

Aritro Shome

+91-9073989733 | aritra.shome.official@gmail.com | [LinkedIn](#) | [GitHub](#) | [Twitter](#) | [Blog](#)

EDUCATION

IIST Shibpur

B.Tech in Information Technology – CGPA 9.15/10.0

Shibpur, West Bengal

Aug 2024 – May 2028

M. P. Birla Foundation Higher Secondary School

ISC (Class 12) 95.3%, ICSE (Class 10) 98.3%

Kolkata, West Bengal

Apr 2010 – May 2024

EXPERIENCE

Research Intern | PyTorch, Hugging Face, Interpretability | AI4Bharat, IIT Madras

Mar 2026 – Present

- Investigating **mechanistic interpretability** and **circuit tracing** of SoTA open models (**Gemma**, **Qwen**) on reasoning tasks posed across **modalities, languages, and model scales**, with a focus on Vision-Language Models (VLMs) on Indic inputs.
- Ran **40,000+ instrumented forward passes** for activation capture and analysis on **multi-billion-parameter models** under constrained hardware (RTX 3090 Ti, 24 GB VRAM); built pre-experiment component-attribution scripts to isolate critical circuits before full runs.
- Engineered **memory-efficient frameworks** for activation patching: disabled fused/attention-kernel optimizations to capture per-layer activations and managed the resulting **non-linear VRAM growth** to keep large models within budget.

AI Engineer Intern | Python, GenAI, Data Pipelines | Sarvam AI

Aug 2025 – Dec 2025

- Built an automated evaluation framework for an **NVIDIA lip-sync model**; the quantitative pipeline detected and resolved a bug causing **~500% variance** in output video file sizes.
- Redesigned the production **dubbing pipeline**: replaced brittle, hard-coded prompt harnesses with **modular YAML instruction blocks** that let model chain-of-thought infer context and produce higher-quality translations; validated the optimal configuration via **A/B testing**.

PROJECTS

ForgeFormer | Mechanistic Interpretability | Writeup, Demo

2026

- Hand-designed an **interpretable 2-layer transformer** that performs two-digit addition with **manually-set weights and no gradient descent**, demonstrating transformer internals (attention + MLP computation) constructed from first principles.

Seven Deadly Sins of Gemma | Steering Vectors, Activation Patching | Writeup

2026

- Replicated Anthropic's persona/emotion-vector discovery methodology to extract **steering vectors** for each of the "seven deadly sins" in **Gemma** on a consumer GPU; verified vector correctness through targeted **steering** and **activation-patching** experiments.

AlponaGen | Synthetic Data, Generative Art | GitHub, Kaggle

Oct 2025

- Engineered a modular Python engine for procedural Bengali folk art using fractal-based layering; published a **13,000-image synthetic dataset (1024×1024)** on Kaggle for generative-AI research, with an extensible OOP style-registration system.

ACHIEVEMENTS

1st – TensorQuest ML Hackathon, Instruo 2025 (IIST Shibpur) – out of **100+ teams** across India.

1st – Brain Dead ML Hackathon, Revelation 2025 (IIST Shibpur) – among **~200 teams** globally; built IPL EDA pipeline, winner-prediction ML models, and a score-based statistical query system.

1st / Best Web App – eXabyte & Hacknoware Hackathons – built *Abler*, a real-time multimodal accessibility app (Apple depth models, TTS/STT, object detection, gesture recognition) for the visually impaired. [Demo](#).

1st Prize – Prof. Brahma Prakash Memorial Material Science Quiz & Paper Presentation 2023 (IGCAR, Kalpakkam) – authored a paper on *Role of AI/ML in Materials Research*; **1st out of 500+ applicants** across India.

INSPIRE Scholarship – Government of India, awarded to the top 1% in the higher-secondary qualifying examination, May 2024.

TECHNICAL SKILLS

Languages: Python, Java, C/C++, JavaScript, HTML/CSS

ML & Frameworks: PyTorch, Hugging Face Transformers, Scikit-Learn, NumPy, Pandas, Matplotlib, LangChain, Flask

Interpretability: Activation patching, circuit tracing, steering/activation-vector extraction, mechanistic interpretability, logit lens

Infra & Tooling: Multi-GPU clusters, distributed & VRAM-constrained training, CUDA, mixed precision, Linux, Git, Docker

Concepts: Transformers, LLMs/VLMs, RL, DSA, OOP, Linear Algebra, Probability & Statistics

Certifications: [Deep Learning Specialization \(deeplearning.ai\)](#), [Sequence Models](#), [CNNs](#), [GenAI Apps with Gemini & Streamlit](#)

WRITING & LEADERSHIP

Technical blog – silicognition.is-a.dev: writeups on solo interpretability and research projects.

AI/ML Co-Lead, Google Developer Groups on Campus (IIST Shibpur) – ran sessions on deep learning and transformer architecture; mentored a junior who secured an internship.