

ADITHYA GIRIDHARAN

Learning From First Principles

AI Engineer | Mechanistic Interpretability

giridharanadithya@gmail.com | +91 89830 98776 | +91 93451 21923 | agilabs.co.in

linkedin.com/in/adithya-giridharan | github.com/IamAGP | github.com/ADITHYAG73 | medium.com/@AdithyaGiridharan

I am a MATHEMATICIAN by heart. I build production-grade GenAI systems, with expertise in AI agents and multi-agent workflows, and I work on mechanistic interpretability. Proficient in Python, LangGraph, Neo4j and AWS Bedrock, with a year and more of daily agentic coding in Claude Code, Cowork and Codex. I reproduce results rather than cite them, and publish what I learn: derivations worked by hand first, then checked against the machine.

RESEARCH & PUBLIC OUTPUT

- **Towards First Principles** — adithyag73.github.io/first_principles. Long-form technical explainers with interactive widgets and measured implementations: *FlashAttention* (Aug 2026), *Towards Free Normalization* (Jul 2026).
- **Interpretability** — **natural language autoencoders**. Running activation-verbaliser and activation-reconstructor inference with per-claim ablation studies over explanations of residual-stream activations; technical write-up in preparation.
- **Fine-tuning for agentic tool use**. Supervised and RL-style fine-tuning of Qwen 3 on Thinking Machines' Tinker API (research credits, Nov 2025) for tool selection and Cypher generation over a Neo4j graph.
- **Writing & competition**. Regular essays on frontier research at medium.com/@AdithyaGiridharan; Kaggle competitor since 2020 — **44 competitions**, best finish top 500 in a playground series.

EXPERIENCE

AI Engineer Jan 2026 – Present

Neo Wealth Asset Management India

- Building the AI backend for **MOR**, a conversational trading platform: LLM-driven agentic workflows that turn natural-language requests into portfolio insight and trade workflows.
- Own agent orchestration, tool and data integration, and conversational state management for real-time financial interactions.

GenAI Engineer Jul 2024 – Jan 2026

Karini AI Pune, India

- Production parts knowledge graph for a Fortune 500 manufacturer — **65K+ nodes, 1.9M relationships, 25K+ SKUs** in Neo4j with constraints, full-text and vector indexes, fed by an AWS Bedrock Titan v2 embedding pipeline.
- Developed **Text2Cypher fine-tuning** for natural-language-to-graph-query translation behind router-based multi-agent workflows.
- Conversational RFP platform for an engineering consultancy: **75–150 page RFPs at \$10–15 each** against a 5.5K-document knowledge base (LangGraph, OpenSearch, PostgreSQL, Textract, DynamoDB auditing).
- GraphRAG talent-acquisition system over **1,500+ resumes** (Textract, Neo4j, OpenSearch, MongoDB) with FastAPI agentic workflows and Traceloop observability.
- Agentic RAG testbed benchmarking embedding models on recall@k, precision@k, NDCG@k and MRR across legal, engineering and product-manual domains; MCP systems for repository analysis and Kaggle automation.

Data Scientist Aug 2023 – Jun 2024

Mercedes-Benz Research & Development India Pune, India

- Computer-vision pipeline for wiper-health diagnostics, classifying dashcam video into streaking vs. skipping defect patterns via transfer learning, with ONNX conversion for production; GraphRAG for smart product data management across automotive components.

Decision Sciences Intern Jun – Nov 2022

HSBC WCR Analytics Bengaluru, India

- Credit-risk analytics on insurance data with RWA impact analysis; climate-financing and sustainability review.

EDUCATION

M.Tech, Data Science & Analytics — IIIT Prayagraj 2021 – 2023 | CGPA 8.88/10

B.E., Instrumentation & Control Engineering — Sri Sairam Engineering College, Anna University 2016 – 2020 | CGPA 9.02/10

Reinforcement Learning Specialization — University of Alberta four-course sequence

TECHNICAL SKILLS

Research	Mechanistic interpretability, sparse autoencoders (SAEs), model fine-tuning (SFT, RL), ablation studies, PyTorch, Hugging Face, Tinker
Agents	AI agents (ReAct), multi-agent workflows, LangGraph, LangChain, MCP, claude-agent-sdk, Anthropic, OpenAI
Agentic coding	Claude Code, Cowork, Codex — 1+ year of daily use
LLM systems	RAG/GraphRAG, Text2Cypher, conversational AI, AWS Bedrock (Claude, Titan v2, Cohere v4)
Databases	Neo4j (constraints, full-text & vector indexes), OpenSearch, PostgreSQL, MongoDB, DynamoDB
Platform	AWS (S3, EC2, Lambda, SageMaker, Bedrock, Textract), Python, FastAPI, Docker, Git, Traceloop/OpenLLMetry