

2025 ACM/IEEE International Symposium on Empirical Software Engineering and Measurement (ESEM 2025)

**Honolulu, Hawaii, USA
2-3 October 2025**

IEEE Catalog Number:
ISBN:

CFP25ENM-POD
979-8-3315-9148-9



Copyright © 2025, IEEE

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:	CFP25ENM-POD
ISBN (Print-On-Demand):	979-8-3315-9148-9
ISBN (Online):	979-8-3315-9147-2

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

TABLE OF CONTENTS

A Defect Taxonomy for Infrastructure as Code: A Replication Study	1
<i>Wendell Oliveira, Filipe Paiva, Thiago Emmanuel Pereira, João Brunet</i>	
Assessing Diversity in Creating Seed Set for Snowballing Search for Systematic Literature Review in Software Engineering	12
<i>Katia Romero Felizardo, Francisco Carlos Souza, Alinne C. Corrêa Souza, Bianca Minetto Napoleão Igor Steinmacher, Marco Gerosa</i>	
Beyond Binary Moderation: Identifying Fine-Grained Sexist and Misogynistic Behavior on GitHub with Large Language Models	23
<i>Tanni Dev, Sayma Sultana, Amiangshu Bosu</i>	
How do Community Smells Influence Self-Admitted Technical Debt in Machine Learning Projects?	35
<i>Shamse Tasnim Cynthia, Nuri Almarimi, Banani Roy</i>	
Interrogative Comments Posed by Review Comment Generators: An Empirical Study of Gerrit	47
<i>Farshad Kazemi, Maxime Lamothe, Shane McIntosh</i>	
Is Diversity a Meaningful Metric in Fairness Testing?	59
<i>Kazuki Funamoto, Takashi Kitamura, Shingo Takada</i>	
Mapping Code Smells and Refactorings Accurately: Insights from an Empirical Study	71
<i>Gautam Shetty, Tushar Sharma</i>	
SESR-Eval: Dataset for Evaluating LLMs in the Title-Abstract Screening of Systematic Reviews	82
<i>Aleksi Huotala, Miikka Kuuttila, Mika Mäntylä</i>	
The Shifting Sands of Toxicity: The Evolving Nature of Interpersonal Challenges in Open Source	94
<i>Sarthak Bharadwaj, Fabio Santos, Bianca Trinkenreich</i>	
Understanding Everything as Code: A Taxonomy and Conceptual Model	105
<i>Haoran Wei, Nazim Madhavji, John Steinbacher</i>	
Where Tests Fall Short: Empirically Analyzing Oracle Gaps in Covered Code	117
<i>Megan Maton, Gregory M. Kapfhammer, Phil McMinn</i>	
Beyond the Job Posting: What Hiring Managers Seek in Entry-Level Software Engineering Candidates	128
<i>Spencer Baloga Loufek, Fabio Santos, Bianca Trinkenreich</i>	
Aggregating Empirical Evidence from Data Strategies Studies: A Case on Model Quantization	139
<i>Santiago del Rey, Paulo Sérgio Medeiros dos Santos, Guilherme Horta Travassos, Xavier Franch Silverio Martínez-Fernández</i>	
Another Systematic Review? A Critical Analysis of Systematic Literature Reviews on Agile Effort and Cost Estimation	150
<i>Henry Edison, Nauman bin Ali</i>	
Can User Feedback Help Issue Detection? An Empirical Study on a One-Billion-User Online Service System	160
<i>Shuyao Jiang, Jiazhen Gu, Wujie Zheng, Yangfan Zhou, Michael R. Lyu</i>	

