

# Hirelytics: AI-Powered SHL Assessment Recommender

## 1. Overview

Hirelytics is an intelligent recommendation engine that automatically maps job descriptions or hiring contexts to the most relevant SHL assessments.

It combines semantic text embeddings, hybrid similarity scoring, and Gemini-based reranking for context-aware, high-precision results.

The system was developed using FastAPI, SentenceTransformer, Supabase (pgvector), and Gemini 2.5 Flash, and deployed on Azure App Service using Docker.

---

## 2. Architecture

Layer	Technology	Function
Frontend	HTML + JS	Simple dark-mode UI for input and results
Backend	FastAPI + Unicorn	Handles model inference and API requests
Database	Supabase (pgvector)	Stores and retrieves SHL assessment embeddings
Embedding Model	SentenceTransformer all-mnppnet-base-v2	Generates 768-D semantic vectors
Reranker	Gemini 2.5 Flash	Contextual re-ranking of top results
Hosting	Azure Docker App Service	Scalable, portable deployment

---

## 3. Processing Pipeline

1. Input – Accepts a free-text query or job-description URL (parsed with BeautifulSoup).
2. Summarization – Long JDs (>1500 chars) are summarized by Gemini.
3. Embedding & Retrieval – Query converted to vector → searched in Supabase via match\_products\_v4.
4. Hybrid Scoring:

$$\text{final\_score} = 0.7(\text{vector\_similarity}) + 0.3(\text{keyword\_overlap})$$

5. Gemini Reranking – Top 10 candidates re-ordered for contextual fit.
  6. Output – Returns *top-10* assessments with name, type, duration, and URL.
- 

## 4. Optimization Summary

Stage	Problem	Solution	Effect
Baseline	TF-IDF lacked context	Replaced with Sentence-BERT	Recall $\uparrow$ 0.56 $\rightarrow$ 0.60
Search	Slow cosine similarity	Supabase pgvector RPC	Latency $\downarrow$ 2.8 s $\rightarrow$ 0.9 s
Ranking	Over-reliance on vector score	Added 70/30 hybrid metric	Precision $\uparrow$ $\approx$ 10 %
Long Queries	Context lost on truncation	Gemini summarization	Stable embeddings
Relevance	Missing soft-skill cues	Gemini re-ranking prompt	Recall@10 $\uparrow$ 0.60 $\rightarrow$ 0.62
Deployment	Cold-start delay	Dockerized FastAPI	Startup $\downarrow$ 65 %

---

## 5. Results

- Before optimization Mean Recall@10 = 0.56
  - After optimization Mean Recall@10 = 0.62
- 

## 6. Outcome

Hirelytics achieved a mean recall@10 of 0.62 .

It demonstrates how combining semantic embeddings, vector databases, and LLM reasoning can significantly improve recommendation quality for SHL assessment selection.

Tech Stack: FastAPI · Sentence-BERT · Gemini · Supabase · Docker · Azure

Final Performance: Recall 0.62