Chenguang Zhu

Contact Information

Meta Platforms, Inc.
1 Hacker Way
Menlo Park, CA 94025, USA

e-mail: cgzhu@utexas.edu
www: chenguang-zhu.github.io

Interests

My interests mainly focus on program analysis techniques with applications in GPU performance optimization, machine learning model efficiency and reliability, automated software testing, and software evolution. I am also interested in developing software engineering techniques for various emerging domains, including AI and smart contracts.

Education

The University of Texas at Austin, Austin, TX, USA

Ph.D., Electrical and Computer Engineering — Aug 2017 – Dec 2022

- Advisor: Sarfraz Khurshid
- Area of Study: Software Engineering

University of Toronto, Toronto, ON, Canada

M.Sc., Computer Science — Sep 2015 – Apr 2017

- Advisor: Marsha Chechik
- Area of Study: Software Engineering

Harbin Institute of Technology, Harbin, China

B.E., Software Engineering — Aug 2011 – Jul 2015

Publications

1. **Chenguang Zhu**, Mengshi Zhang, Xiuheng Wu, Xiufeng Xu, and Yi Li. Client-Specific Upgrade Compatibility Checking via Knowledge-Guided Discovery. In *ACM Transactions on Software Engineering and Methodology (TOSEM)*, 32(4), pages 1–31, 2023.
2. **Chenguang Zhu**, Ye Liu, Xiuheng Wu, and Yi Li. Identifying Solidity Smart Contract API Documentation Errors. In *Proceedings of the 37th IEEE/ACM International Conference on Automated Software Engineering (ASE)*, pages 1–13, 2022.
3. Chengpeng Li, **Chenguang Zhu**, Wenxi Wang, and August Shi. Repairing Order-Dependent Flaky Tests via Test Generation. In *Proceedings of the 44th International Conference on Software Engineering (ICSE)*, pages 1881–1892, 2022.
4. Ripon Saha, Akira Ura, Sonal Mahajan, **Chenguang Zhu**, Linyi Li, Yang Hu, Hiroaki Yoshida, Sarfraz Khurshid, and Mukul Prasad. SapientML: Synthesizing Machine Learning Pipelines by Learning from Human-Written Solutions. In *Proceedings of the 44th International Conference on Software Engineering (ICSE)*, pages 1932–1944, 2022.
5. **Chenguang Zhu**, Ripon Saha, Mukul Prasad, and Sarfraz Khurshid. Restoring the Executability of Jupyter Notebooks by Automatic Upgrade of Deprecated APIs. In *Proceedings of the 36th IEEE/ACM International Conference on Automated Software Engineering (ASE)*, pages 240–252, 2021.
6. Xiuheng Wu, **Chenguang Zhu**, and Yi Li. Diffbase: A Differential Factbase for Effective Software Evolution Management. In *Proceedings of the 29th ACM Joint European Software Engineering Conference and Symposium on the Foundations of Software Engineering (FSE)*, pages 503–515, 2021.
7. **Chenguang Zhu**, Yi Li, Julia Rubin, and Marsha Chechik. GenSlice: Generalized Semantic History Slicing. In *Proceedings of the 36th IEEE International Conference on Software Maintenance and Evolution (ICSME)*, pages 81–91, 2020.

8. Xindong Zhang, **Chenguang Zhu**, Yi Li, Jianmei Guo, Lihua Liu, and Haobo Gu. Precfix: Large-Scale Patch Recommendation by Mining Defect-Patch Pairs. *Proceedings of the 42nd International Conference on Software Engineering (ICSE): Software Engineering in Practice*, pages 41–50, 2020.

9. Xindong Zhang, **Chenguang Zhu**, Yi Li, Jianmei Guo, Lihua Liu, and Haobo Gu. Large-Scale Patch Recommendation at Alibaba. *Proceedings of the 42nd International Conference on Software Engineering: Companion Proceedings (ICSE Companion)*, pages 252–253, 2020.

