

Deployment of AI-Powered Learning Assistant Using Retrieval-Augmented Generation (RAG)

Team members:
1. Harshitha Kapa
2. Abhiram T.S
3. Brian Temu

Project Overview & Milestones

Title: *AI Learning Assistant using RAG-based Q&A System*

- **Objective:** Build an AI assistant that enhances learning by answering document-specific questions using Retrieval-Augmented Generation (RAG).
- **Key Technologies:** Flask (backend), Next.js + TailwindCSS (frontend), ChromaDB (vector database), Railway (deployment).
- **Core Milestones:**
 - Data Collection & Embedding Generation
 - Backend API for document retrieval and response generation
 - Frontend web interface with real-time Q&A
 - Deployed live at: eda.briantemu.dev

Deployment Process and Decisions

Hosting:

Backend: Flask app hosted on [Railway](#)

Frontend: Next.js app deployed via Vercel

- ◆ **Vector Store:** ChromaDB used to store document embeddings locally (future-ready for external DB).
- ◆ **Interaction:** Real-time API integration for querying and responding.

Deployment Challenges:

- ◆ Flask and Next.js coordination
- ◆ Vector Database inference time optimization

GitHub Repository & GUI

GitHub Repository: [🔗 GitHub Link](#)

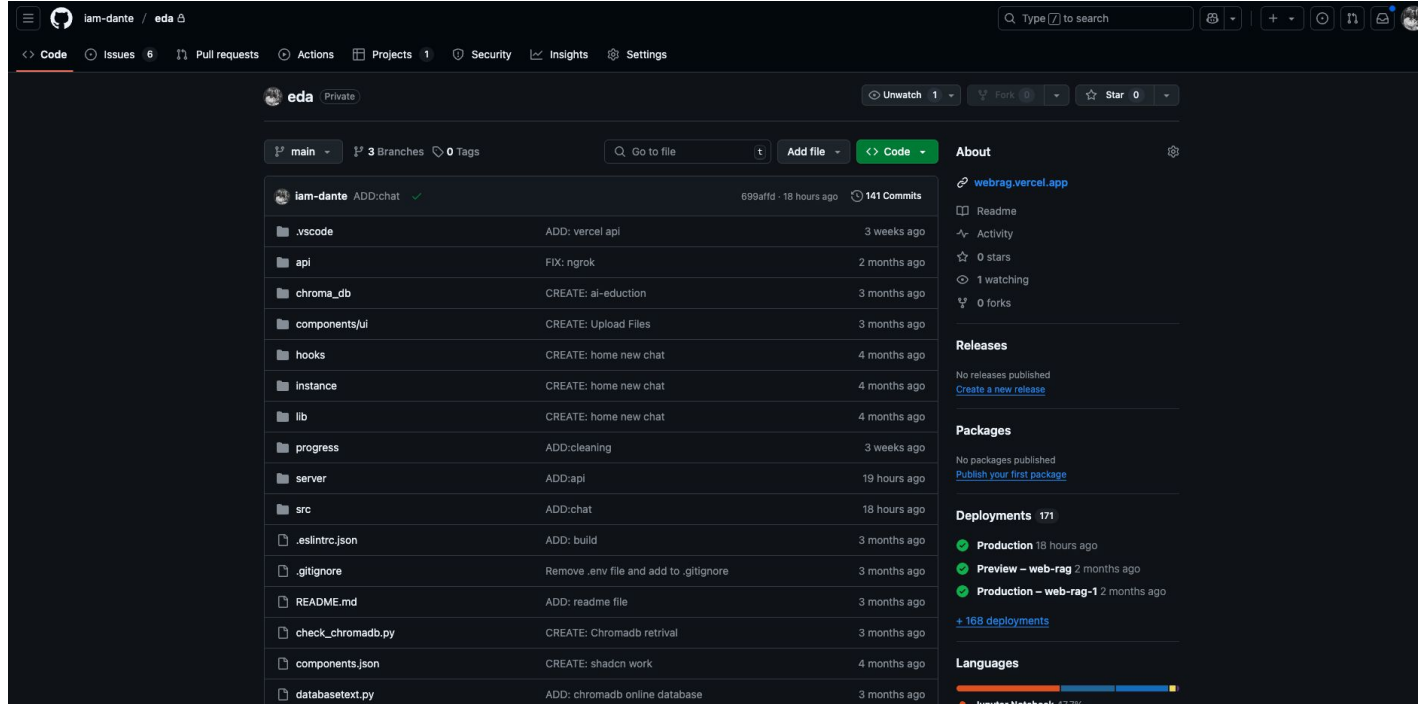
What's Included:

- Well-documented README
- Folder structure: `src/(frontend)` and `server/(backend)`
- Setup instructions for running locally or deploying

GUI Features:

- Clean, responsive chat UI
- Upload documents and ask context-based questions
- Shows real-time responses from RAG system

Github Repo Screenshot



The screenshot shows the GitHub interface for a repository named 'eda' owned by 'iam-dante'. The repository is private and has 141 commits. The main branch is 'main'. The repository contains several files and folders, including .vscode, api, chroma_db, components/ui, hooks, instance, lib, progress, server, src, .eslintrc.json, .gitignore, README.md, check_chromadb.py, components.json, and databasetext.py. The right sidebar shows the repository's metadata, including the web browser 'webbrowser.vercel.app', no releases published, no packages published, and 171 deployments. The languages section shows 'Jupyter Notebook' at 47.7%.

File/Folder	Commit Message	Time Ago
.vscode	ADD: vercel api	3 weeks ago
api	FIX: ngrok	2 months ago
chroma_db	CREATE: ai-eduction	3 months ago
components/ui	CREATE: Upload Files	3 months ago
hooks	CREATE: home new chat	4 months ago
instance	CREATE: home new chat	4 months ago
lib	CREATE: home new chat	4 months ago
progress	ADD:cleaning	3 weeks ago
server	ADD:api	19 hours ago
src	ADD:chat	18 hours ago
.eslintrc.json	ADD: build	3 months ago
.gitignore	Remove .env file and add to .gitignore	3 months ago
README.md	ADD: readme file	3 months ago
check_chromadb.py	CREATE: Chromadb retrival	3 months ago
components.json	CREATE: shadcn work	4 months ago
databasetext.py	ADD: chromadb online database	3 months ago

Future Work and Research Directions

1. Multi-Document and Multi-Modal Support

Expand capabilities to handle multiple documents and add support for .txt, docx, ppt, images, or video transcriptions as learning sources.

2. Real-Time Feedback Loop

Incorporate user feedback to automatically improve model responses or retrain based on question quality and relevance.

THANK YOU