

GSoC 2025

Lighting Talk



PRESENTED BY:

Gopi Trinadh, Maddikunta





Gopi Trinadh, Maddikunta

Who Am I ?

- I'm Gopi Trinadh Maddikunta from Houston, Texas. Pursuing second year of my master's degree in engineering data science at University of Houston.
- Currently an GSoC 2025 Open-source Contributor @ Scala Center and Research Assistant with CS department at UH as well.
- Literally, I don't have any experience with Scala, only thing i got through Daniel's Rock the JVM Courses. Initially, it is hard to turn my idea into implementation later, somewhat better.
- Focuses on RAG Pipelines, Embedding Support and Database indexing and modelling.



Special Thanks to my mentors



Kannupriya Karla



Rory Graves

“Huge thanks to Rory Graves and Kannupriya Karla—you believed in me and helped turn a rough idea into something real. Your clear feedback and steady support made me a better engineer and teammate. Whatever impact EmbedX has, it carries your fingerprint.”

**“Huge thanks to community,
such a great support”**





My Contributions

- Provider-based EmbeddingClient (OpenAI & VoyageAI)
- Config-first static model selection (.env)
- UniversalExtractor/Encoder for PDF, DOCX, XLSX, TXT, URLs
- ModelDimensionRegistry → consistent vector sizes
- Example runner + cosine similarity; structured logs
- Cleanup & dependency fixes (sbt, libs, warnings)