10. Yi Li, **Chenguang Zhu**, Milos Gligoric, Julia Rubin, and Marsha Chechik. Precise Semantic History Slicing Through Dynamic Delta Refinement. *Automated Software Engineering (ASE Journal)*, 26(4), pages 757–793, 2019.

11. **Chenguang Zhu**, Owolabi Legunsen, August Shi, and Milos Gligoric. A Framework for Checking Regression Test Selection Tools. In *Proceedings of the 41st International Conference on Software Engineering (ICSE)*, pages 430–441, 2019.

12. Yi Li, **Chenguang Zhu**, Julia Rubin, and Marsha Chechik. CSlicerCloud: A Web-Based Semantic History Slicing Framework. In *Proceedings of the 40th International Conference on Software Engineering: Companion Proceeedings (ICSE DEMO)*, pages 57–60, 2018.

13. Kaiyuan Wang, **Chenguang Zhu**, Ahmet Celik, Jongwook Kim, Don Batory, and Milos Gligoric. Towards Refactoring-Aware Regression Test Selection. In *Proceedings of the 40th International Conference on Software Engineering (ICSE)*, pages 233–244, 2018.

14. Yi Li, **Chenguang Zhu**, Julia Rubin, and Marsha Chechik. Semantic Slicing of Software Version Histories. *IEEE Transactions on Software Engineering (TSE)*, 44(2), pages 182–201, 2018.

15. Yi Li, **Chenguang Zhu**, Julia Rubin, and Marsha Chechik. FHistorian: Locating Features in Version Histories. In *Proceedings of the 21st International Systems and Software Product Line Conference (SPLC)*, pages 49-58, 2017.

16. **Chenguang Zhu**, Yi Li, Julia Rubin, and Marsha Chechik. A Dataset for Dynamic Discovery of Semantic Changes in Version Controlled Software Histories. In *Proceedings of the 14th International Conference on Mining Software Repositories (MSR)*, pages 523–526, 2017.

17. Yi Li, **Chenguang Zhu**, Julia Rubin, and Marsha Chechik. Precise Semantic History Slicing Through Dynamic Delta Refinement. In *Proceedings of the 31st IEEE/ACM International Conference on Automated Software Engineering (ASE)*, pages 495–506, 2016.

Patents

US Patent 11403304: "Automatically Curating Existing Machine Learning Projects into a Corpus Adaptable for Use in New Machine Learning Projects", August 2, 2022.

Professional Experience

Research Scientist Jan 2023 – Present

Meta Platforms, Inc. Menlo Park, CA, USA

Working in Meta AI Infra.
Topics: GPU Algorithmic Efficiency, Machine Learning Training Performance Optimization

Research Assistant Jan 2020 – Dec 2022

University of Texas at Austin Austin, TX, USA
Advisor: Sarfraz Khurshid

Topics: Program Analysis for AI, API Compatibility, Smart Contract, Test Flakiness

- RELANCER: An automated technique that fixes broken data science programs (Jupyter Notebooks) by upgrading deprecated APIs.

- DocCon: An automated checker that detects inconsistencies between documentations and the corresponding code for Solidity smart contract libraries.
- ODRepair: An automated technique for repairing order-dependent Java flaky tests by generating tests that reset polluted shared states in heap memory.

Software Engineer Intern May 2022 – Aug 2022

Meta Platforms, Inc. New York, NY, USA
Mentor: Mark Kim-Mulgrew
Topic: AI Reliability

- LibCST: Meta's open-source Concrete Syntax Tree (CST) parser and serializer library for Python.
- Automated linters for detecting errors in machine learning programs.

Research Intern May 2020 – Aug 2020

Fujitsu Research of America, Inc. Sunnyvale, CA, USA
Mentors: Ripon Saha, Mukul Prasad

Topics: Dynamic Analysis, Automatic Machine Learning

- An internal Python program analysis framework.
- SapientML: A technique that automatically synthesizes machine learning pipelines by learning from existing human-written solutions.

