing, application writing, paper writing as well as the related review processes.

However, we still lack experience to evaluate the potential pitfalls and downsides of relying on ChatGPT and other AI support for text generation. Hence, this document aims to provide best practice recommendations to ensure scientific rigor while capitalizing on new technological developments. Additionally, it may serve as inspiration what documentation to request when we have to evaluate scientific text (manuscripts, proposals, etc.).

Considerations Before Use of Large Language Models

Before an application of large language models and similar AI tools, we recommend to carefully consider potential negative consequences for:

1. Large language models currently produce reasonably good quality documents. Our observation is, that this output, however, is **not outstanding, novel nor exceptionally creative**. By using such output, one would accept to generate work which of average-quality.
2. Large Language Models reproduce in default settings what is available on the web. Do these systems **foster homogenization** of scientific output or even **risk reinforcing scientific bias**?
3. Producing scientific documents arguably requires time (for us humans). However, the process of writing can help to clarify and sharpen argumentation and improve our conceptual understanding of the subject addressed. **Speeding up the process of text generation may hence undermine output quality**.
4. Personal learning effects: The concise communication of our findings in written text is an essential element of science. The improvement of one's personal writing style is therefore often a lifelong learning goal. AI support during the writing process may strongly affect our long-term learning trajectory – both in positive and negative ways.

Recommended procedures during applications of language models

Once the decision has been made to use a large language model that is directly or indirectly (e.g. by providing a literature synthesis) linked to peer-reviewed publications, funding applications or other scientific documents we recommend a detailed documentation of the process, including:

5. **Documentation of the interaction with the large language model** by saving input and output for transparency reasons and making them publicly available (e.g. as an appendix if appropriate). This includes to also recording the time of application and software version as current AI tools rapidly evolve. However, we fully acknowledge that this documentation is for transparency purposes and cannot be used to attain reproducibility due to the nature of self-evolving large-language models.
6. The **information content** of the produced text needs to be carefully cross-checked with non-AI-generated material. If the AI-generated output contains new information in comparison to the inputted instructions, the documentation of the AI-generated text should also state the references that were used to cross-check the validity of these text sections.
7. **Copy-write considerations** are currently an unclarified issue. A DFG statement (https://www.dfg.de/download/pdf/dfg_im_profil/geschaefsstelle/publikationen/stellungnahmen_papiere/2023/230921_stellungnahme_praesidium_ki_ai.pdf) and several other current recommendations on the use of large language models in science state the necessity to ensure that no intellectual property rights are violated. However, this aim is hardly achievable without a clear understanding of which material is used by the large language machine. A possible mitigation strategy is to use AI models relying on restricted literature pools and listing the respective documents as references in an appendix. However, we acknowledge that this is linked to substantially higher levels of effort and might not be possible in all cases.
8. A large language machine is **not a co-author** of a manuscript or proposal (<https://www.nature.com/articles/d41586-023-00107-z>) but its use and contribution to a certain document has to be clearly acknowledged in appropriate sections (e.g., Method section) of the resulting scientific document.

(3) Disclaimer

These recommendations result from our today's experiences with large language models, our experiences in science in general and department-internal discussions on best-use practices. However, AI-based tools are currently under rapid development, which can partly or fully outdate our recommendations. Thus, the document is prone to revisions and most likely being updated rather sooner than later.

Last update of this policy document: December 2023.