Sumon Biswas

Postdoctoral Researcher at Carnegie Mellon University Software and Societal Systems (S3D), School of Computer Science

Computer science researcher with a goal to innovate in the intersection of Software Engineering (SE) and Artificial Intelligence (AI). With a particular focus on the *SE for AI* area, I improved algorithmic fairness and safety in ML-based software. By adopting both formal and empirical approaches, working towards design, verification, and analysis of learning-enabled software systems for fairness and safety, and aiming to ensure their sustainability.

Research Interest

Software Engineering Verification, Design and Architecture, Empirical Software Engineering Artificial Intelligence Algorithmic Fairness, Auto ML, DNN Verification, ML Technical Debts SE for AI Sustainable AI, Causal Reasoning, Trustworthiness and Safety

Academic Qualifications

Postdoctoral Researcher
Carnegie Mellon University
2022– p.

Advisor: Dr. Eunsuk Kang

Ph.D. in Computer Science

Ames, IA

10wa State University

2016–2022

Iowa State University **Advisor:** Dr. Hridesh Rajan

Dissertation: Understanding and Reasoning Fairness in Machine Learning Pipelines

M.S. in Computer Science
Iowa State University
Ames, IA
2016–2018

Advisor: Dr. Hridesh Rajan

Thesis: Understanding Unfairness and its Mitigation in Open-Source Machine Learning Models

B.S. in Information Technology Dhaka, Bangladesh

Jahangirnagar University **Advisor:** Dr. Shamim Al Mamun

Thesis: Cloud Based Healthcare Application Architecture and Electronic Medical Record Mining

Employments

Carnegie Mellon University Pittsburgh, PA

Postdoc at Software and Societal Systems Department (S3D), School of Computer Science ★ May 2022 − Cont.

Projects: Model based Programming of High-Assurance Systems (DARPA), Sustainable Software Design (NSF)

Iowa State University Ames, IA

Research Assistant, Laboratory of Software Design A May 2018 – May 2022

Projects: Dependable Data-Driven Discovery (NSF TRIPODS), Big-Code Mining & Analysis (NSF), ML Repair

Iowa State University Ames, IA

Teaching Assistant, Department of Computer Science ♠ August 2016 – May 2018

Courses taught: • Software Development Practices • Advanced Programming Techniques in C & C++

2011-2015

Bangladesh University of Business and Technology

Dhaka, Bangladesh

Lecturer, Department of Computer Science & Engineering

January 2016 – July 2016

Courses instructed: Computer and Programming • Structured Programming Language • Pattern Recognition

Publications

Peer Reviewed Conference

The marked (g) authors are graduate students mentored by me. The (*) marked authors have equal contributions.

