Kaustubh Sharma

✉ kaustubh_s@ee.iitr.ac.in | ✉ kaustubhdadhich@gmail.com | ⌂ github.com/kaustubh202/

EDUCATION

• **Indian Institute of Technology Roorkee** — 2023-2027
B.Tech, Electrical Engineering | CGPA: 9.04/10 — Roorkee, India

• **Central Board of Secondary Education (CBSE)** — 2023
Senior Secondary | Class XII: 97.6%, Class X: 98.6% — St. Xavier's School, Jaipur, India

AREAS OF INTEREST

Mechanistic Interpretability, Diffusion Models, ML for Science and Engineering

PUBLICATIONS

• **Explainable AI-Generated Image Forensics: A Low-Resolution Perspective with Novel Artifact Taxonomy**
Proceedings of the International Conference on Computer Vision, APAI Workshop ICCV, 2025
  ◦ Developed a robust, interpretable pipeline for detecting AI-generated images at 32×32 resolution.
  ◦ Introduced a hierarchical artifact taxonomy and an explainability framework.
  ◦ Combined SinSR super-resolution, CLIP-guided artifact detection, and VLM–based explanations.

• **Image-Alchemy: Advancing Subject Fidelity in Personalized Text-to-Image Generation**
DeLTa Workshop, ICLR, 2025
  ◦ Designed a pipeline for personalizing Stable Diffusion XL with high-fidelity novel subject integration.
  ◦ Reduced overfitting and catastrophic forgetting through targeted fine-tuning strategies.
  ◦ Applied LoRA-based training and segmentation-guided Img2Img for realistic subject blending.

EXPERIENCE

• **Research Assistant** — *May 2025 - Ongoing*
P-square Lab, Prof. Parikshit Pareek, Indian Institute of Technology Roorkee
  ◦ Working on a new attention mechanism for Prior Data-Fitted Networks
  ◦ GP Inference time reduction for physical equations solving and scaling of PFNs

• **Incoming Quantitative Research Intern** — *Upcoming*
Goldman Sachs
  ◦ Selected for the quantitative strategist internship involving financial modeling and statistical analysis.

• **Educator at Edufabrica Pvt. Ltd.** — *March 2025*
  ◦ Delivered a workshop lecture for 2 days to 200+ students across India on Generative AI.

OTHER PROJECTS

• **Domain Circuit Discovery in LLMs - Mechanistic Interpretability** — *April 2025 - Ongoing*
Project Website (Ongoing)
  ◦ Investigating domain-specific knowledge emergence in LLaMA 3-3B to locate specialized layers.
  ◦ Evaluated Forward Pass Profiling, Probe Separability, Zero out tests, Hydra Effect and Fine Tuning shifts.
  ◦ Mapped internal representations to identify domain-specific "rooms" across the transformer architecture.

• **Sparsity-Aware Representation Learning for Jet Image Generation via Guided Latent Diffusion** — *2025*
  ⌂ *Github*
  ◦ Developed a sparsity-aware latent diffusion framework to generate High Energy Particles Physics Data.
  ◦ Crafted a custom variational autoencoder with a novel sparsity-focused reconstruction loss.
  ◦ Introduced a mean-pulling mechanism during latent diffusion to minimize reconstruction artifacts.

ACHIEVEMENTS

  ◦ **Micron AI Hackathon** (2025, Micron Technology) — Second Runner Up in micron AI Hackathon.
  ◦ **JEE Advanced** (2023, Govt. of India) — AIR 1624 among 1 lakh+ applicants.
  ◦ **JEE Mains** (2023, Govt. of India) — AIR 1898 among 12 lakh+ applicants.
  ◦ **NTSE Scholar** (2021, Govt. of India) — National scholarship awarded to 2000 out of 9L+ candidates.
  ◦ **KVPY Fellow** (2022, IISc, Govt. of India) — AIR 509 among 2 lakh+ applicants.

ADDITIONAL INFORMATION AND CO-CURRICULARS

Languages: Hindi (Native), English (Advanced), French (Beginner)

Tech Clubs: Member — Data Science Group — IIT Roorkee

Music: Member — Music Section — IIT Roorkee as a pianist

Sports: Member — Swimming team — IIT Roorkee