Anany Sharma

Gainesville Florida
Linkedin: linkedin.com/in/ananyd36

Email: anany.sharma@ufl.edu
Github: github.com/ananyd36

Education

University Of Florida — Gainesville, Florida
Masters in Artificial Intelligence Systems — *Expected Graduation: May 2026*
Coursework: Computer Vision, Natural Language Processing, Machine Learning, Data Structures and Algorithms, Applied Deep Learning, ML for Time Series, EdgeAI, Data Science, Model Deployment, Distributed Computing, Predictive Modeling, Recommendation Systems.

Experience

Graduate AI Research Assistant — Gainesville, Florida
University of Florida — *March 2025 - Present*

- Developed and deployed **edge AI** models on **AHA (AI Hardware Adventure) board** utilizing **5+ sensors** for real-time decision-making applications, achieving **95% accuracy** in data processing.
- Optimized edge AI models through **Edge Impulse** platform, reducing **inference time by 40%** while maintaining model accuracy on **IoT** devices.
- Leading development of comprehensive edge AI curriculum reaching **200+ students** from **K-12 to undergraduate** level, integrating edge AI concepts to equip future innovators with hands-on experience in cutting-edge technologies.

Software Engineer — Noida, India
United Health Group (Optum) — *January 2022 - August 2024*

- Implemented **RAG agents** with **Azure AI Search** and **OpenAI**, achieving **40% faster retrieval** and **50% higher adoption** through **ReactJS/Flask** architecture. Optimized **Azure Blob Storage** for **20%** lower latency.
- Automated claims validation using **Azure Document Intelligence** and **NLP**, reducing processing time by **20%** and improving extraction accuracy by **30%**
- Created real-time **Power BI dashboards** for business KPIs, enabling data-driven decisions and reducing **reporting latency by 40%.**
- Automated **1.5K+ biweekly** financial reports using **MS BI stack**, reducing generation time by **20%** and ensuring business continuity.
- Architected ETL pipelines handling **10M+ monthly claims** using **IBM WTX** on **UNIX AIX** servers, optimizing throughput with IBM message queues service.

Machine Learning Intern — Remote, India
MFIT Technologies — *September 2021 - December 2021*

- Achieved **85% field-extraction accuracy** by developing **NLP-powered** financial data system using **OCR, CRF models**, and spatial modeling for automated transaction monitoring .
- Engineered scalable data extraction system supporting **10+ document formats** by implementing **hybrid NER and layout-aware** framework.
- Built dynamic **CRF-Spatial** model with geometric features enabling robust parsing of financial statements across **4 major banks**.

Skills Summary

- **Languages and Frameworks:** Python (expert), C++ (intermediate), Javascript (intermediate), Dart (intermediate), SQL (expert), NoSQL (intermediate), Django (intermediate), Flask (proficient), FastAPI (proficient), ReactJS (proficient).
- **Technologies:** Azure, Amazon Web Services (AWS), Pytorch, Open AI, Tensorflow, Keras LangChain, LlamaIndex, LangGraph, Scikit-learn, Pinecone, CrewAI, Hugging Face, Supabase, Linux, Git, MSBI(SSIS, SSRS, Power BI), Docker.

Projects

- **SMIRE AI - A Medical Multi-AI Agent system (GenAI)GITHUB:** An AI-powered medical assistant offering services like appointment booking, doctor/clinic search, medical news, consultations, and health management (reports, dosages, insights(**RAG**)) in a single platform. Tech Stack: **NextJS**, **FastAPI**, **Supabase**,**OpenAI/CrewAI**, **PostgreSQL**, **Docker**, **Git**.
- **CORAS - Context-Based Intelligent Knowledge Retrieval System (GenAI)GITHUB:** Built a **Retrieval-Augmented Generation (RAG)** system using **OpenAI embeddings**, **Pinecone** vector indexing, **Flask API**, and **OpenAI Whisper**. Integrated multimodal capabilities to handle both text and audio data. Leveraged **Prometheus** and **Grafana** for monitoring.

Patent Publication

- **"Crowd Detection and a Method Thereof - IIP(2021)" PATENT** : Developed a **YOLOv5-based** multimodal PPE and mask detection system with high-risk crowd detection, leveraging **Python, Flask, OpenCV, Darknet**, and **Docker scalability**. Delivered real-time alerts and optimized performance using **transfer learning** and **mAP evaluation**. Achieved an average precision of **60%** on 'PPE kit' class, **85%** AP on 'mask' class and **80%** on 'No mask' class.