- [1] **Sumon Biswas**, Yining She^g and Eunsuk Kang. Towards Safe ML-based Systems in Presence of Feedback Loops. To appear in *Proceedings of the 1st International Workshop on Dependability and Trustworthiness of Safety-Critical Systems with Machine Learned Components*, **SE4SafeML 2023**, page 1–4, 2023.
- [2] Giang Nguyen^g, **Sumon Biswas** and Hridesh Rajan. Fix Fairness, Don't Ruin Accuracy: Performance Aware Fairness Repair using AutoML. In *Proceedings of the 31st ACM Joint Meeting on European Software Engineering Conference and Symposium on the Foundations of Software Engineering*, **ESEC/FSE 2023**, page 1–13, 2023. [Preprint] [Artifact]
- [3] **Sumon Biswas** and Hridesh Rajan. Fairify: Fairness Verification of Neural Networks. In *Proceedings of the 45th IEEE/ACM International Conference on Software Engineering*, **ICSE 2023**, page 1546-1558, 2023. [DOI] [Preprint] [Artifact]
- [4] Usman Gohar^g, **Sumon Biswas** and Hridesh Rajan. Towards Understanding Fairness and its Composition in Ensemble Machine Learning. In *Proceedings of the 45th IEEE/ACM International Conference on Software Engineering*, **ICSE 2023**, page 1533-1545, 2023. [DOI] [Preprint] [Artifact]
- [5] David OBrien^g, **Sumon Biswas**, Sayem Mohammad Imtiaz, Rabe Abdalkareem, Emad Shihab and Hridesh Rajan. 23 Shades of Technical Debt: An Empirical Study on Machine Learning Software. In *Proceedings of the 30th ACM Joint Meeting on European Software Engineering Conference and Symposium on the Foundations of Software Engineering, ESEC/FSE 2022, page 734-746, 2022. [DOI] [Preprint] [Artifact]*
- [6] **Sumon Biswas**, Mohammad Wardat and Hridesh Rajan. The Art and Practice of Data Science Pipelines: A Comprehensive Study of Data Science Pipelines In Theory, In-The-Small, and In-The-Large. In *Proceedings of The 44th International Conference on Software Engineering*, **ICSE 2022**, page 2091-2103, 2022. [DOI] [Preprint] [Artifact]
- [7] **Sumon Biswas** and Hridesh Rajan. Fair preprocessing: Towards understanding compositional fairness of data transformers in machine learning pipeline. In *Proceedings of the 29th ACM Joint Meeting on European Software Engineering Conference and Symposium on the Foundations of Software Engineering, ESEC/FSE 2021, page 981–993, 2021. [DOI] [Preprint] [Artifact]*
- [8] **Sumon Biswas** and Hridesh Rajan. Do the machine learning models on a crowd sourced platform exhibit bias? an empirical study on model fairness. In *Proceedings of the 28th ACM Joint Meeting on European Software Engineering Conference and Symposium on the Foundations of Software Engineering*, **ESEC/FSE 2020**, page 642–653, 2020. [DOI] [Preprint] [Artifact] [Presentation]
- [9] Sumon Biswas, Md Johirul Islam, Yijia Huang, and Hridesh Rajan. Boa meets python: A boa dataset of data science software in python language. In 2019 IEEE/ACM 16th International Conference on Mining Software Repositories, MSR 2019, page 577–581, 2019. [DOI] [Preprint]
- [10] **Sumon Biswas**, M. S. Kaiser, and S. A. Mamun. Applying ant colony optimization in software testing to generate prioritized optimal path and test data. In 2015 International Conference on Electrical Engineering and Information Communication Technology (ICEEICT), pages 1–6, 2015. [DOI]
- [11] Manan Binth Taj Noor and **Sumon Biswas**. A secure data security infrastructure for small organization in cloud computing. In 2015 International Conference on Electrical Engineering and Information Communication Technology (ICEEICT), pages 1–6, 2015. [DOI]

[12] Sumon Biswas, Anisuzzaman, Tanjina Akhter, M. S. Kaiser, and S. A. Mamun. Cloud based healthcare application architecture and electronic medical record mining: An integrated approach to improve healthcare system. In 2014 17th International Conference on Computer and Information Technology (ICCIT), pages 286–291, 2014. [DOI]

Under Review

- [1] David OBrien^g, **Sumon Biswas**, Sayem Mohammad Imtiaz, Rabe Abdalkareem, Emad Shihab and Hridesh Rajan. Are Prompt Engineering and TODO Comments Friends or Foes? An Evaluation on GitHub Copilot, *Submitted*, pages 1–14, 2023.
- [2] Rangeet Pan, **Sumon Biswas**, Vu Le, Sumit Gulwani and Hridesh Rajan. Can Program Synthesis Suggest Fixes for Deep Learning Crash Bugs? *Submitted*, pages 1-11, 2022.
- [3] Sayem Mohammad Imtiaz^g, **Sumon Biswas**, Shibbir Ahmed and Hridesh Rajan. How Do Deep Learning Models Evolve? A Comprehensive Study of the Evolution of Deep Learning Model Code in Open Source Repositories, *Submitted*, pages 1-12, 2023.
- [4] Rangeet Pan, **Sumon Biswas**, Mohna Chakraborty^g, Breno Dantas Cruz and Hridesh Rajan. An Empirical Study on Bugs Found while Reusing Pre-trained Models in Natural Language Processing, *Submitted*, pages 1-12, 2022.

