

October 12–16, 2026
Munich, Germany

Association for Computing Machinery

Advancing Computing as a Science & Profession

ASE '26

Proceedings of the 41st IEEE/ACM International Conference on

Automated Software Engineering

Edited by:
Stefan Winter, Shiva Nejati, and Justyna Petke

Sponsored by:
ACM SIGSOFT, ACM SIGAI, IEEE CS

Supported by:
Amazon Automated Reasoning, JetBrains, Google, Huawei, SoSy-Lab@LMU

Association for Computing Machinery, Inc.
1601 Broadway, 10th Floor
New York, NY 10019-7434
USA

Copyright © 2026 by the Association for Computing Machinery, Inc (ACM). Permission to make digital or hard copies of portions of this work for personal or classroom use is granted without fee provided that the copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted.

To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permission to republish from: Publications Dept. ACM, Inc.
Fax +1-212-869-0481 or E-mail permissions@acm.org.

For other copying of articles that carry a code at the bottom of the first or last page, copying is permitted provided that the per-copy fee indicated in the code is paid through the Copyright Clearance Center, 222 Rosewood Drive, Danvers, MA 01923, USA.

ACM ISBN: 979-8-4007-2882-2

Cover photo
Title: “New Town Hall at the Marienplatz in Munich, Bavaria, Germany”
Photographer: Dirk Beyer, October 2026.
Cropped from personal photo.

Production: Conference Publishing Consulting
D-94034 Passau, Germany, info@conference-publishing.com

Welcome from the Chairs

On behalf of the conference organizing committee, we welcome you to the 41st edition of the IEEE/ACM International Conference on Automated Software Engineering (ASE 2026). The ASE conference series is the premier research forum focused on automated software engineering. Each year, ASE brings together researchers and practitioners to discuss foundations, techniques, and tools for automated analysis, design, implementation, testing, and maintenance of large software systems.

ASE 2026 will take place from October 12 to October 16, 2026, at the Holiday Inn Munich – City Centre hotel, in Munich, Germany. The main conference will be held on the three days from October 13 to October 15. It features three keynotes, as well as paper and poster presentations from the research, tools and datasets, industry showcase, new ideas and emerging results, and journal-first tracks. The first and the last day of the conference are allocated to workshops, doctoral and new-faculty symposia, and the student research competition.

The goal of ASE is to support developers with automated solutions for the construction and evolution of modern, ever more complex and critical software systems. The conference technical program covers the latest developments in the area of automated software engineering, including research on applications of language models to software engineering, automated program synthesis and repair, mobile-application development, verification and validation, debugging, program analysis, performance testing, green software, security, quality assurance, and different types of software tools and environments.

Keynotes. The program includes three keynote talks given by Michael Pradel from CISPA Helmholtz Center for Information Security, Germany, Joost Noppen from British Telecommunications, UK / King's College London, UK, and Myra B. Cohen from Iowa State University, USA. We express our deepest gratitude to the invited keynote speakers for agreeing to share their experiences and insights.

Submissions. A new record number of submissions was received in almost every track of ASE 2026, with a total of 1 768 papers submitted. This includes 1 304 papers submitted to the main research track, 109 to Tools and Datasets, 187 to Industry Showcase, 93 to NIER, 45 to the Journal-First track, and 30 to the Student Research Competition.

Research track. The research track received a record number of 1 304 submissions, of which, after desk rejections and withdrawals, 1 236 went through a thorough double-anonymous review process. Each submission was reviewed by

at least three Program Committee members and was assigned an area chair. This year we introduced an Early Reject Notification stage. Papers that received only negative scores from the reviewers were rejected at this point. This resulted in 571 early rejections. Authors of the remaining papers were given a chance to respond to reviewers. During the subsequent discussion period, the designated discussion lead and an area chair sought to ensure that each paper was properly discussed and the authors' response was properly taken into account. At the end of the discussion period, 115 submissions were accepted while 150 were given a chance to revise, based on reviewers' comments. After the major-revision round, additional papers were accepted, bringing the total number of accepted papers in the research track to 263.

Our sincere thanks go to Gregory Gay and Thomas Vogel, who served as Review Process Co-chairs and helped administer and monitor the submissions and review processes. We sincerely thank our area chairs, who helped us with the huge amount of reviews and discussions: Aitor Arrieta, Alastair Donaldson, Alberto Bacchelli, Aldeida Aleti, Andrea Stocco, Andreas Zeller, Dániel Varró, Gordon Fraser, Houari Sahraoui, Jie Zhang, Maleknaz Nayebi, Manuel Wimmer, Maria Kechagia, Matias Martinez, Mohammad Reza Mousavi, Myra B. Cohen, Phil McMinn, Ronnie de Souza Santos, Shane McIntosh, Shin Yoo, Sven Apel, Tao Yue, Tim Menzies, and Xin Xia. We also extend our gratitude to our PC members, all named in the Organization section, without whom the review and selection of submissions for our program would not be possible. Additional thanks go to HotCRP for support with the conference submission system and to Conf.ResearchR for support with the web-content management system.

Tools and Datasets track. The tool-demonstration track was renamed this year to "Tools and Datasets", broadening its scope. The track received a large number of high-quality submissions, 63 of which were accepted. Many thanks to the track chairs, Aldeida Aleti and Jie Zhang, and the committee members.

Industry Showcase track. The industry showcase track aims to exhibit the top-quality practical applications of automated software engineering in real-world contexts. A record number of 75 submissions was accepted. Many thanks to the track chairs, Bo-Yuan Huang and Thomas Lemberger, and the committee members for their work on this track.

NIER track. The New Ideas and Emerging Results (NIER) track solicited three types of submissions: groundbreaking research, reflections, and late-breaking advances. From 93 submissions, four papers were desk rejected, and the remaining 89 papers were reviewed by three reviewers. Following online discussions, the program committee accepted 23 papers. We thank the track chairs, Elena Sherman and Kathryn Stolee, and the committee members for their work on this track.

