

# SHL Assessment Recommendation System – Generative AI Intern Assignment

---

## 1. Problem Understanding

The goal of this project is to build a recommendation system that maps recruiter-style natural language queries to relevant SHL assessments. The system must retrieve and rank suitable assessments from the SHL product catalog and return the top recommendations via an API.

---

## 2. Data Collection

Assessment metadata was programmatically collected from the SHL product catalog using a web scraper built with `requests` and `BeautifulSoup`. The scraper extracts assessment names, URLs, test types, remote support, and adaptive support. The final dataset consists of 377 unique assessments.

---

## 3. Data Preprocessing

Scraped data was normalized and stored in CSV format. The provided evaluation datasets required handling non-UTF8 encodings (Windows-1252), which were resolved during ingestion to ensure robust processing.

---

## 4. Recommendation Approach

A heuristic-based recommendation strategy was implemented. Incoming queries are tokenized and matched against assessment names and metadata using keyword overlap and rule-based filtering. Diversity across assessment types (e.g., Ability, Personality, Skills) is maintained to avoid over-specialization. The recommender returns the top-k most relevant assessments.

---

## 5. Evaluation Methodology

System performance was evaluated using **Mean Recall@10**, as specified in the assignment. The provided training dataset contains multiple relevant assessments per query, which were grouped appropriately before evaluation. The baseline system achieved a Mean Recall@10 of **0.0200**, reflecting the difficulty and sparsity of the query-to-assessment mapping.

---

## **6. Limitations and Future Improvements**

While the heuristic approach establishes a correct and interpretable baseline, future improvements may include semantic embeddings (e.g., sentence transformers), learning-to-rank models, and user feedback loops to improve recall.

---

## **7. Conclusion**

This project demonstrates an end-to-end recommendation pipeline including data scraping, preprocessing, recommendation logic, evaluation, and submission generation. The system is modular, extensible, and production-ready for further enhancements.