Aryasomayajula Ram Bharadwaj

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

PROFILE

AI Engineer with six years of experience designing scalable AI systems and agentic workflows that bridge deep learning research and production deployment. Skilled in building LLM evaluation pipelines, developing interpretability experiments, and engineering infrastructure for reliable model reasoning and analysis. Published research on model compression, reasoning efficiency. Known for combining rigorous experimentation with strong systems design to advance trustworthy AI.

CORE COMPETENCIES

- **Research:** LLM evaluation, interpretability
- **Frameworks:** PyTorch, JAX, Transformers, LangGraph, AutoGen
- **Systems:** Docker, Kubernetes, Redis, Kafka, PostgreSQL, Microservices
- **Cloud:** AWS, GCP
- **Languages:** Python, Scala, Java

PROFESSIONAL EXPERIENCE

Associate Technical Architect **Nov 2024 – Present**

Quantiphi Analytics, Bengaluru

- Designed AI agent system for automated issue-severity classification and escalation management.
- Architected and implemented a conversational AI chatbot for a telecom client using LangGraph, improving modularity and reducing response latency by 3×.
- Revamped the RAG ingestion pipeline to improve retrieval accuracy by 20% and ensure consistent evaluation performance.

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

Lossfunk Research Residency

- Selected among 100+ applicants for a 6-week research program in interpretability and evaluation.
- Developed redundancy-aware steering techniques (STU-PID) achieving 32% token reduction and higher reasoning accuracy on GSM8K.

ML Engineer – Innovation & Development Labs **Jun 2019 – Nov 2024**

MuSigma Business Solutions, Bengaluru

- Led development of high-impact systems in LLM operationalization, automated trading, and scalable MLOps.
- Migrated trading infrastructure from bare-metal to Kubernetes with automated hyperparameter search for ARIMA models.
- Rebuilt trade-signal generation from R to Scala (Akka), enabling real-time analytics and adaptive portfolio retraining.
- Designed ML model deployment microservices with CI/CD, blue-green, and canary rollout strategies.

HIGHLIGHTED PROJECT

Wiserank.io – AI-Powered Research Discovery **Jun 2025 – Present**

Creator & Full-Stack Developer

- Built an AI-powered research discovery engine ranking papers by originality and information density.
- Designed full-stack architecture with semantic search, citation generation.
- Reached 100+ active research users within initial launch period.

KEY PROJECTS

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

- Built agentic chatbot handling telecom sales queries using LangGraph.
- Refactored proprietary workflow engine to modular pipelines, cutting code size by 40%.

LLM Agent Platform **2023–2024**

- Designed semi-autonomous data-analysis platform with automated prompt optimization and evaluation metrics.
- Integrated RAG-based reasoning and ensemble evaluation for local LLM deployments.

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

- Migrated backend architecture and integrated near real-time analytics using Akka and Scala.
- Added retraining and visualization modules for portfolio metrics and latency diagnostics.

ML Model Operationalization Platform **2019–2021**

- Automated retraining workflows for image models with multi-environment deployment via CI/CD.
- Developed APIs for model serving and notebook-based microservice integration.

RESEARCH & PUBLICATIONS

Scaling Laws for LLM-Based Data Compression **Jul 2025**

Lead Investigator — Discovered scaling laws linking language model's capacity to compression efficiency across modalities.

Steering Token Usage with PID Control **Jun 2025**

Lead Investigator — Introduced redundancy-aware steering method that improved reasoning efficiency and reduced token usage.

Understanding Hidden Computations in Transformer Language Models **Aug 2024**

Lead Investigator — Explored how transformer internals encode multi-step reasoning in chain-of-thought tasks.

AWARDS & RECOGNITION

Winner – AI Alignment Awards **Jul 2023**

Recognized among 118 global entries for work on goal misgeneralization and AI safety evaluation.

Honorable Mention – Eliciting Latent Knowledge (ARC) **Mar 2022**

Acknowledged for novel solutions in latent knowledge elicitation.

Bronze Medal – Build-on-Redis Hackathon **Feb 2021**

Developed private code search using CodeBERT embeddings and Redis Stack.

Excellence Awards (8×) – MuSigma Business Solutions **2019–2024**

EDUCATION

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

GMR Institute of Technology, Andhra Pradesh