# Abhay Raj

 BarryByte | Leetcode
 Abhay Raj | Portfolio | abhayraj.12667@gmail.com | +91 6200731408

## EDUCATION

2023 - present    BS+MS Computer Science at **Scaler School of Technology, Bangalore**

## WORK EXPERIENCE

**Data Science Intern @ Motion Education**                              *May 2025 – Sept 2025 — Bangalore, India*

- **Architected and deployed** a production-grade, full-stack **RAG chatbot system** serving as the primary **AI counselor for 55+ educational centers nationwide**, processing **1000+ daily conversations** with secure authentication, lead capture, and real-time streaming responses using **FastAPI, MongoDB, and React**.
- **Built enterprise-scale data ingestion pipeline** using **dual-crawler architecture** (**Selenium**) to extract and process **150+ URLs**, implementing **intelligent text cleaning, metadata enrichment**, and **batch processing with Pinecone vector embeddings**, reducing processing time by **60%**.
- **Engineered LLM orchestration** with **GPT-4 prompt chains** for **multi-entity query processing, intent recognition, and lead extraction**, implementing **JSON schema validation, context-aware query rewriting**, and **multi-location filtering** with alias mapping for enhanced retrieval accuracy.
- **Developed responsive React frontend** with **custom hooks architecture** , **real-time streaming responses**, and **modular component design**, while implementing **comprehensive logging pipeline** capturing full retrieval context in **MongoDB** for debugging and analytics.

## PROJECTS

**Lecture Notes Pipeline**                                          github.com/BarryByte/lecture-notes-pipeline

- **Built a 100% offline, privacy-first pipeline** that transforms lecture videos into structured Markdown notes using **FFmpeg, Whisper, and Ollama LLM**.
- **Engineered intelligent caching** based on video file hash to skip re-transcription, reducing repeat processing time by **90%**.
- **Optimized for CPU-only systems** with configurable Whisper and LLM models, processing 2-hour lectures in  44 minutes.

**Flashmob**                                                                      github.com/BarryByte/flashmob

- **Built a real-time flashcard app** using **React**, Express.js, and Firebase for collaborative study sessions.
- **Integrated Google Gemini 2.0 Flash API** to auto-generate Q&A pairs from uploaded materials, saving users time.
- **Enabled PDF export and Gmail-based deck sharing** for group learning and accessibility.

**GPT Teaching Assistant**                                          github.com/BarryByte/gpt-teaching-assistant

- **Built a Socratic-method DSA tutor** using **Gemini 2.5 Flash Lite**, featuring **3-level adaptive hinting** that guides users through problem-solving without giving direct solutions.
- **Developed multi-platform scrapers** (LeetCode GraphQL, Codeforces HTML) to fetch problem metadata, constraints, and hints for **real-time contextual tutoring**.
- **Implemented production-grade DevOps** with full **CI/CD pipeline** (Ruff, Bandit, Trivy, Pytest), **Docker containerization**, and **Kubernetes manifests** for scalable deployment.

## SKILLS

| | |
|---|---|
| **Languages** | Python, JavaScript, SQL |
| **AI/ML** | LangChain, Pinecone, OpenAI, Ollama, Whisper, Scikit-learn, XGBoost |
| **Backend** | FastAPI, MongoDB, REST APIs, WebSockets, Firebase |
| **Frontend** | React.js, Tailwind CSS, Vite |
| **Tools** | Docker, Git, GitHub Actions, Playwright, BeautifulSoup, Pandas, NumPy |