Aryasomayajula Ram Bharadwaj

ram.bharadwaj.arya@gmail.com | +91-9108832338 | Bengaluru, India | github.com/rokosbasilisk

PROFESSIONAL SUMMARY

Senior ML Engineer with 6+ years developing production AI systems and leading engineering teams. Background in AI safety research with published work on model interpretability and LLM inference optimization. Experienced architecting scalable solutions for conversational AI and agentic workflows. Strong expertise in MLOps, distributed systems, and technical leadership.

EDUCATION

Bachelor of Technology - Electronics and Communications Engineering 2015 - 2019
GMR Institute of Technology, Andhra Pradesh

SKILLS

Programming: Python, Scala, Java, R, SQL
ML/AI: PyTorch, JAX, Transformers, LangChain, AutoGen
DevOps: Docker, Kubernetes, CI/CD, GCP
Systems: Redis, Kafka, Microservices

PROFESSIONAL EXPERIENCE

Associate Technical Architect - MLOps Nov 2024 - Present
Quantiphi Analytics, Bengaluru

- Designed and implemented an AI agent system for automated replies for a telecom client's sales chatbot
- Experimented with LangGraph agent architectures to improve controllability and modularity in conversational flows
- Revamped and automated the scheduled RAG data ingestion pipeline and improved retrieval accuracy by 20%
- Improved retriever latency from 6 seconds to 1.4 seconds through extensive code refactoring

AI Resident - Lossfunk AI Residency April 2025 - May 2025
Lossfunk AI Residency (Remote)

- Selected among 100 applicants for elite 6-week intensive AI residency program with 10 researchers
- Developed STU-PID (Steering Token Usage with PID Control), a novel activation steering technique to reduce redundant reasoning tokens in LLMs
- Conducted experiments on reasoning models like DeepSeek-R1-Distill-Qwen-1.5B with dynamic activation interventions, achieving 32% token reduction and improved reasoning accuracy on GSM8K benchmark
- Published research findings: Steering Token Usage with PID Control

Senior Developer - Innovation & Development Labs June 2019 - Nov 2024
Musigma Business Solutions, Bengaluru

- Led development of multiple high-impact projects in LLM operationalization, automated trading, and MLOps
- Specialized in backend development, DevOps, and ML engineering across various domains
- Received multiple recognitions including Impact Awards and Star Performer of the Team

KEY TECHNICAL PROJECTS

LLM Agent Platform Dec 2023 - Nov 2024

Team Lead, ML Engineer, Backend Developer

- Designed and implemented semi-autonomous data analysis platform using Microsoft AutoGen framework
- Developed automated prompt optimization strategies and integrated RAG support
- Built evaluation framework using ensemble of locally hosted LLMs
- Successfully deployed for two major clients, assisted client teams in creating use cases

High-Velocity Trading Platform 2021 - Dec 2023

Team Lead, Backend Developer, DevOps

- Refactored backend code and migrated deployment from bare-metal to Kubernetes
- Implemented automated hyperparameter search for ARIMA models
- Rewrote legacy trade-signal generation from R to Scala using Akka framework
- Enabled near real-time metric calculation and portfolio visualization

ML Model Operationalization Platform 2019 - 2021

Backend Developer

- Developed automatic retraining pipelines for image classification models
- Designed Java microservices for creating and serving Jupyter notebooks
- Engineered ML model deployment service with canary and blue-green deployment strategies

RESEARCH & PUBLICATIONS

Scaling Laws for LLM-Based Data Compression July 2025

Lead Author

Investigated scaling laws for data-compression capabilities of LLMs on text, image, and speech modalities.

Steering Token Usage with PID Control June 2025

Lead Author

Novel technique reducing computational overhead in LLMs through activation steering with 32% token reduction on GSM8K.

Understanding Hidden Computations in Transformer Language Models August 2024

Lead Author

Investigated internal mechanisms of chain-of-thought reasoning and developed interpretability methods for LLM reasoning.

SELECTED PERSONAL PROJECTS

Wiserank.io - AI-Powered Research Discovery June 2025

Creator & Full-Stack Developer

- Built research paper search engine with relevance and 'creativity' ranking algorithms
- Implemented automatic citation generation feature for uploaded manuscripts.
- Deployed scalable architecture serving researchers with enhanced paper discovery and citation workflows

AWARDS & RECOGNITIONS

AI Alignment Awards - Winner July 2023

AI Safety Research Competition

Selected among 118 global entries for winning research proposal on "goal misgeneralization" in AI systems.

Honorable Mention - Eliciting Latent Knowledge March 2022

Alignment Research Center

Recognized for innovative approach to open research problem in AI interpretability and knowledge extraction.

Bronze Medal - Build-on-Redis Hackathon February 2021

Redis Labs

Developed text-to-code search tool using CodeBERT embeddings and Redis Stack for private repository indexing.

Excellence Awards (8x) 2019 - 2023

Musigma Business Solutions

6 SPOT (Star Performer) awards and 2 Impact Awards for technical leadership, innovation, and delivery excellence.