Abhinav Rajeev Kumar

🌐 abhinavrkumar.com ✉ abku2504@gmail.com ⌂ github.com/MajorTimberWolf ☛ Google Scholar

Education

May 2026 Aug 2022	**SRM Institute of Science & Technology** *B.Tech in Computer Science and Engineering*	**Chennai, India**

Experience

Present June 2025	**Lossfunk [🌐]** *Research Intern*	**Bangalore, India**

Investigating sycophancy in language models using mechanistic interpretability; specifically whether persona vectors steer circuits causing sycophantic behavior. Also Built an agentic research mentor system for undergraduate guidance.

October 2025 May 2025	**Darko Marinov's Research Group, UIUC [🌐]** *Undergraduate Researcher	Advisor: Prof. Darko Marinov*	**Remote**

Engineered LLM workflows for semantic evaluation on software mutants, classifying code changes as equivalent. Developed analysis scripts to identify "fishy" mutants by extending BugFarm ICSE heuristics.

May 2023 May 2024	**Google (Gemini Team)** *Software Engineer Intern*	**Remote**

Refined inappropriate question detection using XAI/LIME (20% FP reduction). Applied chain-of-thought prompting for multi-step reasoning (24% F1 improvement). Also worked with symbolic reasoning for complex mathematical problems like Lorentz ODE proofs.

Publications   [* = Equal Contribution]

[2025]	**Evaluation of LLM Agents for Research Mentorship** [PDF] Abhinav Rajeev Kumar, Dhruv Trehan, Paras Chopra *AI For Scientific Research Workshop, AAAI 2025*	**[AAAI-W]**
[2025]	**Analyzing Quantization in Multilingual Transformer Models for Indic Translation** [PDF] Praneeth Kasiraju*, Abhinav Rajeev Kumar*, Ramprasath M, Manikandan N *IEEE Conference on Computational Methods for Sound Science (CMSS) 2025*	**[IEEE]**

Select Projects

Joker-LLM   May '25 - Jun '25
Application Project for Lossfunk | [Technical Writeup]   [🌐]

> Implemented PlanSearch for generating "funny" jokes by using multi-dimensional evaluations and bias mitigations via LLM-as-a-judge methodology.

RAO: Recursive Agent Orchestrator   Jan '25 - Mar '25
Research Project | [Technical Writeup]   [🌐]

> Developed an agentic system that spawns and manages LLM-driven agents via a master-delegator model in the form of layers, built for complex multi-agent workflows.

ReportEase   Jan '24 - Feb '24
Third Place Overall @ Rajasthan Police Hackathon | Developed by Team Gallants [🌐]

> Developed an automated legal insight platform using Mistral LLM fine-tuned on BNS & BNSS datasets to analyze FIRs and generate insightful reports.

> Incorporated few-shot technique with adaptive prompts to enhance LLM performance on Hindi text classification tasks, addressing limited labeled data.

Surge Protector   Feb '23 - Mar '23
Best Track using AI/ML @ Stanford TreeHacks   [🌐]

> Built a drone and dashboard system capable of detecting stampedes by marking heads in crowds using PyTorch CNN and generating heatmaps with Matplotlib and OpenCV.