Md Sifat Hossain

+880-1828-102576 sifatb910@gmail.com LinkedIn GitHub Google Scholar Website

Research Interests

LLM-Based Code Generation and Automated Reasoning · Multi-Agent Systems for Autonomous Software Engineering · Benchmarking and Evaluation of Foundation Models · AI Alignment and Reinforcement Learning from Human Feedback (RLHF)

Education

University of Dhaka January 2020 – February 2025

Bachelor of Science in Computer Science and Engineering (CGPA: 3.13/4.00)

- **Thesis:** *A Hybrid LLM Feedback Framework for Automated Competitive Programming Workflows*
  Proposed a novel test-driven iterative benchmarking framework integrating multiple LLMs (OpenAI o3-mini, DeepSeek, Qwen) with Codeforces-based validation to evaluate and improve automated code generation and error correction in competitive programming contexts.
  Advisor: Md. Fahim Arefin, Lecturer, Dept. of CSE, University of Dhaka.
- **Relevant Coursework:** Data Structures and Algorithms, Object-Oriented Programming, Software Design Patterns, Artificial Intelligence, Probability & Statistics, Theory of Computation, Machine Learning, Natural Language Processing, Compiler Design, Operating Systems, Database Management Systems, Software Engineering

Publications

[J1] Anika Tabassum*, **Md Sifat Hossain***, Md. Fahim Arefin, Tariqul Islam, and Tarannum Shaila Zaman.
"A-ProS: Towards Reliable Autonomous Programming Through Multi-Model Feedback."
ACM Transactions on Software Engineering and Methodology (**TOSEM**), 2026.
`DOI: 10.1145/3820891` · `[PDF]` · `[Project Page]` · `[Code]`
*Equal contribution. Autonomous agentic framework separating solution generation (GPT-4/GPT-5) from specialized debugging feedback (DeepSeek-R1, Llama-3.3, Codestral) via persistent multi-model feedback loops. Evaluated on 367 ICPC and Codeforces problems; stateful refinement achieves 2.2–2.3× greater gains than stateless baselines and reduces error repetition by 2.9–3.5×.

[C1] **Md Sifat Hossain**, Anika Tabassum, Md. Fahim Arefin, and Tarannum Shaila Zaman.
"LLM-ProS: Analyzing Large Language Models' Performance in Competitive Problem Solving."
In *Proceedings of the 1st ACM International Workshop on Large Language Models for Code* (**LLM4Code**), co-located with ICSE 2025 – 47th International Conference on Software Engineering, Ottawa, Canada, pp. 80–87.
`DOI: 10.1109/LLM4Code66737.2025.00015` · `[PDF]` · `[Code]`
Novel evaluation framework benchmarking five state-of-the-art LLMs on 166 ICPC World Finals problems (2011–2024), revealing that o1 models significantly outperform general-purpose models due to chain-of-thought reasoning and iterative refinement capabilities.

Research Experience

Research Assistant August 2023 – Present

Data Mining Research Lab, Dept. of CSE, University of Dhaka

- **Supervisors:** Md. Fahim Arefin, Lecturer, Dept. of CSE, University of Dhaka; in collaboration with Tarannum Shaila Zaman, Dept. of Information Systems, University of Maryland, Baltimore County (UMBC).
- Designed and implemented **LLM-ProS**, a novel evaluation framework for benchmarking LLM performance on ICPC World Finals problems. Curated a dataset of 166 problems (2011–2024), developed automated submission pipelines via Codeforces Gym, and conducted systematic analysis of five state-of-the-art models across correctness, resource utilization, and chain-of-thought reasoning metrics. Published at **ICSE 2025 (LLM4Code Workshop)**.
- Extended this work into **A-ProS**, an autonomous multi-model agentic framework that separates solution generation from specialized debugging feedback under a 2×3 factorial design. Developed the full orchestration pipeline, Codeforces browser automation (Selenium + Playwright), verdict capture system, and SQLite logging infrastructure. Conducted ablation studies on persistent vs. stateless context and trust calibration analysis (ECE) across critic models. Accepted at **ACM TOSEM 2026**.
- Completed undergraduate thesis titled *"A Hybrid LLM Feedback Framework for Automated Competitive Programming Workflows,"* proposing an iterative test-driven benchmarking pipeline integrating OpenAI o3-mini with specialist LLMs (DeepSeek, Qwen) for error diagnosis and code refinement via Codeforces-based validation.

RLHF Data Researcher & Pod Lead July 2025 – December 2025

Turing Enterprises Inc. *Remote*

- Contributed to **Reinforcement Learning from Human Feedback (RLHF)** data creation initiatives supporting large-scale AI model alignment research, focusing on dataset quality, consistency, and annotation methodology for code and reasoning tasks.
- Led a team of 10, establishing quality control protocols that ensured annotation reliability across diverse task domains.
- Coordinated cross-functional team activities and maintained consistency standards critical to downstream model training, directly supporting AI alignment objectives.

