

Aptitude Search

Applied AI Workflow Engineering for Capability Discovery

A multi-stage LLM + embedding pipeline that turns messy human text into schema-validated, explainable recommendations – with career search as the working domain.

ARCHITECTURE



Why This Matters

The project explores how AI systems can transform ambiguous human information into structured, explainable recommendations through controlled workflows rather than unconstrained generation.

ENGINEERING SIGNALS

- Multi-stage AI workflow architecture
- Schema-constrained LLM extraction + validation
- Embedding-based semantic matching (pgvector)
- Occupation retrieval over vector store
- Orchestrated LLM + tool-calling stages
- Reliability: schema checks, URL grounding

Multi-stage, schema-validated pipeline – natural fit for graph orchestration (e.g. LangGraph): staged nodes, structured state, tool-backed discovery.

Domain = Example

Career opportunity discovery is the application surface. The engineering problem is capability extraction, semantic grounding, and controlled synthesis.

PRODUCT FLOW

Resume → profile → role plan → verified openings

IMPLEMENTATION DETAIL

LLMS

Staged chat completions for profile extraction, role-family planning, and result synthesis – each bound to a JSON schema.

EMBEDDINGS / VECTORS

Aptitude profiles embedded and matched against O*NET occupation vectors in Postgres (pgvector) for semantic role grounding.

ORCHESTRATION

Python orchestrates stage order, web discovery tools, aptitude-fit ranking, and post-synthesis URL filtering to observed sources.

Controlled AI workflows over unconstrained generation.

Built as a working FastAPI + React system: multi-stage LLM pipeline, embedding retrieval, schema validation, and tool-assisted discovery – demonstrating applied AI engineering in practice.

RELEVANCE

**Applied AI systems
for capability discovery**