Journal-first track. The journal-first track for ASE 2026 accepted 43 relevant papers from several top software-engineering journals, including Automated Software Engineering (Springer), Transactions on Software Engineering (IEEE), Transactions on Software Engineering and Methodology (ACM), Empirical Software Engineering (Springer), and Journal of Systems and Software (Elsevier). We thank the track chairs, Michael Schwarz and Vesal Vojdani, as well as the committee members.

Workshops, Doctoral and New-Faculty Symposia, and Student Research Competition. ASE 2026 additionally features a record number of 19 workshops, as well as doctoral and new-faculty symposia, and a student research competition. We would like to sincerely thank the organizers of these events: Tegan Brennan and Malte Mues (Doctoral Symposium Chairs), Marcel Böhme (New-Faculty Symposium Chair), Andrea Stocco and Dominik Winterer (Student Research Competition Chairs), whose committee accepted 21 of 30 submissions, as well as the organizers and their committees of all 19 workshops.

Poster Session. ASE 2026 has a poster session with 126 posters throughout the conference. Every poster will be displayed for one day. In addition, we arranged for authors who could not attend the conference to have their work displayed on posters. We would like to thank Daniel Baier for organizing the poster session.

Artifact Evaluation and Badges. There were 108 submissions for artifact evaluation, consisting of 73 submissions from the Research track, 27 from the Tools and Datasets track, and 8 from the Industry Showcase track. Of those, 61 were awarded the Reusable badge and 21 the Functional badge. Many thanks to the co-chairs, Jürgen Cito and Ben Hermann, and their committee, for their work.

Contractually Agreed Publication Timeline. In the past, ASE authors have occasionally suffered from late (up to a year after the conference) publications of their results, even though rapid dissemination of research results is more important than ever. To improve the situation, the ASE Steering Committee organized the selection of the proceedings-processing vendor, in which a committee evaluated bids from various vendors. Bids were requested on the basis of a bidding document that defined a timeline and a penalty for late publication, which incentivizes production vendors to adhere to the timeline. The timeline and the penalty rule became part of the contract. This approach was successful: the full proceedings were published in the ACM Digital Library on October 7, 2026, five days before the start of the conference.

Open Science. ACM has transitioned to full open access, and the proceedings of ASE 2026 are the first ASE volume that is fully open access. We also would like to

highlight two changes to the publication process this year, designed to ensure the openness, trustworthiness, and reliability of the research published at ASE.

Mandatory Data-Availability Statement. First, we required all authors for all tracks to provide a "Data-Availability Statement (DAS)" as the final section before the references. In principle, all materials that led to the results in the paper were required to be made available to readers. If materials could not be shared, the circumstances had to be explained in the Data-Availability Statement. Accepted submissions that failed to meet this requirement were removed from the proceedings. A total of 159 papers were supplemented by long-term archived artifacts with DOIs, were linked to them in the ACM DL, and thus received the Available badge.

DOIs in References. Second, we strongly encouraged including DOIs for all referenced publications or other materials that have been assigned an identifier. The use of DOIs ensures that references can be reliably traced to their source. References become verifiable, and the paper and its claims become more trustworthy. Readers can directly access the referenced material using the links provided in the paper. Citation trackers can also easily find the references. DOIs also make it easier to identify hallucinated references.

Thank You. Finally, we would like to thank all authors for contributing their results to ASE 2026, and we look forward to making the conference, together with all participants, what it should be: a vibrant meeting and networking event on software-engineering research. We would like to thank Thomas Lemberger (Sponsoring Chair) and all our financial supporters for their contributions to the conference: Amazon Automated Reasoning, JetBrains, Google, and Huawei. Furthermore, we thank Marian Lingsch-Rosenfeld and his team of 27 student volunteers for the local organization, Isabella Graßl for taking care of DEI, Shuyin Ouyang for maintaining the web page, and Lars Grunske and David Lo for the MIP selection.

We would like to conclude with a very special and heartfelt thank you to Dirk Beyer for his exceptional dedication and unwavering support of ASE 2026. Throughout the organization of the conference, Dirk was extraordinarily generous with his time, experience, and assistance. We are deeply grateful for everything he did, often going far beyond what could reasonably be expected, to support the organizing committee and the ASE community.

We wish all participants inspiring talks, fruitful discussions, a lively exchange of ideas, new research connections, and a pleasant stay in beautiful Munich — the state capital of Bavaria and Northern gateway to the Alps.

Herzlich Willkommen!

Stefan Winter (General Chair)
Shiva Nejati and Justyna Petke (Program Co-chairs)

Munich,
October 7, 2026

ASE 2026 Organization

Organizing Committee

General Chair
Stefan Winter (LMU Munich, Germany)

Program Co-chairs
Shiva Nejati (University of Ottawa, Canada)
Justyna Petke (University College London, UK)

Tools and Datasets Co-chairs
Aldeida Aleti (Monash University, Australia)
Jie M. Zhang (Mistral AI, France and King's College London, UK)

Industry Co-chairs
Bo-Yuan Huang (Intel, USA)
Thomas Lemberger (Technische Hochschule Deggendorf, Germany)

NIER Co-chairs
Elena Sherman (Boise State University, USA)
Kathryn Stolee (North Carolina State University, USA)

Student Research Competition Co-chairs
Andrea Stocco (Technical University of Munich, fortiss, Germany)
Dominik Winterer (University of Manchester, UK)

Artifact-Evaluation Co-chairs
Jürgen Cito (TU Wien, Austria and MIT, USA)
Ben Hermann (University of Stuttgart, Germany)

New-Faculty Symposium Chair
Marcel Böhme (CISPA Helmholtz Center for Information Security, Germany)

