2024 IEEE/ACM 21st International Conference on Mining Software Repositories (MSR 2024)

Lisbon, Portugal
15-16 April 2024

IEEE Catalog Number: CFP2478C-POD
ISBN: 979-8-3503-6398-2

Copyright © 2024, Association for Computing Machinery (ACM)
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:	CFP2478C-POD
ISBN (Print-On-Demand):	979-8-3503-6398-2
ISBN (Online):	979-8-4007-0587-8
ISSN:	2574-3848

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

2024 IEEE/ACM 21st International Conference on Mining Software Repositories (MSR)

MSR 2024

Table of Contents

Message from the MSR 2024 General, Program, and Junior PC Chairs ..... **xvi**
Message from the MSR 2024 Data and Tool Showcase Track Co-Chairs ..... **xx**
Message from the MSR 2024 Industry Track Co-Chairs ..... **xxi**
Message from the MSR 2024 Mining Challenge Co-Chairs ..... **xxii**
Message from the MSR 2024 Registered Reports Track Co-Chairs ..... **xxv**
Message from the MSR 2024 Tutorials Track Co-Chairs ..... **xxvii**
Message from the MSR 2024 Vision and Reflection Track Co-Chairs ..... **xxviii**
Organizing Committee ..... **xxix**
Program Committee ..... **xxxi**

Ecosystems, Reuse and APIs & Tutorials

Thirty-Three Years of Mathematicians and Software Engineers: A Case Study of Domain Expertise and Participation in Proof Assistant Ecosystems 1
Gwenyth Lincroft (Northeastern University), Minsung Cho (Northeastern University), Katherine Hough (Northeastern University), Mahsa Bazzaz (Northeastern University), and Jonathan Bell (Northeastern University)

Boosting API Misuse Detection via Integrating API Constraints from Multiple Sources 14
Can Li (Nanjing University of Aeronautics and Astronautics, China), Jingxuan Zhang (Nanjing University of Aeronautics and Astronautics, China), Yixuan Tang (Nanjing University of Aeronautics and Astronautics, China), Zhuhang Li (Nanjing University of Aeronautics and Astronautics, China), and Tianyue Sun (Nanjing University of Aeronautics and Astronautics, China)

Availability and Usage of Platform-Specific APIs: A First Empirical Study 27
Ricardo Job (IFPB) and Andre Hora (UFMG)

AndroLibZoo: A Reliable Dataset of Libraries Based on Software Dependency Analysis 32
Jordan Samhi (CISPA Helmholtz Center for Information Security, Germany), Tegawendé F. Bissyandé (University of Luxembourg, Luxembourg), and Jacques Klein (University of Luxembourg, Luxembourg)

Goblin: A Framework for Enriching and Querying the Maven Central Dependency Graph 37
Damien Jaime (SAP, LIP6, Sorbonne University), Joyce El Haddad (Université Paris Dauphine-PSL, CNRS, LAMSADE), and Pascal Poizat (Sorbonne Université, CNRS, LIP6)

Dataset: Copy-based Reuse in Open Source Software 42
Mahmoud Jahanshahi (University of Tennessee Knoxville, USA) and Audris Mockus (University of Tennessee Knoxville, USA; Vilnius University, Lithuania)

Mining Our Way Back to Incremental Builds for DevOps Pipelines 48
Shane McIntosh (University of Waterloo)

Defects, Bugs and Issues

Enhancing Performance Bug Prediction Using Performance Code Metrics 50
Guoliang Zhao (IBM, Canada), Stefanos Georgiou (simpleTechs), Ying Zou (Department of Electrical and Computer Engineering, Queen's University), Safwat Hassan (Faculty of Information, University of Toronto), Derek Truong (IBM, Canada), and Toby Corbin (IBM, United Kingdom)

An Investigation of Patch Porting Practices of the Linux Kernel Ecosystem 63
Xingyu Li (University of California, Riverside, USA), Zheng Zhang (University of California, Riverside, USA), Zhiyun Qian (University of California, Riverside, USA), Trent Jaeger (University of California, Riverside, USA), and Chengyu Song (University of California, Riverside, USA)

