Viresh Duvvuri

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

AI Engineer specializing in production AI systems with Model Context Protocol (MCP), context-aware chatbots, and multi-agent orchestration. 5+ years developing GenAI solutions including AI assistants serving 200+ daily users, explainable AI features, and intelligent task automation. Expert in LangChain, RAG, and deploying transparent, user-facing AI experiences on AWS that improve workflow efficiency by 50-80%.

Skills

AI/ML & Orchestration: Model Context Protocol (MCP), LangChain, LangGraph, Multi-Agent Systems, Agentic AI, RAG, Context Engineering, Prompt Engineering, Explainable AI, Model Evaluation, MLOps, LLMs, Embeddings, Vector Search, Human-in-the-Loop

Programming & Backend: Python, JavaScript, TypeScript, Node.js, FastAPI, Flask, SQL, PostgreSQL, Redis, C++, NumPy, Pandas

Cloud & Deployment: AWS, Azure, Docker, Kubernetes, CI/CD, API Design, Monitoring, Performance Tuning, Scalability

Frontend & UX: React, Streamlit, Interactive Dashboards, Data Visualization, User-Facing AI

Data & Analytics: Time-Series Analysis, Data Processing, Knowledge Graphs, Forecasting, Analytics Pipelines, Data Integration

Work Experience

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

- Architected multi-agent AI system using LangChain and Model Context Protocol (MCP) for smart grid analytics, where specialized agents coordinate complex workflows and maintain long-lived context across user interactions, implementing explainable AI with confidence scoring and transparent decision reasoning that reduced analyst workflows by 70% within 2 months
- Built context-aware AI assistant that proactively surfaces recommendations, automates data retrieval, and provides interactive analytics through natural language queries over structured grid data, deployed on AWS (Lambda, DynamoDB) with observability tracking quality metrics, latency, and user satisfaction across 50-100 daily queries
- Implemented production AI features with human-in-the-loop validation, model evaluation pipelines, and MLOps best practices, establishing robust error handling and graceful degradation patterns that achieve 99%+ uptime and accelerated deliverables by 60% through rapid iteration

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

- Built and deployed context-aware AI chatbot for automated diagnostic analysis, integrating foundation model APIs (Ollama, Llama 3.2) with RAG and vector embeddings, implementing explainable AI features (source citations, confidence scoring) that provide transparent insights to users, serving 200+ daily queries in production with 80% reduction in analysis time
- 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 AI Assistant with MCP
- Problem: Power grid analysts needed intelligent automation for complex workflows with transparent explainable decisions

- Solution: Architected multi-agent system using LangChain and Model Context Protocol where specialized agents coordinate tasks and maintain context across interactions. Implemented RAG with vector search, confidence scoring for transparent AI decisions, and interactive analytics dashboards. Deployed on AWS with MLOps pipelines
- Impact: 70% reduction in analyst workflows, 50-100 daily queries, 99% uptime with production-ready explainable AI

AI Diagnostic Chatbot for Engineering Teams
- Problem: Manual system diagnostics taking hours of expert analysis creating bottlenecks
- Solution: Built production context-aware chatbot with React frontend and Python Flask backend, integrated foundation model APIs with RAG for natural language queries. Implemented explainable AI features with source citations and confidence scoring
- Impact: Transformed hours into minutes, 200+ daily users in production, 80% time reduction

AI Travel Planning Agent
- Problem: Travel planning requiring hours of research across fragmented sources
- Solution: Developed AI agent using Claude and LangChain with real-time web search integration for personalized itinerary generation
- Impact: End-to-end demonstration of AI agent development and conversational AI patterns