Developer Prompts in Practice: An Empirical Study of Bias, Security, and Optimization	172
<i>Dhia Elhaq Rzig, Dhruba Jyothi Paul, Kaiser Pister, Jordan Henkel, Foyzul Hassan</i>	
Exploring Engagement in Hybrid Meetings	184
<i>Daniela Grassi, Fabio Calefato, Darja Smite, Nicole Novielli, Filippo Lanubile</i>	
Exploring the Jupyter Ecosystem: An Empirical Study of Bugs and Vulnerabilities	196
<i>Wenyuan Jiang, Dian Pressato, Harsh Darji, Thibaud Lutellier</i>	
Go-Oracle: Automated Test Oracle for Go Concurrency Bugs	208
<i>Foivos Tsimpourlas, Chao Peng, Carlos Rosuero, Ping Yang, Ajitha Rajan</i>	
Is LLM-Generated Code More Maintainable & Reliable Than Human-Written Code?	219
<i>Alfred Santa Molison, Marcia Moraes, Glaucia Melo, Fabio Santos, Wesley K. G. Assunção</i>	
One Size Does Not Fit All: How to Organize Hybrid Work in Agile Software Development?	231
<i>Fateme Broomandi, Emily Laue Christensen, Maria Paasivaara</i>	
On the Harmfulness of Test Smells in Manual System Testing: A Controlled Experiment	242
<i>Gabriela Soares, Vanessa Santos, Márcio Ribeiro, Luana Martins, Valeria Pontillo, Manoel Aranda Rohit Gheyi, Ivan Machado, Fabio Palomba</i>	
Perspectives, Needs and Challenges for Sustainable Software Engineering Teams: A FinServ Case Study	253
<i>Satwik Ghanta, Peggy Gregory, Gül Çalıklı</i>	
Secret Breach Detection in Source Code with Large Language Models	264
<i>Md Nafiu Rahman, Sadif Ahmed, Zahin Wahab, S M Sohan, Rifat Shahriyar</i>	
SIExVulTS: Sensitive Information Exposure Vulnerability Detection System Using Transformer Models and Static Analysis	275
<i>Kyler Katz, Sara Moshtari, Ibrahim Mujhid, Mehdi Mirakhorli, Derek Garcia</i>	
Using Voting and Stacking Ensemble Techniques to Optimize Software Requirements Classification	287
<i>María-Isabel Limaylla-Lunarejo, Nelly Condori-Fernandez, Miguel R. Luaces</i>	
We Know What You're Looking For: Recommendation for Large-Scale Open Source Software	299
<i>Xing Cui, Jingzheng Wu, Xiang Ling, Tianyue Luo</i>	
What About Our Bug? A Study on the Responsiveness of NPM Package Maintainers	311
<i>Mohammadreza Saeidi, Ethan Thoma, Raula Gaikovina Kula, Gema Rodríguez-Pérez</i>	
When Domains Collide: An Activity Theory Exploration of Cross-Disciplinary Collaboration	323
<i>Zixuan Feng, Thomas Zimmermann, Lorenzo Pisani, Christopher Gooley, Jeremiah Wander, Anita Sarma</i>	
A Fully Automated Agent for End-to-End Code Translation and Validation	335
<i>Eray Erer, Aysun Bozanta, Turgay Aytac, Ayse Basar</i>	
An Empirical Investigation into Maintenance of Load Testing Scripts	342
<i>Ibuki Nakamura, Kosei Horikawa, Brittany Reid, Yutaro Kashiwa, Hajimu Iida</i>	
A Preliminary Assessment of SLR's Reliance on Preprints, in the Area of LLMs4SE	349
<i>Sarah Buckley, Abdul Razzaq, Michael English</i>	
A Vision for Debiasing Confirmation Bias in Software Testing via LLM	355
<i>Iflaah Salman, Muhammad Waseem, Vladimir Mandić, Rasanjana Dhanushkha De Alwis</i>	