CrashJS: A NodeJS Benchmark for Automated Crash Reproduction 75
Philip Oliver (Victoria University of Wellington, New Zealand), Jens Dietrich (Victoria University of Wellington, New Zealand), Craig Anslow (Victoria University of Wellington), and Michael Homer (Victoria University of Wellington, New Zealand)

An Empirical Study on Just-in-time Conformal Defect Prediction 88
Xhulja Shahini (paluno, University of Duisburg-Essen)

Fine-Grained Just-In-Time Defect Prediction at the Block Level in Infrastructure-as-Code (IaC) 100
Mahi Begoug (ETS Montreal, University of Quebec Montreal, QC, Canada), Moataz Chouchen (ETS Montreal, University of Quebec Montreal, QC, Canada), Ali Ouni (ETS Montreal, University of Quebec Montreal, QC, Canada), Eman Abdullah AlOmar (Stevens Institute of Technology, USA), and Mohamed Wiem Mkaouer (University of Michigan, Ann Arbor, USA)

TrickyBugs: A Dataset of Corner-case Bugs in Plausible Programs 113
Kaibo Liu (Peking University, China), Yudong Han (Peking University, China), Yiyang Liu (Peking University, China), Zhenpeng Chen (University College London, United Kingdom), Jie M. Zhang (King's College London, United Kingdom), Federica Sarro (University College London, United Kingdom), Gang Huang (Peking University, China), and Yun Ma (Peking University, China)

GitBug-Java: A Reproducible Benchmark of Recent Java Bugs 118
André Silva (KTH Royal Institute of Technology), Nuno Saavedra (INESC-ID/IST, University of Lisbon), and Martin Monperrus (KTH Royal Institute of Technology)

P3: A Dataset of Partial Program Patches 123
Dirk Beyer (LMU Munich, Germany), Lars Grunske (Humboldt-Universität zu Berlin, Germany), Matthias Kettl (LMU Munich, Germany), Marian Lingsch-Rosenfeld (LMU Munich, Germany), and Moeketsi Raselimo (Humboldt-Universität zu Berlin, Germany)

BugsPHP: A dataset for Automated Program Repair in PHP 128
Dushan Pramod (University of Moratuwa, Sri Lanka), Thushann De Silva (University of Moratuwa, Sri Lanka), Udith Thabrew (University of Moratuwa, Sri Lanka), Ridwan Shariffdeen (National University of Singapore, Singapore), and Sandareka Wickramanayake (University of Moratuwa, Sri Lanka)

AW4C: A Commit-Aware C Dataset for Actionable Warning Identification 133
Zhipeng Liu (Chongqing University, China), Meng Yan (Chongqing University, China), Zhipeng Gao (Zhejiang University, China), Dong Li (Chongqing University, China), Xiaohong Zhang (Chongqing University, China), and Dan Yang (Chongqing University, China)

Predicting the Impact of Crashes Across Release Channels 138
Suhaib Mujahid (Mozilla Corporation, Canada), Diego Elias Costa (Concordia University, Canada), and Marco Castelluccio (Mozilla Corporation, United Kingdom)

Zero Shot Learning based Alternatives for Class Imbalanced Learning Problem in Enterprise Software Defect Analysis 140
Sangameshwar Patil (TCS Research and Dept of CSE, IIT Madras, India) and B. Ravindran (Dept. of CSE and RBC DSAI, IIT Madras, India)

Mining Challenge

ChatGPT Chats Decoded: Uncovering Prompt Patterns for Superior Solutions in Software Development Lifecycle 142
Liangxuan Wu (Huazhong University of Science and Technology), Yanjie Zhao (Huazhong University of Science and Technology), Xinyi Hou (Huazhong University of Science and Technology), Tianming Liu (Monash University), and Haoyu Wang (Huazhong University of Science and Technology)

Write me this Code: An Analysis of ChatGPT Quality for Producing Source Code 147
Konstantinos Moratis (Aristotle University of Thessaloniki), Themistoklis Diamantopoulos (Aristotle University of Thessaloniki), Dimitrios-Nikitas Nastos (Aristotle University of Thessaloniki), and Andreas Symeonidis (Aristotle University of Thessaloniki)

