

ADYA PRASAD

Applied AI Researcher | Mechanistic Interpretability | AI Safety

✉ adyaprasad.info@gmail.com

📍 Uttar Pradesh, India

🌐 <https://adya-prasad.github.io/>

EXPERIENCE

Applied AI Research Intern | *BaseThesis Lab, Bangalore*

May 2026 - Present

- Working on research: **Evaluating Chain-of-Thought Unfaithfulness from Residual Stream Trajectories and linear probe.**
- Initiated and co-led an empirical research project investigating whether a model's internal residual-stream representations faithfully align with its verbalized chain-of-thought during reasoning task.
- Built a diversified evaluation dataset spanning mathematics, reasoning, history, and AI safety for robust probe training.
- Scored AUROC 0.85, demonstrating residual-stream trajectories provide useful signals for detecting unfaithful reasoning

Open Source ML Contributor | *HuggingFace: PEFT and TRL*

Mar 2026 - Present

- PEFT: Conduct an **benchmarking experiment on different PEFT methods on conversational AI** designed an evaluation pipeline to evaluation LLaMA-Adapter method. My experiment improve the benchmark accuracy from **21.2%** to **37.7%**.
- Contributed a fix to the PEFT library ensuring `ensure_weight_tying` correctly respects model configuration `tie_word_embeddings` during serialization.

Full Stack Engineer | *NowTreat Healthcare*

Jun 2023 - Nov 2025

- Ownership of official website and integrate a **patient post treatment care and booking service**, along with feedback loop using WordPress and JavaScript to support post-treatment guidance, and improve patient engagement.
- Along with marketing team and founder, designed an marketing funnel for better user experience and engagement and converting online traffic to potential patents. Collectively achieve **38%** more engagement and **26%** more footfall.

PROJECTS

Cupid: Agentic Social Media Content Generation Platform | Multi Agent Orchestration

- Engineered a production grade **AI agents, swarm architecture experimental platform (AI Suite) for diverse and user personalized content creations** using LangGraph, FastAPI, API Services, PostgreSQL, Redis, ChromaDB, and Docker.
- Designed a multi-stage safety and response evaluation pipeline combining heuristic validation, policy checks, and supervisor-agent verification to filter user inputs and validate LLM outputs before responding to user.
- Implemented hybrid persistent user memory, personalized context management, retrieval-augmented research, and intelligent ranking agents for platform-specific content generation.

AushadhiNet-GATv2 DDI Model | Predicting Drug-Drug Interactions in Polypharmacy | Hackathon Winner

- A **graph attention network** version 2 architecture ai model for predicting drug-drug interactions using multi-view molecular representations. Consist of dual-head classification for interaction detection and mechanism prediction across **86 interaction classes**
- Model predict drug combination with probability scores and classifies interaction types. **Awarded 1st Place** at the *Bye to Beat Hackathon*.

PUBLICATION

AushadhiNet-GATv2: Graph Attention Networks for Drug-Drug Interaction Prediction in Polypharmacy (Preprint)

TECHNICAL SKILLS

- Languages:** Python, TypeScript, R lang, JavaScript, SQL
- ML:** PyTorch, Transformers, Hugging Face, PEFT, Mechanistic Interpretability, Graph Neural Networks, Reinforcement Learning, Retrieval-Augmented Generation, Post Training, and Empirical Evaluation.
- Tools:** Apache Ecosystem (Kafka, Spark, and Hadoop), Hugging Face Ecosystem, Docker, Redis, MongoDB, n8n
- Libraries:** LangGraph, TransformerLens, NLTK, Numpy, Pandas, Matplotlib, Langchain, SQLAlchemy, ChromaDB

EDUCATION

- B.Tech(Computer Science & Engineering) | *Veer Bahadur Singh Purvanchal University, Affiliated College*
- Harvard CS50:** Computer Science, Introduction to Cybersecurity, Programming with R, Artificial Intelligence with Python
- Master Diploma in Computer Application | *Amritkosh Infotech (2 years)*
- Stanford CS229 (Machine Learning), Stanford CS230 (Deep Learning), Stanford CS224R (Deep Reinforcement Learning)** | Online Certification

EXTRACURRICULAR ACTIVITIES

- Technology Job Simulation | Deloitte Australia:** Develop a program to normalise different telemetry JSON data of IOT machine (Simple ETL pipelines)
- 200+ algorithmic problems solved** across LeetCode, HackerRank, and Codeforces