Doctoral Symposium Co-chairs
Tegan Brennan (Stevens Institute of Technology, USA)
Malte Mues (Bergische Universität Wuppertal, Germany)

Journal-First Co-chairs
Michael Schwarz (National University of Singapore, Singapore)
Vesal Vojdani (University of Tartu, Estonia)

Review-Process Co-chairs
Gregory Gay (Chalmers University of Technology and University of Gothenburg, Sweden)
Thomas Vogel (Humboldt-Universität zu Berlin, Germany)

Open-Science Chair
Stefan Winter (LMU Munich, Germany)

Sponsoring Chair
Thomas Lemberger (Technische Hochschule Deggendorf, Germany)

Student-Volunteer Chair
Marian Lingsch-Rosenfeld (LMU Munich, Germany)

Web Chair
Shuyin Ouyang (King's College London, UK)

Poster Chair
Daniel Baier (LMU Munich, Germany)

Diversity, Equity, and Inclusion Chair
Isabella Graßl (Technical University of Darmstadt, Germany)

MIP Co-chairs
Lars Grunske (Humboldt-Universität zu Berlin, Germany)
David Lo (Singapore Management University, Singapore)

ASE 2026 Research Papers: Program Committee

Justyna Petke (University College London, UK), Co-chair
Shiva Nejati (University of Ottawa, Canada), Co-chair

Gregory Gay (Chalmers University of Technology and University of Gothenburg, Sweden), Review-Process Co-chair
Thomas Vogel (Humboldt-Universität zu Berlin, Germany), Review-Process Co-chair

Aldeida Aleti (Monash University, Australia), Area Chair
Sven Apel (Saarland University, Germany), Area Chair
Aitor Arrieta (Mondragon University, Spain), Area Chair
Alberto Bacchelli (University of Zurich, Switzerland), Area Chair
Myra Cohen (Iowa State University, USA), Area Chair
Ronnie de Souza Santos (University of Calgary, Canada), Area Chair
Alastair F. Donaldson (Imperial College London, UK), Area Chair

Gordon Fraser (University of Passau, Germany), Area Chair
Maria Kechagia (National and Kapodistrian University of Athens, Greece), Area Chair
Matias Martinez (Universitat Politècnica de Catalunya, Spain), Area Chair
Shane McIntosh (University of Waterloo, Canada), Area Chair
Phil McMinn (University of Sheffield, UK), Area Chair
Tim Menzies (North Carolina State University, USA), Area Chair
Mohammad Reza Mousavi (King's College London, UK), Area Chair
Maleknaz Nayebi (York University, Canada), Area Chair
Houari Sahraoui (Université de Montréal, Canada), Area Chair
Andrea Stocco (Technical University of Munich, fortiss, Germany), Area Chair
Daniel Varro (Linköping University / McGill University, Sweden), Area Chair
Manuel Wimmer (JKU Linz, Austria), Area Chair
Xin Xia (Zhejiang University, China), Area Chair
Shin Yoo (KAIST, South Korea), Area Chair
Tao Yue (Beihang University, China), Area Chair
Andreas Zeller (CISPA Helmholtz Center for Information Security, Germany), Area Chair
Jie M. Zhang (Mistral AI, France and King's College London, UK), Area Chair

Ahmad Abdellatif (University of Calgary, Canada)
Shamsa Abid (Lahore University of Management Sciences, Pakistan)
Rui Abreu (University of Porto, Portugal)
Sallam Abualhaija (University of Luxembourg, Luxembourg)
Shivali Agarwal (IBM, India)
Nazareno Aguirre (University of Rio Cuarto/CONICET, Argentina and Guangdong Technion-Israel Institute of Technology, China)
Iftekhar Ahmed (University of California at Irvine, USA)
Amin Alipour (University of Houston, USA)
Eman Abdullah AlOmar (Stevens Institute of Technology, USA)
Gabin An (Korea University, South Korea)
Paolo Arcaini (Simula Research Laboratory, Norway)
Jinsheng Ba (Chinese University of Hong Kong, Shenzhen, China)
Thomas Bach (SAP)
Hamid Bagheri (University of Nebraska-Lincoln, USA)
Guangdong Bai (City University of Hong Kong, Hong Kong SAR China)
Lingfeng Bao (Zhejiang University, China)
Sébastien Bardin (CEA List - Université Paris-Saclay, France)
Luciano Baresi (Politecnico di Milano, Italy)
Ezio Bartocci (Technische Universität Wien, Austria)
Jonathan Bell (Northeastern University, USA)
Oussama Ben Sghaier (Queen's University, Canada)
Matteo Biagiola (USI Lugano, Switzerland and University of St. Gallen, Switzerland)
Tegawendé F. Bissyandé (University of Luxembourg, Luxembourg)
Marcel Böhme (CISPA Helmholtz Center for Information Security, Germany)

