2025 IEEE/ACM Second International Conference on AI Foundation Models and Software Engineering (Forge 2025)

Ottawa, Ontario, Canada
27-28 April 2025

IEEE Catalog Number: CFP25VO9-POD
ISBN: 979-8-3315-0212-6

Copyright © 2025 by the Institute of Electrical and Electronics Engineers, Inc.
All Rights Reserved

Copyright and Reprint Permissions: Abstracting is permitted with credit to the source. Libraries are permitted to photocopy beyond the limit of U.S. copyright law for private use of patrons those articles in this volume that carry a code at the bottom of the first page, provided the per-copy fee indicated in the code is paid through Copyright Clearance Center, 222 Rosewood Drive, Danvers, MA 01923.

For other copying, reprint or republication permission, write to IEEE Copyrights Manager, IEEE Service Center, 445 Hoes Lane, Piscataway, NJ 08854. All rights reserved.

****** This is a print representation of what appears in the IEEE Digital Library. Some format issues inherent in the e-media version may also appear in this print version.***

IEEE Catalog Number: CFP25VO9-POD
ISBN (Print-On-Demand): 979-8-3315-0212-6
ISBN (Online): 979-8-3315-0211-9

Additional Copies of This Publication Are Available From:

Curran Associates, Inc
57 Morehouse Lane
Red Hook, NY 12571 USA
Phone: (845) 758-0400
Fax: (845) 758-2633
E-mail: curran@proceedings.com
Web: www.proceedings.com

2025 IEEE/ACM Second International Conference on AI Foundation Models and Software Engineering (Forge)

Forge 2025

Table of Contents

Message from the Chairs ..... **ix**
Organizing Committee ..... **xi**
Program Committee ..... **xii**

Session1: FM for Code Generation

PerfCodeGen: Improving Performance of LLM Generated Code with Execution Feedback 1
Yun Peng (The Chinese University of Hong Kong, China), Akhilesh Deepak Gotmare (Salesforce Research, Singapore), Michael R. Lyu (The Chinese University of Hong Kong, China), Caiming Xiong (Salesforce Research, Singapore), Silvio Savarese (Salesforce Research, Singapore), and Doyen Sahoo (Salesforce Research, Singapore)

RepoHyper: Search-Expand-Refine on Semantic Graphs for Repository-Level Code Completion 14
Huy N. Phan (FPT Software AI Center, Viet Nam), Hoang N. Phan (Nanyang Technological University, Singapore), Tien N. Nguyen (The University of Texas at Dallas, USA), and Nghi D. Q. Bui (FPT Software AI Center, Viet Nam)

SoTaNa: An Open-Source Software Engineering Instruction-Tuned Model 26
Ensheng Shi (Xi'an Jiaotong University, China), Yanlin Wang (Sun Yat-sen University, China), Fengji Zhang (Microsoft Research Asia, China), Bei Chen (Microsoft Research Asia, China), Hongyu Zhang (Chongqing University, China), Yanli Wang (Sun Yat-sen University, China), Daya Guo (Sun Yat-sen University, China), Lun Du (Microsoft Research Asia, China), Shi Han (Microsoft Research Asia, China), Dongmei Zhang (Microsoft Research Asia, China), and Hongbin Sun (Xi'an Jiaotong University, China)

Automated Codebase Reconciliation using Large Language Models 38
Aneri Gandhi (University of Toronto, Canada), Sanjukta De (Advanced Micro Devices, Canada), Marsha Chechik (University of Toronto, Canada), Vinay Pandit (Advanced Micro Devices, Canada), Max Kiehn (Advanced Micro Devices, Canada), Matthieu Chan Chee (Advanced Micro Devices, Canada), and Yonas Bedasso (Advanced Micro Devices, Canada)

AI-Powered, But Power-Hungry? Energy Efficiency of LLM-Generated Code 49
Lola Solovyeva (University of Twente, The Netherlands), Sophie Weidmann (University of Twente, The Netherlands), and Fernando Castor (University of Twente, The Netherlands)

