Viresh Duvvuri

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

AI Engineer with 5+ years developing and deploying machine learning and generative AI solutions that drive real-world impact at scale, specializing in designing, training, and deploying complex models for production applications. Proven expertise collaborating cross-functionally with product managers, engineers, and domain experts to translate AI concepts into practical solutions, with strong track record communicating technical concepts to non-technical stakeholders and producing thorough documentation. Deep experience with Python, ML frameworks (PyTorch/TensorFlow), cloud deployment (AWS), and ethical AI design ensuring performance, scalability, and accuracy through comprehensive evaluation frameworks. Passionate about staying current with AI advancements, integrating emerging methods, and sharing knowledge to enable teams while maintaining focus on responsible AI use that delivers measurable impact.

Skills

AI/ML Development: PyTorch, TensorFlow, ML Frameworks, Model Training & Deployment, Generative AI, LLMs (GPT-4, Claude, Ollama, Llama), Multi-Agent Systems, LangChain, RAG Pipelines, NLP, Prompt Engineering

Programming & Engineering: Python, Java, TypeScript, JavaScript, SQL, C++, FastAPI, Flask, React, NumPy, Pandas, Software Engineering, API Development, Library Development

Cloud & MLOps: AWS, Azure, Docker, Kubernetes, CI/CD Pipelines, Cloud Deployment, Model Serving, Scalable Infrastructure, DevOps Tooling, Production Deployment, Real-Time Systems

Ethical AI & Evaluation: Fairness, Accountability, Data Privacy, Model Evaluation Frameworks, Performance Monitoring, Accuracy Tracking, Human-in-the-Loop Validation, Golden Set Testing, Observability

Cross-Functional Collaboration: Technical Communication, Non-Technical Stakeholder Management, Documentation, Knowledge Sharing, Product Management Collaboration, Content Expert Collaboration

Work Experience

Grid CoOperator — Seattle, WA

AI Engineer — *03/2025 - Present*

- Developed and deployed generative AI solution integrating LLMs (GPT-4, Claude) with multi-agent orchestration for educational analytics use case, designed and trained AI models to automate complex analysis workflows for non-technical users (power grid analysts), reduced manual effort by 70% within 2 months through iterative model development and cross-functional collaboration with stakeholders to translate requirements into practical AI applications that enhanced productivity for domain experts
- Designed, trained, and deployed ML models ensuring performance, scalability, and accuracy through comprehensive evaluation frameworks tracking model metrics (accuracy, latency, reliability) across production workloads serving 50-100 daily queries, implemented cloud infrastructure on AWS using Python (FastAPI) with DevOps tooling for deployment and monitoring, built observability systems validating 40% improvement in AI accuracy through rigorous testing and human-in-the-loop validation ensuring ethical AI performance
- Collaborated cross-functionally with product teams, engineers, and domain experts to integrate AI effectively into operational workflows, communicated complex AI concepts clearly to non-technical stakeholders through demos and thorough documentation, shared knowledge across teams about generative AI applications and limitations, stayed current with AI advancements (MCP, function calling, RAG techniques) and integrated new methods into products to drive continuous innovation

Freefly Systems — Woodinville, WA

Senior Software Engineer — *11/2021 - 10/2025*

- Developed and deployed AI-powered diagnostic solution from prototype to production serving 200+ daily queries, integrated foundation model APIs (Ollama, Llama) with ML frameworks for real-time technical analysis, designed evaluation frameworks ensuring ethical AI performance including accuracy and reliability monitoring, reduced troubleshooting time by 80% through AI-powered automation while enabling non-technical engineering teams to leverage AI capabilities through natural language interfaces
- Contributed to enterprise-scale platform codebases implementing features and system optimizations, managed software integration projects from planning through production release in high-velocity environment
- Led release management for mission-critical platforms coordinating testing phases from alpha through production deployment, troubleshooting integration issues, executing comprehensive testing protocols with cross-functional teams
- Built automated systems to process complex technical data and identify system failures, developing support tools that streamlined operations in data-intensive environment

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 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

Production GenAI System for Educational Analytics | Grid CoOperator

- Problem: Organization needed AI solution to enhance learning and productivity outcomes for analysts, requiring development of ML models that translate complex data queries into actionable insights for non-technical users, ensuring ethical AI design with fairness, accountability, and data privacy while delivering measurable impact on user productivity and learning
- Solution: Designed, trained, and deployed generative AI system using PyTorch/TensorFlow patterns with LLM integration (GPT-4, Claude), implemented multi-agent orchestration with LangChain for scalable model architecture, built comprehensive evaluation frameworks ensuring model performance and ethical AI considerations (accuracy, fairness, data privacy), deployed on AWS cloud platform with DevOps tooling for production serving, created thorough documentation and knowledge-sharing materials for team enablement
- Impact: Successfully deployed production AI solution serving 50-100 daily queries with 99%+ uptime, enhanced user productivity by 70% through AI-driven automation and learning support, validated 40% improvement in analysis accuracy through rigorous evaluation frameworks and human-in-the-loop testing, demonstrated passion for advancing technology that contributes to user learning outcomes through measurable improvements in productivity and decision-making capabilities

ML Infrastructure for Technical Education | Freefly Systems

- Problem: Engineering teams needed AI solution for technical education and troubleshooting, requiring deployment of ML models that enable non-technical users to learn complex diagnostic processes through natural language interaction while ensuring ethical AI performance through comprehensive monitoring and evaluation
- Solution: Developed and deployed AI-powered solution integrating foundation model APIs (Ollama, Llama) with ML frameworks, implemented evaluation frameworks for ethical AI including performance monitoring and accuracy tracking, created natural language interface enabling educational self-service for technical teams, deployed with DevOps practices ensuring scalability and reliability
- Impact: Delivered production AI system serving 200+ daily queries, reduced learning curve from hours to minutes (80% reduction) through AI-powered educational interface, enabled self-service learning for technical teams, demonstrated commitment to advancing technology for educational outcomes

AI Agent Prototyping for Continuous Learning | Personal

- Problem: Needed continuous experimentation with emerging AI frameworks and techniques to stay current with AI advancements and validate new methods for integration into production applications, demonstrating passion for learning and advancing AI capabilities
- Solution: Built AI agent using Claude API with LangChain, implemented function calling and API integration, developed structured output patterns, applied prompt engineering techniques, created interfaces for experimentation and validation of emerging AI approaches
- Impact: Demonstrated commitment to staying current with AI advancements through hands-on experimentation, validated emerging techniques for production integration, showcased passion for advancing AI capabilities and continuous learning mindset