Rishi Raj Mullur

📞 +1 (206) 731-9171 ✉ rishirm@uw.edu in linkedin.com/in/rishimullur

Education

University of Washington, Global Innovation Exchange — Seattle, WA
Master of Science in Computer Science [Technology Innovation] — *Sep. 2023 – March 2025*
(Courses: Deep Learning, Programming for Digital Interfaces, Planning & Managing Hardware/Software Dev)

National Institute of Technology, Warangal — Warangal, India
Bachelor of Technology in Mechanical Engineering — *Aug. 2015 – May 2019*

Experience

AI/ML Software Engineering Intern — June. 2024 – Current
2ndChair AI — *Seattle, WA*
- Engineered and optimized core AI systems for Retrieval Augmented Generation (RAG), integrating Large Language Models (LLMs) such as GPT-4 and Claude using PyTorch and Hugging Face, improving law firms' information retrieval efficiency.
- Designed and implemented custom algorithms/methods for AI hallucination mitigation by leveraging new parsing architectures and prompt engineering, achieving a 60% improvement in the precision of legal summaries.
- Engineered data and document parsing pipelines with Python and SQL, improving contextual data extraction for downstream generation by over 45%
- Leveraged advanced techniques such as reflexion, chain-of-thought, planning and memory chaining in agentic systems to enable pseudo-autonomous task execution with tools for legal use cases.

Summer Research Collaborator — July. 2024 – Current
Allen Institute for Artificial Intelligence — *Seattle, WA*
- Architecting a cutting-edge data acquisition system for forestry research, integrating LiDAR, IMU, and high-resolution imaging with a Raspberry Pi 5 platform.
- Developing sophisticated Python software to fuse multi-sensor data streams, implementing advanced algorithms for real-time processing, storage, and visualization in challenging forest environments.
- Laying groundwork for ML-enabled wildfire risk estimation by ensuring high-quality, consistent data, suitable for CV model training.

Associate Product Manager — Jan. 2023 – Aug 2023
ClearFeed Inc. (Sequoia-backed SaaS Startup) — *Bangalore, India*
- Led the successful launch of LLM-powered support and Q&A features, integrating GPT-3.5 based models to enhance user experience and automate customer support, resulting in a 50% reduction in response time.
- Collaborated with ML engineers to implement AI-driven features, resulting in a 70% user growth and 55%+ increase in ARR through improved user engagement and retention.
- Utilized data analytics tools to process and analyze over 1 million user interactions, informing product decisions and measuring the impact of AI-powered features on user satisfaction.

Security Researcher — March. 2022 – December 2022
Freelance — *Bangalore, India*
- **No.1 Worldwide ranked** independent security researcher in 2022 for the Twitter Bug Bounty Program.
- Conducted security assessments of popular AI/LLM products, identifying potential vulnerabilities and privacy concerns.

Software Project Engineer — June 2019 – December 2020
Wipro Technologies — *Bangalore, India*
- Engineered and implemented high-performance, scalable 5G core network software protocols (PFCP, NGAP, GTPU, NAS) using C++ and Python, adhering to 3GPP standards and optimizing for low latency and high throughput in telecommunication systems.

Projects

Tale:AI Speech Device | *Python, ReactJS* — Accepted to Dempsey Startup Competition Investment Round
- Developed a contextual speech-completion system powered by LLMs, reducing aphasia patients' communication time by over 65% through real-time language processing.

Decoding Prompting for LLMs (Research) — Poster Presentation for CSE599 - Advanced Topics in Computer Science
- Conducted research on the limitations of large language models (LLMs) like GPT-4, Claude 3 Opus, and GPT-3.5 Turbo in solving problems requiring domain-specific knowledge, advanced reasoning, and critical thinking skills.
- Curated an adversarial dataset containing 30 multi-domain problems and conducted a user study to investigate the effectiveness of prompting strategies in enabling LLMs to solve these problems.

Skills

AI/ML: PyTorch, Hugging Face Transformers, OpenAI API, Anthropic API, LLM fine-tuning, Transfer Learning, RAG, Transformer Models, LLM Fine-Tuning, Prompt Engineering, Hyperparameter Tuning
Data: Pandas, NumPy, SQL (Postgres), Matplotlib
Cloud: GCP, AWS, Firebase

Languages: Python, C++, Go, JavaScript, SQL, R, Swift
DevOps/Cloud: Docker, AWS (EC2, S3), GCP (Vertex AI, Cloud Functions), Firebase, CI/CD Pipelines
Tools: Git, Docker, Jupyter, VS Code, Android Studio, Jira, Figma
Product: A/B testing, Google Analytics, Mixpanel, Product Roadmapping, User Research Methods