Antu Saha

Ph.D. Candidate in Computer Science, William & Mary • Available from June 2027
asaha02@wm.edu • antu-saha.github.io • Google Scholar • GitHub • LinkedIn

SUMMARY

AI for Software Engineering (AI4SE) researcher who builds and evaluates **LLM-based tools** and **AI coding agents** for bug reporting, bug localization, and program repair. Designs **benchmarks and controlled experiments** for agents (over 7,500 runs), builds tools that ground LLMs in **app execution data**, and analyzes why agents succeed or fail. Six peer-reviewed papers at top SE venues (*e.g.*, ICSE and ISSTA) and eight more under review (two received major revisions), including two on AI coding agents.

SKILLS

- **AI coding agents & evaluation**: running and evaluating coding agents (CODEX, CLAUDE, QWEN, GEMINI), agent trajectory and failure analysis, benchmark and dataset construction, LLM-as-a-judge
- **LLMs & ML**: GPT, CLAUDE, LLAMA, BERT, and multimodal models; prompting, embeddings, and fine-tuning (LLMs, ROBERTA, CLIP, BLIP, SBERT)
- **Software engineering**: dynamic analysis of Android apps, GUI testing, bug reproduction and localization, program repair, code review, issue tracking (GitHub, Bugzilla), repository mining, empirical studies (qualitative analysis, developer interviews, user studies)
- **Languages & tools**: Python, Java, C++, R, SQL, Bash; PyTorch, Hugging Face, scikit-learn, TensorFlow; LangGraph, FastAPI, Redis; Docker, Kubernetes, Git; Android emulators, UiAutomator

RESEARCH PROJECTS

Graduate Research Assistant, William & Mary (advisor: Dr. Oscar Chaparro) Jun 2023 – Present

- Led and contributed to projects with researchers at the University of Central Florida, George Mason University, the University of Minnesota, and the University of Dhaka.
- **Bug report content and AI coding agents** (lead; under review at FSE'27) Designed the first controlled study of agent repair for Android app bugs, with over 7,500 repair attempts by three agents (CODEX, CLAUDE, QWEN) on about 50 bugs in over a dozen report configurations. Steps to reproduce helped all agents, UI hierarchies were the most useful GUI information, and code localization cut the median repair cost by over 30%. *Built:* a harness that runs agent CLIs headless in sandboxed app copies, validates patches with UI tests on Android emulators, and tracks tokens and cost per attempt; mixed-effects models in R.
- **Agent patch analysis** (co-author; under review at FSE'27) Validated agent patches and analyzed the root causes of failures across hundreds of attempts. About three-quarters of attempts passed the tests and one in seven matched the developer's fix. Nearly all other passing patches were valid alternative fixes.
- **BUGSCRIBE: LLM-based bug report enhancement** (lead; under review at MSR'27) paper artifact Developed a technique that grounds LLM reasoning in app execution data to enhance incomplete bug reports. Improved steps-to-reproduce quality by up to 84.7% and observed/expected behavior quality by up to 35.2%, and raised an LLM agent's bug reproduction rate by up to 57.7%. *Built:* a five-stage LLM pipeline (GPT and CLAUDE APIs) over app execution graphs with per-phase token and runtime tracking, and a bug reproduction runner on Mobilerun, an open-source LLM framework for Android devices, that records each run's trajectory and device logs.
- **BURT++: conversational bug-reporting agent** (lead; in preparation) code A user describes a bug in plain language, and the agent asks targeted follow-up questions grounded in the app's GUI graph until each report element is resolved. *Built:* a full-stack agent with React, FastAPI, LangGraph, Redis, and Docker Compose, with per-run logs and an LLM-as-a-judge evaluation against human ground truth.
- **ASTROBR: assessing reproduction steps** (co-first author; ICPC'25) paper artifact Combined LLMs with dynamic app analysis to assess steps to reproduce and identify missing steps, outperforming the state of the art by 25.2% and 71.4%.
- **Buggy UI localization from bug descriptions** (lead; ISSTA'24) paper artifact Introduced a new task and built a benchmark of 228 bug descriptions from 39 Android apps; evaluated text and multimodal models (*e.g.*, SBERT, CLIP), and used the localized screens to rank buggy code up to 12% more accurately.
- **Issue resolution in practice: Mozilla Firefox** (lead; ICSE'25; W&M research grant) paper dataset Led a team of seven researchers that analyzed 356 Firefox issue reports, built a model of the resolution process and a catalog of 47 resolution patterns, and validated the findings with Mozilla developers.
- **Developer tools**: INCEPTOR, a dashboard for exploring coding agent trajectories (ICSE'27 demo, under review; designed the tool, its annotation codebook, and its evaluation) live demo, and SPRINT, an LLM-powered GitHub assistant for issue report management (MSR'25) paper code

- **Finding solutions in issue discussions** (lead; major revision at EMSE) paper code Compared traditional ML models, pre-trained language models, and LLMs in 68 configurations on 4,917 annotated comments; a fine-tuned LLM found 83.8% of solution-related comments, and a fine-tuned RoBERTa model was nearly as accurate and took minutes rather than days to fine-tune. *Built:* QLoRA fine-tuning of Llama-3 (4-bit, Hugging Face PEFT) as Kubernetes GPU jobs, and Llama-3 70B prompting with Ollama.

