Aaryan Ajay Sharma

: Aaryan Sharma
: aaryanajaysharma

: aaryan.s@research.iiit.ac.in
Google Scholar
: aaryanajaysharma.github.io

Education

- **International Institute of Information Technology, Hyderabad** — Hyderabad, IN
  B.Tech. (Hons.) and MS by Research in Computer Science and Engineering, CGPA: 8.38 — *July 2022 – July 2026*
  - Advisor: Dr. Ankit Gangwal

Experience

- **Applied AI Researcher** — Hyderabad, IN
  ServiceNow — *January 2026 - June 2026*
  - Working on benchmarking content moderation in enterprise-level agentic systems, with a focus on long-context multi-modal agentic systems and reasoning-oriented content moderation models.

- **Research Intern** — Hyderabad, IN
  Infosys — *July 2025 - October 2025*
  - Worked on red teaming of LLMs via membership inference and data extraction attacks, along with investigating privacy implications of model merging.

- **Undergraduate Researcher** — Hyderabad, IN
  Center for Security, Theory & Algorithmic Research (CSTAR), IIIT Hyderabad — *April 2023 - Present*
  - Investigated ownership verification and watermarking of Deep Neural Networks (DNNs) by implementing state-of-the-art methods. Guided by Dr. Ankit Gangwal.
  - Worked on watermarking Graph Neural Networks (GNNs) for Link Prediction using PyTorch Geometric, evaluating across multiple architectures and real-world datasets. Work under review at TMLR. Advised by Dr. Charu Sharma.

- **Research Assistant** — Hyderabad, IN
  CSTAR, IIIT Hyderabad — *April 2024 - January 2025*
  - Worked on multiple ML security problems such as Explainable AI (XAI), explanation-aware backdoor attacks, and adversarial attacks, culminating in a poster at a CORE:A conference (acceptance rate: 18%).

Publications

Ankit Gangwal and **Aaryan Ajay Sharma*** (2025a). "Merge Now, Regret Later: The Hidden Cost of Model Merging Is Adversarial Transferability". In: *arXiv, arXiv:2509.23689*. Under Review - AsiaCCS 2026.

Ankit Gangwal and **Aaryan Ajay Sharma*** (2025b). "POSTER: Investigating Transferability of Adversarial Examples in Model Merging". In: *Proceedings of the 20th ACM Asia Conference on Computer and Communications Security*, pp. 1812–1814.

VSP Bachina*, Ankit Gangwal, **Aaryan Ajay Sharma***, and Charu Sharma (2026). "GENIE: Watermarking Graph Neural Networks for Link Prediction". In: *Transactions on Machine Learning Research*.

Honours and Awards

- **Dean's List Award.** Awarded the Deans List twice, reserved for top 5% of the students.
- **Dean's Research Award.** For publishing at an international conference while being an undergraduate.
- **JEE Main 2020 examination.** Top 1.78% among 1.2 million students.

* denotes first author/equal contribution. Authors ordered alphabetically by last name.

Projects

- **Fine-Tuning for Summarization**: Implemented Prompt Tuning, Low-Rank Adaptation (LoRA), and Traditional Fine-Tuning for GPT-2 small on summarization tasks. **Tech Stack/Concepts:** PyTorch, Hugging Face, Transformers, Fine tuning, LoRA.
- **Adversarial evaluation of LIME for Hindi text**: Adapted the XAIFooler attack method for Hindi by developing a sequential perturbation algorithm that generates adversarial explanations while preserving semantic integrity and prediction stability. Fine-tuned IndicBERT and XLM-RoBERTa models on Hindi datasets, showcasing the applicability of adversarial techniques to low-resource languages.**Tech Stack/Concepts:** PyTorch, Hugging Face, Transformers.
- **Quantization and Model Compression**: Applied whole-model and selective component quantization, integrated bitsandbytes for 8-bit, 4-bit quantization, and NF4 nonlinear quantization. Evaluated trade-offs in efficiency and accuracy. **Tech Stack/Concepts:** PyTorch, Quantization, Bitsandbytes, Nonlinear Quantization.
- **Neural Languange Modelling**: Developed and implemented multiple language models using PyTorch, including 5-gram neural networks, RNNs, LSTMs, and Transformer-based architectures. Evaluated models based on perplexity scores and optimized hyperparameters to improve performance. **Tech Stack**: PyTorch, TorchText, Numpy, NLTK.
- **Multi-Layer Perceptron from scratch**: Implemented a customizable MLP with support for various activations, optimizers, & training modes. Built forward/backpropagation in NumPy. **Tech Stack/Concepts:** Computational Graphs, Matrix Calculus.
- **Metagenomic Binning using Graph Neural Networks**: Applied Graph Representation Learning (GRL) to enhance metagenomic binning by analyzing DNA fragments in assembly graphs. Built upon the RepBin framework, systematically testing alternative methods to evaluate their impact on binning performance using the Sim-5G dataset. **Tech Stack/Concepts:** PyTorch Geometric, Contrastive Learning.
- **Role of Feedback in Sensorimotor Tasks**: Investigated how real-time performance feedback affects sensorimotor learning and control across repeated trials. Applied Power transformation for normality and used Student's t-tests to reveal significant performance differences under feedback conditions. **Tech Stack/Concepts:** R, Python, Pandas, Experiment Design, Null Hypothesis Significance Testing.

Programming Skills

- **Technologies**: PyTorch, NumPy, Scikit-learn, PyTorch Geometric, Pandas, Docker, mpi4py, Hadoop MapReduce, AWS, Git, React, Node, Strapi **Languages**: Python, C/C++, Javascript, Bash, SQL, Java

Teaching Roles

- **Teaching Assistant – Systems Network Security (Spring 2024)**
- **Teaching Assistant – Advanced Computer Networks (Fall 2024, 2025)**
- **Teaching Assistant – Blockchain and Web3 Development (Fall 2024, 2025)**
- **Teaching Assistant – Introduction to Information Security (Spring 2025)**
- **Teaching Assistant – Computer Systems I (Division of Flexible Learning, IIIT-H)**
- **Teaching Assistant – Computer Systems II (Division of Flexible Learning, IIIT-H)**

Assisted in designing and grading assignments and exams, delivering tutorials, and mentoring student projects across multiple undergraduate courses.

Relevant Course Work

Topics in Deep Learning (Stanford CS224W: Machine Learning with Graphs), Computer Vision, Advanced Natural Language Processing, Advanced Algorithms, Open Quantum Systems and Thermodynamics (Advanced Linear Algebra & Probability), Statistical Methods in AI (Stanford CS229: Machine Learning), Behavioral Research Statistical Methods, Behavioral Research Experiment Design, Introduction to Information Security, Performance Modeling of Computer Systems (Queueing Theory), Probability & Statistics, Algorithm Analysis and Design, Distributed Systems, Automata Theory, Data and Applications.