Honors & Awards

0	Distinguished Reviewer : Selected as <i>Distinguished Reviewer</i> for the journal ACM Transactions on Software Engineering and Methodology (TOSEM) ☑	2022
0	Research Excellence Award: Awarded by Iowa State University recognizing outstanding research	2022
0	ACM CAPS Award : Awarded by ACM SIGSOFT to attend in-person ICSE 2023 conference held at Pittsburgh, PA	2022
0	Publication Award : Awarded by Computer Science Department at Iowa State University for publication in the top-tier venues in consecutive years	2021
0	ACM CAPS Award: Awarded by ACM SIGSOFT to attend ESEC/FSE conference	2021
0	Panelist at ESEC/FSE′20: Selected as a panelist in the session on Fairness at ESEC/FSE 2020 ☑	2020
0	PLMW Scholar in PLDI'19 : Awarded by Programming Language Mentoring Workshop (PLMW) at the PLDI conference at Phoenix, AZ	2019
0	Professional Advancement Grants : Awarded by Graduate and Professional Student Senate at Iowa State University	2019
0	ACM Travel Grant : Awarded to attend IMA Workshop organized by Institute for Mathematics & its Applications (IMA), University of Minnesota \Box	2019

Services

Journal Reviewer.

Publons Web of Science reviewer profile

- o **TOSEM**: Serving in the *Board of Distinguished Reviewer* for ACM Transactions on Software Engineering and Methodology (TOSEM) ☑ 2021-p
- \circ **TSE**: Serving as the reviewer for IEEE Transactions on Software Engineering (TSE) \square 2021-p
- o **EMSE**: Serving as the reviewer for Springer Journal Empirical Software Engineering (EMSE) ☑ 2022-*p*

Conference Program Committee (PC).....

- o ICSE'24 Technical Track: Served in the program committee of ICSE 2024 Technical Track ☑ 2023
- o **ASE'23 Technical Track**: Served in the program committee of ASE 2023 Technical Track ☑ 2023
- o ASE'23 Industry Showcase: Served in the Industry Showcase track of ASE 2023, Luxembourg 🗗 2023

0	ASE'22 - Industry Showcase : Served in the Industry Showcase track of ASE 2022, Oakland, MI ☑	2022
0	ASE'22 - LBR: Served in the Late Breaking Results (LBR) track of ASE 2022, Oakland, MI ☑	2022
0	ESEC/FSE′22 - DS : Serving in the Program Committee of Doctoral Symposium track of ESEC/FSE 2022, Singapore ☑	2022
0	MSR'21 - Technical Track : Shadow PC member of the Technical Track of International Conference on Mining Software Repositories (MSR'21) held at Madrid, Spain ☑	2021
0	OOPSLA'21 - AEC: PC Member of Artifact Evaluation Committee (AEC) at ACM SIGPLAN conference on Object-Oriented Programming, Systems, Languages & Applications (OOPSLA) held at Chicago, IL ☑	2021
C	Organizing Committee	
0	WiDS Ambassador : Served as the Ambassador of Women in Data Science (WiDS) organized by Stanford University. I organized and promoted the WiDS event at Iowa State University as part of the annual WiDS Worldwide Conference	2022
0	SPLASH'21 (Hybrid): Served as the Accessibility Chair of the ACM SIGPLAN conference on Systems, Programming, Languages, and Applications (SPLASH 2021) with OOPSLA and REBASE 🗷	2021
0	Web Chair : Served as the web chair of the Midwest Big Data Summer School, organized by Iowa State University, May 17-20, 2021 ☑	2021
0	SPLASH'20 (Virtual): Served as the Accessibility Chair of the ACM SIGPLAN conference on Systems, Programming, Languages, and Applications (SPLASH 2020) with OOPSLA, ECOOP and REBASE 🗷	2020
F	Research Experience	

DARPA V-SPELLS

Carnegie Mellon University | 2022 - 2024

Verified Security and Performance Enhancement of Large Legacy Software

Working in the synergetic DARPA project to design compositional DSLs and check safety and liveness of Unmanned Aerial Vehicles (UAV). In particular, building model extraction technique from legacy source code and leverage that to verify drone systems, e.g., Ardupilot, MAVSDK.

Research Grants

NSF EAGER

Carnegie Mellon University | 2022 – p

Towards a Design Methodology for Software-Driven Sustainability 🗷

While the impact of hardware sustainability is well understood, research on software sustainability is still in an early stage. Through this grant, we aim to propose a framework for sustainability as a *first-class software quality attribute*. Currently, building a simulation platform to generate sustainable design strategies for ML-enabled systems.