Markus Borg (CodeScene, Sweden)
Kevin Borgolte (Ruhr University Bochum, Germany)
Lei Bu (Nanjing University, China)
Alexandra Bugariu (University of Zurich, Switzerland)
Tevfik Bultan (University of California at Santa Barbara, USA)
Nimrod Busany (Traigent | Tel-Hai University, Israel)
Haipeng Cai (University at Buffalo, SUNY, USA)
Yan Cai (Institute of Software at Chinese Academy of Sciences, China)
Radu Calinescu (University of York, UK)
José Pablo Cambronero (Google, USA)
Matteo Camilli (Politecnico di Milano, Italy)
Jialun Cao (Hong Kong University of Science and Technology, China)
Gemma Catolino (University of Salerno, Italy)
Shalini Chakraborty (University of Bayreuth, Germany)
Sudipta Chattopadhyay (University of Missouri - Kansas City, USA)
Bihuan Chen (Fudan University, China)
Boqi Chen (University of Ottawa, Canada)
Chunyang Chen (TU Munich, Germany)
Junjie Chen (Tianjin University, China)
Xiangping Chen (Sun Yat-sen University, China)
Zhenbang Chen (National University of Defense Technology, China)
Mingfei Cheng (Singapore Management University, Singapore)
Xiao Cheng (Macquarie University, Australia)
Shing-Chi Cheung (Hong Kong University of Science and Technology, China)
Shaiful Chowdhury (University of Manitoba, Canada)
Jane Cleland-Huang (University of Notre Dame, USA)
Michael Coblenz (University of California, San Diego, USA)
Benoit Combemale (Inria, Univ Rennes, CNRS, IRISA, France)
Vittorio Cortellessa (University of L'Aquila, Italy)
Diego Elias Costa (Concordia University, Canada)
Marcelo d'Amorim (North Carolina State University, USA)
Nancy Day (University of Waterloo, Canada)
Jose Luis de la Vara (Universidad de Castilla - La Mancha, Spain)
Renzo Degiovanni (Luxembourg Institute of Science and Technology, Luxembourg)
Giovanni Denaro (University of Milano - Bicocca, Italy)
Antinisca Di Marco (University of L'Aquila, Italy)
Massimiliano Di Penta (University of Sannio, Italy)
Davide Di Ruscio (University of L'Aquila, Italy)
Claudio Di Sipio (Johannes Kepler University Linz, Austria)
Yangruibo Ding (University of California, Los Angeles, USA)
Zishuo Ding (Hong Kong University of Science and Technology (Guangzhou), China)
Hyunsook Do (University of North Texas, USA)
Wensheng Dou (Institute of Software Chinese Academy of Sciences, China)

Xueying Du (Fudan University, China)
Gregory J. Duck (National University of Singapore, Singapore)
Robert Dyer (University of Nebraska-Lincoln, USA)
Michael D. Ernst (University of Washington, USA)
Naser Ezzati-Jivan (Brock University, Canada)
Saad Ezzini (King Fahd University of Petroleum and Minerals, Saudi Arabia)
Chunrong Fang (Nanjing University, China)
Sidong Feng (Chinese University of Hong Kong, Shenzhen, China)
João F. Ferreira (University of Porto, Portugal and INESC-ID, Portugal)
Filomena Ferrucci (Università di Salerno, Italy)
Antonio Filieri (AWS, USA and Imperial College London, UK)
Emily First (Rutgers University, USA)
Bernd Fischer (Stellenbosch University, South Africa)
Marcelo F. Frias (University of Texas at El Paso, USA)
Thomas Fritz (University of Zurich, Switzerland)
Davide Fucci (Blekinge Institute of Technology, Sweden)
Tommaso Fulcini (Politecnico di Torino, Italy)
Khouloud Gaaloul (University of Michigan - Dearborn, USA)
Alessio Gambi (Austrian Institute of Technology, Austria)
Zhipeng Gao (Zhejiang University, China)
Simos Gerasimou (Cyprus University of Technology, Cyprus)
Sona Ghahremani (Hasso Plattner Institute, University of Potsdam, Germany)
Taher A. Ghaleb (Trent University, Canada)
Yaroslav Golubev (JetBrains Research, Serbia)
Rahul Gopinath (University of Sydney, Australia)
Alessandra Gorla (IMDEA Software Institute, Spain)
Sebastian Götz (Technische Universität Dresden, Germany)
Alex Groce (Northern Arizona University, USA)
Iris Groher (Johannes Kepler University, Linz, Austria)
Lars Grunske (Humboldt-Universität zu Berlin, Germany)
Shenghui Gu (University of Ottawa, Canada)
Xiaodong Gu (Shanghai Jiao Tong University, China)
Jianmei Guo (East China Normal University, China)
Huong Ha (RMIT University, Australia)
Mohammad Hamdaqa (Polytechnique Montreal, Canada)
Carol Hanna (University College London, UK)
Dan Hao (Peking University, China)
Mark Harman (Meta Platforms, UK and UCL, UK)
Foyzul Hassan (University of Michigan at Dearborn, USA)
Ningyu He (Hong Kong Polytechnic University, Hong Kong)
Hadi Hemmati (York University, Canada)
José Antonio Hernández López (University of Murcia, Spain)
Abbas Heydarnoori (Bowling Green State University, USA)

Robert Hierons (University of Sheffield, UK)
Michael Hilton (Carnegie Mellon University, USA)
Renáta Hodován (University of Szeged, Hungary)
Jaemin Hong (UNIST, South Korea)
Shin Hong (Chungbuk National University, South Korea)
Qiang Hu (Tianjin University, China)
Xing Hu (Zhejiang University, China)
Dong Huang (University of Hong Kong, Hong Kong SAR China)
Yihao Huang (National University of Singapore, Singapore)
Marieke Huisman (University of Twente, Netherlands)
Nargiz Humbatova (Università della Svizzera italiana, Switzerland)
Yintong Huo (Singapore Management University, Singapore)
Soha Hussein (Ain Shams University, Egypt)
Fuyuki Ishikawa (National Institute of Informatics, Japan)
Gunel Jahangirova (King's College London, UK)
Pooyan Jamshidi (University of South Carolina, USA)
Jason Jaskolka (Carleton University, Canada)
Jiajun Jiang (Tianjin University, China)
Nan Jiang (Purdue University, USA)
Yanjie Jiang (Tianjin University, China)
Yanyan Jiang (Nanjing University, China)
Yu Jiang (Tsinghua University, China)
Zhen Ming (Jack) Jiang (York University, Canada)
Zhi Jin (Peking University, Wuhan University, China)
Robbert Jongeling (University of York, UK)
Nafiseh Kahani (Carleton University, Canada)
Eunsuk Kang (Carnegie Mellon University, USA)
Hong Jin Kang (University of Sydney, Australia)
Sungmin Kang (NUS, Singapore)
Gregory Kapfhammer (Allegheny College, USA)
Kadiray Karakaya (Amazon Web Services, UK)
Oliver Karras (TIB - Leibniz Information Centre for Science and Technology, Germany)
Timo Kehrer (University of Bern, Switzerland)
Narges Khakpour (Newcastle University, UK)
Raffi Khatchadourian (CUNY Hunter College, USA)
Djamel Eddine Khelladi (CNRS, IRISA, University of Rennes, France)
Lwin Khin Shar (Singapore Management University, Singapore)
Foutse Khomh (Polytechnique Montréal, Canada)
Fitsum Kifetew (Fondazione Bruno Kessler, Italy)
Dongsun Kim (Korea University, South Korea)
Jinhan Kim (Università della Svizzera italiana, Switzerland)
Kisub Kim (DGIST, South Korea)
Mijung Kim (UNIST, South Korea)

