

Mounika Yalamarty

Phone #: 4255204900 | E-mail: mounika.yrs@gmail.com

LinkedIn: <https://www.linkedin.com/in/sumithra-mounika-yalamarty/> | Github: <https://github.com/monyal25>

PROFESSIONAL EXPERIENCE

Software Engineer AI/ML- Microsoft, Remote

Feb 2025 - Present

- Designed and contributed to Microsoft's LLM platform (Project Sophia) using Python, LangChain, OpenAI Evals, and Azure ML, building deployment-ready pipelines with evaluation, monitoring, and regression testing for production reliability.
- Defined evaluation metrics (accuracy, coherence, helpfulness) with AI/ML teams and automated prompt regression testing pipelines, ensuring version stability and high-quality outputs.
- Applied MLOps best practices including model versioning, CI/CD (GitHub Actions), containerization (Docker/Kubernetes), and monitoring (Prometheus/Grafana) to deliver scalable, production-grade AI systems.
- Refined prompt strategies by simulating diverse inputs, analyzing LLM performance, and optimizing outputs for enterprise staffing automation use cases.

Research Assistant - Santa Clara University, California

Dec 2024 - Feb 2025

- Developed multi-agent models for multicultural text-to-image generation using AltDiffusion, leveraging a novel 9,000-image dataset spanning five countries, three age groups, two genders, 25 historical landmarks, and five languages to enhance cultural diversity and representation. <https://arxiv.org/pdf/2502.15972>

Senior Software Engineer - PropertyGuru, Singapore

July 2021 - March 2023

- Built scalable micro-frontends and FastAPI-based microservices in Python/MERN Stack as part of the (Hive UI) initiative, modernizing legacy systems for reusability and high performance.
- Built and optimized CI/CD pipelines with Jenkins, Docker, and Kubernetes, enabling rapid, reliable deployment of AI/ML services
- Developed real-time streaming APIs for property listing and recommendation workflows, supporting 8K+ concurrent users with low-latency responses.
- Delivered 35% faster response times and 40% higher lead conversions while collaborating with leadership in a startup-like environment to shape system design and scalability.

Software Engineer 2 - Illumina, Singapore

June 2020 - June 2021

- Worked as a full-stack developer using C# and .NET to build secure microservices with OAuth 2.0 via Keycloak and K3s, enhancing authentication and system scalability.
- Refactored the LRM Framework and implemented Cypress tests, reducing manual QA by 60% and accelerating release cycles for genomic applications.

Software Engineer Intern - Land Transport Authority, Singapore

January 2020 - May 2020

- Led a project analyzing travel patterns of Singapore's 6 million residents using GPS and multi-sensory data. Developed AI models to predict transportation modes, built a real-time feedback module, and optimized AWS architecture with Lambda and CI/CD, reducing manual analysis time by 20% and enabling real-time transportation insights.

Software Engineer - Arcesium(D.E.Shaw Group), Hyderabad

June 2016 - June 2019

- Developed APIs to streamline complex post-trade activities and designed new features based on client requirements.
- Facilitated smooth client onboarding and gained expertise in financial domains like swaps, derivatives, and rates..

EDUCATION

Master's in Computer Science	Santa Clara University	GPA: 3.8
Master's of Technology in Software Engineering	National University of Singapore	GPA: 3.9
Bachelor's of Engineering, Computer Science	Osmania University	GPA: 4

SKILLS/COURSES

Programming Languages: Python, Java, C++, C, PHP, JavaScript, C#, React, TypeScript, HTML/CSS.

Databases: MongoDB, PostgreSQL, MySQL, Redis, DynamoDB

AI/ML: PyTorch, TensorFlow, Transformers, LangChain, OpenAI Evals, MLflow, Multi-Agent Systems, Prompt Engineering

Cloud & DevOps: AWS, Azure ML, GCP, Docker, Kubernetes, Terraform, Helm, Jenkins, GitHub Actions, Kafka, Kinesis, Prometheus/Grafana