Md Sifat Hossain

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

Research Interests

Large Language Models for Code Generation and Automated Reasoning · Agentic AI Systems and Multi-Model Feedback Frameworks · Autonomous Software Engineering · Competitive Programming as LLM Evaluation Benchmark · 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.
- **Relevant Coursework:** Data Structures and Algorithms (C++), Object-Oriented Programming (Java), Software Design Patterns (Java), Artificial Intelligence (Python), Probability & Statistics, Theory of Computation, Machine Learning, Natural Language Processing, Compiler Design, Operating Systems, Database Management Systems, Software Engineering

Publications

Md Sifat Hossain, Anika Tabassum, Md. Fahim Arefin, Tarannum Shaila Zaman.
"LLM-ProS: Analyzing Large Language Models' Performance in Competitive Problem Solving"
LLM4Code 2025 Workshop, ICSE 2025 – 47th International Conference on Software Engineering, Ottawa, Canada. pp. 80–87
`arXiv:2502.04355`
Proposed LLM-ProS, a novel evaluation framework benchmarking five state-of-the-art LLMs (GPT-4o, Mistral Large, Llama-3.1-405B, o1-mini, o1-preview) 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.

Anika Tabassum*, **Md Sifat Hossain***, Md. Fahim Arefin, Tariqul Islam, Tarannum Shaila Zaman.
"A-ProS: Towards Reliable Autonomous Programming Through Multi-Model Feedback"
Accepted at **ACM Transactions on Software Engineering and Methodology (TOSEM)**, 2026.
`arXiv:2605.18073 [Project Page] [Code]`
*Equal contribution. Introduced A-ProS, an autonomous agentic framework that separates solution generation (GPT-4/GPT-5) from specialized debugging feedback (DeepSeek-R1, Llama-3.3, Codestral) using 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×.

Research Experience

Research Assistant *January 2024 – February 2025*
Data Mining Research Lab *Dept. of Computer Science and Engineering, University of Dhaka*

- **Supervisor:** Md. Fahim Arefin, Lecturer, Dept. of CSE, University of Dhaka; in collaboration with Prof. 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. It got 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 researchers to design and evaluate high-quality training data pipelines, 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.

Industry Experience

Therap Services LLC *April 2025 – Present*
Software Engineer *Dhaka, Bangladesh*

- Developed and maintained 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.
- Built and containerized 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 for clients using **React.js** and **FastAPI**, significantly reducing average page load times through performance optimization.
- Architected **CI/CD pipelines** with automated testing and deployment workflows on **AWS** (EC2, S3, RDS), accelerating release cycles while maintaining production code quality.

Projects

Research & Data Projects

LLM Benchmark Data Pipeline *Python, Selenium, BeautifulSoup, SQLite3*

- *Research question: How can ICPC World Finals problems be systematically collected, cleaned, and structured as a reproducible LLM evaluation benchmark?* Engineered the full data pipeline underlying the **LLM-ProS** and **A-ProS** publications — scraping 166 ICPC problems, normalizing LaTeX/HTML content, and structuring problem components (statements, I/O specs, constraints, sample tests) into a consistent format for automated LLM prompt construction and verdict logging.
- Built SQLite-backed storage for per-attempt metadata (verdict, runtime, memory, iteration) and extended the pipeline to support Codeforces contest problems (200 additional problems), forming the 367-problem benchmark used in A-ProS.

TikTok Scraper *Python, Selenium, Requests, BeautifulSoup, SQLite3*

- *Research question: How can large-scale social media metadata be collected efficiently for downstream NLP and content analysis?* Engineered a scraper extracting video descriptions and author metadata for specified keywords and tags, enabling structured data analysis across **5000+ videos**.
- Used Selenium for dynamic content rendering, BeautifulSoup for HTML parsing, and SQLite3 for structured storage and query — the same scraping architecture later applied in the LLM benchmark pipeline.

Software Projects

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

- Architected a multi-role event management platform with pessimistic locking mechanisms to guarantee transactional consistency and prevent concurrent booking conflicts at scale.
- Engineered real-time event filtering with asynchronous data retrieval, improving search responsiveness and overall user experience.

JobGenie *React.js, FastAPI, MongoDB*

- Developed a job search platform with automated CV generation and personalized job matching, integrating live job scraping to fetch and rank relevant listings.

OyeAmigo *Kotlin, Android SDK*

- Built a personality-based social networking Android app with optimized null-safe Kotlin architecture, reducing crash rates and improving runtime stability.

Technical Skills

Languages: Python, Java, C++, JavaScript, Kotlin
Frameworks & Libraries: Spring Boot, Hibernate, FastAPI, React.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

Achievements

- **Zelf Hackathon 2.0**: Received **Honorable Mention** in the Scraping Engineer track.
- **Competitive Programming**: Solved 1000+ problems across Codeforces (max. rating 1603, *Expert*), CodeChef (4★), and AtCoder; ICPC Dhaka Regional top-35 (2023, 220+ teams) and top-49 (2024, 309 teams); 5th place BUET Inter-University Programming Contest 2023 (102 teams); Samsung R&D BD Coding Contest 2024 Final Round qualifier (rank 55/908); Meta Hacker Cup 2024 Round 2 qualifier (rank 2166 globally).