

- Key Features:**
- Hybrid Search (Vector + Keyword)
  - Cross-Encoder Reranking
  - Streaming Responses
  - Conversation Memory
  - 100% Local Processing

**Tech Stack:**

Frontend: Next.js + React  
 Backend: FastAPI  
 Vector DB: ChromaDB  
 Embeddings: Sentence Transformers  
 LLM: Llama 3.1 (Ollama)

# Studio Tours AI Assistant - High-Level Architecture