Jacques Klein (University of Luxembourg, Luxembourg)
Masanari Kondo (Kyushu University, Japan)
Vladimir Kovalenko (JetBrains Research, Netherlands)
Anil Koyuncu (Bilkent University, Turkey)
Srđan Krstić (ETH Zurich, Switzerland)
Hongyu Kuang (Nanjing University, China)
Sandeep Kuttal (North Carolina State University, USA)
Wing Lam (George Mason University, USA)
Maxime Lamothe (Polytechnique Montreal, Canada)
Van-Hoang Le (University of Luxembourg, Luxembourg)
Jaekwon Lee (Kangwon National University, South Korea)
Nian-Ze Lee (National Taiwan University, Taiwan)
Seongmin Lee (University of California, Los Angeles, USA)
Livia Lestingi (DEIB, Politecnico di Milano, Italy)
Ding Li (Peking University, China)
Heng Li (Polytechnique Montréal, Canada)
Jia Li (Tsinghua University, China)
Jianwen Li (East China Normal University, China)
Li Li (Beihang University, China)
Shaohua Li (Chinese University of Hong Kong, Hong Kong SAR China)
Tianlin Li (Beihang University, China)
Yi Li (Nanyang Technological University, Singapore)
Yuekang Li (UNSW Sydney, Australia)
Zhenhao Li (York University, Canada)
Yun Lin (Shanghai Jiao Tong University, China)
Hui Liu (Beijing Institute of Technology, China)
Jiawei Liu (University of Illinois at Urbana-Champaign, USA)
Mingwei Liu (Sun Yat-Sen University, China)
Xiao Liu (Deakin University, Australia)
Xueqing Liu (Stevens Institute of Technology, USA)
Yang Liu (Nanyang Technological University, Singapore)
Yepang Liu (Southern University of Science and Technology, China)
Francesca Lonetti (CNR-ISTI, Italy)
Xiapu Luo (Hong Kong Polytechnic University, China)
Chen Lyu (Shandong Normal University, China)
Michael Lyu (CUHK, Hong Kong)
Sam Malek (University of California at Irvine, USA)
Ravi Mangal (Colorado State University, USA)
Shahar Maoz (Tel Aviv University, Israel)
Michaël Marcozzi (Université Paris-Saclay, CEA, List, France)
Leonardo Mariani (University of Milano-Bicocca, Italy)
Darko Marinov (University of Illinois Urbana-Champaign, USA)
Lina Marsso (Polytechnique Montreal, Canada)

Kristóf Marussy (Budapest University of Technology and Economics, Hungary)
Antonio Mastropaolo (William and Mary, USA)
Anastasia Mavridou (KBR / NASA Ames Research Center, USA)
Alexandra Mendes (University of Porto, Portugal)
Hector Menendez (King's College London, UK)
Ruijie Meng (CISPA Helmholtz Center for Information Security, Germany)
Claudio Menghi (University of Bergamo, Italy and McMaster University, Canada)
Amin Milani Fard (New York Institute of Technology, Canada)
Raffaela Mirandola (Karlsruhe Institute of Technology, Germany)
Kevin Moran (University of Central Florida, USA)
Austin Mordahl (University of Illinois Chicago, USA)
Elisa Yumi Nakagawa (University of São Paulo, Brazil)
Daye Nam (University of California, Irvine, USA)
Tien N. Nguyen (University of Texas at Dallas, USA)
Pengyu Nie (University of Waterloo, Canada)
Feifei Niu (Graz University of Technology, Austria)
Jianwei Niu (University of Texas at San Antonio, USA)
Nan Niu (University of North Florida, USA)
Yannic Noller (Ruhr University Bochum, Germany)
Bentley Oakes (Polytechnique Montréal, Canada)
Hakjoo Oh (Korea University, Republic of Korea)
Mitchell Olsthoorn (Delft University of Technology, Netherlands)
Ali Ouni (Ecole de Technologie Superieure, Canada)
Fabio Palomba (University of Salerno, Italy)
Shidong Pan (New York University, USA)
Annibale Panichella (Delft University of Technology, Netherlands)
Jihyeok Park (Korea University, South Korea)
Owain Parry (University of Edinburgh, UK)
Michele Pasqua (University of Verona, Italy)
Dan Pei (Tsinghua University, China)
Sven Peldszus (Chalmers University of Technology and University of Gothenburg, Sweden)
Chao Peng (Tencent, China)
Yun Peng (Fudan University, China)
Dietmar Pfahl (University of Tartu, Estonia)
Denys Poshyvanyk (William & Mary, USA)
Michael Pradel (CISPA Helmholtz Center for Information Security, Germany)
Alexander Pretschner (TU Munich, Germany)
Geguang Pu (East China Normal University, China)
Binhang Qi (National University of Singapore, Singapore)
Masud Rahman (Dalhousie University, Canada)
Ajitha Rajan (University of Edinburgh, UK)
Georges Aaron Randrianaina (University of Waterloo, Canada)
Sanjai Rayadurgam (University of Minnesota, USA)