Quality Assessment of ChatGPT Generated Code and their Use by Developers 152
Mohammed Latif Siddiq (University of Notre Dame), Lindsay Roney (University of Notre Dame), Jiahao Zhang (University of Notre Dame), and Joanna C. S. Santos (University of Notre Dame)

Analyzing Developer Use of ChatGPT Generated Code in Open Source GitHub Projects 157
Balreet Grewal (University of Alberta, Canada), Wentao Lu (University of Alberta, Canada), Sarah Nadi (University of Alberta, Canada), and Cor-Paul Bezemer (University of Alberta, Canada)

How I Learned to Stop Worrying and Love ChatGPT 162
Piotr Przymus (Nicolaus Copernicus University in Toruń Poland), Mikołaj Fejzer (Nicolaus Copernicus University in Toruń Poland), Jakub Narębski (Nicolaus Copernicus University in Toruń Poland), and Krzysztof Stencel (University of Warsaw Poland)

Can ChatGPT Support Developers? An Empirical Evaluation of Large Language Models for Code Generation 167
Kailun Jin (York University, Canada), Chung-Yu Wang (York University), Hung Viet Pham (York University), and Hadi Hemmati (York University)

The role of library versions in Developer-ChatGPT conversations 172
Rachna Raj (Concordia University, Canada) and Diego Elias Costa (Concordia University, Canada)

AI Writes, We Analyze: The ChatGPT Python Code Saga 177
Md Fazle Rabbi (Idaho State University), Arifa Islam Champa (Idaho State University), Minhaz Zibran (Idaho State University), and Md Rakibul Islam (Lamar University)

ChatGPT in Action: Analyzing Its Use in Software Development 182
Arifa Islam Champa (Idaho State University), Md Fazle Rabbi (Idaho State University), Costain Nachuma (Idaho State University), and Minhaz F. Zibran (Idaho State University)

Chatting with AI: Deciphering Developer Conversations with ChatGPT 187
Suad Mohamed (Belmont University, USA), Abdullah Parvin (Belmont University, USA), and Esteban Parra (Belmont University, USA)

Does Generative AI Generate Smells Related to Container Orchestration?: An Exploratory Study with Kubernetes Manifests 192
Yue Zhang (Auburn University), Rachel Meredith (Auburn University), Wilson Reeves (Auburn University), Julia Coriolano (Federal University of Pernambuco), Muhammad Ali Babar (University of Adelaide), and Akond Rahman (Auburn University)

On the Taxonomy of Developers' Discussion Topics with ChatGPT 197
Ertugrul Sagdic (Lamar University, United States), Arda Bayram (Lamar University, United States), and Md Rakibul Islam (Lamar University, United States)

How to Refactor this Code? An Exploratory Study on Developer-ChatGPT Refactoring Conversations 202
Eman AlOmar (Stevens Institute of Technology), Anushkrishna Venkatakrishnan (Rochester Institute of Technology), Mohamed Wiem Mkaouer (University of Michigan-Flint), Christian Newman (Rochester Institute of Technology), and Ali Ouni (ETS Montreal, University of Quebec)

Analyzing Developer-ChatGPT Conversations for Software Refactoring: An Exploratory Study 207
Omkar Sandip Chavan (Rochester Institute of Technology), Divya Dilip Hinge (Rochester Institute of Technology), Soham Sanjay Deo (Rochester Institute of Technology), Yaxuan Olivia Wang (Rochester Institute of Technology), and Mohamed Wiem Mkaouer (University of Michigan-Flint)

How Do Software Developers Use ChatGPT? An Exploratory Study on GitHub Pull Requests 212
Moataz Chouchen (ETS Montreal, University of Quebec, Canada), Narjes Bessghaier (ETS Montreal, University of Quebec, Canada), Mahi Begoug (ETS Montreal, University of Quebec, Canada), Ali Ouni (ETS Montreal, University of Quebec, Canada), Eman Abdullah AlOmar (Stevens Institute of Technology, USA), and Mohamed Wiem Mkaouer (University of Michigan-Flint, USA)

