

BRAHMANAIDU MUCHUKUNTLA

Machine Learning & Associate AI Engineer

Bangalore, India — +91 8688064974 — brahmanaidu2172@gmail.com

[LinkedIn Profile](#) — [GitHub Profile](#) — [Portfolio Website](#)

Professional Summary

Computer Science graduate specializing in Machine Learning, similarity search, and generative AI workflows. Leverages internship experience in building CLIP-based visual search pipelines, fine-tuning vector retrieval, and deploying cloud microservices. Combines a strong mathematical foundation with backend engineering to build scalable, low-latency AI solutions.

Technical Expertise

Programming: Python (Advanced), SQL (PostgreSQL), JavaScript, C + +, Go
AI & Machine Learning: PyTorch, Hugging Face, LangChain, RAG Systems, FAISS, LLMs, CrewAI
Backend & Cloud: FastAPI, Node.js, AWS (Lambda, S3, SQS), Docker, Terraform
Databases: PostgreSQL, pgvector, MongoDB Atlas, Pinecone (Vector DB), Redis

Professional Experience

Zithara.AI (Propel5000) — Machine Learning Engineer Intern Jan 2025 – Apr 2025 — Hyderabad, India

- Architected a Visual Search Engine for an e-commerce platform, enabling sub-100ms vector retrieval across commercial catalog SKUs.
- Built an OpenAI CLIP-based image embedding pipeline generating 512-dimensional features for visual catalog mapping.
- Implemented MongoDB Atlas Vector Search using HNSW indexing and cosine similarity, eliminating lookup bottlenecks.
- Integrated async FastAPI microservices with Docker, connecting vector search pipelines to Tailwind frontends.

Education

B.Tech in Computer Science & Engineering — Annamacharya Institute of Technology & Sciences Apr 2025 — 7.9 CGPA

Engineering Projects

Production-Grade RAG System — Python, FastAPI, pgvector, LLMs, Docker May 2026 – Aug 2026

- Engineered an end-to-end RAG platform for document QA using hybrid semantic/keyword search, cross-encoder reranking, and contextual chunking.
- Built source citation mechanisms to reduce hallucinations and an automated evaluation pipeline measuring retrieval accuracy.
- Containerized and deployed scalable REST APIs using Docker and FastAPI for document ingestion and QA retrieval.

Multimodal Healthcare Diagnostic Predictor — Python, PyTorch, EfficientNet-B0 Jan 2026 – Apr 2026

- Developed a 7-class skin-lesion classification system using the HAM10000 dataset, fusing dermoscopic images with patient metadata.
- Engineered EfficientNet-B0 encodings, categorical embeddings, late-feature fusion, and lesion-aware data splitting.
- Built an end-to-end ML pipeline with Grad-CAM/SHAP interpretability, utilizing config-driven experiments and Git version control.

Multi-Agent Autonomous Research Assistant — CrewAI, Python, LLM APIs May 2025 – Dec 2025

- Architected a multi-agent AI workflow using CrewAI to autonomously research, extract, analyze, and synthesize web information.
- Implemented specialized agents with function calling and state management for automated research and report generation.
- Reduced research effort by 3 hours per task while improving consistency and scalability across workflows.

Key Achievements

- Algorithmic Problem-Solving:** Solved 250+ Data Structures and Algorithms problems; ranked in top 10% of college hackathons.
- Open Source Contributions:** Authored accepted bug fixes and documentation updates for major Python AI repositories (e.g., LangChain, Hugging Face) on GitHub.