# Building Effective Search system

**Exploring Project Options**

In this segment, we present you with two exciting project options to choose from. These options provide diverse opportunities to apply your newfound skills and creativity, ensuring a rewarding learning experience. Whether you wish to craft your custom AI project or choose the project provided by us, both avenues promise valuable insights and contributions to search systems using LLMs. Let's explore these options in detail.

**Option 1: Build Your Own Project (BYOP)**

Under the "Build Your Own Project" option, you can conceive, design, develop, and implement your own AI-driven project. This is an exciting opportunity to explore your passion for AI and create a search system that can impact any domain you choose. Here are some project ideas to spark your imagination:

* **Email Search AI:**Develop a generative search system for emails that helps organisation find and validate past decisions, stratagies, and data in a huge corpus of email threads. Here is a [dataset](https://www.kaggle.com/datasets/marawanxmamdouh/email-thread-summary-dataset) from Kaggle that might be useful for this endeavour.
* **Fashion Search AI:** Create a generative search system capable of searching a plethora of product descriptions to find and recommend appropriate choices against a user query. You may use this [Myntra dataset](https://www.kaggle.com/datasets/djagatiya/myntra-fashion-product-dataset) on Kaggle to build such a system.

As you work on your custom project, remember to document your progress meticulously. Create comprehensive documentation that outlines your project's goals, the AI model used, data sources, key design decisions, and any challenges you encountered.

*NOTE: Remember, you can source data for your project from publicly available datasets from platforms like Kaggle or even collect data yourself. The key is choosing a project that aligns with your interests and objectives and has a meaningful application in the real world.*

**Option 2: Build Mr.HelpMate AI**

If you prefer a structured project with a clear starting point, you can build a project in the insurance domain, similar the project you saw in the "retrieval augmented generation" session. The goal of the project will be to build a robust generative search system capable of effectively and accurately answering questions from a policy document.

You will be using a single long life insurance policy document for this project. The PDF for the document can be downloaded below.

**Group Member Life Insurance Policy**

**Download**

The project should implement all the three layers effectively. It will be key to try out various strategies and experiments in various layers in order to build an effective search system. Let's explore what you need to do in each of the layers, and the possible experimentations that you can perform based on various choices.

1. **The Embedding Layer:**The PDF document needs to be effectively processed, cleaned, and chunked for the embeddings. Here, the choice of the chunking strategy will have a large impact on the final quality of the retrieved results. So, make sure that you try out various stratgies and compare their performances.  
     
   Another important aspect in the embedding layer is the choice of the embedding model. You can choose to embed your chunks using the OpenAI embedding model or any model from the [SentenceTransformers library](https://huggingface.co/sentence-transformers" \t "_blank) on HuggingFace.
2. **The Search Layer:**Here, you first need to design at least 3 queries against which you will test your system. You need to understand and skim through the document, and accordingly come up with some queries, the answers to which can be found in the policy document.  
     
   Next, you need to embed the queries and search your ChromaDB vector database against each of these queries. Implementing a cache mechanism is also mandatory.  
     
   Finally, you need to implement the re-ranking block, and for this you can choose from a range of [cross-encoding models](https://huggingface.co/cross-encoder) on HuggingFace.
3. **The Generation Layer:** In the generation layer, the final prompt that you design is the major component. Make sure that the prompt is exhaustive in its instructions, and the relevant information is correctly passed to the prompt. You may also choose to provide some few-shot examples in an attempt to improve the LLM output.

Make sure you try out different strategies and models in each layer, and you might be surprised with the variety of top answers retrieved and generated by your system.

**Note:** You need to test your system with at least 3 self-designed queries that can help you gauge the quality of outputs delivered by your system. Also, as a check, you need to submit the screenshots of the outputs along with the code and documentation. The details of the screenshots that you need to share are provided below.

1. **Top 3 Results from the Search Layer:**You need to share 3 screenshots against 3 self-designed queries that clearly showcase the top 3 results/chunks retrieved from the search layer. Please share 1 screenshot for each query and its output.
2. **Final Generated Answer from the Generation Layer:** Here, you are required to share 3 screenshots of the same 3 queries with the final output generated by the LLM in the generation layer. Please share one screenshot for each query and its output.  
     
   So, overall, you will need to share 6 screenshots.

Both the project options offer exciting opportunities to apply your AI skills and creativity. Whether you build your own project or Mr.HelpMate AI, you'll gain valuable experience in AI search systems and contribute to the development of innovative AI solutions.

**Evaluation Rubrics**

As you delve into the next phase of your AI journey, it's crucial to outline the evaluation criteria that will guide the assessment of your projects. Whether you build your own project (BYOP) or Mr.HelpMate AI, these assessment criteria provide a fair and comprehensive assessment framework for evaluation.

|  |  |  |
| --- | --- | --- |
| **Criteria** | **Weight** | **Build Your Own Project (BYOP) / Mr.HelpMate AI** **- Evaluation Parameters** |
| Overall System Design | 10% | - Innovation and creativity in system design - Search system's architecture, workflow, and implementation |
| Embedding Layer | 25% | - Effectiveness in processing the text data - Application of an effective and optimal chunking strategy - Appropriate choices of embedding models and proper implementation of embeddings for all chunks |
| Search Layer | 30% | - Quality of the search results - Implementation of cache - Selection and implementation of a re-ranker |
| Generative Layer | 10% | - Quality of the prompt and final answers |
| Query Search | 10% | - Performance of the whole system against 3 self-designed queries - Screenshots of the outputs of the search layer and the generative layer against each of the 3 queries |
| Documentation | 15% | - Comprehensive documentation detailing project goals, data sources, design choices, and challenges faced |