Research Intern Jun 2019 – Jan 2020

Nanyang Technological University Singapore, Singapore
Advisor: Yi Li

Topics: Software Evolution, Patch Recommendation, Java Bytecode Analysis

- GenSlice: A generalized semantic history slicing framework that integrates several history slicing techniques under a uniform lens and provides a systematic approach for comparisons and analyses.
- Diffbase: a framework for extracting, storing, and supporting efficient querying and manipulating differential facts to support various program analysis tasks.
- Precfix: a framework that mines bug-fix patterns from large scale codebases, then utilizes these patterns to detect defects and recommend patches.

Research Assistant Aug 2017 – May 2019

University of Texas at Austin Austin, TX, USA
Advisor: Milos Gligoric

Topics: Regression Testing, Refactoring, Bug Detection

- RTSCheck: A framework for automatically finding bugs in regression test selection tools.
- Reks: A refactoring-aware regression test selection technique for Java that saves developers' testing time by only selecting tests affected by behavioral changes.

Research Assistant Sep 2015 – May 2017

University of Toronto Toronto, ON, Canada
Advisor: Marsha Chechik

Topics: Java Bytecode Analysis, Code History Analysis

- Definer: A technique relying on dynamic analysis to extract Java code changes related to a high-level functionality.
- DoSC: A dataset for benchmarking software analysis techniques that dynamically discover semantic changes.

- FHISTORIAN: A technique that establishes feature relationship graphs based on software version-controlled history.

Research Intern Jul 2016 – Sep 2016

Carnegie Mellon University Mountain View, CA, USA
Advisors: Arie Gurfinkel, Temesghen Kahsai

Topic: Formal Verification

- SeaHorn-ICE: An LLVM-based invariant generation technique that utilizes machine learning approaches to synthesize invariants for verification.

Software Engineer Intern Nov 2014 – Feb 2015

Tencent Shenzhen, China

Developed a message-exchanging module of the QQMusic iOS App.

Undergraduate Research Assistant May 2014 – Jun 2015

Harbin Institute of Technology Harbin, China
Advisors: Tiantian Wang, Peijun Ma

Topic: Code Clone Detection

- A system that relies on static analysis to identify copy-pasted code in software systems.

Professional Service

Reviewer	IEEE Transactions on Software Engineering (TSE)
Reviewer	Journal of Systems and Software (JSS)
Program Committee Member	ICST 2023
Artifact Evaluation Committee Member	ASE 2022
External Reviewer	ICSE 2022, FASE 2022
Program Committee Member	ICST 2022
External Reviewer	ASE 2021, FSE 2021, FASE 2021
Program Committee Member	ICST 2021
External Reviewer	ASE 2020, ISSRE 2020
Program Committee Member	ICST 2020
External Reviewer	ASE 2019, ISSTA 2019, ICST 2019, ICSE 2019
External Reviewer	ASE 2018

Awards

UT Austin Graduate School Professional Development Award Dec 2022
- Total value: USD $400

ACM SIGSOFT CAPS Travel Grant Oct 2022
- Total value: USD $200

ACM SIGSOFT Best Artifact Award Aug 2021
- The 29th ACM Joint European Software Engineering Conference and Symposium on the Foundations of Software Engineering (FSE 2021)

Best Artifact Award Sep 2020
- The 36th IEEE International Conference on Software Maintenance and Evolution (ICSME 2020)

UT Austin Graduate School Fellowship 2017 – 2020
- Total value: USD $36,000

National Scholarship 2013 – 2014
- Awarded by the Ministry of Education of China, to the student ranked first in the department
- Received the award in two consecutive academic years

Gold Medal in the 16th National Robot-soccer Championships Jul 2014
- Awarded by the Chinese Association for Artificial Intelligence, CAAI

