

SIDDHARTHA MAHAJAN

+91-7599-063750 ◇ Delhi, India ◇ siddhartha-mahajan.github.io 

siddharthamahajan03@gmail.com  ◇ [Siddhartha-Mahajan](#)  ◇ [siddharthamahajan03](#) 

EDUCATION

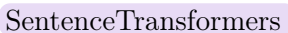
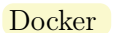
Bachelor of Technology (B.Tech.), Cluster Innovation Centre, University of Delhi 2021–2025
Study Areas: IT & Mathematical Innovations, Computational Biology **8.8/10**
Relevant Courses: Calculus, Data Structures & Algorithms, Probability, AI/ML, DBMS (SQL), Statistics, Linear Algebra, Computational Biology. [Visit Website](#) 

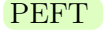
SKILLS

Programming Languages	Python, Java, Julia, C, R, SQL, Lean 4
Data Analysis and ML	PyTorch, scikit-learn, NumPy, Pandas, SciPy, Keras, Transformers, SentenceTransformers, LLMs, RAG, LoRA, GRPO
Backend, Cloud and Data	AWS, PostgreSQL, FastAPI, Flask, Docker, ChromaDB, Pinecone
Software and Tools	Git, Jupyter, L ^A T _E X, Mathlib, MATLAB, Mathematica, Selenium, BeautifulSoup
Scientific Computing	NetworkX, Matplotlib, Modeller, ChimeraX, GROMACS

WORK EXPERIENCE

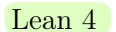
Software Engineer, RedGraphs March 2026 – Present
Applied AI, NLP, and Relationship Intelligence   
Build NLP and backend pipelines that extract company relationships from regulatory filings, financial disclosures, and public information. Develop entity resolution and evaluation workflows to improve precision, recall, and evidence traceability. Work with AWS and PostgreSQL on backend performance, observability, and production reliability.

AI Consultant, Amogha AI Technologies November 2025 – June 2026
Applied NLP, Conversational AI   
Worked on NLP pipelines and conversational AI workflows using RAG, ChromaDB, and Docker for a voice-first mental health support application, with a focus on experimentation and evaluation.

Research Associate, SBILab, IIIT-Delhi August 2025 – February 2026
Parameter-Efficient Fine-Tuning and ECG Classification   
Worked with Prof. Anubha Gupta on ECG-based CVD subclass classification using a custom ViT trained on over 10 million ECGs. Explored parameter-efficient fine-tuning methods, including LoRA extensions and custom adapters.

Machine Learning Intern, Beyond Exams May 2022 – November 2022
Developed a model to classify YouTube videos as educational/non-educational and into subject subcategories using a self-scraped dataset. Built data collection and cleaning workflows with the YouTube V3 API, BeautifulSoup, Selenium, Python, and Excel; trained models in ML.NET (C#). 

RESEARCH EXPERIENCE

Research Fellow, Lossfunk March 2026 – August 2026
AI-Assisted Mathematics and Formal Verification   
Worked on AI-assisted mathematical research across combinatorics, integer sequences, and formal verification. Co-authored research refining bounds for Neil Sloane's lollipop problem, developed exact results and computational certificates for OEIS sequences, and formalized mathematical results in Lean 4/Mathlib with an emphasis on reproducibility and verifiability. [Preprint](#)  [OEIS Work](#)  [Lean Code](#) 

RESEARCH EXPERIENCE (CONTINUED)

Research Assistant, SBILab, IIIT-Delhi

March 2025 – August 2025

ECG-based Subclass Classification for Cardiovascular Diseases

Transformers

Signal Processing

Deep Learning

Worked with Prof. Anubha Gupta on ECG-based subclass classification for CVD. Also worked on a review of risk calculators for Multiple Myeloma staging.

Research Intern, IISc Physics Dept.

June 2024 – July 2024

Predicting Binding Affinities in Antibody-Antigen Binding

ChimeraX

Modeller

GROMACS

Under the guidance of Prof. Prabal Maiti at the Department of Physics, IISc, conducted research on predicting binding affinities in antibody-antigen binding. Employed ChimeraX, Modeller, and GROMACS to analyze protein structures and simulate molecular interactions.

Predicting Efficacy of Antiseizure Medication Treatment with Machine Learning Algorithms in North Indian Population

2023–2024

Mahima Kaushik, Siddhartha Mahajan, Bibhu Biswal et al.

scikit-learn

Pandas

Data Analysis

Epilepsy Research, Volume 205, September 2024. ISSN: 0920-1211 [10.1016/j.eplepsyres.2024.107404](https://doi.org/10.1016/j.eplepsyres.2024.107404) [↗](#)

Conducted predictive modeling for anti-epileptic drug outcomes using patient data. Utilized six machine learning algorithms for prediction and classification. Achieved over 70% accuracy in predicting drug responses.

Student Researcher, CIC, DU

2023–2024

Disease–Disease SNP Network Analysis and Community Detection

NetworkX

Matplotlib

PyVis

Analyzed disease networks using SNP data. Built graphs with NetworkX, applied Louvain and SLPA for community detection, and visualized results using Matplotlib and PyVis.

PROJECTS

CAISc 2026 Verifier API

FastAPI

PostgreSQL

Docker

Built an automated verification and scoring API for five problems in mathematics, computer science, and computational biology for the CAISc 2026 Verifiable Problems Track. Implemented constraint validation, scoring, and percentile rankings with FastAPI, PostgreSQL, and Docker. [GitHub Repo](#) [↗](#)

Reinforcement Learning Pipeline

LoRA

GRPO

Qwen

Applied supervised fine-tuning and GRPO with LoRA on Qwen using MedMCQA. Implemented reward shaping for answer correctness and structured outputs, and evaluated baseline, SFT, and GRPO checkpoints. [GitHub Repo](#) [↗](#)

PROFESSIONAL ACTIVITIES

Organizing Committee, CAISc 2026 – Conference for AI Scientists

2026

Helped organize CAISc 2026, a research conference focused on AI-assisted scientific discovery. Contributed to the Verifiable Problems Track and supporting infrastructure for evaluating submissions across problems in mathematics, computer science, and computational biology. [Conference Website](#) [↗](#)

Mentor/Speaker – DYOD Induction Program, Jammu University

September 2023

Served as a mentor and speaker for the inaugural Design Your Own Degree (DYOD) cohort. Mentored students on project development, presentations, creative problem solving, and interdisciplinary learning.