Xiaoxue Ren (Zhejiang University, China)
Filippo Ricca (DIBRIS, Università di Genova, Italy)
Vincenzo Riccio (University of Udine, Italy)
Cedric Richter (University of Luxembourg, Luxembourg)
Jan Oliver Ringert (Bauhaus-University Weimar, Germany)
Romain Robbes (CNRS, LaBRI, University of Bordeaux, France)
Martin P. Robillard (McGill University, Canada)
Banani Roy (University of Saskatchewan, Canada)
Chanchal K. Roy (University of Saskatchewan, Canada)
Guido Salvaneschi (University of St. Gallen, Switzerland)
Jordan Samhi (University of Luxembourg, Luxembourg)
Jesús Sánchez Cuadrado (Universidad de Murcia, Spain)
Italo Santos (University of Hawai'i at Mānoa, USA)
Jaydeb Sarker (University of Nebraska at Omaha, USA)
Simone Scalabrino (University of Molise, Italy)
Patrizia Scandurra (University of Bergamo, Italy)
Gian Luca Scoccia (Gran Sasso Science Institute, Italy)
Alexander Serebrenik (Eindhoven University of Technology, Netherlands)
Weiyi Shang (University of Waterloo, Canada)
Ridwan Salihin Shariffdeen (SonarSource, Singapore)
Dongdong She (Hong Kong University of Science and Technology, China)
August Shi (University of Texas at Austin, USA)
Qingkai Shi (Nanjing University, China)
Donghwan Shin (University of Sheffield, UK)
Jiho Shin (York University, Canada)
Seung Yeob Shin (University of Luxembourg, Luxembourg)
Xujie Si (University of Toronto, Canada)
Stefan Sobernig (Vienna University of Business and Economics, Austria)
Jeongju Sohn (Kyungpook National University, South Korea)
Ezekiel Soremekun (Singapore University of Technology and Design, Singapore)
Romina Spalazzese (Malmö University, Sweden)
Tim Storer (University of Glasgow, UK)
Yanqi Su (Technical University of Munich, Germany)
Allison Sullivan (University of Texas at Arlington, USA)
Jun Sun (Singapore Management University, Singapore)
Ruoxi Sun (CSIRO, Australia)
Weisong Sun (Nanyang Technological University, Singapore)
Zeyu Sun (Institute of Software, Chinese Academy of Sciences, China)
Yiming Tang (Rochester Institute of Technology, USA)
Yutian Tang (University of Glasgow, UK)
Leopoldo Teixeira (Federal University of Pernambuco, Brazil)
Valerio Terragni (University of Auckland, New Zealand)
Yongqiang Tian (Monash University, Australia)

Yuan Tian (Queen's University, Canada)
Abhishek Tiwari (University of Southern Denmark, Denmark)
Saeid Tizpaz-Niari (University of Illinois Chicago, USA)
Catia Trubiani (Gran Sasso Science Institute, Italy)
Uraz Cengiz Turker (University of Lancaster, UK)
Gias Uddin (York University, Canada)
Mahsa Varshosaz (IT University of Copenhagen, Denmark)
Andreea Vescan (Babes-Bolyai University, Romania)
Andreas Vogelsang (University of Duisburg-Essen, Germany)
Neil Walkinshaw (University of Sheffield, UK)
Yao Wan (Huazhong University of Science and Technology, China)
Zhiyuan Wan (Zhejiang University, China)
Bo Wang (Beijing Jiaotong University, China)
Chao Wang (University of Southern California, USA)
Dong Wang (Tianjin University, China)
Guancheng Wang (Lero, University of Limerick, Ireland)
Haoyu Wang (Huazhong University of Science and Technology, China)
Huiyan Wang (Nanjing University, China)
Jingyi Wang (Zhejiang University, China)
Junjie Wang (Institute of Software at Chinese Academy of Sciences, China)
Shangwen Wang (National University of Defense Technology, China)
Shaowei Wang (University of Manitoba, Canada)
Shuai Wang (Hong Kong University of Science and Technology, Hong Kong)
Song Wang (York University, Canada)
Wenwen Wang (University of Georgia, USA)
Xiaoyin Wang (University of Texas at San Antonio, USA)
Yanlin Wang (Sun Yat-sen University, China)
Ying Wang (Northeastern University, China)
Andrzej Wąsowski (IT University of Copenhagen, Denmark)
Nimmi Weeraddana (University of Calgary, Canada)
Lili Wei (McGill University, Canada)
Shiyi Wei (University of Texas at Dallas, USA)
Yuxiang Wei (University of Illinois at Urbana-Champaign, USA)
Jinfeng Wen (Beijing University of Posts and Telecommunications, China)
Ming Wen (Huazhong University of Science and Technology, China)
Seunghoon Woo (Korea University, South Korea)
Franz Wotawa (Technische Universität Graz, Austria)
Lu Xiao (Stevens Institute of Technology, USA)
Tao Xiao (Kyushu University, Japan)
Xusheng Xiao (Arizona State University, USA)
Tao Xie (Peking University, China)
Xiaofei Xie (Singapore Management University, Singapore)
Xiaoyuan Xie (Wuhan University, China)

