

Viresh Duvvuri

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

AI Engineer with 5+ years designing and building intelligent multi-agent systems that autonomously perform complex reasoning, analysis, and decision-support tasks in enterprise environments. Proven expertise architecting AI workflows combining LLMs, tools, and retrieval mechanisms (RAG) into reliable, explainable outcomes. Strong track record developing evaluation frameworks, orchestrating agent collaboration through LangChain, and translating business problems into scalable AI capabilities that delivered 70-80% efficiency improvements. Deep software fundamentals in Python, APIs, and data handling with passion for making AI truly useful in real business ecosystems.

Skills

Agentic AI & Multi-Agent Systems: LangChain, LangGraph, Multi-Agent Orchestration, Autonomous Reasoning, Agent Collaboration, Tool Integration, Function Calling, Model Context Protocol (MCP), Agent-to-Agent (A2A)

LLM Frameworks & Architectures: OpenAI APIs (GPT-4), Anthropic Claude, Ollama, Llama, Hugging Face, Prompt Engineering, Prompt Optimization, Few-Shot Learning, Context Engineering

AI Workflows & Retrieval: RAG Pipelines, Vector Databases (FAISS, Pinecone), Vector Search, Retrieval Mechanisms, Semantic Search, Structured Outputs, API-based Reasoning

Evaluation & Control: Model Evaluation Frameworks, Golden Set Testing, Synthetic Data Generation, Human-in-the-Loop, Performance Monitoring, Accuracy Tracking, Traceability, Observability

Programming & Systems: Python, JavaScript, TypeScript, SQL, C++, FastAPI, Flask, React, NumPy, Pandas, System Architecture, API Design, Data Handling

Cloud & Infrastructure: AWS, Azure, Docker, CI/CD Pipelines, REST APIs, Web Services, Deployment, Monitoring, Scalability

Work Experience

Grid CoOperator

Seattle, WA

AI Engineer

03/2025 - Present

- Designed, built, and scaled intelligent multi-agent system that autonomously performs complex reasoning and analysis tasks for enterprise analytics workflows, architecting specialized agents (SQL generation, context retrieval, web search) that collaborate through LangChain orchestration with Model Context Protocol (MCP) integration, reducing manual effort by 70% within 2 months through autonomous decision-support capabilities
- Architected AI workflows combining LLMs (GPT-4, Claude), tool integration (function calling for SQL/APIs), and retrieval mechanisms (RAG with FAISS vector databases) into reliable, explainable, and repeatable outcomes, established comprehensive evaluation and control frameworks tracking agent accuracy, performance, and traceability across 50-100 daily queries with observability dashboards and logging
- Continuously refined agent behavior and orchestration through experimentation, system-level tuning, and real-world feedback from stakeholders, collaborated closely with product teams and customers to translate business problems into scalable AI capabilities, demonstrating deep understanding of how AI can simulate reasoning and collaboration within enterprise business ecosystems

Freefly Systems

Woodinville, WA

Senior Software Engineer

11/2021 - 10/2025

- Built intelligent diagnostic agent system integrating foundation model APIs (Ollama, Llama 3.2) with RAG semantic search for autonomous reasoning about complex technical failures, developed evaluation frameworks and prompt engineering strategies for reliable outcomes, served 200+ daily queries with continuous performance optimization based on real-world system behavior
- 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

Software Engineer - R&D

York, PA

08/2018 - 05/2020

- Developed prototype software for 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

Education

Washington State University

Master of Science Computer Science

Pullman, WA

01/2015 - 01/2017

GITAM University

Bachelor of Technology Information Technology

Visakhapatnam, India

01/2011 - 01/2015

Projects

GridCOP: Intelligent Multi-Agent System for Enterprise Analytics | Grid CoOperator

- Problem: Enterprise analytics required autonomous AI system capable of complex reasoning, analysis, and decision-support across database querying, context retrieval, and insight generation, needing agents that could collaborate to deliver actionable intelligence in dynamic business environment
- Solution: Designed and built intelligent multi-agent system using LangChain orchestration where specialized autonomous agents (SQL reasoning agent, context retrieval agent, web search agent) collaborate through tool integration and function calling. Architected AI workflows combining LLMs (GPT-4, Claude) with retrieval mechanisms (RAG pipeline with FAISS vector databases) and APIs for reliable, explainable outcomes. Developed comprehensive evaluation and control frameworks including golden set testing, accuracy tracking, performance monitoring, and human-in-the-loop validation. Deployed on AWS with FastAPI, observability tools, and traceability logging
- Impact: Successfully scaled multi-agent system serving 50-100 daily queries with 99%+ uptime and 70% efficiency improvement, demonstrated autonomous reasoning capabilities through 40% accuracy improvement validated by evaluation framework, continuously refined agent behavior through experimentation and real-world feedback, showcased deep understanding of AI collaboration and autonomy within enterprise ecosystems through stakeholder engagement and iterative system-level tuning

Intelligent Diagnostic Agent System | Freely Systems

- Problem: Technical troubleshooting required autonomous AI agent capable of reasoning about complex system failures, analyzing logs, and providing actionable diagnostic insights without human intervention for every query
- Solution: Built intelligent agent system integrating foundation model APIs (Ollama, Llama 3.2) with RAG pipeline for semantic search and retrieval, developed prompt engineering strategies for autonomous reasoning, implemented evaluation frameworks for accuracy and reliability, architected full-stack application with React frontend and Python Flask backend
- Impact: Deployed production agent system serving 200+ daily queries, reduced troubleshooting time by 80% through autonomous analysis and decision-support, continuously refined agent behavior through real-world feedback and experimentation, demonstrated ability to make AI truly useful in high-stakes technical environment

AI Travel Planner Agent | Personal

- Problem: Needed rapid prototyping of intelligent agent demonstrating autonomous reasoning, tool integration, and real-time data retrieval capabilities for multi-step planning tasks
- Solution: Built AI agent using Claude 3.5 Sonnet API with LangChain orchestration, implemented function calling for tool integration with DuckDuckGo Search API enabling autonomous information retrieval and reasoning, developed prompt engineering techniques for multi-step task execution
- Impact: Demonstrated end-to-end intelligent agent development showcasing autonomous reasoning, tool orchestration, and adaptability to emerging AI technologies, validated agentic architecture patterns through hands-on experimentation