**Approach Overview:**

**1. Data Crawling and Collection**

- The data was scraped from **SHL's Products Catalog** using **Playwright.**
- Each test entry's structured metadata (e.g., *Name, Description, Test Type, Job Levels, Assessment Length*) was extracted and saved in a normalized format using **Pandas**.

**2. Data Representation and Storage**

- Data was stored in **MongoDB Atlas.**
- Text embeddings were generated using **SentenceTransformer** with the `'all-MiniLM-L6-v2'` model from the **Hugging Face Transformers**. These embeddings were stored alongside the original documents.

**3. Semantic Embeddings and Vector Search**

- User queries were **refined and restructured** using **Google's Gemini 1.5 Pro** LLM via the **google.generativeai** Python client. The LLM extracted fields such as `Assessment Length`, `Languages`, `Test Type`, and created structured prompts.
- The refined queries were then embedded using **SentenceTransformer.**
- For search, I used **MongoDB's vectorSearch operator**, performing **approximate nearest neighbor search** using Cosine Similarity on the precomputed embeddings.

**4. Query Expansion and Multi-Skill Search**

- A second Gemini prompt was used to **extract core skills** from unstructured job descriptions.
- For each skill, a separate semantic search was run and the results merged and deduplicated using Python's set logic and sorted by **vector similarity score**.

**5. Stack and Integration**

- **Backend**: Built using **FastAPI**.
- **Frontend**: A simple interactive demo was developed using **Streamlit.**

**6. Evaluation and Metrics**

- Evaluated using **Accuracy**, **Mean Recall@3**, and **Mean Average Precision (MAP@3)** — against the TEST set.
- The achieved accuracy was 60%.

**Technologies and Libraries Used**

- **Playwright** – Web scraping
- **MongoDB Atlas** – Scalable cloud database
- **SentenceTransformers** – Semantic embeddings
- **google.generativeai (Gemini 1.5 Pro)** – LLM-based query refinement
- **FastAPI** – Backend API
- **Streamlit** – Frontend demo UI