SEEMANTH RAJU KURAPATI

+91 89195 68249 | seemanth.kurapati@gmail.com
linkedin.com/in/kurapati-seemanth-raju | github.com/seemanth-raju

PROFESSIONAL SUMMARY

AI / Generative AI Engineer with experience in Python-based backend systems, LLM-integrated applications, and real-time inference pipelines. Skilled in building production-ready AI services including GenAI voicebots, multimodal systems, and cloud-hosted microservices with low latency and high reliability.

TECHNICAL SKILLS

Languages: Python, SQL, JavaScript
Frameworks: FastAPI, Flask, PyTorch, TensorFlow, LangChain, RAG
Databases: PostgreSQL, MySQL, Firebase (Cloud Firestore)
Cloud & DevOps: GCP, Azure, Docker, Linux
AI / ML: LLMs, Generative AI, NLP, Deep Learning, Time Series, Computer Vision, Reinforcement Learning, Transfer Learning
Tools: Git, Docker, ReactJS, Pandas, NumPy, Scikit-learn, OpenAI, Gemini, LlamaIndex
Practices: RESTful APIs, Microservices, Telephony/IVR, Unit Testing, CI/CD basics, Agile

PROFESSIONAL EXPERIENCE

Purview India Consulting and Services LLP — *Hyderabad, India*
AI Engineer — *Jul 2024 – Present*

- Built **production-grade AI services** including a telephony GenAI voicebot, smart glasses backend, multimodal query systems, and a virtual receptionist, focusing on **low-latency Python backends** and reliable deployment
- Designed and implemented an end-to-end **telephony GenAI voicebot** for customer and loan support, integrating telephony APIs with **streaming ASR, LLM, and TTS** for natural phone conversations with **sub-second turn latency**
- **AI Inference System for Amazon Warehouse Automation:** Built a scalable, on-premise inference pipeline for vehicle navigation using **200+ camera feeds**, with **FastAPI microservices** and PostgreSQL handling **5,000+ detections per hour**
- **SEVA Smart Glasses Backend:** Developed Python backend for AR features (scene description, **OCR, object detection, LLM Q&A**), maintaining **˜95% uptime** under 1,000+ requests/hour
- **Multimodal "Talk to Me" Mode:** Integrated **LLMs with RAG** and web search APIs, reaching **90%+ query resolution** and supporting multi-user sessions
- **Virtual Receptionist Backend:** Built an **LLM-powered receptionist** using **LangChain** for natural-language office visitor interaction serving **200+ users/day**, with fallback-to-human flows
- Optimised backend performance across services, reducing **latency by ˜30%** and improving API reliability, contributing to **98% satisfaction** and **40% fewer escalations**

IBM — *India*
Data Analyst Intern — *Jul 2023 – Aug 2023*

- Cleaned and preprocessed **enterprise datasets**, improving downstream **model accuracy by ˜15%** through better handling of missing data, noise, and inconsistencies
- Performed **EDA** using **Pandas, Seaborn, and Matplotlib** to uncover trends and anomalies for internal decision-making
- Automated reporting and transformation tasks with **Python scripts**, boosting **pipeline efficiency by ˜25%**

PROJECTS

PCAP Flow Analysis and Anomaly Detection Pipeline — *Jul 2025 – Aug 2025*

- Built a **Python pipeline** to process **1GB+ PCAP files** and extract millions of network flows with **Scapy**
- Applied **K-Means clustering** to segment traffic and detect **SYN flood and suspicious UDP patterns**, reducing manual investigation scope by **˜95%** through automated rare-cluster flagging

EDUCATION

JNTUH University College of Engineering — *Hyderabad, India*
B.Tech, Computer Science and Engineering — *Dec 2020 – Jun 2024*

ACHIEVEMENTS

Led a team at national-level Krithathon to build a CT-scan based lung cancer prediction model, achieving runner-up position.