Bachelor's Thesis: Surveying Scientific Literature with LLMs

In this thesis you will experiment with large language models as aides for writing scientific survey papers. A focus will be laid on the identification of relevant prior work and the extraction of information from it.

As a starting point you will use our 2023 article "Machine learning in computational literary studies" and experiment with replicating the selection steps we performed as well as extracting information from the specific articles. You will use LLMs to create an updated version of the data for the paper (with more recent publications). Additionally you will consider integrating survey papers from other domains and testing the approach on them.

This will serve to answer the main research questions: **Are LLMs capable of aiding in the creation of high-quality scientific survey papers?**
As part of the thesis you will derive a set of multiple more specific questions that you seek to answer in the thesis.

Prerequisite Skills:
- Experience with Python
- Knowledge of evaluation metrics
- Initial experience prompting LLMs
- Experience with data extraction and transformation
- Good grasp of (written) English

Related work:
Unfortunately, for now, you will also need to perform your own (manual) literature work and find papers with related approaches. As a starting point consider this work, currently in the process of being peer-reviewed: https://openreview.net/forum?id=heeJqQXKg7

If you are interested contact Hans Ole Hatzel.