

The State of AI Powered Search and Retrieval — Frank Liu, MongoDB (prev Voyage AI)



AI Engineer
247K subscribers

Subscribe

16



Share

Ask

Download



667 views Jun 27, 2025 [AIEWF 2025 Complete Playlist](#)

In this talk, we examine the state-of-the-art in AI-powered search and retrieval. We detail techniques for enhancing performance beyond base embedding models, including hybrid search, reranking strategies, query decomposition and document enrichment, the use of domain-specific and fine-tuned embeddings, custom data processing pipelines (ETL), and contextualized chunking methods.

About Frank Liu

Frank Liu is Staff Product Manager at MongoDB. He has over a decade of industry experience in machine learning and hardware engineering and presents at major industry events like the Open Source Summit Open Data Science Conference. In his spare time, he enjoys experimenting with and training models. Frank holds MS and BS degrees in Electrical Engineering from Stanford University.

About Voyage AI



Most accurate and cost-effective embedding models and rerankers for
RAG and semantic search

Available via the Voyage AI API, AWS marketplace, Azure marketplace

A quick refresher

Real-world applications

What's to come

A quick refresher

What is AI-powered search?



Finds **related concepts**, even without identical wording

Understands the **user's intent**, e.g. "my best friend is sick"

Can perform some **reasoning** and/or **instruction-following**

RAG is a popular use case



What is Voyage AI's mission?

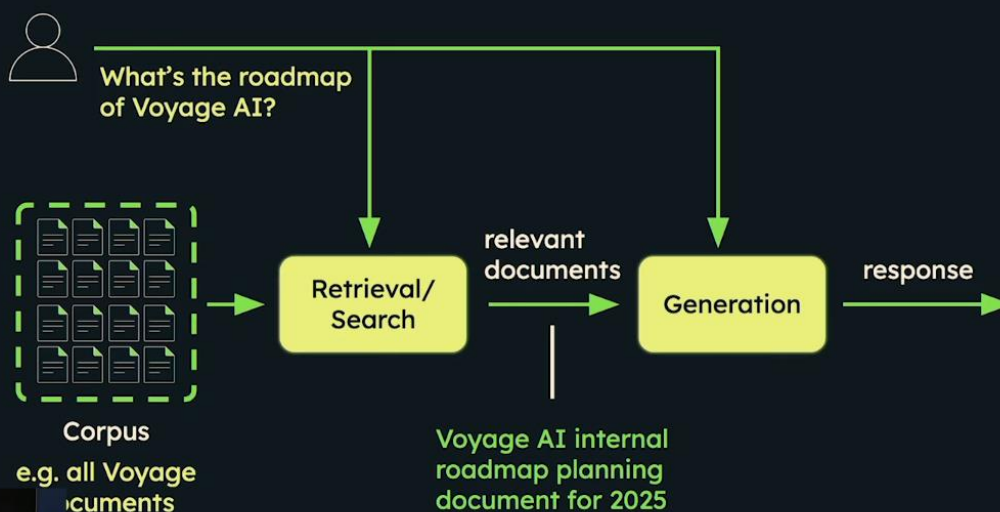
Voyage AI's mission is to provide advanced AI-powered solutions that enable organizations to **optimize decision-making, enhance operational efficiency...**

Voyage AI's mission is to help build the very best **Retrieval Augmented Generation (RAG)** and **semantic search applications...**

