# Viresh Duvvuri

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

Full-stack AI Engineer who ships production AI systems fast. Currently building multi-agent AI applications and RAG-based products using LangChain, Python, and React that deliver 50-80% efficiency gains in production. Recent work includes deploying agentic workflows serving 200+ daily users and building end-to-end AI features from architecture to deployment. Strong ownership mentality: independently designed, built, and shipped AI-powered tools from concept to production in high-velocity environments. Experience spans LLM integrations, vector search, agent frameworks, and scalable backend systems on AWS.

## Skills

**AI & LLM Engineering:** Agentic AI, LangChain, LangGraph, Multi-Agent Systems, MCP (Model Context Protocol), RAG, Vector Search, FAISS, Pinecone, LLM APIs, OpenAI, Anthropic Claude, Ollama, Prompt Engineering, Context Engineering, Model Evaluation, deepeval, LangSmith

**Full-Stack Development:** React, JavaScript, TypeScript, Python, FastAPI, Flask, Node.js, REST APIs, GraphQL, Frontend Frameworks, Backend Systems

**Cloud & DevOps:** AWS, GCP, Azure, Docker, CI/CD, Deployment, Monitoring, Scalability, Performance Optimization

**Data & Systems:** SQL, Data Pipelines, ETL, NumPy, Pandas, C++, System Architecture, API Design, Production Systems

## 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: Multi-Agent Smart Grid Analytics System**

- Problem: Power grid analysts needed automated database querying and intelligent insights to understand complex data patterns beyond basic visualizations
- Solution: Built end-to-end multi-agent system using LangChain orchestration and MCP where specialized agents coordinate tasks, implemented RAG with FAISS vector search, monitored agent decision quality using LangSmith tracing and deepeval achieving 0.85+ context precision, deployed on AWS with full observability and error handling
- Impact: Shipped production-ready AI system that enhanced analyst productivity by 70%, built data pipeline to collect and annotate 500+ user queries for continuous improvement, demonstrated rapid iteration from prototype to production in 2 months

**AI-Powered Diagnostic Tool (Production)**

- Problem: Manual system analysis taking hours of expert time, creating bottlenecks in product development and customer support
- Solution: Independently shipped full-stack application with React frontend and Python Flask backend, integrated LLM APIs (Ollama, Llama 3.2) for real-time log processing, implemented RAG pipeline for context-aware analysis with prompt engineering and evaluation
- Impact: Transformed hours-long expert analysis into minutes, deployed to production serving 200+ daily active users, demonstrated full ownership from requirements gathering through production deployment and iteration

**AI Travel Planner Agent**

- Problem: Manual travel planning requiring hours of research across multiple sources with inconsistent information
- Solution: Built AI agent using Claude 3.5 Sonnet, LangChain, and Streamlit with real-time web search integration via DuckDuckGo API for personalized itinerary generation
- Impact: Shipped working prototype in days demonstrating rapid development velocity, learned agentic workflow patterns and API integration techniques through iterative building