PRANEETH SAI CHUNCHU

+1 (659) 346-6106 | praneethsaichunchu@gmail.com | **Linkedin** | **Portfolio**

EXPERIENCE

AI Intern & Co-Founding Team Member - LifeSizeAgents.ai, USA   **Mar 2025 - Present**

- Built backend pipelines with Python, LangChain, CrewAI, and vector databases, powering LLM AI agents and improving processing speed by 80%, serving 4000+ queries/day.
- Integrated highly interactive 3D avatar models (.glb) into a React + Three.js frontend with backend APIs, enabling real-time life-size avatars that enhanced user engagement by 70% and delivered immersive experiences.
- Developed RAG pipelines with embeddings, semantic search, and knowledge graphs, increasing contextual recommendation accuracy by 75% for 200+ users.
- Optimized multi-agent orchestration, prompt tuning, and embedding models for scalable, low-latency AI, reducing response time from 5s to 1s.

ML Engineer - KaarTech Solutions, India   **Jan 2023 – Dec 2023**

- Built and deployed production-ready ML models (classification, NLP, clustering) using TensorFlow, PyTorch, scikit-learn, serving real-time business applications with 90% uptime.
- Automated training, evaluation, and deployment with MLflow, FastAPI, and Docker, reducing release cycle time by 50%.
- Containerized ML workflows and orchestrated with Kubernetes, enabling scalable, low-latency inference (<500ms) under heavy traffic for multiple applications.
- Implemented monitoring and alerting frameworks using Prometheus & Grafana, detecting data drift and performance issues in real-time, improving production reliability by 40%.

Research Assistant - GVPCE, India   **June 2022 – Nov 2022**

- Designed and trained Convolutional Neural Networks (CNNs) with transfer learning (ResNet, EfficientNet) for face detection and recognition, enabling automated student attendance tracking with 90%+ accuracy.
- Optimized models using quantization, pruning, and TensorRT acceleration, reducing inference latency by 65% and allowing real-time video analytics on classroom camera sensors.
- Boosted CNN accuracy using OpenCV preprocessing for robust recognition in varied lighting and occlusion.

PROJECTS

End-to-End MLOps Pipeline for Predictive Maintenance of Turbines.   **June 2025 – Aug 2025**

- Developed modular 6-stage MLOps pipeline with Airflow from ETL to prediction.
- Integrated MLflow for experiment tracking, model registry, automated hyperparameter tuning, and model versioning.
- Containerized with Docker, automated CI/CD via GitHub Actions, and served models through FastAPI.
- Monitored data, drift, and model performance with Grafana and Prometheus.
- Project Link - Github Code

Attendance Management System Using CNN and HOG   **May 2024 – July 2024**

- Developed a real-time web application with Python, Django, HTML/CSS, and JavaScript to manage user records & monitor attendance.
- Integrated OpenCV and CNNs for accurate facial recognition and real-time tracking.
- Containerized the application using Docker and deployed it on AWS EKS for scalable cloud orchestration.
- Configured a custom domain for public access, demonstrating full-stack development and cloud deployment expertise.
- Project Link - Github Code

CERTIFICATIONS &ACHIEVEMENTS

1. AWS Certified ML Engineer - Assosiate
2. Databricks Certified Associate Data Engineer
3. Complete MLOps Bootcamp by Udemy
4. ML in Production by Andrew Ng
5. AWS Certified Cloud Practitioner - Foundational
6. Solved 350+ Python (DSA) problems and 70+ SQL problems on LeetCode, earning 7+ LeetCode badges.   **Leetcode Profile**

SKILLS

Software Development & Cloud: Python, Java, C, JavaScript, Django, Flask, FastAPI, React.js, React Three.js, AWS, Databricks, Docker, Kubernetes, Git, Jenkins, Airflow, Lambda, Glue, MySQL, PostgreSQL, MongoDB.

Machine Learning & AI: TensorFlow, PyTorch, Scikit-learn, Hugging Face, OpenAI APIs, BERT, GPT, Dlib, OpenCV, MLflow, DVC, Prometheus, Grafana.

EDUCATION

Master of Science, Computer Science | GPA: 3.87 / 4.0   **Jan 2024 – 2025**
University of Alabama at Birmingham

Bachelor of Technology, Electronics and Communication Engineering | GPA: 8.8 / 10   **June 2019 – April 2023**
Gayatri Vidya Parishad College of Engineering, Visakhapatnam, India

Relevant Course Work: Data Structures and Algorithms, Database Systems, Machine Learning, Deep Learning, Cloud Computing, Data Mining, Advanced Web Applications, Advance Algorithms and Apps.