Investigating the Utility of ChatGPT in the Issue Tracking System: An Exploratory Study 217
Joy Krishan Das (University of Saskatchewan, Canada), Saikat Mondal (University of Saskatchewan, Canada), and Chanchal K. Roy (University of Saskatchewan, Canada)

Enhancing User Interaction in ChatGPT: Characterizing and Consolidating Multiple Prompts for Issue Resolution 222
Saikat Mondal (University of Saskatchewan, Canada), Suborno Deb Bappon (University of Saskatchewan, Canada), and Chanchal K. Roy (University of Saskatchewan, Canada)

DevGPT: Studying Developer-ChatGPT Conversations 227
Tao Xiao (Nara Institute of Science and Technology, Japan), Christoph Treude (The University of Melbourne, Australia), Hideaki Hata (Shinshu University, Japan), and Kenichi Matsumoto (Nara Institute of Science and Technology, Japan)

Software Quality

Not all Dockerfile Smells are the Same: An Empirical Evaluation of Hadolint Writing Practices by Experts 231
Giovanni Rosa (University of Molise), Simone Scalabrino (University of Molise), Gregorio Robles (Universidad Rey Juan Carlos), and Rocco Oliveto (University of Molise)

Supporting High-Level to Low-Level Requirements Coverage Reviewing with Large Language Models 242
Anamaria-Roberta Preda (Johannes Kepler University), Christoph Mayr-Dorn (Johannes Kepler University), Atif Mashkoor (Johannes Kepler University), and Alexander Egyed (Johannes Kepler University)

On the Executability of R Markdown Files 254
Md Anaytul Islam (Lakehead University)

APIstic: A Large Collection of OpenAPI Metrics 265
Souhaila Serbout (Software Institute (USI)) and Cesare Pautasso (Software Institute (USI))

Improving Automated Code Reviews: Learning from Experience 278
Hong Yi Lin (The University of Melbourne), Patanamon Thongtanunam (The University of Melbourne), Christoph Treude (Singapore Management University), and Wachiraphan Charoenwet (The University of Melbourne)

Multi-faceted Code Smell Detection at Scale using DesigniteJava 2.0 284
Tushar Sharma (Dalhousie University, Canada)

SATDAUG - A Balanced and Augmented Dataset for Detecting Self-Admitted Technical Debt 289
Edi Sutoyo (University of Groningen, The Netherlands) and Andrea Capiluppi (University of Groningen, The Netherlands)

Curated Email-Based Code Reviews Datasets 294
Mingzhao Liang (The University of Melbourne, Australia), Wachiraphan Charoenwet (The University of Melbourne, Australia), and Patanamon Thongtanunam (The University of Melbourne, Australia)

TestDossier: A Dataset of Tested Values Automatically Extracted from Test Execution 299
Andre Hora (UFMG)

Greenlight: Highlighting TensorFlow APIs Energy Footprint 304
Saurabhsingh Rajput (Dalhousie University, Canada), Maria Kechagia (University College London, United Kingdom), Federica Sarro (University College London, United Kingdom), and Tushar Sharma (Dalhousie University, Canada)

Mobile Apps

Automating GUI-based Test Oracles for Mobile Apps 309
Kesina Baral (CQSE America, USA), Jack Johnson (University of Minnesota, USA), Junayed Mahmud (University of Central Florida, USA), Sabiha Salma (George Mason University, USA), Mattia Fazzini (University of Minnesota, USA), Julia Rubin (The University of British Columbia Vancouver, Canada), Jeff Offutt (University of Albany, USA), and Kevin Moran (University of Central Florida, USA)

Global Prosperity or Local Monopoly? Understanding the Geography of App Popularity 322
Liu Wang (Beijing University of Posts and Telecommunications), Conghui Zheng (Beijing University of Posts and Telecommunications), Haoyu Wang (Huazhong University of Science and Technology), Xiapu Luo (The Hong Kong Polytechnic University), Gareth Tyson (The Hong Kong University of Science and Technology), Yi Wang (Beijing University of Posts and Telecommunications), and Shangguang Wang (Beijing University of Posts and Telecommunications)