NSF TRIPODS

Iowa State University | 2019 – 2022

D4: *Dependable Data-Driven Discovery* **∠**

Contributed to writing the grant proposal and thereafter conducted research towards the goal of this institute-level grant under HDR TRIPODS by NSF. I along with my advisor (PI of the grant), <u>launched</u> the interdisciplinary data science hub at Iowa State.

NSF CNS - Boa

Iowa State University | 2022 – 2024

Enhancing Infrastructure for Studying Software and its Evolution at a Large Scale 🗷

By enabling program analysis on machine learning programs and Jupyter notebooks, I contributed to the grant – <u>Boa</u>: Enhancing Infrastructure for Studying Software and its Evolution at a Large Scale. Built the Python language support for Boa for large-scale software repository mining and analysis.

Facebook Probability and Programming Award

Iowa State University | 2020 – 2021

Big code assisted neural architecture search

Contributed to writing the grant proposal and thereafter my dissertation was partially supported by this industry grant. My work on compositional reasoning of ML pipeline (FSE'21) was funded by the grant.

Research Mentoring.

2022 - p Yining She (he/him)

Carnegie Mellon University (Ph.D. student)

Mentored from his beginning of Ph.D. on ensuring AI fairness and safety in the long-term, from ideation to formal modeling and evaluation. Also, guided him to build the Monte-Carlo simulation framework that enabled sensitivity analysis of properties that endure over time. A position paper on the topic has been accepted in the FSE'23 Workshop – SE4SafeML.

2021 – 2022 Giang Nguyen (he/him)

Iowa State University (Ph.D. student)

Mentored him as a senior Ph.D. student on the research on bias mitigation using AutoML techniques. I envisioned and guided him to leverage Bayesian multi-objective optimization to achieve fairness-accuracy trade-off, which has been accepted in ESEC/FSE'23.

2022 Ruchira Manke (she/her)

Iowa State University (Ph.D. student)

Mentored her on the research on verification of neural networks. I guided her to identify exact research gap and how to harness existing tools towards formal modeling and certification.

2020 - 2022 David OBrien (he/him)

University of Northern Iowa (B.S.) → Iowa State University (Ph.D. student)

First, mentored him as an undergraduate REU student on mining software repositories and analyzing ML programs. His research progress inspired and helped to get admitted into the Ph.D. program. I mentored him further as a Ph.D. student on *analyzing technical debts in machine learning systems* which was accepted in ESEC/FSE 2022, and a followup work is submitted in ICSE 2024.

2021 Mohna Chakraborty (she/her)

Iowa State University (Ph.D. student)

Mentored her on understanding the issues developers face with NLP-based software, and analyzing related bugs from StackOverflow posts. She published a student research competition short-paper under the mentorship in ESEC/FSE 2021 on the *bugs found in pretrained NLP models* [DOI]. Another followup work on the topic Reuse Bugs in NLP Models is under submission.

2020 – 2022 Usman Goher (he/him)

Iowa State University (Ph.D. student)

Mentored him on empirical software engineering research, collecting benchmark, designing experiments, and answering research questions related to *SE for AI*. Especially, I came up with the research direction on Ensemble Fairness and mentored him on that, which has been accepted in ICSE 2023.

2019 – 2020 Senior Design Project Team

Department of Electrical and Computer Engineering, Iowa State University

Mentored a five-member senior design team for their two semesters long research project – "Analysis of GitLab projects using Boa". Team members are: Diego Realpe (team leader), Adrian Hamill, Benjamin Carland, Megan Miller and Yi-Hsien Tan. The goal of the project was to mine source code and metadata from GitLab using the Boa infrastructure. [Project website]

Research Highlights

- 2022 **Data Science Pipeline:** My research explained how data science pipelines are designed in theory and practice. ☑
- 2021 Fair ML: My research identified prevalence of unfairness in machine learning models.
- 2020 **Dependability:** My association and contribution towards the NSF TRIPODS grant has been highlighted. ☑

Teaching Experience

Teaching Assistant

COMS 327 - Advanced Programming Techniques

Class size: \sim 230

Undergraduate level, Required CS course, Iowa State University

Fall 2016, Spring 2017, Fall 2019

Topics Differences between managed (Java) and unmanaged languages (C/C++) • Design and build large programs from specification • Memory management in C and C++ • Templates and standard library • Concurrent and network programming

