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 agentic workflows, interpretability, and AI safety. Six years of experience in ML infrastructure, workflow automation, and large-model analysis. Published 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**

Lead Investigator

- 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**

Lead Investigator

- 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**

Lead Investigator

- Analyzed an open problem underlying interpretability of chain-of-thought reasoning.
- Designed probing techniques revealing latent computation patterns in transformer intermediate layers.

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.

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**

Quantiphi Analytics, Bengaluru

- Designed AI agent system for automated issue-severity classification and escalation management.
- Architected and implemented a conversational AI-agent chatbot for answering sales-related queries at a major telecom company, refactoring legacy systems to a leaner modular implementation.
- Revamped the scheduled RAG ingestion pipeline to improve retrieval accuracy and reduce time-to-first-token by 3×.

AI Resident – Lossfunk AI Residency **May 2025**

Lossfunk Research Residency

- Selected among 100 applicants for elite 6-week intensive AI residency program with 10 researchers.
- Developed redundancy-aware steering techniques (STU-PID) achieving 32% token reduction and improved reasoning accuracy on GSM8K.

ML Engineer – Innovation & Development Labs **Jun 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 multiple domains.
- Refactored backend code and migrated trading deployments from bare-metal to Kubernetes with automated hyperparameter search for ARIMA models.
- Rewrote legacy trade-signal generation system from R to Scala (Akka framework), enabling near-real-time metric computation and visualization.
- Designed ML model deployment microservices with canary and blue-green deployment strategies.

KEY PROJECTS

Conversational Sales Chatbot using AI Agents **2024 – Present**

- Designed and implemented an AI-agent-based system for automated replies for a telecom client's sales chatbot.
- Refactored codebase from a proprietary workflow engine to LangGraph, reducing code size and maintenance complexity.

LLM Agent Platform **2023–2024**

- Designed a semi-autonomous data-analysis platform using AutoGen.
- Integrated automated prompt optimization strategies and RAG evaluation support with ensembles of locally hosted LLMs.

High-Velocity Trading Platform **2021–2023**

- Refactored backend and enabled migration from legacy R-based signal engine to Scala using Akka.
- Added automated retraining and portfolio metrics modules with visualization dashboards.

ML Model Operationalization Platform **2019–2021**

- Built microservice-based pipelines for retraining and deploying ML models with CI/CD integration.
- Automated image-classification retraining workflows with configurable deployment templates.

EDUCATION

Bachelor of Technology – Electronics and Communications Engineering **2015–2019**

GMR Institute of Technology, Andhra Pradesh