GuiEvo: Automated Evolution of Mobile Application GUIs 335
Sabiha Salma (George Mason University, USA), SM Hasan Mansur (George Mason University, USA), Yule Zhang (George Mason University, USA), and Kevin Moran (University of Central Florida, USA)

Comparing Apples to Androids: Discovery, Retrieval, and Matching of iOS and Android Apps for Cross-Platform Analyses 348
Magdalena Steinböck (TU Wien, Austria), Jakob Bleier (TU Wien, Austria), Mikka Rainer (CISPA Helmholtz Center for Information Security, Germany), Tobias Urban (Institute for Internet Security, Germany), Christine Utz (CISPA Helmholtz Center for Information Security, Germany), and Martina Lindorfer (TU Wien, Austria)

Keep Me Updated: An Empirical Study on Embedded JavaScript Engines in Android Apps 361
Elliott Wen (The University of Auckland), Jiaxiang Zhou (The Hong Kong Polytechnic University), Xiapu Luo (The Hong Kong Polytechnic University), Giovanni Russello (The University of Auckland), and Jens Dietrich (Victoria University of Wellington)

Large Language Model vs. Stack Overflow in Addressing Android Permission Related Challenges 373
Sahrima Jannat Oishwee (University of Saskatchewan), Natalia Stakhanova (University of Saskatchewan), and Zadia Codabux (University of Saskatchewan)

DATAR: A Dataset for Tracking App Releases 384
Yasaman Abedini (Sharif University of Technology, Iran), Mohammad Hadi Hajihosseini (Sharif University of Technology, Iran), and Abbas Heydarnoori (Sharif University of Technology, Iran)

AndroZoo: A Retrospective with a Glimpse into the Future 389
Marco Alecci (University of Luxembourg), Pedro Jesús Ruiz Jiménez (University of Luxembourg), Kevin Aliix (Independent researcher), Tegawendé F. Bissyandé (University of Luxembourg), and Jacques Klein (University of Luxembourg)

Machine learning for Software Engineering

Whodunit: Classifying Code as Human Authored or GPT-4 generated- A case study on CodeChef problems 394
Oseremen Joy Idialu (University of Waterloo, Canada), Noble Saji Mathews (University of Waterloo, Canada), Rungroj Maipradit (University of Waterloo, Canada), Joanne M. Atlee (University of Waterloo, Canada), and Meiyappan Nagappan (University of Waterloo, Canada)

GIRT-Model: Automated Generation of Issue Report Templates 407
Nafiseh Nikeghbal (Independent Researcher, Germany), Amir Hossein Kargaran (Ludwig Maximilian University, Gremany), and Abbas Heydarnoori (Bowling Green State University, USA)

MicroRec: Leveraging Large Language Models for Microservice Recommendation 419
Ahmed Saeed Alsayed (University of Wollongong, Australia), Hoa Khanh Dam (University of Wollongong, Australia), and Chau Nguyen (University of Wollongong, Australia)

PeaTMOSS: A Dataset and Initial Analysis of Pre-Trained Models in Open-Source Software 431
Wenxin Jiang (Purdue University, USA), Jerin Yasmin (Queen's University, Canada), Jason Jones (Purdue University, USA), Nicholas Synovic (Loyola University Chicago, USA), Jiashen Kuo (Purdue University, USA), Nathaniel Bielanski (Purdue University, USA), Yuan Tian (Queen's University, Canada), George K. Thiruvathukal (Loyola University Chicago, USA), and James C. Davis (Purdue University, USA)

Data Augmentation for Supervised Code Translation Learning 444
Binger Chen (Technische Universität Berlin), Jacek Golebiowski (Amazon AWS), and Ziawasch Abedjan (Leibniz Universität Hannover)

On the Effectiveness of Machine Learning-based Call Graph Pruning: An Empirical Study 457
Amir M. Mir (Delft University of Technology), Mehdi Keshani (Delft University of Technology), and Sebastian Proksch (Delft University of Technology)

