

Tech Stack explanation

- A list of the chosen technologies used for the developing the application, with explanation of their pros and cons.

Category	Chosen Tech stack	Alternative	Pros	Cons
Database	PostgreSQL + pgvector	FAISS, Pinecone, ChromaDB	<ul style="list-style-type: none">- SQL based database (can store meta data)- Supports hybrid search (text + vector)- Easy to deploy- Familiar technology	<ul style="list-style-type: none">- Limited scalability(effective up to ~1M vectors)- Not optimized for large-scale ANN(approximate nearest neighbor) search
Backend	FastAPI	Flask, Django	<ul style="list-style-type: none">- Fast development with type hints- Async support- Built in data validation- Familiar technology	<ul style="list-style-type: none">- No built-in admin panel or authentication tools- Missing built-in tools for rate limiting, task queues, monitoring
Frontend	Next.js (React)	Vite + React, SvelteKit	<ul style="list-style-type: none">- File-based routing- React ecosystem- Familiar technology	<ul style="list-style-type: none">- Larger bundle size and complexity
Deployment	Docker compose	Kubernetes, AWS Lambda + API Gateway	<ul style="list-style-type: none">- Multi-service coordination (frontend, backend, database)- Reproducible Environments- Easy to deploy in different envs- Familiar technology	<ul style="list-style-type: none">- Limited scalability- No built-in autoscaling- No built-in load balancing