Without retrieval

With retrieval

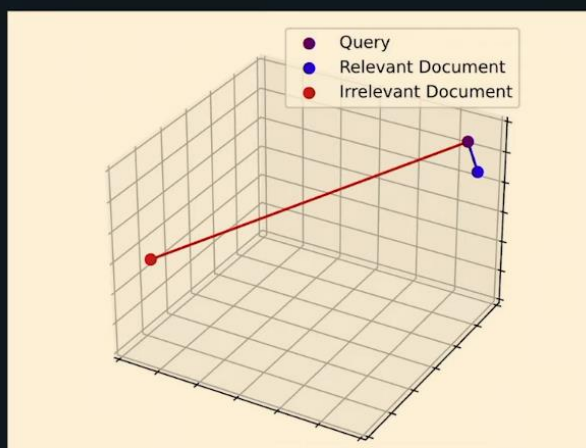
RAG is a popular use case



Embedding quality is a core component



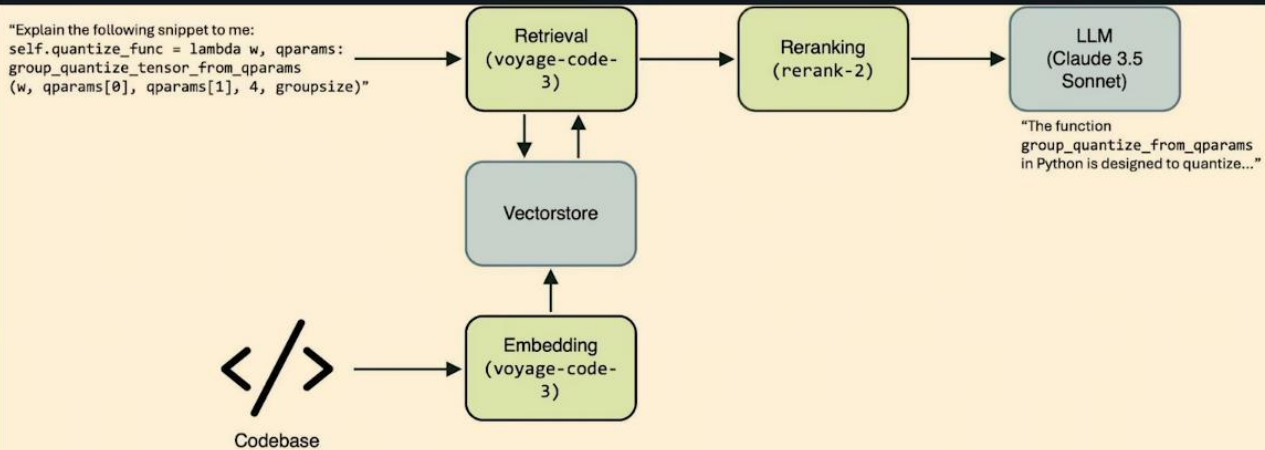
Query



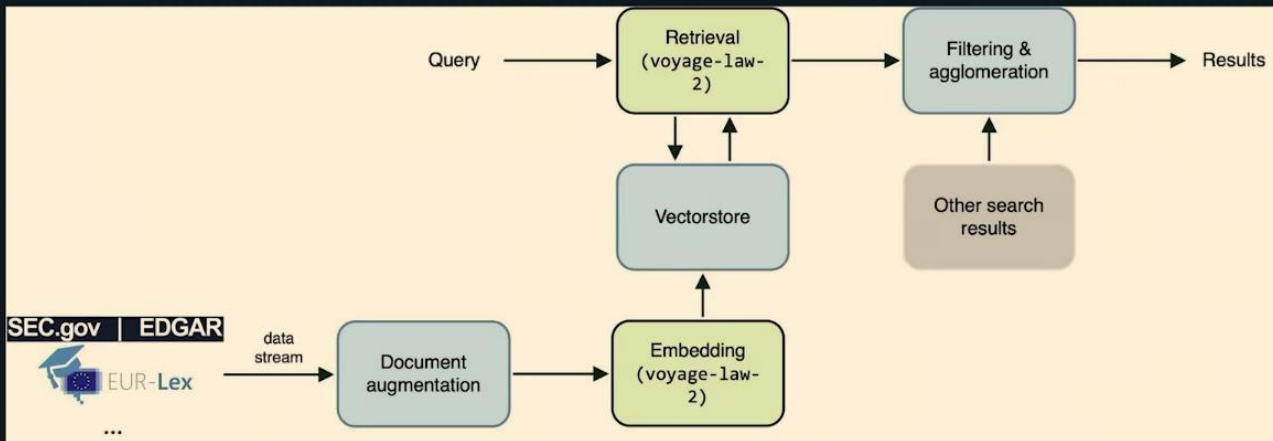
ements

Real-world applications

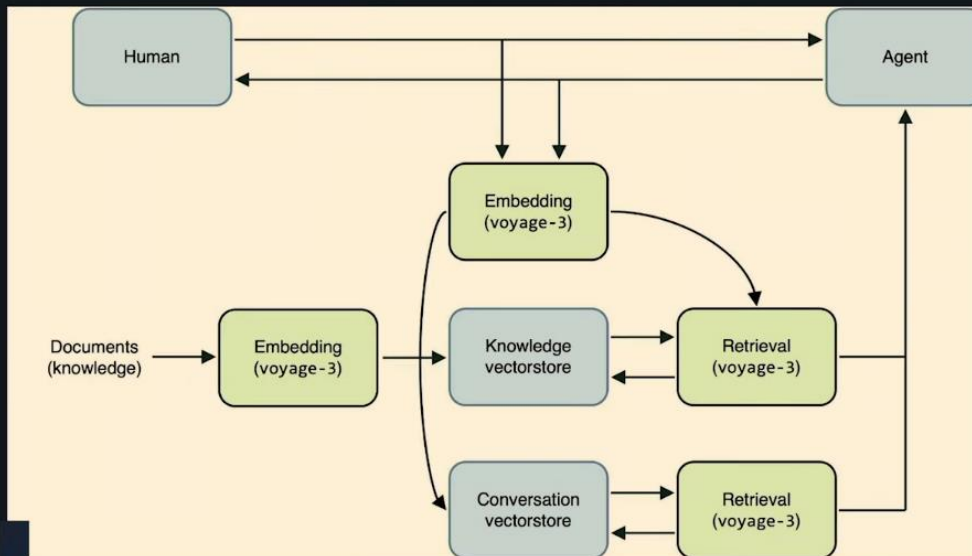
Chat with your codebase



Application-specific (legal) retrieval



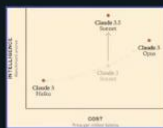
Conversational/agentic retrieval



What's to come