Chang Xu (Nanjing University, China)
Hui Xu (Fudan University, China)
Jingling Xue (UNSW Sydney, Australia)
Sinem Getir Yaman (University of York, UK)
Meng Yan (Chongqing University, China)
Jinqiu Yang (Concordia University, Canada)
Zhou Yang (University of Alberta, Canada and Alberta Machine Intelligence Institute, Canada)
Daphne Yao (Virginia Tech, USA)
Kundi Yao (Ontario Tech University, Canada)
Peisen Yao (Zhejiang University, China)
He Ye (University College London, UK)
Jooyong Yi (UNIST, South Korea)
Jiongchi Yu (Nanyang Technological University, Singapore)
Shengcheng Yu (Technical University of Munich, Germany)
Tingting Yu (University of Connecticut, USA)
Yijun Yu (Open University, UK)
Fiorella Zampetti (University of Sannio, Italy)
Vadim Zaytsev (University of Twente, Netherlands)
Juan Zhai (University of Massachusetts at Amherst, USA)
Cen Zhang (Microsoft, USA)
Hongyu Zhang (Chongqing University, China)
Jian Zhang (Nanyang Technological University, Singapore)
Man Zhang (Beihang University, China)
Tao Zhang (Macau University of Science and Technology, Macao SAR China)
Tianyi Zhang (Purdue University, USA)
Ting Zhang (Monash University, Australia)
Wuqi Zhang (MegaETH, Hong Kong SAR China)
Xiaoyu Zhang (Nanyang Technological University, Singapore)
Xueling Zhang (Rochester Institute of Technology, USA)
Yakun Zhang (Harbin Institute of Technology (Shenzhen), China)
Ziqi Zhang (University of Illinois Urbana-Champaign, China)
Dehai Zhao (Zhejiang University, China)
Jianjun Zhao (Kyushu University, Japan)
Yanjie Zhao (Huazhong University of Science and Technology, China)
Han Zheng (EPFL, Switzerland)
Hao Zhong (Shanghai Jiao Tong University, China)
Chenguang Zhu (University of Texas at Austin, USA)

ASE 2026 Tools and Datasets: Program Committee

Aldeida Aleti (Monash University, Australia), Co-chair
Jie M. Zhang (Mistral AI, France and King's College London, UK), Co-chair

Brad Alexander (University of Adelaide, Australia)
Baharin Aliashrafi Jodat (University of Ottawa, Canada)
Preethu Rose Anish (TCS Research, India)
Majid Babaei (University of the Fraser Valley, Canada)
Scott Barnett (Deakin University, Australia)
Roman Bögli (University of Bern, Switzerland)
Andrea Bombarda (University of Bergamo, Italy)
Jialun Cao (Hong Kong University of Science and Technology, China)
Simin Chen (George Mason University, USA)
Songqiang Chen (Hong Kong University of Science and Technology, China)
Cesar Cornejo (University of Bergamo, Italy)
Hetong Dai (University of Waterloo, Canada)
Santanu Dash (University of Surrey, UK)
Cristina David (University of Bristol, UK)
Ronnie de Souza Santos (University of Calgary, Canada)
Gianmarco De Vita (Università della Svizzera italiana, Switzerland)
Karine Even-Mendoza (King's College London, UK)
Naser Ezzati-Jivan (Brock University, Canada)
Sakina Fatima (Fujitsu Research, Canada)
Yu Gao (Institute of Software, Chinese Academy of Sciences, China)
Aayush Garg (Luxembourg Institute of Science and Technology, Luxembourg)
Jingzhi Gong (King's College London, UK)
Hashini Gunatilake (Monash University, Australia)
Simón Gutiérrez Brida (University of Rio Cuarto, Argentina and CONICET, Argentina)
Carol Hanna (University College London, UK)
Andre Hora (UFMG, Brazil)
Soneya Binta Hossain (University of Texas at Dallas, USA)
Dong Huang (University of Hong Kong, Hong Kong SAR China)
Mingyue Jiang (Zhejiang Sci-Tech University, China)
Nafiseh Kahani (Carleton University, Canada)
Timo Kehrer (University of Bern, Switzerland)
Ahmed Khanfir (University of Manouba, Tunisia)
Mugdha Khedkar (Heinz Nixdorf Institute, Paderborn University, Germany)
Lwin Khin Shar (Singapore Management University, Singapore)
Yunho Kim (Hanyang University, South Korea)
Thomas Laurent (Lero@Trinity College Dublin, Ireland)
Yan Lei (Chongqing University, China)
Qingyuan Liang (Peking University, China)
Sherlock A. Licorish (University of Otago, New Zealand)
Yepang Liu (Southern University of Science and Technology, China)
Deyun Lyu (National Institute of Informatics, Japan)
Fuchen Ma (Tsinghua University, China)

Shin Nakajima (National Institute of Informatics, Japan)
Noor Nashid (University of British Columbia, Canada)
Tien N. Nguyen (University of Texas at Dallas, USA)
Lam Nguyen Tung (Monash University, Australia)
Ali Ouni (Ecole de Technologie Superieure, Canada)
Shuyin Ouyang (King's College London, UK)
Luan Pham (University of New South Wales, Australia)
Valeria Pontillo (Gran Sasso Science Institute, Italy)
Pablo Ponzio (University of Rio Cuarto, Argentina and CONICET, Argentina)
Giovanni Quattrocchi (University of Milan, Italy)
Aurora Ramírez (University of Córdoba, Spain)
Xiaoxue Ren (Zhejiang University, China)
José Raúl Romero (University of Cordoba, Spain)
Sergio Segura (Universidad de Sevilla, Seville, Spain)
Faiz Ali Shah (University of Tartu, Estonia)
Yi Song (Wuhan University, China)
Ezekiel Soremekun (Singapore University of Technology and Design, Singapore)
Maria Spichkova (RMIT University, Australia)
Andrea Stocco (Technical University of Munich, fortiss, Germany)
Chang-ai Sun (University of Science and Technology Beijing, China)
Zeyu Sun (Institute of Software, Chinese Academy of Sciences, China)
Shuncheng Tang (Wuhan University of Technology, China)
Christoph Treude (Singapore Management University, Singapore)
Alexi Turcotte (CISPA, Germany)
Anthony Ventresque (Trinity College Dublin, Ireland and Lero, Ireland)
Guancheng Wang (Lero, University of Limerick, Ireland)
Jiyuan Wang (Tulane University, USA)
Wentao Wang (Oracle America, USA)
Huayao Wu (Nanjing University, China)
Ying Xiao (King's College London, UK)
Congying Xu (Hong Kong University of Science and Technology, China)
Juyeon Yoon (Korea Advanced Institute of Science and Technology, South Korea)
José Antonio Zamudio Amaya (CISPA Helmholtz Center for Information Security, Germany)
Haonan Zhang (University of Waterloo, Canada)
Qian Zhang (University of California at Riverside, USA)
Ting Zhang (Monash University, Australia)
Yakun Zhang (Harbin Institute of Technology (Shenzhen), China)
Zhenya Zhang (Kyushu University, Japan)
Xi Zheng (Macquarie University, Australia)
Hengcheng Zhu (Hong Kong University of Science and Technology, Hong Kong SAR China)

