

Saumya Chaturvedi

saumyachat03@gmail.com | 9667730960 | linkedin.com/in/schaturv | github.com/shollercoaster

Education

Netaji Subhas University of Technology, BTech in Electronics and AI Jan 2022 - May 2025

- GPA: 8.4/10.0
- **Coursework:** Computer Architecture, Computer Networks, Computer Programming, Optimization in AI ML, Data Structures and Algorithms, Machine Learning, Deep Learning, Mathematics of ML, Computer Vision

Publications

SQL-of-Thought: Multi-agentic Text-to-SQL with Guided Error Correction Sept 2025

Saumya Chaturvedi, Aman Chadha, Laurent Bindschaedler

- Accepted at the **NeurIPS 2025 Workshop DL4C (Deep Learning for Code)**
- Introduces a multi-agent NL2SQL framework that combines chain-of-thought planning with a taxonomy-guided correction loop, achieving state-of-the-art 91.69% execution accuracy on Spider benchmarks.

LoRACode: LoRA Adapters for Code Embeddings March 2025

Saumya Chaturvedi, Aman Chadha, Laurent Bindschaedler

- Accepted at the **ICLR 2025 Workshop DL4C (Deep Learning for Code)**
- Introduces a novel method for finetuning code embedding models on code retrieval based on text and code queries, using low rank adapters, achieving up to 86.69% improvements on Mean Reciprocal Rank using less than 2% of trainable parameters.

Experience

Software Engineer, American Express, India Sept 2025 - Present

- Developing a **Data Profiler** Engine to calculate statistics of records affected before and after data deletion and anonymization operations. Also developing a smart scheduler to efficiently schedule multiple records for single ingestion/deletion request.
- **Tools & technologies:** Python, FastAPI, Jenkins, Kafka, GCP, Data Warehouse, Git, PostGreSQL, Agile, CI/CD

Software Engineer Intern, American Express, India Jan 2025 - July 2025

- Engineered a fully automated security remediation system for fetching, resolving and upgrading vulnerable Maven dependencies in multiple large-scale Spring Boot microservices by parsing direct and transitive nested pom.xml trees and optimizing resolution graph traversal for faster CI pipelines, leading to **80% deduction** in manual efforts.
- Contributed to internal documentation (sequence diagrams, deployment workflows, and API architecture), and co-developed a colleague engagement survey management tool with an analytics dashboard within 48 hours during a global internal hackathon.
- **Tools & technologies:** Java, Spring Boot, Maven, Jenkins, Git, PostGreSQL, Agile, CI/CD

Research Scholar, Max Planck Institute for Software Systems (MPI-SWS), Germany May 2024 - Dec 2024

- Created custom datasets and wrote scripts to finetune various LLMs using LoRA (Low Rank Adaptation) adapters to answer questions about context-encoding and parallelization of prompt processing, to create efficient Compound AI Systems. Also increased cross-domain output accuracy by 30-40%.
- Conducted extensive research on DNN/LLM Inference Serving Systems analyze inter-token latency, GPU utilization, and throughput, optimizing token scheduling and batching strategies for high-throughput, low-latency LLM serving.
- Worked on and wrote my first research paper LoRACode: LoRA Adapters for Code Embeddings as first author.
- **Tools and skills used:** Docker, Anaconda, PyTorch, Python, Llama model, Adapters, PEFT, Transformers, Huggingface, SciPy, Remote Access Control, Unix, Academic Research, Scientific Writing, WandB

Software Engineering STEP Intern, Google – India

May 2023 - July 2023

- Created a new feature from scratch in the credit vertical of the GPay app, worked in both frontend and backend layers.
- Finalised the overall technical design, created the experimental feature highlighting analytics of all user interactions, rendered the interactive frontend UI for the whole feature.
- Instantiated web framework nodes and wrote the backend API, while writing tests of every unit affected in different layers of production environment.
- **Tools & technologies:** Java, Javascript, Typescript, Soy, SCSS, PostgreSQL, Version Control.

Projects and Certifications

BlueDot Impact Technical AI Safety Course

[Link](#)

- Enrolled in intensive version of the BlueDot Technical AI Safety Course. Participated in 3/3 classes, finished coursework, 3 weeks left to completion.

ARENA Course

[Github](#)

- Notes and solutions of Mechanistic Interpretability exercises in the ARENA Course through online material.
- Created and led an online study group to discuss learnings, challenges and explain concepts from the course in teacher-student pairs.

"Scit": Git from Scratch in Scala

[Gitlab](#), [Documentation](#)

- Built a lightweight Git-like version control system in Scala with support for init, add, commit, log, status, and restore commands via a custom CLI interface.
- Implemented content-addressable storage using SHA-1 hashing and a hierarchical .scit/objects structure, along with custom commit history and staging area using nested Maps.
- Generated optimized JAR binaries using sbt-assembly, and handled low-level file I/O and state recovery for reliable version tracking and file restoration.

Conducted lectures on Quantum Computing

[Github](#)

- Implemented reading notes, summary, whiteboard illustrations, mathematical proofs and lecture presentations, did chapterwise presentations for 200 students in my WE program classes as part of peer-teaching.
- **Languages & concepts learnt:** LaTeX Beamer, Typesetting, Formal Logic, Mathematical Set Theory, Turing Machines, Halting Problem, Consciousness, Quantum Computing

Leadership

President, Axiom: Philosophical Society of NSUT

May 2023 - Jun 2024

- President of University-wide Philosophy Club, Founder of Axiom Research.
- Have organized Special Interest Groups on different disciplines in philosophy, philosophy JAM events, flagship Intra-University events, Alumni Mixer day as well as movie-dissection and book-reading sessions.

Skills and Achievements

- **Languages:** Python, Java, JavaScript, Typescript, C, C++ , MySQL, PostgreSQL, HTML, Bash, Regex
- **Technologies and Tools:** Flask, Django, FastAPI, Springboot, PyTorch, Docker, Git, Linux, LaTeX, OpenCV, Tensorflow, qBraid, Microsoft Azure
- Selected in top 0.9% candidates out of 28,000+ applicants all over India for the Women Engineers Mentorship program on 100% scholarship and additional cash stipend of 1 lakh rupees.
- Scored **116/120 in TOEFL**, appeared in March 2024.
- **Harvard Project for Asian and International Relations (HPAIR) Conference 2024 Delegate** at Harvard University
- Winner of **GHC (Grace Hopper Celebration) Scholarship 2023**
- Winner of **SIH (Smart India Hackathon) 2023**, University Round