

Take Home Assessment

Motivation:

In this Take-Home Assessment, we're looking to create a small project that will emulate the type of work new hires on our team may do - namely working with and interpreting the output of major LLMs, as well as basic MLOPs work.

Project Description:

You are tasked with using a language model of your choice to write a book report. In the parent directory to this file, you'll find 3 different books that are somewhat similar. Your job is to ingest the text from each book into a language model and comparatively analyze how each work deals with the theme of social isolation. What are the authors' points of view on this subject, and what parts of the novel corroborate these claims? Your output will be a 5 paragraph book report (also generated by a language model) that states a clear thesis statement, makes clear arguments based on the content of each novel, and accurately cites sections of each novel, culminating in a concluding paragraph to summarize the arguments.

At no point should you, the human, have to read the text or write the review - every aspect from ingestion to analysis to comparison to writing the report should be done by a model. This also shouldn't take more than 4-5 hours of work.

Considerations:

Here are some questions you may want to consider as you go about designing your solution.

- How should you ingest the different file types for the books? Are there any general modules out there, or should you implement your own parser for any reason?
- Once ingested, what language model(s) should you use to process and analyze the books? What is the context length of the model you use, and will you be able to fit the entire book into context? If not, how should you break down the text?
- Should the model you use for analysis be the same model you use for writing the report?
- What prompts should you use for summarizing/analyzing the texts? What about generating the report? Should it all be 1 prompt, or should different parts of the report be generated by different prompts (or even models)?
- Outside of the LLMs themselves, what other tools should you use in this process? Vector DBs, indexers, RAG tooling, etc?

None of the above considerations should have a right answer - you're welcome to choose whichever tradeoff you think makes the most sense.

Take Home Assessment

Deliverables:

Your deliverables are (1) a document output containing the final generated book review (you are welcome to use any format you'd like; .txt, .pdf, .word, etc.) and (2) any and all scripts or modules you wrote to generate the review. These files should be added back into this directory, so that we can run the script(s) to generate a similar output should we want to.

We have no particular preference for programming languages used, libraries used, or even which models you use. You are welcome to use whatever you think is most appropriate or most comfortable for you.

Rubric:

Your solution will be evaluated on a few dimensions:

Correctness:

There is no right analysis for the report, but we will evaluate if you've generated coherent text and largely cohere to our standard of a 5 paragraph essay with a thesis, argument, and conclusion. Additionally, we'll if you successfully and accurately cite parts from the books in the report without hallucinating excerpts.

Cleanliness of Code:

Is your code legible and easy to follow? Is it relatively modular and recomposable for future work? We don't require all of the code to be in one file or multiple, but depending on your architecture, you may find one to be more clean than another style.

Documentation:

You aren't expected to have a separate documentation file for this, but simple comments around functions to explain your thinking will be valuable.

Usage of AI tools:

Are you using different model providers for different parts of the process? Are you making use of any RAG frameworks, vector databases, or chaining frameworks? There is no requirement to use any/all of these, but having a point of view on which (if any) to use will be valuable.

Prompt Engineering:

For any given model, what specific prompt do you use to accomplish your task (summarization/analysis/etc)? How do you evaluate which prompts perform better than others?

Further Questions:

Take Home Assessment

If you have any questions or comments on the assignment, you can reach out to nbaronia@tower-research.com