Viresh Duvvuri

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

Senior AI Software Engineer

Senior AI Software Engineer specializing in prototype development and production AI systems, with 7+ years building scalable ML solutions from concept to deployment. Proven track record developing AI/ML prototypes that scale to production in small team environments, delivering 70-80% efficiency improvements through rapid iteration, MLOps practices, and AWS infrastructure. Strong full-stack capability with autonomous project ownership from initial concept through production deployment.

Skills

Programming: Python, C++, JavaScript, Java (learning), SQL, FastAPI, Flask, React

AI/ML & Prototyping: Machine Learning, Prototype Development, MLOps, Model Development, NLP, Model Deployment, Model Evaluation

Cloud & MLOps: AWS, MLOps, CI/CD, Docker, Monitoring, Performance Tuning, Scalability

APIs & Integration: FastAPI, Flask, REST APIs, WebSocket, API Design, LLM Integration

Work Experience

Grid CoOperator — Seattle, WA
AI Engineer — *03/2025 - Present*

- Developed and scaled AI prototype from concept to production multi-agent system in 2 months, implemented MLOps practices including evaluation frameworks and monitoring, delivered 70% efficiency improvement through rapid iteration and testing with real users
- Built production-ready AI system on AWS infrastructure with API backend using FastAPI, established monitoring tracking quality metrics, latency, and cost optimization across 50-100 daily queries
- Independently led technical implementation collaborating with business stakeholders, designed evaluation frameworks and testing pipelines, provided technical guidance on AI capabilities and best practices

Freefly Systems — Woodinville, WA
Senior Software Engineer — *11/2021 - 10/2025*

- Independently designed and developed AI-powered diagnostic tool from prototype to production serving 200+ daily users, built full-stack solution with React and Flask with LLM integration, delivered 80% productivity improvement through autonomous project ownership
- 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
- Built automated systems to process complex technical data and identify system failures, developing knowledge base enhancements and support tools that streamlined operations

Lumenier — Sarasota, FL
Drone Software Developer — *07/2020 - 10/2021*

- 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 — *08/2018 - 05/2020*

- 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 — *01/2015 - 01/2017*

GITAM University — Visakhapatnam, India
Bachelor of Technology Information Technology — *01/2011 - 01/2015*

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 with 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 evaluation and testing pipelines for production-ready AI systems with robust error handling through rapid iteration

Production System Optimization Tool

- Problem: Manual system analysis taking hours of expert time, creating bottlenecks in product development and customer support resolution
- Solution: Built full-stack application with React frontend, Python Flask backend, integrated foundation model APIs with Ollama and Llama 3.2 for real-time log processing and interactive analysis using prompt engineering and model evaluation
- Impact: Transformed expert analysis from hours to minutes, deployed to production serving 200+ daily queries with significant performance improvements through rapid iteration and continuous optimization

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, LangChain, Streamlit, and DuckDuckGo Search API for personalized itinerary generation using prompt engineering techniques
- Impact: Demonstrated end-to-end AI application development, learned conversational AI patterns and real-time data integration techniques through iterative development