

Building Decarbonization Coalition (BDC), Winter 2025

The Building Decarbonization Coalition (BDC) is a non-profit with the goal to help the U.S. transition to climate-friendly energy technologies. Thermal Energy Networks (TEN) are a specific technology which uses geothermal energy to produce heating/cooling. Currently, to evaluate the success of TEN pilot projects, one must look across various PDF documents in a government database. To reduce the friction in evaluating the success of these projects, the BDC tasked the team with creating an LLM chatbot to answer questions using relevant documents.

The chatbot relies on a few key components: a web-scraper to download the PDFs, a retrieval model for providing relevant context, an open-source LLM, and a command-line interface. The team improved upon a rudimentary prototype of the chatbot. Improvements included enabling pre-processing to work with images and tables, state-of-the-art context retrieval techniques, and LLM fine-tuning.

As a result of our improvements, we found a 20% increase in the semantic similarity between model-generated answers and our ‘gold-standard’ responses.¹ Future work could include techniques to reinforce LLM response structure and building out the front-end for a web-based app.



Chart 1. The framework of the chatbot prototype, pre-improvements.

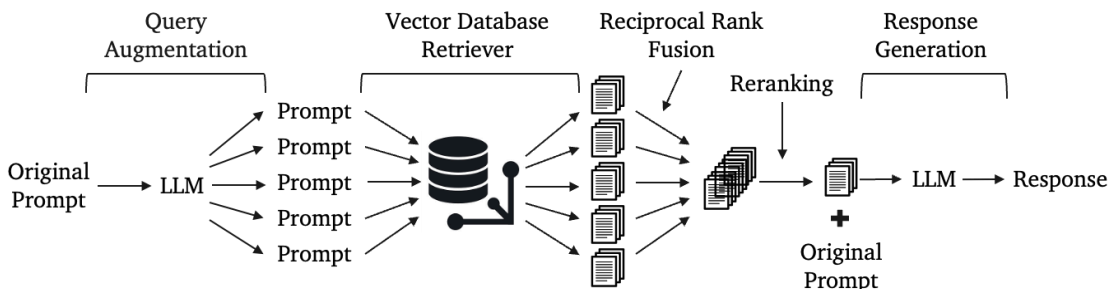


Chart 2. The end-to-end framework of our improved chatbot.

¹Performance benchmarking for this project was captured by the BERT F1 score metric.