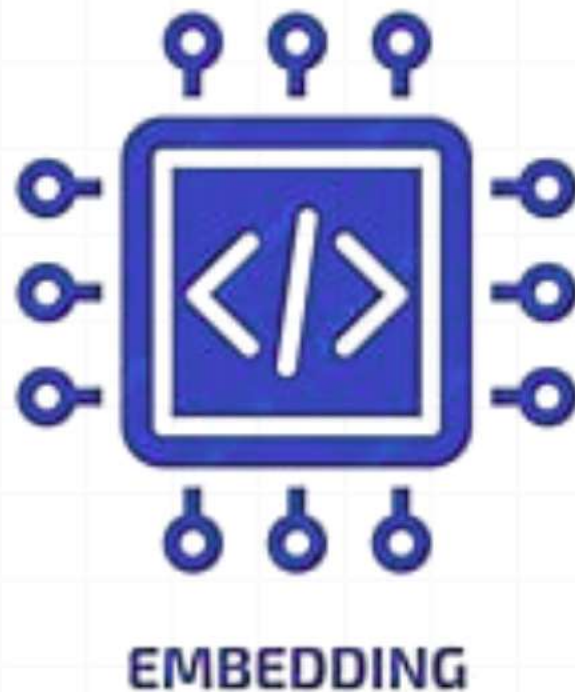
Closed
9

Open
3

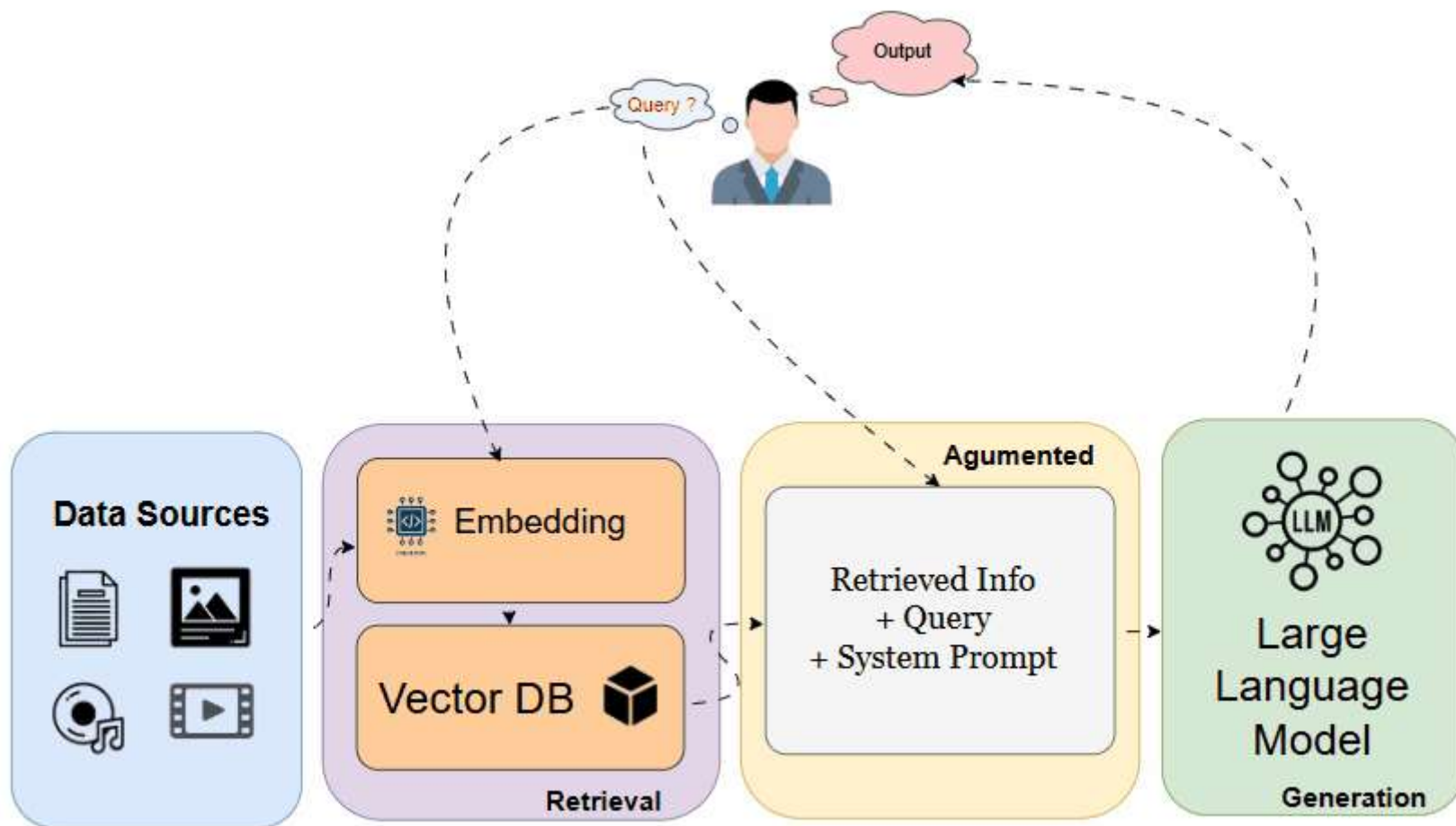
Merged
3

Impact: That matters

- Embedding Support is an initial step for entire process, which helps for better retrieval.
- Embedding engine named Embedx, is supporting as modular which helps to converting any data in vectors.
- Those vectors are in form easy storing in VectorDB and as well as retrieval.
- Which helps for Faster, reproducible Embedding vectors.



Retrieval Augmented Generation - Architecture



Scheduled PRs

DBx Core

- Postgres schema for chunks + embeddings (pgvector column)
- Batch upsert, migrations, connection/config hygiene
- Health checks & simple metadata (JSONB) model

DBx Mid

- pgvector setup & extension enablement
- Distance metrics (cosine / L2 / inner product) selectable
- Index choices: IVFFlat (and HNSW if available in your version)
- Top-k search SQL; metadata pre-filters (JSONB) before similarity

DBx Full

- CLI tools: ingest, rebuild, reindex, validate
- Benchmarks & EXPLAIN ANALYZE guides; basic dashboards
- Partitions/retention for scale; backups/ops notes
- Docs + sample apps for local, reproducible runs

Purpose of RAG

- Connects to your data: PDFs, notes, wikis, DB exports at answer time.
- Reduces mistakes: Model reads retrieved passages, not just memory.
- Stays current: Re-index new content; no fine-tune needed.
- Explainable: Answers come with sources and context.
- Efficient: Smaller prompts; fast top-k via FAISS.



Thank you



Google
Summer of Code



LLM4S



scalacenter

Connect with me!