Machine Learning Research Intern August 2024 – September 2024

Brainwave Matrix Solutions *Remote, India*

- Developed a fraud detection model applying anomaly detection and supervised learning techniques on imbalanced datasets, achieving 85% precision. Investigated model behavior under class imbalance and evaluated trade-offs between precision and recall in high-stakes classification settings.
- Automated model training and deployment pipelines using Docker and Jenkins, enabling reproducible ML experimentation and continuous integration of model updates.

Research Software & Artifacts

LLM Benchmark Data Pipeline · [Code] Python, Selenium, BeautifulSoup, SQLite3

- Engineered the full data pipeline underlying the LLM-ProS and A-ProS publications: scraped 166 ICPC World Finals problems, normalized LaTeX/HTML content, and structured problem components into a consistent format for automated LLM prompt construction and verdict logging. Extended to 200+ Codeforces problems, forming the 367-problem A-ProS benchmark.
- Built SQLite-backed storage for per-attempt metadata (verdict, runtime, memory, iteration count), enabling reproducible experimental analysis across all model configurations.

TikTok Scraper · [Code] Python, Selenium, Requests, BeautifulSoup, SQLite3

- Engineered a scraper extracting video descriptions and author metadata for specified keywords and tags, enabling structured analysis across 5000+ videos. Same scraping architecture later reused in the LLM benchmark pipeline.

Industry Experience

Therap BD Ltd April 2025 – Present
Software Engineer *Dhaka, Bangladesh*

- Develop and maintain scalable features for Therap's EHR SaaS platform (used across 50 US states) using Java, Spring, Hibernate, JSP, and Oracle DB, with a focus on correctness and reliability under HIPAA compliance constraints.
- Build and containerize full-stack modules using ReactJS and Docker, deployed on WebLogic Server, contributing to platform stability and consistent delivery across multiple environments.

Zeroxa DT March 2023 – June 2024
Software Engineer (Part-time) *Remote, London, UK*

- Built and deployed scalable web applications using React.js and FastAPI; architected CI/CD pipelines with automated testing on AWS (EC2, S3, RDS).

Honors & Awards

- **Zelf Hackathon 2.0** – Honorable Mention, Scraping Engineer track.
- **ICPC Dhaka Regional Onsite** – Top-35 of 220+ teams (2023); Top-49 of 309 teams (2024).
- **BUET Inter-University Programming Contest 2023** – 5th place among 102 university teams.
- **Samsung R&D BD Coding Contest 2024** – Final Round qualifier (rank 55 of 908 in Round 1).
- **Meta Hacker Cup 2024** – Round 2 qualifier (global rank 2166).
- **NCPC 2023** – 44th of 198 teams.
- **Competitive Programming** – Solved 1000+ problems across Codeforces (max. rating 1603, *Expert*), CodeChef (4★, max. rating 1921), AtCoder, and LeetCode.

Technical Skills

Languages: Python, Java, C++, JavaScript, TypeScript, Kotlin
Frameworks & Libraries: Spring Boot, Hibernate, FastAPI, React.js, Next.js, JSP, Selenium, Playwright, Docker, Nginx
Databases: PostgreSQL, Oracle DB, MongoDB, SQLite
Research & ML Tools: PyTorch, HuggingFace Transformers, OpenAI API, Pandas, NumPy, Jupyter, LaTeX
Developer Tools: Git, Linux, Bash, Jenkins, JUnit5, WebLogic Server

Service & Leadership

- **Vice-President, ICT and Graphics**, Notre Dame English Club, Notre Dame College, Dhaka.
  Organized the 6th National English Carnival (10,000+ participants, 30+ events); previously co-ordinated the 5th edition (5,000+ participants). Led judge coordination, external communications, and sponsorship outreach.
- **Debater**, Gregorian Debating Club. Participated in school-level debate competitions.

Selected Software Projects

Smart Event Ticketing System Java, Spring MVC, Hibernate/JPA, PostgreSQL, JSP

- Multi-role event management platform with pessimistic locking for transactional consistency; real-time event filtering with asynchronous data retrieval.

JobGenie · [Code] React.js, FastAPI, MongoDB

- Job search platform with automated CV generation and personalized job matching via live job scraping.

OyeAmigo · [Code] Kotlin, Android SDK

- Personality-based social networking Android app with null-safe Kotlin architecture.

References

Md. Fahim Arefin fahim@cse.du.ac.bd
Lecturer, Dept. of Computer Science and Engineering, University of Dhaka
Supervised my undergraduate thesis and both publications (LLM-ProS, A-ProS).

Tarannum Shaila Zaman zamant@umbc.edu
Associate Professor, Dept. of Information Systems, University of Maryland, Baltimore County (UMBC)
Research collaborator on LLM-ProS and A-ProS.