

Building Decarbonization Coalition (BDC)

The BDC aims to help the U.S. transition to climate-friendly energy technologies. Thermal Energy Networks (TENs) are a key focus due to their efficiency over traditional HVAC systems that rely on air heating/cooling. Currently, NYC's TEN pilot projects have publicly available data stored in the NY Department of Public Service database.

However, the data's unorganized structure makes it difficult to evaluate the success of TEN's, causing frustration for stakeholders. To address this, BDC tasked our team with creating an LLM chatbot to answer questions about the pilot projects and provide relevant document references.

Our system development included four components:

1. A web scraper to convert PDFs from the database into a parsable format for embeddings.
2. Nvidia RAG model for embedding creation and retrieval.
3. Llama Instruct model for integration with the RAG model.
4. Using LangChain to combine these components into a Command Line Interface (CLI).

We completed an initial CLI that supports reliable information retrieval from relevant documents in the database. The next steps should include fine-tuning for more accurate and detailed responses.

Model Flow Diagram of Chatbot w/ Example Use Case