Press

Diffbase: A Differential Factbase for Effective Software Evolution Management

- NTU News: https://bit.ly/2Zde01B

GenSlice: Generalized Semantic History Slicing

- NTU News: https://bit.ly/3dgrrC1

Precfix: Large-Scale Patch Recommendation by Mining Defect-Patch Pairs

- Alibaba Cloud: https://bit.ly/3gWXWHT
- Sohu News (Chinese): https://bit.ly/3dhUcyh
- Zhihu (Chinese): https://bit.ly/3xWpPFs

Invited Talks

Python Static Analysis: Performance Improvement and Applications in Machine Learning Domain Jul 2022
Meta Platforms, Inc., New York, NY, USA

Towards Reliable Software Evolution May 2022
Microsoft, Virtual

Towards Reliable Software Evolution May 2022
Guest talk, Nanyang Technological University, Singapore, Singapore

Restoring the Executability of Jupyter Notebooks by Automatic Upgrade of Deprecated APIs Nov 2021
IEEE/ACM International Conference on Automated Software Engineering (ASE), Virtual

Artefacts Plenary: Relancer Nov 2021
IEEE/ACM International Conference on Automated Software Engineering (ASE), Virtual

Handling Nondeterminism in Data Science Applications Sep 2021
Guest lecture, The University of Texas at Austin, Austin, TX, USA

GenSlice: Generalized Semantic History Slicing Sep 2020
IEEE International Conference on Software Maintenance and Evolution (ICSME), Adelaide, Australia

A Framework for Checking Regression Test Selection Tools Jul 2019
Nanyang Technological University, Singapore, Singapore

Automated Analysis and Tooling for Supporting Software Evolution Jun 2018
Harbin Institute of Technology, Harbin, China

Teaching Experience

Teaching Assistant

Department of Electrical and Computer Engineering, University of Texas at Austin

- ECE 360T: Software Testing Sep 2022 – Dec 2022
  Held office hours and graded student work. Topics include testing process, unit, integration and system testing, manual and automatic techniques for generation of test inputs and validation of test outputs, and coverage criteria, and focus on functional testing.
- ECE 382V: Software Testing in the Era of Nondeterminism Aug 2021 – Dec 2021
  Presented guest lectures, held office hours, guided course projects, and graded student work. Topics include test flakiness, regression testing, performance testing, and concurrency.
- ECE 382V: Algorithmic Foundations for Software Systems Jun 2021 – Aug 2021
  Held office hours, prepared exam materials, and graded student work. Topics include discrete math basics, theoretical foundations and implementation aspects of algorithms.
- ECE 360C: Algorithms Jan 2021 – May 2021
  Held office hours, prepared assignment materials, and graded student work. Topics include proof-based reasoning about algorithms, asymptotic complexity analysis, and algorithm design.
- ECE 461L: Software Engineering and Design Lab Aug 2020 – Dec 2020
  Taught weekly tutorials, held office hours, and graded student work. Topics include advanced software engineering tools, design patterns, and advanced Python programming.
- ECE 360T/ECE 382C-16: Software Testing Jan 2020 – May 2020
  Held office hours and graded student work. Topics include software testing, symbolic execution, and model checking.

Department of Computer Science, University of Toronto

- CSC 410: Software Testing and Verification — Sep 2016 – Dec 2016
  Taught weekly tutorials, held office hours, and graded student work. Topics include software testing, theorem proving, model checking, and static analysis.
- CSC 207: Software Design — Jan 2016 – Apr 2016
  Taught lab section, held office hours, and graded student work. Topics include Android application development, Java programming, the software design basics, and version control.
- CSC 108: Introduction to Programming — Sep 2015 – Dec 2015
  Graded student work and exams. Topics include Python programming and coding styles.

Technical Skills

Programming: Python, Java, Jupyter Notebooks, Solidity, Shell, C, C++
Productivity: Emacs, LaTeX, Git, Mercurial, Docker, JetBrains, VS Code
Operating Systems: Linux, Mac OS, Microsoft Windows