Leveraging GPT-like LLMs to Automate Issue Labeling 469
Giuseppe Colavito (University of Bari), Filippo Lanubile (University of Bari), Nicole Novielli (University of Bari), and Luigi Quaranta (University of Bari)

Development: practices and humans

Exploring the Effect of Multiple Natural Languages on Code Suggestion Using GitHub Copilot..... 481
Kei Koyanagi (Kyushu University, Japan), Dong Wang (Kyushu University, Japan), Kotaro Noguchi (Kyushu University, Japan), Masanari Kondo (Kyushu University, Japan), Alexander Serebrenik (Eindhoven University of Technology, Netherlands), Yasutaka Kamei (Kyushu University, Japan), and Naoyasu Ubayashi (Kyushu University, Japan)

A Four-Dimension Gold Standard Dataset for Opinion Mining in Software Engineering 487
Md Rakibul Islam (Lamar University, USA), Md Fazle Rabbi (Idaho State University), Yungeun Jo (Lamar University, USA), Arifa Champa (Idaho State University, USA), Ethan Young (Lamar University, USA), Camden Wilson (Lamar University, USA), Gavin Scott (Lamar University, USA), and Minhaz Zibran (Idaho State University, USA)

Opening the Valve on Pure-Data: Usage Patterns and Programming Practices of a Data-Flow Based Visual Programming Language 492
Anisha Islam (University of Alberta), Kalvin Eng (University of Alberta), and Abram Hindle (University of Alberta)

The PIPr Dataset of Public Infrastructure as Code Programs 498
Daniel Sokolowski (University of St. Gallen, Switzerland), David Spielmann (University of St. Gallen, Switzerland), and Guido Salvaneschi (University of St. Gallen, Switzerland)

A Dataset of Microservices-based Open-Source Projects 504
Dario Amoroso d'Aragona (Tampere University), Alexander Bakhtin (University of Oulu, Finland), Xiaozhou Li (University of Oulu, Finland), Ruoyu Su (University of Oulu, Finland), Lauren Adams (Baylor University USA), Ernesto Aponte (Universidad del Sagrado Corazón, USA), Francis Boyle (Baylor University, USA), Patrick Boyle (Baylor University, USA), Rachel Koerner (Baylor University, USA), Joseph Lee (University of Richmond, USA), Fangchao Tian (University of Oulu, Finland), Yuqing Wang (University of Helsinki, Finland), Jesse Nyyssölä (University of Helsinki, Finland), Ernesto Quevedo (Baylor University, USA), Shahidur Md Rahaman (Baylor University, USA), Amr S. Abdelfattah (Baylor University, USA), Mika Mäntylä (University of Helsinki, Finland), Tomas Cerny (SIE, University of Arizona, USA), and Davide Taibi (University of Oulu, Finland)

SensoDat: Simulation-based Sensor Dataset of Self-driving Cars 510
Christian Birchler (Zurich University of Applied Sciences & University of Bern, Switzerland), Cyrill Rohrbach (University of Bern, Switzerland), Timo Kehrer (University of Bern, Switzerland), and Sebastiano Panichella (Zurich University of Applied Sciences, Switzerland)

Incivility in Open Source Projects: A Comprehensive Annotated Dataset of Locked GitHub Issue Threads 515
Ramtin Ehsani (Drexel University, USA), Mia Mohammad Imran (Virginia Commonwealth University, USA), Robert Zita (Elmhurst University, USA), Kostadin Damevski (Virginia Commonwealth University, USA), and Preetha Chatterjee (Drexel University, USA)

A Dataset of Atoms of Confusion in the Android Open Source Project 520
Davi Tabosa (Federal University of Ceará), Oton Pinheiro (Federal University of Ceará), Lincoln Rocha (Federal University of Ceará), and Windson Viana (Federal University of Ceará)

PlayMyData: a curated dataset of multi-platform video games 525
Andrea D'Angelo (University of l'Aquila), Claudio Di Sipio (University of l'Aquila), Cristiano Politowski (Universitè de Montreal), and Riccardo Rubei (University of l'Aquila)

