Viresh Duvvuri

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

Software Engineer

Software Engineer specializing in backend systems and AI integration, with 5+ years building scalable APIs and distributed services that leverage foundational AI models. Expert in designing microservices architecture, backend infrastructure, and AI orchestration through rapid prototyping and test-and-learn experimentation. Proven track record deploying production systems on AWS with 50-80% efficiency improvements.

SKILLS

Backend & Infrastructure: Python, FastAPI, Flask, REST APIs, Microservices Architecture, Distributed Systems, Backend Development, API Design, System Integration, Cloud Infrastructure (AWS, Azure), Docker, Kubernetes, CI/CD Pipelines, Monitoring, Scalability

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

AI Integration: LangChain, LangGraph, Agentic AI, Multi-Agent Systems, MCP (Model Context Protocol), RAG, Vector Databases (FAISS, Pinecone), Foundation Model APIs (OpenAI, Claude, Hugging Face, Ollama), Prompt Engineering, Model Evaluation

Data & DevOps: Data Integration, Data Processing, SQL, Snowflake, AWS Services, Deployment Pipelines, Performance Tuning, Observability, Technical Documentation

---

WORK EXPERIENCE

Grid CoOperator | AI Engineer • Full-time

03/2025 - Present | Seattle, WA

- Architected scalable backend system with microservices architecture where AI agents communicate via REST APIs and message queues, deployed on AWS with containerized services, observability monitoring, and cost tracking, serving 50-100 daily queries with 99%+ uptime
- Designed and implemented backend services integrating foundational AI models (LangChain, Claude, GPT-4) into enterprise workflows, building data processing pipelines and API endpoints that reduced analyst workflows by 70% within 2 months through test-and-learn experimentation
- Built production infrastructure with CI/CD pipelines, automated deployment workflows, comprehensive logging and monitoring dashboards, accelerating feature deliverables by 60% within first quarter while maintaining system reliability and performance

Freefly Systems | Senior Software Engineer • Full-time

11/2021 - 10/2025 | Woodinville, WA

- Built full-stack application with Python Flask backend, REST APIs, and React frontend integrating foundation model APIs (Ollama, Llama 3.2) for real-time log analysis, deployed to production serving 200+ daily queries with automated monitoring and evaluation metrics
- Designed scalable backend architecture with API endpoints, data processing pipelines, and third-party AI service integration, reducing manual analysis workflows by 80% within 3 months through rapid prototyping and iterative development
- Enhanced distributed flight control systems with microservices architecture, implemented CI/CD pipelines and automated testing frameworks, improving deployment efficiency by 60% over 6 months across engineering divisions

Lumenier | Software Engineer - Embedded Systems • Full-time

07/2020 - 10/2021 | Sarasota, FL

- Implemented backend software using C++ and optimized data structures for real-time embedded systems, enabling autonomous capabilities within 8 weeks through algorithm optimization and efficient data ingestion pipelines
- Architected testing frameworks and system design patterns, reducing implementation issues by 30% within 3 months while improving operational efficiency by 45% across multiple deployment environments

York Exponential | Software Engineer - R&D • Full-time

08/2018 - 05/2020 | York, PA

- Developed backend systems for Human Machine Interface using Python and ROS2 message-passing architecture, reducing operator programming complexity by 50% within 4 months through API design and workflow automation
- Built autonomous robot prototype integrating computer vision and machine learning pipelines from requirements gathering to production deployment, demonstrating end-to-end system development capabilities

---

EDUCATION

Master of Science in Computer Science Washington State University | Pullman, WA, USA • 01/2015 - 01/2017

Bachelor Of Technology in Information Technology GITAM University | Visakhapatnam, India • 01/2011 - 01/2015

PROJECTS

GridCOP: Enterprise AI Analytics Platform | Grid CoOperator

- **Challenge:** Power grid analysts needed automated database querying and intelligent insights delivered through scalable backend infrastructure supporting complex data patterns and concurrent user requests
- **Solution:** Architected multi-agent backend system using microservices architecture where specialized AI agents coordinate via REST APIs and LangChain orchestration. Implemented data processing pipelines with RAG and vector search (FAISS), integrated multiple foundational model providers, built comprehensive API layer with authentication and rate limiting, deployed on AWS with auto-scaling and observability
- **Impact:** Reduced analyst workflows by 70% through intelligent automation. Built test-and-learn experimentation framework enabling rapid iteration on AI features. Implemented human-in-the-loop evaluation pipelines, monitoring dashboards, and comprehensive API documentation for cross-team integration

Production Log Analysis System | Freefly Systems

- **Challenge:** Manual system analysis creating bottlenecks requiring scalable backend solution to process large log volumes and deliver real-time insights through AI-powered analysis
- **Solution:** Designed Flask backend with REST API endpoints, integrated foundation model APIs (Ollama, Llama 3.2) for log processing, built data ingestion pipelines handling multiple log formats, implemented caching layer and asynchronous processing for performance optimization
- **Impact:** Transformed analysis from hours to minutes, deployed to production cloud infrastructure serving 200+ daily queries. Built monitoring and evaluation systems tracking API performance, model quality metrics, and system reliability. Documented API specifications and integration patterns for team knowledge sharing

AI Travel Planning Service | Personal

- **Solution:** Built backend service using LangChain and Claude 3.5 Sonnet API, Streamlit frontend consuming REST endpoints, integrated DuckDuckGo Search API for real-time data retrieval, demonstrating AI service integration patterns and API design through test-and-learn development