SINGISETTI CHITRAKSH

chitraksh456@gmail.com • 9398770578 • Linkedin • Github

Professional Experience

AI Engineer — Orion Innovation, Chennai, India *Nov 2024 – Present*

Client Project — GenAI Financial Document Intelligence

- Developed a **GenAI document intelligence** solution for financial document Q&A, cutting analyst document-search time by 40%.
- Built end-to-end **RAG pipelines** using embeddings, semantic search, and an **OpenSearch** vector database, improving retrieval recall by 25%.
- Designed **agentic AI workflows** in **LangGraph** for multi-step orchestration, retrieval, and tool integration, cutting manual follow-ups by 30%.
- Applied **prompt engineering**, few-shot prompting, and context grounding, improving accuracy by 20% and cutting hallucinations by 35%.
- Built **LLM evaluation workflows** for relevance, recall, faithfulness, and consistency, raising faithfulness by 15%.
- Used **AWS S3, EC2, and Lambda** for scalable document processing and LLM inference, lowering latency by 25%.

Nexus — Enterprise GenAI Knowledge Assistant

- Developed **OI Nexus**, an internal **AI chatbot** for employee queries on HR policies, hybrid-work, and leave, cutting query resolution time by 40% and repetitive HR tickets by 30%.
- Built and optimized **RAG pipelines** using **embeddings** and **vector search** over enterprise document repositories, improving retrieval accuracy by 25% and reducing irrelevant results by 35%.
- Engineered **document ingestion, chunking, and indexing** workflows for large internal knowledge bases, processing 10,000+ documents while reducing indexing latency by 20%.
- Refined response quality via **prompt engineering**, retrieval tuning, and evaluation, increasing answer accuracy by 15% and user satisfaction by 20%.
- Partnered with HR and product stakeholders to define requirements and ship features, driving a 30% adoption increase in the first quarter.

Software Engineer Intern — Liquidmind AI, Bangalore, India *Jul 2024 – Nov 2024*

- Developed and deployed backend **Flask APIs** to integrate a **WhatsApp bot** for business automation workflows, reducing manual message-handling effort by 30%.
- Built an **OCR pipeline** using **Document AI** to extract structured information from unstructured documents, improving data extraction accuracy and cutting manual data-entry time by 40%.
- Collaborated with senior engineers to test and refine API endpoints prior to release, improving system reliability and reducing post-deployment bug reports.

Projects

Plant Disease Detection Using Feature Fusion

Tools: Python, TensorFlow, Keras, NumPy, Pandas, Scikit-learn, OpenCV, Jupyter

- Designed a **deep learning** model fusing spatial (RGB) and spectral (**hyperspectral**) features for plant disease classification, achieving **97% classification accuracy** — a **5% improvement** over single-modality baseline models.
- Processed **hyperspectral plant image data** (spectral band selection, normalization, feature extraction) and augmented **5,000+ images** with **OpenCV** and **NumPy**, improving generalization and reducing overfitting by 15%.
- Implemented **feature fusion** of handcrafted descriptors (**HOG**) with **ResNet50** deep features, improving robustness under varying lighting conditions.
- Co-authored the findings for publication at an **IEEE conference**, contributing novel multi-modal fusion techniques to agricultural disease-detection research.

MicroManus — Deep Research AI Agent

Tools: Next.js, React, TypeScript, Node.js, Supabase (PostgreSQL, OAuth), Stripe, SerpAPI, OpenAI API, Anthropic Claude API, AES-256-GCM, Vercel, Agentic AI

- ***Live:*** ***Micromanus*** — Architected and shipped MicroManus, a full-stack **deep research AI agent** with a visible agent reasoning loop, giving users real-time transparency into the model's multi-step research and tool-use process.
- Implemented social login via **OAuth** and secure authentication, alongside usage-based billing, using **Supabase** and **Stripe**, enabling production-ready onboarding and metered monetization.
- Designed a **bring-your-own-key (BYOK)** model support system, encrypting user-provided API keys with **AES-256-GCM** for secure, self-managed access to multiple LLM providers.
- Integrated **SerpAPI** for live web search grounding and deployed on **Vercel**, improving factual accuracy of research outputs and delivering a fully live, publicly accessible product.

Technical Skills

Languages: Python, C

ML/DL: TensorFlow, Keras, PyTorch, Scikit-learn, Pandas, NumPy, Hugging Face Transformers.

GenAI & LLMs: RAG, LangChain, LangGraph, LlamaIndex, OpenAI API, Anthropic Claude API, Prompt Engineering, LLM Evaluation (RAGAS), Fine-tuning (LoRA, QLoRA), vLLM, Ollama.

Agentic AI: AI Agents, MCP, Function/Tool Calling, Structured Outputs, Model Context Engineering.

Data & Vector DBs: FAISS, Chroma DB, Pinecone, OpenSearch, MySQL.

Cloud & DevOps: AWS, Docker, Kubernetes, CI/CD, GitHub Actions, FastAPI, Git, Jupyter.

Education

B.Tech — Computer and Communication Engineering *2020 – 2024*

Amrita Vishwa Vidyapeetham, Coimbatore, India

Certifications & Publications

- AWS Certified Cloud Practitioner
- IEEE Conference Publication: Plant Disease Detection Using Feature Fusion