ARYAN D HARITSA

AI / ML Researcher & Engineer · B.Tech CSE Final Year · Bengaluru, India
+91 7483723182 · aryanharitsa@gmail.com · github.com/Aryanharitsa · linkedin.com/in/aryan-haritsa-60292925b
Research interests: mechanistic interpretability · alternative encoder architectures for dense retrieval · mesa-optimization in LLMs

EDUCATION

PES University — Bengaluru, India *2022 – 2026*
B.Tech, Computer Science & Engineering · CGPA 8.00 / 10
Pre-University: Sadhana PU (94.4%) · SSLC: SRKVS (99.4%, **5th rank in state**)

RESEARCH & PROJECTS

Mesa-Optimization Probe *github.com/Aryanharitsa/My-AI-Journey/mesa-probe* *Apr 2026 – Present*
Mechanistic interpretability research · PyTorch, TransformerLens, einops

- Investigating whether small transformers trained on in-context linear regression implement gradient-descent-like circuits internally; reproducing the von Oswald et al. (2023) baseline as the starting point.
- Applying activation patching, attention-pattern analysis, and linear probes to identify whether intermediate representations encode OLS, ridge, or GD-iterate solutions; targeting a 4–6 page technical writeup.

Encoder Archaeology *github.com/Aryanharitsa/My-AI-Journey/encoder-archaeology* *Apr 2026 – Present*
Dense retrieval architecture comparison · BEIR, FAISS, sentence-transformers, Mamba

- Benchmarking transformer (BERT-base, MiniLM, GTE-small), state-space (Mamba Retriever), and recurrent / convolutional encoders on NFCorpus, SciFact, and FiQA — mapping the latency-vs-nDCG@10 Pareto frontier under realistic batch sizes.
- Running attention-head pruning and token-position-shuffle experiments to isolate where transformer attention provides genuine retrieval value versus wasted compute; producing an opinionated technical note.

Project Ramanujan — AIMO Prize 3 Solo Submission *github.com/Aryanharitsa/My-AI-Journey/project-ramanujan* *Aug – Sep 2025*
Solo Kaggle submission · GPT-OSS-120B, vLLM, Harmony reasoning format

- Scored **36 / 50** in a 3-week sprint against international teams in the AI Mathematical Olympiad Prize 3 competition on Kaggle.
- Built the full inference stack from scratch: GPT-OSS-120B served via vLLM with Harmony reasoning, N=8 majority voting, and tool-integrated reasoning (TIR) for adaptive compute allocation across problem difficulty.

TITAN — Trustless Identity & Transaction Authentication Network *2025 – 2026*
Capstone project · Paper submitted to SeCrypt 2026

- End-to-end KYC / AML platform combining AI-powered document forensics, Zero-Knowledge Proofs for private on-chain identity verification, and a GNN-based AML engine for real-time fraud detection with off-chain scalable storage.

EXPERIENCE

Nagravision Kudelski (Kudelski IoT) Bengaluru, India *Sep 2025 – Present*
AI / ML Engineering Intern · Project Rick — Recover internal chatbot

- Building an agentic internal chatbot for **Recover**, Kudelski IoT's flagship asset-tracking product. Deployed as a Microsoft Teams bot serving the engineering and operations teams.
- Static knowledge layer: AWS Bedrock + Knowledge Bases indexed over Recover's architecture and product specs. Dynamic layer: live S3 → Glue → Athena pipeline for real-time asset-telemetry queries. Stack: Bedrock, Knowledge Bases, S3, Athena, Glue, ECS, ECR, Terraform, FastAPI.

August AI Remote, India *Jun – Aug 2025*
AI / Backend Intern

- Contributed to production deployment of **Tata 1mg's** conversational healthcare chatbot *'grace'*, and built appointment-booking agentic workflows for Manipal Group clinical clients.
- Designed API schemas and adapter architecture for multi-provider LLM integration; built an internal LLM Playground for company-specific model testing and fine-tuning.

TECHNICAL SKILLS

RESEARCH	PyTorch · TransformerLens · vLLM · HuggingFace · FAISS · BEIR · einops
LANGUAGES	Python · C++ · TypeScript / JavaScript · SQL
CLOUD & INFRA	AWS (Bedrock, Knowledge Bases, S3, Athena, Glue, ECS, ECR) · Terraform · Docker · CI / CD
BACKEND	FastAPI · Node.js · PostgreSQL · React
AREAS	Mechanistic interpretability · Mesa-optimization · Dense retrieval · RAG · Mathematical reasoning
OPEN SOURCE	AML & Kannada–English datasets (Kaggle / HF) · Tokenization + finetuning PRs to LLaMA-3.2B, Mistral
CERTIFICATIONS	Advanced Math for ML · IBM Machine Learning · IBM Deep Learning · Generative AI · Advanced DSA