EDUCATION

Ph.D., Computer Science, William & Mary, Virginia, USA — Jan 2022 – May 2027 (expected)
- Dissertation proposal: Studying and Automating Issue Resolution to Ensure Software Quality
- CGPA: 3.96/4.00

B.Sc., CSE, Rajshahi University of Engineering & Technology, Bangladesh — Mar 2013 – Dec 2017
- 3rd among 117 students (CGPA: 3.82/4.0)

SELECTED PUBLICATIONS

Peer-Reviewed — 82 citations on Google Scholar

- [ICSE'25] **A. Saha** and O. Chaparro, "Decoding the Issue Resolution Process in Practice via Issue Report Analysis: A Case Study of Firefox."
- [ISSTA'24] **A. Saha**, Y. Song, J. Mahmud, Y. Zhou, K. Moran, and O. Chaparro, "Toward the Automated Localization of Buggy Mobile App UIs from Bug Descriptions."
- [ICPC'25] J. Mahmud*, **A. Saha***, O. Chaparro, K. Moran, and A. Marcus, "Combining Language and App UI Analysis for the Automated Assessment of Bug Reproduction Steps." ***(*co-first authors)***
- [ICSME'26] A. Adnan, M. Rahman, **A. Saha**, and O. Chaparro, "Rethinking Issue Resolution for AI/ML Systems."
- [MSR'25] A. Adnan, **A. Saha**, and O. Chaparro, "Sprint: An Assistant for Issue Report Management."
- [ICSE'26 DS] **A. Saha**, "Studying and Automating Issue Resolution for Software Quality."

Under Review

- [EMSE] **A. Saha**, M. Sun, and O. Chaparro, "Evaluating Language Model Applications for Identifying Solution-Related Content in Issue Report Discussions." ***(major revision)***
- [TOSEM] M. Sun, **A. Saha**, N. De Silva, A. Mastropaolo, and O. Chaparro, "Recovering Fine-Grained Code Change Rationale from Multiple Software Artifacts." ***(major revision)***
- [FSE'27] **A. Saha**, S. Pandey, A. Marcus, K. Moran, and O. Chaparro, "An Empirical Study of Bug Report Content for AI Coding Agents on Mobile Apps."
- [FSE'27] S. Pandey, J. Wu, J. Mahmud, N. De Silva, **A. Saha**, A. K. Dipongkor, O. Chaparro, M. Fazzini, and K. Moran, "An Analysis of AI Agent Patches for Mobile App Bugs."
- [MSR'27] **A. Saha**, A. K. Dipongkor, S. Bennett, K. Moran, A. Marcus, and O. Chaparro, "Automatically Enhancing the Quality of Android App Bug Reports."
- [ICSE'27 Demo] M. Hasan, **A. Saha**, K. Moran, and O. Chaparro, "INCEPTOR: A Trajectory Exploration Tool for Program Repair Agents."
- [ICSE'27 Demo] T. Idobe, M. Sun, **A. Saha**, N. De Silva, A. Mastropaolo, and O. Chaparro, "ArgusVS: Recovering and Exploring Code Change Rationale in the IDE."
- [MSR'27] M. Hasan, **A. Saha**, and O. Chaparro, "SE AI Agent Index: A Dataset of Software Engineering Practices for Agentic AI Systems."

TEACHING & MENTORING

- Teaching assistant at W&M (Data Structures, Software Development, Algorithms); four years as a lecturer (instructor of record) in Bangladesh, teaching programming and machine learning.
- Mentored 18 students, including six Ph.D. and two M.Sc. students at W&M.
- Graduate International Peer Leader at W&M, helping new international graduate students settle in.

SERVICE & AWARDS

- Program committee member, MSR'26; session chair, ICSE'26; reviewer for ICSE, ICSME, ESEC/FSE, SANER, ICPC, EMSE, and ACM Computing Surveys.
- Presented papers at ICSME'26, ICSE'25, ICPC'25, MSR'25, and ISSTA'24.
- Grants: W&M Graduate Research Seed Grant (2024); NSF and ACM SIGSOFT CAPS travel grants.
- Student volunteer at ICSE (2024–2026), ESSEC'24, and ICSME'23.
- Web chair, W&M CS Symposium for Graduate Studies (2023).
- International Student Achievement Award, William & Mary (2025).
- Best Poster Award, W&M CS Symposium for Graduate Studies (2023 and 2024).