Keynote and Tutorial

Questioning the Questions We Ask About the Impact of AI on Software Engineering 530
Margaret Anne Storey (University of Victoria, Canada)

Process automation & DevOps and Tutorial I

Learning to Predict and Improve Build Successes in Package Ecosystems 531
Harshitha Menon (Lawrence Livermore National Laboratory), Daniel Nichols (University of Maryland), Abhinav Bhatele (University of Maryland), and Todd Gamblin (Lawrence Livermore National Laboratory)

The Impact of Code Ownership of DevOps Artefacts on the Outcome of DevOps CI Builds 543
Ajiromola Kola-Olawuyi (University of Waterloo), Nimmi Rashinika Weeraddana (University of Waterloo), and Meiyappan Nagappan (University of Waterloo)

A Mutation-Guided Assessment of Acceleration Approaches for Continuous Integration: An Empirical Study of YourBase 556
Zhili Zeng (University of Waterloo), Tao Xiao (Nara Institute of Science and Technology), Maxime Lamothe (Polytechnique Montréal), Hideaki Hata (Shinshu University), and Shane McIntosh (University of Waterloo)

Cohort Studies for Mining Software Repositories 569
Nyyti Saarimaki (University of Luxembourg), Sira Vegas (Universidad Politécnica de Madrid), Valentina Lenarduzzi (University of Oulu), Davide Taibi (University of Oulu), and Mikel Robredo (University of Oulu)

Software Evolution & Analysis

Unveiling ChatGPT's Usage in Open Source Projects: A Mining-based Study 571
Rosalia Tufano (Università della Svizzera italiana), Antonio Mastropaolo (Università della Svizzera italiana), Federica Pepe (University of Sannio), Ozren Dabić (Università della Svizzera italiana), Massimiliano Di Penta (University of Sannio), and Gabriele Bavota (Università della Svizzera italiana)

DRMiner: A Tool For Identifying And Analyzing Refactorings In Dockerfile 584
Emna ksontini (University of Michigan - Flint, USA), Aycha Abid (Oakland University, USA), Rania Khalsi (University of Michigan - Flint, USA), and Marouane Kessentini (University of Michigan - Flint, USA)

A Large-Scale Empirical Study of Open Source License Usage: Practices and Challenges 595
Jiaqi Wu (The State Key Laboratory of Blockchain and Data Security, Zhejiang University), Lingfeng Bao (The State Key Laboratory of Blockchain and Data Security, Zhejiang University), Xiaohu Yang (The State Key Laboratory of Blockchain and Data Security), Xin Xia (Huawei), and Xing Hu (The State Key Laboratory of Blockchain and Data Security)

Analyzing the Evolution and Maintenance of ML Models on Hugging Face 607
Joel Castaño (Universitat Politècnica de Catalunya), Silverio Martínez-Fernández (Universitat Politècnica de Catalunya), Xavier Franch (Universitat Politècnica de Catalunya), and Justus Bogner (Vrije Universiteit Amsterdam)

On the Anatomy of Real-World R Code for Static Analysis 619
Florian Sihler (Ulm University, Germany), Lukas Pietzschmann (Ulm University, Germany), Raphael Straub (Ulm University, Germany), Matthias Tichy (Ulm University, Germany), Andor Diera (Ulm University, Germany), and Abdelhalim Dahou (GESIS - Institute for the Social Sciences, Germany)

Encoding Version History Context for Better Code Representation 631
Huy Nguyen (The University of Melbourne, Australia), Christoph Treude (Singapore Management University, Singapore), and Patanamon Thongtanunam (The University of Melbourne, Australia)

CodeLL: A Lifelong Learning Dataset to Support the Co-Evolution of Data and Language Models of Code 637
Martin Weyssow (DIRO, Université de Montréal), Claudio Di Sipio (University of L'Aquila), Davide Di Ruscio (University of L'Aquila), and Houari Sahraoui (DIRO, Université de Montréal)

