

Rishi RajMullur

📍 San Francisco Bay Area, CA

☎ +1 (206) 731-9171

✉ rishi.rmullur@gmail.com

🌐 linkedin.com/in/rishimullur

Summary

Software Engineer with 4+ years of experience delivering robust, scalable systems across AI, productivity, and data infrastructure domains. Former founder and AI/ML engineer. Proven ability to ship high-impact features. Experienced in designing AI-enabled systems under real-world constraints, with cross-functional collaboration across product, engineering, and research teams.

Experience

Software Engineer

Kanary Inc

May 2025 – June 2025

Bellevue, WA

- Engineered intelligent matching pipelines for high-volume data systems, cutting false positives by 50%.
- Built LLM-powered browser automation tools for web scraping and data extraction, improving data collection efficiency.
- Led algorithmic improvements using heuristics and statistical tuning, increasing engagement and accuracy.

Software Engineer Intern - AI Systems

2ndChair AI

June 2024 – December 2024

Seattle, WA

- Designed and scaled modular RAG pipelines using Python, FastAPI, and Hugging Face for legal document retrieval.
- Implemented performance-optimized parsing pipelines with SQL and async processing, improving throughput by 45%.
- Built agent-based task orchestration with memory and reflexion modules, reducing redundant computation.
- Created LLM evaluation and debugging tools for token-level latency and accuracy tracking in production.

Lead Software Engineer/CTO

Scryorb Technologies

Jan. 2021 – Mar. 2022

Bangalore, India

- Architected distributed backend systems with Go and GCP to support real-time short-video ingestion and playback.
- Built analytics pipelines for creator activity and engagement using PostgreSQL and Firebase event streams.
- Implemented JWT-based auth gateway; enforced least-privilege IAM on GCP.

Software Security Researcher

Freelance

Mar. 2022 – Dec. 2022

Bangalore, India

- Ranked #1 globally in Twitter Bug Bounty 2022; discovered critical security flaws in production apps used by billions.

Associate Product Manager (Technical)

ClearFeed Inc.

Jan. 2023 – Aug 2023

Bangalore, India

- Drove development and launch of GPT-3.5-powered support automation, reducing response time by 50% and boosting ARR by 55% through collaboration with engineering and analytics teams.

Project Engineer

Wipro Technologies

June 2019 – December 2020

Bangalore, India

- Engineered and optimized 5G core network protocols (PCF, NGAP, NAS) in C++ for telecom, ensuring low-latency and high-throughput performance.
- Collaborated in a multi-team agile environment to design modular, reusable components aligned with 3GPP specs.
- Developed test harnesses in Python for simulating mobile core scenarios and validating protocol behavior under edge-case loads.
- Documented and presented technical design specifications to internal review boards and contributed to cross-site engineering knowledge sharing.

Education

University of Washington, Global Innovation Exchange

Master of Science in Computer Science [Technology Innovation]

(Courses: Deep Learning, Programming for Digital Interfaces, Planning & Managing Hardware/Software Dev)

Seattle, WA

Sep. 2023 – March 2025

National Institute of Technology, Warangal

Bachelor of Technology in Mechanical Engineering

Warangal, India

Aug. 2015 – May 2019

Projects

ClaudePlaysPokémon: Human-AI Gameplay System | Creator & Presenter

2025

- Enabled human coaching through live text input, demonstrating collaborative decision-making between LLM and human user.
- Presented AI agent which plays Pokemon Red demo and architecture overview to 150+ attendees at exclusive AI event, showcasing agentic control, strategy abstraction, and emergent behavior.
- Featured as main stage talk at AI Tinkerers Seattle, highlighting innovative human-AI collaboration techniques.

Decoding Prompting for LLMs (Research) | Poster, CSE599 Advanced CS Topics

2024

- Benchmarked GPT-4, Claude, GPT-3.5 on adversarial tasks (30+ examples) to evaluate prompt strategies.
- Published best practices based on user study across reasoning prompt types.

Skills

Languages: Python, C++, Go, JavaScript, SQL, R

AI/ML: PyTorch, Transformers (HF), OpenAI/Anthropic APIs, RAG, Prompt Engineering, LLM Evaluation

Systems: REST APIs, Data Pipelines, Docker, CI/CD

Cloud: GCP (Vertex AI), AWS (S3, EC2), Firebase

Infra: Load Balancing, Caching, API Rate Limiting, WebSockets

Testing: Pytest, Unit Testing, Integration Testing

Databases: PostgreSQL, MongoDB, Redis

Messaging: Kafka, Pub/Sub, Celery

Monitoring: Grafana, Sentry