HyRACC: A Hybrid Retrieval-Augmented Framework for More Efficient Code Completion 61
Chuanyi Li (Nanjing University, China), Jiwei Shang (Nanjing University, China), Yi Feng (Nanjing University, China), and Bin Luo (Nanjing University, China)

OptCodeTrans: Boost LLMs on Low-Resource Programming Language Translation 67
Jianbo Lin (Nanjing University, China), Yi Shen (Nanjing University, China), Chuanyi Li (Nanjing University, China), Changan Niu (Nanjing University, China), and Bin Luo (Nanjing University, China)

SwiftEval: Developing a Language-Specific Benchmark for LLM-Generated Code Evaluation 73
Ivan Petrukha (MacPaw, Ukraine), Yana Kurliak (MacPaw, Ukraine), and Nataliia Stulova (MacPaw, Ukraine)

SE Arena: An Interactive Platform for Evaluating Foundation Models in Software Engineering..... 78
Zhimin Zhao (Queen's University, Canada)

Session 2: FM for Software Quality Assurance & Testing

Augmenting Large Language Models with Static Code Analysis for Automated Code Quality Improvements 82
Seyed Moein Abtahi (Ontario Tech University) and Akramul Azim (Ontario Tech University)

Benchmarking Prompt Engineering Techniques for Secure Code Generation with GPT Models 93
Marc Bruni (University of Applied Sciences and Arts Northwestern Switzerland, Switzerland), Fabio Gabrielli (University of Applied Sciences and Arts Northwestern Switzerland, Switzerland), Mohammad Ghafari (Technische Universität Clausthal, Germany), and Martin Kropp (University of Applied Sciences and Arts Northwestern Switzerland, Switzerland)

ELDetector: An Automated Approach Detecting Endless-Loop in Mini Programs 104
Nan Hu (Xi'an Jiaotong University), Ming Fan (Xi'an Jiaotong University), Jingyi Lei (Xi'an Jiaotong University), Jiaying He (Xi'an Jiaotong University), and Zhe Hou (China Mobile System Integration Co.)

Testing Refactoring Engine via Historical Bug Report Driven LLM 113
Haibo Wang (Concordia University, Canada), Zhuolin Xu (Concordia University, Canada), and Shin Hwei Tan (Concordia University, Canada)

Vulnerability-Triggering Test Case Generation from Third-Party Libraries 125
Yi Gao (Zhejiang University), Xing Hu (Zhejiang University), Zirui Chen (Zhejiang University), Tongtong Xu (Nanjing University), and Xiaohu Yang (Zhejiang University)

Microservices Performance Testing with Causality-Enhanced Large Language Models 136
Cristian Mascia (Università degli Studi di Napoli Federico II, Italy), Antonio Guerriero (Università degli Studi di Napoli Federico II, Italy), Luca Giamattei (Università degli Studi di Napoli Federico II, Italy), Roberto Pietrantuono (Università degli Studi di Napoli Federico II, Italy), and Stefano Russo (Università degli Studi di Napoli Federico II, Italy)

MaRV: A Manually Validated Refactoring Dataset 141
Henrique Nunes (Federal University of Minas Gerais, Brazil), Tushar Sharma (Dalhousie University, Canada), and Eduardo Figueiredo (Federal University of Minas Gerais, Brazil)

PyResBugs: A Dataset of Residual Python Bugs for Natural Language-Driven Fault Injection 146
Domenico Cotroneo (University of Naples Federico II, Italy), Giuseppe De Rosa (University of Naples Federico II, Italy), and Pietro Liguori (University of Naples Federico II, Italy)

The Heap: A Contamination-Free Multilingual Code Dataset for Evaluating Large Language Models 151
Jonathan Katzy (Delft University of Technology, The Netherlands), Razvan Mihai Popescu (Delft University of Technology, The Netherlands), Arie van Deursen (Delft University of Technology, The Netherlands), and Maliheh Izadi (Delft University of Technology, The Netherlands)

Session 3: Collaborative Software Development

AgileCoder: Dynamic Collaborative Agents for Software Development Based on Agile Methodology 156
Minh Huynh Nguyen (FPT Software AI Center, Vietnam), Thang Phan Chau (FPT Software AI Center, Vietnam), Phong X. Nguyen (FPT Software AI Center, Vietnam), and Nghi D. Q. Bui (FPT Software AI Center, Vietnam)

