Viresh Duvvuri

Seattle, WA | +1-509-964-5469 | vireshduvvuri@gmail.com | linkedin.com/in/viresh-duvvuri

AI Engineer building production AI systems using LangChain, MCP, Python, and LLM integrations (OpenAI, Anthropic, Ollama). Developed multi-agent systems and RAG-based tools that improved operational efficiency by 50-80%, serving 200+ daily users in production. Experience spans GenAI prototyping, model evaluation, prompt engineering, and full-stack development. 5+ years shipping AI agents from concept to production across robotics and cloud platforms.

Skills

Programming: Python, JavaScript, TypeScript, C++, SQL, FastAPI, Flask, React, NumPy, Pandas, OOP

AI/ML & Model Integration: Agentic AI, LangChain, LangGraph, Multi-Agent Systems, MCP (Model Context Protocol), OpenAI APIs, Anthropic (Claude), Ollama, Llama, RAG, Context Engineering, Prompt Engineering, Model Evaluation, MLOps, GenAI, FAISS, Pinecone, Vector Search, PyTorch, TensorFlow, Scikit-learn, Model Deployment

Cloud & Infrastructure: AWS, Azure, API Design, Deployment, DevOps, Docker, Kubernetes, Monitoring, Performance Tuning, Scalability

Data & Analytics: Data Integration, Data Processing, Data Science, Knowledge Graph, Feature Engineering, Operational Efficiency

Work Experience

Grid CoOperator — Seattle, WA
AI Engineer — *Mar 2025 - Present*

- Developed AI-enabled data processing system using LangChain, Python, and SQL databases from concept to deployment, reducing analyst research workflows by 70% within 2 months through intelligent query generation
- Built scalable backend service with API architecture handling 50-100 daily queries, ensuring reliable performance for real-time smart grid data analysis and operational decision support
- Implemented automated report generation pipeline accelerating stakeholder deliverables by 60% within first quarter, eliminating manual documentation processes for utility operations

Freefly Systems — Woodinville, WA
Senior Software Engineer — *Nov 2021 - Oct 2025*

- Independently designed and built AI-powered diagnostic tool using Python and modern LLM frameworks (Ollama, Llama 3.2) from requirements to production, serving 200+ daily queries
- Built automated systems to process complex technical data and identify system failures, developing knowledge base enhancements and support tools that streamlined operations
- Contributed to drone platform codebases implementing new features and optimizations for flight control systems and payload integration across multiple product lines, managed software integration projects from planning through release
- Led release management for drone platforms overseeing testing phases from alpha through production deployment, coordinating firmware updates and executing comprehensive testing protocols with cross-functional teams

Lumenier — Sarasota, FL
Drone Software Developer — *Jul 2020 - Oct 2021*

- Wrote embedded code in C++ to integrate LiDAR and optical flow sensors for obstacle avoidance and position holding with/without GPS under various lighting conditions
- Collaborated with open-source flight control software maintainers for integration, testing, and deployment of autonomous flight algorithms, prototyped innovative features like toss-to-launch for product roadmap development

York Exponential — York, PA
Software Engineer - R&D — *Aug 2018 - May 2020*

- Developed prototype software for in-house autonomous surveillance mobile robots using ROS2, SLAM, and computer vision technologies
- Built Human Machine Interface for Universal Robot welding applications using Python and Kivy framework, implemented multi-robot control systems with platform independence

Education

Washington State University — Pullman, WA
Master of Science Computer Science — *Jan 2015 - Jan 2017*

GITAM University — Visakhapatnam, India
Bachelor of Technology Information Technology — *Jan 2011 - Jan 2015*

Projects

GridCOP: Smart Grid Analytics Agent

- Problem: Power grid analysts needed automated database querying and intelligent insights to understand complex data patterns beyond basic visualizations
- Solution: Developed A2A multi-agent system using LangChain orchestration and MCP where specialized agents coordinate tasks through prompt engineering strategies, implemented RAG and vector search (FAISS) for intelligent querying, implemented model evaluation frameworks to monitor quality and cost metrics, deployed on AWS with observability and logging
- Impact: Enhanced analyst productivity by 70% through AI co-pilot that augments domain experts with automated workflows, implemented human-in-the-loop (HIL) evaluation and testing pipelines for production-ready AI systems with robust error handling through rapid iteration

Automated Drone Diagnostics System

- Problem: Manual analysis of drone flight logs required deep expertise and took hours per incident, limiting support team scalability and slowing customer issue resolution
- Solution: Developed full-stack GenAI application (React + Flask) using Ollama and Llama 3.2, implemented RAG to retrieve relevant log sections and recommend fixes based on similar past failures
- Impact: Automated 80% of routine diagnostics, serving 200+ daily queries, enabling support team to handle 3x more cases with faster resolution times

AI Travel Planner Agent

- Problem: Manual travel planning requiring hours of research across multiple sources with inconsistent and outdated information
- Solution: Built AI agent using Claude 3.5 Sonnet and LangChain for personalized itinerary generation, experimented with prompt chaining and context handling techniques
- Impact: Demonstrated rapid prototyping and end-to-end AI application development through scrappy iteration and continuous tinkering