Cognitive Biases in Software Engineering: Debiasing Through Reconceptation	362
<i>Heidi Hietala, Burak Turhan</i>	
Contextual Code Retrieval for Commit Message Generation: A Preliminary Study	369
<i>Bo Xiong, Linghao Zhang, Chong Wang, Peng Liang</i>	
Contribution History as a Key Feature in OSS Task Recommendation: An LLM-Based Empirical Study	376
<i>Md Abdul Hannan, Mohammad Habibullah Rakib, Khondaker Masfiq Reza, Fabio Santos</i>	
Dealing with SonarQube Cloud: Initial Results from a Mining Software Repository Study	383
<i>Sabato Nocera, Davide Fucci, Giuseppe Scanniello</i>	
Exploring Large Language Models for Analyzing and Improving Method Names in Scientific Code	390
<i>Gunnar Larsen, Carol Wong, Anthony Peruma</i>	
Exploring the Evidence-Based SE Beliefs of Generative AI Tools	397
<i>Chris Brown, Jason Cusati</i>	
How Small is Enough? Empirical Evidence of Quantized Small Language Models for Automated Program Repair	404
<i>Kazuki Kusama, Honglin Shu, Masanari Kondo, Yasutaka Kamei</i>	
Identifier Name Similarities: An Exploratory Study	411
<i>Carol Wong, Mai Abe, Silvia De Benedictis, Marissa Halim, Anthony Peruma</i>	
“Is It Responsible?” Emerging Results on Comparing Guardrails for Harm Mitigation in LLM-Enhanced Software Applications	418
<i>Manoel Verissimo Dos Santos Neto, Valdemar Vicente Graciano Neto, Arlindo Rodrigues Galvão Filho Mohamad Kassab, Edson Oliveira Jr</i>	
Robust or Overfitted? Investigating the Generalization of Pretrained Models in Requirement Classification	425
<i>Farha Kamal, Md Rakibul Islam</i>	
ROSE: Transformer-Based Refactoring Recommendation for Architectural Smells	432
<i>Samal Nursapa, Anastasiya Samuilova, Alessio Bucaioni, Phuong T. Nguyen</i>	
Toward Real-Time Intrusion Detection for Autonomous Vehicles: A Vision for Deep Learning-Based Security Frameworks	439
<i>Damiano Torre, Amirpasha Javid</i>	
When Retriever Meets Generator: A Joint Model for Code Comment Generation	446
<i>Tien P. T. Le, Anh M. T. Bui, Huy N. D. Pham, Alessio Bucaioni, Phuong T. Nguyen</i>	
Invisible Risks, Visible Code: A Vision for Understanding Ethical Debt in AI-Based Coding	453
<i>Dina Salah</i>	
Empirical Insights into Microservice Language Heterogeneity in Practice	458
<i>Amr S. Abdelfattah, Tomas Cerny, Marwa Elsayed</i>	
Exploring LLMs for Stakeholder-Specific Insight Generation from Software Contracts	467
<i>Jyoti S Shukla, Aditya Kahol, Mohit Chaudhary, Preethu Rose Anish</i>	

From Assessment to Enhancement of Pull Requests at Scale: Aligning Code Reviews with Developer Competencies Using Large Language Models	478
<i>Luca Mariotto, Christian Medeiros Adriano, René Eichhorn, Daniel Burgstahler, Holger Giese</i>	
Rethinking Code Review Workflows with LLM Assistance: An Empirical Study	488
<i>Fannar Steinn Aðalsteinsson, Björn Borgar Magnússon, Mislav Milicevic, Adam Nirving Davidsson Chih-Hong Cheng</i>	
Project OSCAR: Promoting CrOss-Cutting Digital Skills Through Europe-Wide Non-Conventional LeARning Experiences	498
<i>Ilenia Fronza, Tommi Mikkonen</i>	
Secure Software Engineering Through Sensible AutoMation (SESAM)	502
<i>Davide Fucci</i>	
Threat Modeling for Large Language Model-Integrated Applications (Thremolia)	505
<i>Felix Viktor Jedrzejewski, Oleksandr Adamov, Davide Fucci</i>	
A Proposal on an AI-Based Framework for Software Defect Detection Using Multimodality in Software Industries	508
<i>Shrabanti Kundu, Deepti Mishra, Alok Mishra</i>	
Human Factors in Industrial Software Engineering – A Theory-Based Empirical Study Focused on Challenges in the Context of Software Quality and Cyber Resilience	514
<i>Philipp A. Mueller</i>	
How Do Programmers Evaluate AI-Generated Code?	522
<i>Samuli Määttä</i>	
Explainable AI for Identifying and Managing Test Debt in Automated Testing	528
<i>Mahsa Radnejad</i>	