Arnesh Banerjee

Machine Learning and AI Engineer

arneshbanerjee24@gmail.com | arnesh24 | ArneshBanerjee | portfolio

Experience

Indian Institute of Technology Kharagpur **May 2026 – Present**

Summer Research Intern *IIT Kharagpur, India*

- Assembled a leakage-controlled multimodal dataset of 1,206 samples across 6 molecular and clinical modalities from the public TCGA head-and-neck cohort, then reformulated the task from binary detection to 10-class tumor typing, selecting the best of 12 model families per modality to reach 0.91 balanced accuracy.
- Designed a Qwen3-32B LLM orchestrator fusing 6 reliability-weighted model predictions with retrieval-augmented grounding and a deterministic safety layer, and diagnosed a training-set leakage flaw that had inflated 3 modalities to a spurious 1.00 accuracy.

Jadavpur University **Nov 2025 – May 2026**

Research Intern *Remote*

- Designed a hybrid CNN-Transformer segmentation model with a novel differentiable Chan-Vese level-set loss and benchmarked it against 4 established architectures on the DMR-IR dataset (357 images, 119 patients), reaching 0.97 Dice under patient-stratified 5-fold cross-validation.
- Discovered an annotation-quality ceiling by proving 5 architectures of widely differing capacity converge to statistically indistinguishable Dice (0.97, p ¿ 0.05) under pseudo-label supervision, validating the finding with Wilcoxon testing, bootstrap confidence intervals, and label-noise robustness experiments.

New Jersey Institute of Technology **June 2025 – Nov 2025**

Summer Research Intern *Newark, USA*

- Constructed a curated safety dataset of 2,500 prompt-response pairs spanning benign, context-sensitive, and jailbreak categories, and fine-tuned the final 6 layers of LLaMA-2-7B-chat-hf into a semantic cost model reaching 92% alignment with human safety judgments.
- Benchmarked the cost model on the external XS-Test suite, scoring 0.91 to 0.96 versus 0.07 to 0.32 for Safe-RLHF, and demonstrated via 1,000-prompt and 1,172-pair human-preference evaluations that it is 5x more efficient at flagging unsafe responses, winning 60% of head-to-head comparisons (+70 Elo) against the strongest baseline, SACPO.

Technical Skills

Languages: Python, C++, Go, Rust, TypeScript, JavaScript, SQL, Java
Frameworks & Tools: PyTorch, TensorFlow, Scikit-learn, FastAPI, Next.js, React, Docker, Git

Projects

Kimi K3 For All | Python, PyTorch, CUDA/Triton **Code**

- Built a from-scratch training pipeline for Kimi K3, a 2.8T-parameter Mixture-of-Experts model released with inference-only code, by fixing **4** undocumented bugs blocking gradient flow through its router and experts.
- Trained a **1.27**B-parameter (**364**M active) Kimi K3 model end-to-end on **4.6**B Wikipedia tokens on a single H100, reaching validation perplexity **13.89** (down from **55.6**) across **18** pause/resume cycles.

Open Source

OpenCV | C++ **PRs**

- Fixed a pointer overflow bug in 16u image-codec conversion helpers, restored in-place support for `extractChannel`, and resolved a `boundingRect` crash on boolean masks across 5 merged PRs to a computer-vision library used by millions of applications.

Zed | Rust **PRs**

- Added Gemini 3.6 Flash provider support and shipped 3 Jupyter notebook editor improvements, including cell deletion and single-file worktree support, to a fast-growing open-source code editor.

Kubescape | Go **PRs**

- Fixed a registry memory-growth bug via idle eviction and resolved IPv6 CIDR and parsing errors across 6 merged PRs to this CNCF Kubernetes security platform, spanning its core scanner, storage, and admission-policy repos.

Education

Heritage Institute of Technology

B.Tech in Computer Engineering (Data Science), CGPA: 8.9 *2023 – 2027*