

Rohit Marathe

📍 Atlanta, GA | ✉️ rohitmarathe1995@gmail.com | 🔗 [linkedin.com/in/rohit-marathe-ucf](https://www.linkedin.com/in/rohit-marathe-ucf) | 🌐 rpm360.github.io

Professional Summary

Software Engineer specializing in Agent Orchestration and GenAI Evaluation Pipelines. Proven track record of scaling AI agents for high-traffic Fintech environments (Flash Sales) and architecting High-Availability Multi-LLM Systems.

Technical Skills

AI Agents & Orchestration: Multi-Agent Systems (Bedrock), Function Calling, MCP Server.

LLM Ops & Evals: LLM-as-a-Judge, Langfuse, Dynatrace, RAG, Prompt Engineering, Fine-tuning.

Models & Data: GPT-4, Claude 3 (Opus/Haiku), AWS Bedrock, Llama 3, DynamoDB.

Core Engineering: Python, Node.js, TypeScript, AWS, Docker, Kubernetes, Terraform, CI/CD, Claude Code, Git Copilot.

Experience

Software Engineer I (AI & Fintech)

Rocket Mortgage

Jan 2024 – Present

Atlanta, GA

- Engineered a Multi-Agent Orchestration framework using AWS Bedrock, intelligently routing complex workflows and FAQ queries to specialized sub-agents, improving resolution accuracy by 20%.
- Built an automated LLM-as-a-Judge evaluation pipeline and integrated Langfuse for granular observability, ensuring zero regression during high-traffic Flash Sales events.
- Developed an AI-driven product to evaluate banker performance by analyzing banker-client interactions, deriving actionable quality metrics that directly influenced coaching strategies.
- Developed an autonomous Incident Resolving Agent utilizing RAG pipelines, Function Calling, and MCP Server to diagnose system alerts and execute remediation scripts without human intervention.
- Architected a Multi-LLM Failover System with dynamic switching (e.g., OpenAI, Anthropic), ensuring 99.9% uptime under peak concurrency through rigorous load testing and rate-limiting guardrails.

Software Engineer (Promoted from Technical Intern)

Kore.ai

June 2020 – Oct 2023

Orlando, FL

- Architected and deployed 20+ conversational bots using Kore.ai SDKs, leveraging Few-Shot Learning to boost NLP intent recognition accuracy by 11% over baseline models.
- Engineered an Unstructured-to-Structured Data Pipeline (GraphRAG precursor) to parse complex PDFs for Knowledge Graphs, reducing node generation time by 2X.
- Achieved a 95% F-1 score for an insurance bot handling 8,000+ daily requests by implementing advanced entity extraction and sentiment analysis engines.
- Designed "Agent Transfer" logic to gracefully handle out-of-domain queries, significantly improving Net Promoter Score (NPS) and customer retention.

Software Engineer

TIBCO Software Inc.

July 2017 – Jan 2019

Pune, India

- Enhanced legacy SQL query performance and implemented a semantic caching system, resulting in 3X faster API response times for data-heavy endpoints.
- Built fault-tolerant web applications using Java and JMS, ensuring high availability for enterprise-grade financial systems.

Projects

KG Generator Tool V2

- Engineered a scalable Knowledge Graph utility handling millions of entities to streamline FAQ integration and structural rebuilds.

Personal Tweet Classification (NLP Deep Learning)

- Implemented a Deep Learning NLP model to classify personal vs. non-personal tweets with high accuracy.

Education

University of Central Florida

Master of Science in Computer Science (Minor in AI)

Aug 2019 – May 2021

Orlando, FL

Maharashtra Institute of Technology

Bachelor of Engineering in Information Technology

Aug 2013 – June 2017

Pune, India