Responsibilities Assist students with large programming projects and assignments • Debugging and pair programming • Prepare and grade tests • Teach usage of the tools and technologies

Technologies GDB, Valgrind, Ncurses, Build systems, e.g., Make

COMS 309 - Software Development Practices

Class size: \sim 250

Undergraduate level, Required CS course, Iowa State University

Fall 2017, Spring 2018, Fall 2018

Topics Develop complex software in a team: from idea to release • Software development criteria: client-server architecture, relational database, multi-user setting, concurrent features e.g., online chat • Using SE tools, IDE, source-control e.g., Git • SE lifecycle

Responsibilities Weekly lecture (1 hour) in a class section of \sim 30 students on design pattern, version-control, server configuration, database design, etc. • Prepare and record screencasts • Supervise 4-member teams developing software projects throughout the semester

Screencasts Create short videos for the tools and technologies used in the SE lifecycle • Example videos: develop first Android app, how to host project to Linux server, Using MySQL Workbench for to remote database server, etc.

Project Supervised 8 out of ∼55 teams in each semester • Team under my supervision won best project supervision: award in Spring 2017 and Fall 2018 • Weakly meeting: explain software requirements and deliverables, solve problems • Technologies: Android SDK, Spring Boot, .Net, RESTful API, Angular JS, MySQL, Java Socket, MVC, etc.

Project grading Evaluate demonstrations • UI design • Code review: functionality, quality, bugs, etc. • Teamwork • Weakly project reporting

Supervised Fall'18: Business QR, CookBuddy, ISU Service, Twenty One, Dog Matcher, Project X (best project projects: award), Movie App, Campus Connections

Spring'18: HabiTracker, Smart Art, Image Guesser, CyBike, Next Generation 911 (best project award), CyChat, Time Flies

Fall'17: Battle of the Worlds, Pre Park, Run Samurai, CyDisc, Songusoid, Cute and Fluffy

Lecturer

Guest LecturerClass size: \sim 40Iowa State UniversityOctober 2023

Course Responsible AI: Risk Management in Data-Driven Discovery (Graduate course)

Lecture title Assessing and Ensuring Fairness in ML Models

Topics covered Understanding societal bias and AI fairness • Fairness metrics and mitigation • Fairness verification • Sustaining fairness

Primary Instructor
Bangladesh University of Business and Technology

Class size: \sim 50 *January* 2016 - *July* 2016

Responsibilities Lecture • Design courses • Create tests • Conduct programming lab sessions

Courses **1. Computer and Programming Concepts:** Instruct two sections: computer science major and instructed non computer science major

2. Structured Programming Language: Instruct theory and lab sessions

3. Pattern Recognition: Instructor for the seniors in computer science major

Selected Software Tools

Fairify 2023

https://github.com/sumonbis/Fairify

Released first SMT based formal verification tool for Neural Networks (NN) to verify individual fairness properties. The tool takes an imperative NN and given property to perform modular verification. Fairify also supports targeted verification, given user specified population subset and tolerance level (ϵ -fairness).

FairPreprocessing 2022

https://github.com/sumonbis/FairPreprocessing

Developed the first method to measure compositional bias of preprocessing stages and data transformers in ML pipeline. The repository implements four novel group fairness metrics to identify biased components.

DS/Python Miner

https://boa.cs.iastate.edu

Built the large-scale *data science* software repository mining framework using Boa. Leveraged DLTK Python parser to enable program analysis on more than 3 billion AST nodes in the initial dataset. Later mined several datasets listed here, which include mining of Jupyter Notebooks, and were used in several research projects.

Fair AutoML 2022

https://github.com/giangnm58/Fair-AutoML

This is the first tool proposed to leverage AutoML for unfairness repair. While the existing technique mitigate bias in specific scenarios, this tool achieves better fairness-accuracy tradeoff, which shows robustness over various datasets.

DS Pipeline 2021

https://github.com/sumonbis/DS-Pipeline

The artifact contains pipeline architecture benchmark for data science software. The pipelines in-the-small are extracted automatically by program analysis on Kaggle programs. Pipelines in-the-large are extracted manually from a mined GitHub dataset.

