

Improving Accessibility and Reproducibility by Guiding Large Language Models

Thursday 18 September 2025 11:20 (20 minutes)

Research data repositories store numerous entries of research data, to among other advantages one goal is allowing to store us all data to reproduce experiments.

Working with large corpora of texts is made significantly easier with Large Language Models.

However, Large Language Models are trained for general purposes and are not finetuned for the data originating from different kinds of projects.

But the creators of such texts have an expert viewpoint on the data.

Therefore, we propose to leverage the expert viewpoints of creators to obtain better answers from a Large Language Model.

When creating an entry for the Research Data Repository, the creators have the possibility to add a so-called interpretation prompt.

The interpretation prompt contains their expert viewpoint and be of any textual form to guide the Large Language Model to interpret the project-specific data.

In particular, the interpretation prompt may contain instructions on how to reproduce experiments right inside the LLM invocation.

Afterward, the interpretation prompt is prepended to the query of the Large Language Model.

In our examples, we show how the interpretation prompt helps to receive more tailored answers.

Authors: MARWITZ, Florian (Universität Hamburg); GEHRKE, Marcel (Universität Hamburg)

Presenter: MARWITZ, Florian (Universität Hamburg)