

Saikat Chakraborty

PHONE: +1 (434) 242-1306

E-MAIL : saiكاتc@cs.columbia.edu

ABOUT ME

I am a Ph.D. candidate at the Computer Science department at Columbia University, New York. My research area is *Programming Language Processing (PLP)* – a coalescence between Software Engineering and Machine Learning. More specifically, my research is motivated by building tools and techniques that reduces software engineering practitioners' burden. My primary research focus is divided into two part – (i.) source code understanding, (ii.) source code generation. I have developed tools for program comprehension, code generation, code translation across PLs, vulnerability detection, and modeling program changes/program repair. My future research interest is to push the boundary of machine understanding of source code and building tools for improving the software development pipeline(i.e., building better developer assisting tools) with Machine Learning and Deep Learning.

EDUCATION

Expected graduation - May 2022	Ph.D. in Computer Science Columbia University, New York, NY, USA Area: Artificial Intelligence for Software Engineering. Expertise: Source Code Analysis, Deep Learning, Natural Language Processing, Neural Machine Translation. Advisor: Dr. Baishakhi Ray.
--------------------------------	---

WORK EXPERIENCE

JANUARY 2019 -till date	Research Assistant, Arise Lab, Columbia University, New York, NY. Working in AI4SE (Artificial Intelligence for Software Engineering) sub-group.
SUMMER 2021	Software Engineer Intern at Facebook Inc., Remote. Probability (bigcode) team. Worked in Bigcode team for designing and development of source code diff model. Such model initiates first step towards automating code review process and improved the performance of regression prediction and prediction of different code review metrics.
SUMMER 2019	Software Engineer Intern at Google LLC., Sunnyvale, CA. BinEval team. Worked in designing ML based tool for analyzing security and privacy. Designed models for identifying embedded malicious code in cloud documents.
SUMMER 2017	Research Intern at Fujitsu Laboratories of America, Sunnyvale, CA Worked with AI based fault localization. Extracted subtle information from auxiliary sources to improve the performance of fault localization and program repair.
OCTOBER 2014 - August 2016	Lecturer Ahsanullah University of Science and Technology, Dhaka, Bangladesh Courses taught : Compilers (both theory and lab), Digital system design (both theory and lab).

SELECTED PROJECTS

CODIT [5]	A tree based hierarchical NMT tool for learning frequent code change patterns. Tree based modeling technique guarantees syntactic correctness of the edited code. This is an industry collaboration with Microsoft Research, Cambridge.
PLBART [4]	A large scale pretrained model for multiple programming languages. PLBART is trained on several hundred millions source code in Java and Python and technical natural languages from stackoverflow.
REDCODER [3]	A framework combining code search and source code synthesis. Given a summary of programmer intention, REDCODER relevant source code and adapts those code based on developers' need.

MODIT [2]	A multi-modal framework for source code editing. MODIT accounts for code edit context and developers' intention for editing to generate precise edited code.
BOOST [1]	A source code understanding pretrained model that learns to reason about the functional properties of the code. This is an industry collaboration with IBM Research.

SELECTED PUBLICATIONS

- [1] [Contrastive Learning for Source Code with Structural and Functional Properties](#), Y. Ding, L. Burratti, S. Pujar, A. Morari, B. Ray, **S. Chakraborty**, under review [[authors' preprint](#)].
- [2]* [On Multi-Modal Learning of Editing Source Code](#), **S. Chakraborty**, B. Ray, Accepted to be published in The 36th IEEE/ACM International Conference on Automated Software Engineering.
- [3] [Retrieval Augmented Code Generation and Summarization](#), MDR. Parvez, WU. Ahmad, **S. Chakraborty**, B. Ray, K. Chang, Findings of The 2021 Conference on Empirical Methods in Natural Language Processing (EMNLP-findings), 2021.
- [4]* [A Unified Pre-training for Program Understanding and Generation](#), WU. Ahmad[§], **S. Chakraborty**[§], B. Ray, K. Chang, Annual Conference of the North American Chapter of the Association for Computational Linguistics (NAACL), 2021. [§] Co-first authors.
- [5]* [CODIT: Code Edits with Tree Based Machine Translation](#), **S. Chakraborty**, Y. Ding, M. Allamanis, B. Ray, in IEEE Transactions on Software Engineering, 2020.
- [6]* [A Transformer-based Approach for Source Code Summarization](#) (short paper), WU. Ahmad, **S. Chakraborty**, B. Ray, K. Chang, 58th Annual Meeting of the Association for Computational Linguistics (ACL) 2020.
- [7] [Toward Optimal Selection of Information Retrieval Models for Software Engineering Tasks](#), MM. Rahman, **S Chakraborty**, G. Kaiser, B. Ray, 19th International Working Conference on Source Code Analysis and Manipulation (SCAM) 2019.
- [8]* [Building Language Models for Text with Named Entities](#), R. Parvez, **S. Chakraborty**, B. Ray, K. Chang, 56th Annual Meeting of the Association for Computational Linguistics (ACL) 2018.
- [9]* [Deep Learning based Vulnerability Detection: Are We There Yet?](#) **S. Chakraborty**, R. Krishna, Y. Ding, B. Ray, recommended minor-revision by IEEE Transaction of Software Engineering.
- [10] [Which similarity metric to use for software documents?: a study on information retrieval based software engineering tasks](#). Md. Rahman, **S. Chakraborty**, and B. Ray, **Poster** at Proceedings of the 40th International Conference on Software Engineering: Companion Proceedings. ACM, 2018.

SERVICES

Reviewer	IEEE Transaction of Software Engineering (TSE). ACM Transactions on Software Engineering and Methodology (TOSEM). IEEE Software.
PC member	MSR Mining Challenge, 2021. Workshop on Natural Language Processing for Programming (NLP4Prog), 2021.
Session Chair	The ACM Joint European Software Engineering Conference and Symposium on the Foundations of Software Engineering (ESEC/FSE) 2021. Sessions chaired : Program Repair, Code Recommendation.
Leadership	Secretary , Association of Bangladeshi Students at UVa. (2017-2018). Founding Vice President , Engineering Students' Association of Bangladesh, (2013- 2015). Organizer , International Engineering Innovation Summit Bangladesh, 2015.

*Top tier publications.