Enhancing Pull Request Reviews: Leveraging Large Language Models to Detect Inconsistencies Between Issues and Pull Requests 168
Ali Tunahan Işık (Bilkent University, Turkey), Hatice Kübra Çağlar (Bilkent University, Turkey), and Eray Tüzün (Bilkent University, Turkey)

Session 4: Human-AI Collaboration & Legal Aspects of using FM

Extracting Fix Ingredients using Language Models 179
Julian Aron Prenner (Free University of Bozen-Bolzano, Italy) and Romain Robbes (Univ. Bordeaux, France)

CodeFlow: Program Behavior Prediction with Dynamic Dependencies Learning 192
Cuong Chi Le (FPT Software AI Center, Viet Nam), Nhat Hoang Phan (Nanyang Technological University, Singapore), Nhat Huy Phan (FPT Software AI Center, Viet Nam), Tien N. Nguyen (University of Texas at Dallas, USA), and Nghi D. Q. Bui (FPT Software AI Center, Viet Nam)

Addressing Specific and Complex Scenarios in Semantic Parsing 204
Yu Wang (Xi'an Jiaotong University, China), Ming Fan (Xi'an Jiaotong University, China), and Ting Liu (Xi'an Jiaotong University, China)

Skill Over Scale: The Case for Medium, Domain-Specific Models for SE 212
Manisha Mukherjee (Carnegie Mellon University, USA) and Vincent J. Hellendoorn (Carnegie Mellon University, USA)

Resource-Efficient & Effective Code Summarization 224
Saima Afrin (William & Mary, USA), Joseph Call (William & Mary, USA), Khai-Nguyen Nguyen (William & Mary, USA), Oscar Chaparro (William & Mary, USA), and Antonio Mastropaolo (William & Mary, USA)

How Developers Interact with AI: A Taxonomy of Human-AI Collaboration in Software Engineering 236
Christoph Treude (Singapore Management University, Singapore) and Marco A. Gerosa (Northern Arizona University, USA)

So what if I used GenAI? - Legal Implications of Using Cloud-based GenAI in Software Engineering Research 241
Gouri Ginde (University of Calgary)

Evaluating the Ability of GPT-4o to Generate Verifiable Specifications in VeriFast 246
Marilyn Rego (Purdue University, USA), Wen Fan (Purdue University, USA), Xin Hu (University of Michigan, USA), Sanya Dod (Purdue University, USA), Zhaorui Ni (Purdue University, USA), Danning Xie (Purdue University, USA), Jenna DiVincenzo (Purdue University, USA), and Lin Tan (Purdue University, USA)

Towards Generating App Feature Descriptions Automatically with LLMs: the Setapp Case Study .. 252
Yevhenii Peteliev (MacPaw, Ukraine), Ivan Synytsia (MacPaw, Ukraine), and Nataliia Stulova (MacPaw, Ukraine)

A Comprehensive Study of Bug Characteristics on Foundation Language Models 257
Junxiao Han (Hangzhou City University, China), Guanqi Wang (Zhejiang University, China), Jiakun Liu (Singapore Management University, Singapore), Lingfeng Bao (Zhejiang University, China), Xing Hu (Zhejiang University, China), Jinling Wei (Hangzhou City University, China), and Shuiguang Deng (Zhejiang University, China)

Cyber-Attack Detection and Localization for SCADA System of CPSs 269
Dan Li (Sun Yat-Sen University, China), Junnan Tang (Sun Yat-Sen University, China), Shunyu Wu (Sun Yat-Sen University, China), Zibin Zheng (Sun Yat-Sen University, China), and See-Kiong Ng (National University of Singapore, Singapore)

Testing Android Third Party Libraries with LLMs to Detect Incompatible APIs 280
Tarek Mahmud (Texas State University, USA), Bin Duan (University of Queensland, Australia), Meiru Che (Central Queensland University, Australia), Anne Ngu (Texas State University, USA), and Guowei Yang (University of Queensland, Australia)

Author Index **293**