

Aryasomayajula Ram Bharadwaj

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

PROFILE

Research-focused AI Engineer working at the intersection of large-scale systems, interpretability, and AI safety. Six years of experience in ML infrastructure, agentic workflow design, and large-model analysis. Published multiple papers on representation scaling, activation steering, and reasoning efficiency. Skilled at bridging deep research with deployable production systems.

CORE COMPETENCIES

- **Research:** Mechanistic interpretability, scaling laws, activation steering, representation alignment, model awareness
- **Frameworks:** PyTorch, JAX, Transformers, LangGraph, AutoGen
- **Systems:** Docker, Kubernetes, Redis, Kafka, PostgreSQL, Microservices
- **Cloud:** AWS, GCP
- **Languages:** Python, Scala, Java

RESEARCH & PUBLICATIONS

- | | |
|--|-----------------|
| Scaling Laws for LLM-Based Data Compression | Jul 2025 |
| <i>Lead Investigator</i> | |
| • Derived universal power-law scaling between model size and compression efficiency across text, image, and speech domains. | |
| Steering Token Usage with PID Control | Jun 2025 |
| <i>Lead Investigator</i> | |
| • Proposed a redundancy-aware activation steering method that reduced reasoning token usage by 32% while improving GSM8K accuracy. | |
| • Released open-source implementation STU-PID. | |
| Understanding Hidden Computations in Transformer Language Models | Aug 2024 |
| <i>Lead Investigator</i> | |
| • Analyzed an open problem underlying interpretability of chain-of-thought reasoning. | |
| • Designed probing techniques revealing latent computation patterns in transformer intermediate layers. | |

AWARDS & RECOGNITION

- | | |
|---|------------------|
| Winner – AI Alignment Awards | Jul 2023 |
| Selected among 118 global entries for work on goal misgeneralization and alignment failure modes. | |
| Honorable Mention – Eliciting Latent Knowledge (ARC) | Mar 2022 |
| Recognized for novel approach to eliciting implicit model knowledge. | |
| Bronze Medal – Build-on-Redis Hackathon | Feb 2021 |
| Developed private code search engine using CodeBERT embeddings and Redis Stack. | |
| Excellence Awards (8×) – MuSigma Business Solutions | 2019–2023 |

PROFESSIONAL EXPERIENCE

- | | |
|--|---------------------------|
| Associate Technical Architect – Platform | Nov 2024 – Present |
| <i>Quantiphi Analytics, Bengaluru</i> | |
| • Architected AI agent infrastructure for automated issue-severity classification and escalation prediction, integrating RAG and reasoning-evaluation pipelines. | |
| • Built conversational sales-intelligence chatbot for a major telecom client; refactored proprietary framework to LangGraph, cutting code size 40% and latency 3×. | |
| • Designed modular data-ingestion layer for scheduled RAG pipelines, boosting retrieval precision 20%. | |
| AI Resident – Lossfunk AI Residency | May 2025 |
| <i>Lossfunk Research Residency</i> | |

- One of 10 selected from 100+ applicants for six-week research residency in interpretability.
- Developed redundancy-aware steering methods for reducing chain-of-thought length in reasoning models.

ML Engineer – Innovation & Development Labs

Jun 2019 – Nov 2024

MuSigma Business Solutions, Bengaluru

- Led multi-team initiatives in LLM operationalization, automated trading, and scalable MLOps.
- Migrated high-velocity trading systems from bare-metal to Kubernetes; enabled automatic hyperparameter search for ARIMA models.
- Re-engineered trade-signal generator from R to Scala (Akka), enabling near-real-time metrics and adaptive portfolio retraining.
- Built microservice-based ML deployment framework with blue-green and canary rollout strategies.

KEY PROJECTS

LLM Agent Platform (LangGraph)

2023–2024

- Designed semi-autonomous data-analysis platform with prompt optimization feature.
- Added support for building and orchestrating AI agents in with minimal coding.

High-Velocity Trading Platform

2021–2023

- Implemented asynchronous signal pipelines, model serving, and portfolio visualization dashboards.

ML Model Operationalization Framework

2019–2021

- Developed automatic retraining workflows for CV models and multi-environment deployment pipelines.

INDEPENDENT PROJECTS

Wiserank.io – AI-Powered Research Discovery

Jun 2025

Creator & Full-Stack Developer

- Built research-paper search engine ranking works by originality and citation impact; integrated automatic reference generation and semantic vector search.
- Achieved 100+ active research users in first few months.

EDUCATION

Bachelor of Technology – Electronics and Communications Engineering

2015–2019

GMR Institute of Technology, Andhra Pradesh