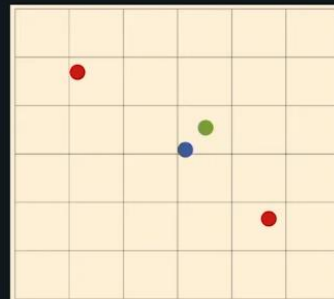
The future is multimodal

"Frontier intelligence at 2x the speed"



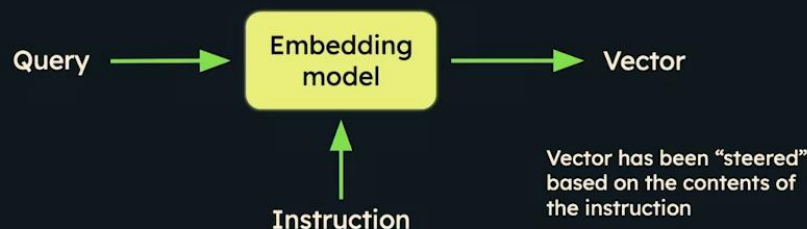
"Strong LLMs"

voyage-multimodal-3

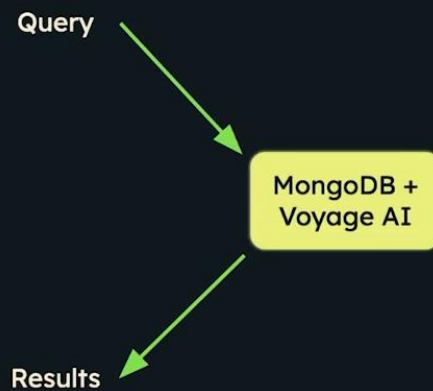
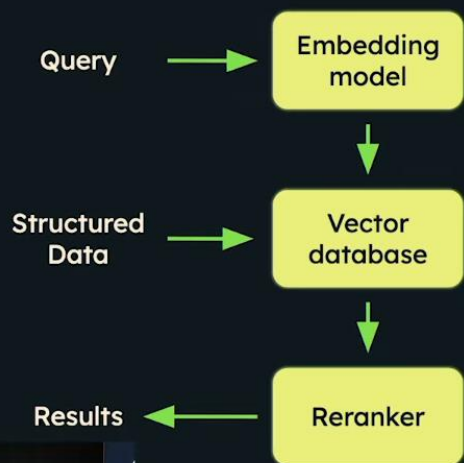


"Pikachu is a small, yellow Pokemon"

Instruction-tuned embeddings & rerankers



Agent-native database



Thank you for
your time.

VOYAGE AI By MongoDB



VoyageAI
Linkedin



MongoDB
Linkedin