Bidirectional Paper-Repository Tracing in Software Engineering 642
Daniel Garijo (Universidad Politécnica de Madrid), Miguel Arroyo (Universidad Politécnica de Madrid), Esteban González (Universidad Politécnica de Madrid), Christoph Treude (University of Melbourne), and Nicola Tarocco (CERN)

DistilKaggle: A Distilled Dataset of Kaggle Jupyter Notebooks 647
Mojtaba Mostafavi Ghahfarokhi (Sharif University of Technology, Iran), Arash Asgari (Sharif University of Technology, Iran), Mohammad Abolnejadian (Sharif University of Technology, Iran), and Abbas Heydarnoori (Sharif University of Technology, Iran)

Estimating Usage Of Open Source Projects 652
Sophia Vargas (Google), Georg Link (Bitergia), and JaYoung Lee (Google)

Process automation & DevOps and Tutorial II

Options Matter: Documenting and Fixing Non-Reproducible Builds in Highly-Configurable Systems 654
Georges Aaron Randrianaina (Univ Rennes, CNRS, Inria, IRISA), Djamel Eddine Khelladi (Univ Rennes, CNRS, Inria, IRISA), Olivier Zendra (Univ Rennes, CNRS, Inria, IRISA), and Mathieu Acher (Univ Rennes, CNRS, Inria, IRISA, IUF)

How do Machine Learning Projects use Continuous Integration Practices? An Empirical Study on GitHub Actions 665
João Helis Bernardo (Federal University of Rio Grande do Norte, Brazil), Daniel Alencar da Costa (University of Otago, New Zealand), Sérgio Queiroz de Medeiros (Federal University of Rio Grande do Norte, Brazil), and Uirá Kulesza (Federal University of Rio Grande do Norte, Brazil)

A dataset of GitHub Actions workflow histories 677
Guillaume Cardoen (University of Mons, Belgium), Tom Mens (University of Mons, Belgium), and Alexandre Decan (University of Mons, Belgium)

gawd: A Differencing Tool for GitHub Actions Workflows 682
Pooya Rostami Mazrae (University of Mons, Belgium), Alexandre Decan (University of Mons, Belgium), and Tom Mens (University of Mons, Belgium)

RABBIT: A tool for identifying bot accounts based on their recent GitHub event history 687
Natarajan Chidambaram (University of Mons), Tom Mens (University of Mons), and Alexandre Decan (University of Mons)

Security and Vision & Reflection

Quantifying Security Issues in Reusable JavaScript Actions in GitHub Workflows 692
Hassan Onsori Delicheh (University of Mons, Belgium), Alexandre Decan (University of Mons, Belgium), and Tom Mens (University of Mons, Belgium)

What Can Self-Admitted Technical Debt Tell Us About Security? A Mixed-Methods Study 704
Nicolás E. Díaz Ferreyra (Hamburg University of Technology), Mojtaba Shahin (RMIT University), Mansooreh Zahedi (The University of Melbourne), Sodiq Quadri (Hamburg University of Technology), and Riccardo Scandariato (Hamburg University of Technology)

Are Latent Vulnerabilities Hidden Gems for Software Vulnerability Prediction? An Empirical Study 716
Triet Huynh Minh Le (The University of Adelaide, Australia), Xiaoning Du (Monash University, Australia), and Muhammad Ali Babar (The University of Adelaide, Australia)

MalwareBench: Malware samples are not enough 728
Nusrat Zahan (North Carolina State University), Philipp Burckhardt (Socket, Inc), Mikola Lysenko (Socket, Inc), Feross Aboukhadijeh (Socket, Inc), and Laurie Williams (North Carolina State University)

Hash4Patch: A Lightweight Low False Positive Tool for Finding Vulnerability Patch Commits 733
Simone Scalco (Università degli Studi di Trento, Italy) and Ranindya Paramitha (Università degli Studi di Trento, Italy)

MegaVul: A C/C++ Vulnerability Dataset with Comprehensive Code Representations 738
Chao Ni (Zhejiang University), Liyu Shen (Zhejiang University), Xiaohu Yang (Zhejiang University), Yan Zhu (Zhejiang University), and Shaohua Wang (Zhejiang University)

Author Index **743**