Talks

- 1. **DARPA PI Meeting**: Presented "Software Architecture for Unmanned Aerial Vehicles (UAV): Virtual Automatic AADL Model Extraction and Reasoning" in the DARPA V-SPELLS PI Meeting, 2023
- 2. **ICSE'23**: Presented the technical track paper entitled "Fairify: Fairness Verification of Neural Networks" at the 45th ACM/IEEE ICSE conference, 2023
- 3. **ICSE'23**: Presented the technical track paper entitled "*Towards Understanding Fairness and its Composition in Ensemble ML*" at the 45th ACM/IEEE ICSE conference, 2023
- 4. **Invited Talk**: Invited by Dr. Wallapak Tavanapong (Professor and Chair of Broadening Participation, Iowa State University) to present my research on "Fairness Verification and Sustainability of AI Systems", 2023
- 5. **ICSE'22**: Presented technical track paper entitled "The Art and Practice of Data Science Pipeline" *Pittsburgh, PA* at the 44th ACM/IEEE ICSE conference, 2022
- 6. **Invited Talk**: Presented my research in the <u>CREATE SE4AI</u> group participated by Concordia University, Polytechnique Montreal, Queen's University, and University of Alberta, entitled "Understanding and Reasoning Fairness of ML-Based Software" (February 2022)
- 7. **NSF PI Meeting**: Presented my work on "Modular DNN Verification" in the NSF TRIPODS PI wirtual meeting chaired by Dr. Lenore J. Cowen from Tufts University, December 2021
- 8. **ESEC/FSE'21**: Presented technical track paper entitled "Fair Preprocessing" at the 29th ACM Athens, Greece ESEC/FSE, 2021
- 9. **ESEC/FSE'20**: Presented technical track paper entitled "An empirical study on ML Model *Sacramento, CA* Fairness" at the 28th ACM ESEC/FSE, 2020
- 10. **TADS Presentation**: Presented my research on "Fairness of Machine Learning Models" in front of Theoretical and Applied Data Science (TADS) Group of D4 Institute, Iowa State University, 2020 ☑

Virtual

- 11. MSR'19 Presentation: Data showcase and paper presentation entitled "Boa meets python: A boa Montreal, Canada dataset of data science software in Python language" in Mining Software Repository Conference (MSR), 2019 ☑
- 12. Panelist @ ESEC/FSE'20: Selected as a panelist in the session on Fairness at ESEC/FSE 2020 🗗 Sacramento, CA

Outreach Activities

• Research Outreach : Mentored Zihe Zhao (she/her), an undergraduate student from R University (later MS student at CMU), in the ICSE 2023 Student Mentoring Workshop (SMe on <i>graduate study and research career trajectory</i> . https://conf.researchr.org/committee/icse-2023/icse-2023-student-mentoring-workshop-nergian-	W)
 SIGPLAN Blog: Co-authored the ACM SIGPLAN blog to share the experience of organizin premier hybrid conference. https://blog.sigplan.org/2022/08/25/hybrid-splash-2021-retrospective 	g a 2022
• Lunch-n-Learn : Organized the weekly interdisciplinary research discussion for the whole year towa State Theoretical and Applied Data Science (TADS) group.	ear 2020-21
• MS Project Mentor: Mentored Raghul Durairaj Gobi (he/him) on his Masters Project on usage of Boa framework and dataset.	the 2021
• K-12 Outreach : Organized and served as the judge of the K-12 computational project comption at Iowa State University.	eti- 2018

Professional Activities

• Summer School: Completed professional development program in Midwest Big Data Summer	Ames, Iowa
School 🗹	
• IMA Workshop: Attended 4-day workshop on "Recent Themes in Resource Tradeoffs: Privacy,	Minneapolis
Fairness and Robustness" organized by Institute for Mathematics & its Applications (IMA),	·
University of Minnesota, 2019 [7]	

o Turing Lecture: I was invited and attended the Turing Lecture at FCRC 2019 by Geoffrey Phoenix, AZ Hinton and Yann LeCun on "The Deep Learning Revolution" ☑

Affiliation

• ACM: Member of Association for Computing Machinery (ACM)	2019-22
• SIGSOFT & SIGPLAN: Member of the ACM SIGSOFT and ACM SIGPLAN	2020-22
IEEE Computer Society: Member of the IEEE Computer Society	2023
o IEEE Student Branch: Served as the vice chair of IEEE Student Branch, Jahangirnagar University	2015–16