ASE 2026 Industry Showcase: Program Committee

Bo-Yuan Huang (Intel, USA), Co-chair
Thomas Lemberger (Technische Hochschule Deggendorf, Germany), Co-chair

Nadia Alshahwan (Meta Platforms)
Luca Berardinelli (Johannes Kepler University Linz, Austria)
Tingting Bi (University of Melbourne, Australia)
Priyam Biswas (Intel)
Marek Chalupa (Zeroth Research)
Simin Chen (George Mason University, USA)
Zhenyu Chen (Nanjing University, China)
Chun Yong Chong (Monash University Malaysia, Malaysia)
Siwei Cui (Amazon Web Services)
Peng Di (Kunlunxin, China and UNSW Sydney, Australia)
Daniel Dietsch (Qt Group, Germany)
Andreas Gaiser (Qt Company)
Alessio Gambi (Austrian Institute of Technology, Austria)
Aayush Garg (Luxembourg Institute of Science and Technology, Luxembourg)
Gaurav Ghosh (Google)
Yaroslav Golubev (JetBrains Research, Serbia)
Martin Gruber (BMW Group, Germany)
Tim Henderson (Google, USA)
Dhruva Juloori (Uber Technologies)
Soomin Kim (Ewha Womans University, South Korea)
Anil Koyuncu (Bilkent University, Turkey)
Jihyun Lee (Jeonbuk National University, South Korea)
Valentina Lenarduzzi (University of Southern Denmark, Denmark)
Fernanda Madeiral (Universidade Federal de Pernambuco, Brazil)
Louis Mandel (IBM, USA)
Muhammad Numair Mansur (Amazon Web Services, Germany)
Ramya Jayaram Masti (Ampere Computing)
Rohit Mehra (Accenture Labs, India)
Nikita Mehrotra (Microsoft Research)
Adriana Meza Soria (MIT-IBM Watson AI Lab, USA)
Manjiri Moghe (Coinbase)
Malte Mues (Bergische Universität Wuppertal, Germany)
Ganesh Nalawade (RED HAT, USA)
Feifei Niu (Graz University of Technology, Austria)
Helena Holmström Olsson (Malmö University, Sweden)
Jihun Park (Chungnam National University, South Korea)
Kannan Parthasarathy (MontyCloud)
Chao Peng (Tencent, China)

Ivan Perez (KBR @ NASA Ames Research Center, USA)
Ilham Qasse
Daniel Ramos (Uber)
Bhala Ranganathan (Microsoft)
Nikitha Rao (Google DeepMind, USA)
Gabriela Cunha Sampaio (Meta)
Saad Shafiq (University of Southern California, USA)
August Shi (University of Texas at Austin, USA)
Andreas Stahlbauer (CQSE, Germany)
Philipp Straubinger (CQSE, Germany)
Ralf Teusner (Technische Hochschule Brandenburg, Germany)
Saksham Thakur (Google, USA)
Ayse Tosun (Istanbul Technical University, Turkey)
Sakshi Udeshi (Nanyang Technological University, Singapore)
Qinying Wang (EPFL, Switzerland)
Markus Weninger (JKU Linz, Austria)
Daoyuan Wu (Lingnan University, Hong Kong)
Geunseok Yang (Hankyong National University, South Korea)
Rahul Yedida (LexisNexis)
Yanjie Zhao (Huazhong University of Science and Technology, China)

ASE 2026 Industry Showcase (2nd Cycle): Program Committee

Baharin Aliashrafi Jodat (University of Ottawa, Canada)
Priyam Biswas (Intel)
Andrea Bombarda (University of Bergamo, Italy)
Manuel Carrasco (Imperial College London, UK)
Daniel Dietsch (Qt Group, Germany)
Yaroslav Golubev (JetBrains Research, Serbia)
Guifo Fodjo A Yvan (EFREI, France)
Dmytro Humeniuk (Polytechnique Montréal, Canada)
Dhruva Juloori (Uber Technologies)
Lwin Khin Shar (Singapore Management University, Singapore)
Shaohua Li (Chinese University of Hong Kong, Hong Kong SAR China)
Weixi Ma (Meta, USA)
Jacopo Maltagliati (University of Milano-Bicocca, Italy)
Malte Mues (Bergische Universität Wuppertal, Germany)
Ganesh Nalawade (RED HAT, USA)
Darsh Shah (Apple, USA)
Weiyi Shang (University of Waterloo, Canada)
Abhishek Kumar Singh (Meta, USA)
Andrea Stocco (Technical University of Munich, fortiss, Germany)

Xiaoyang Sun (University of Leeds, UK)
Mohammad A. Tayebi (Simon Fraser University, Canada)
Jiongchi Yu (Nanyang Technological University, Singapore)

ASE 2026 Industry Showcase: Emergency Reviewers

Shivali Agarwal (IBM, India)
Baharin Aliashrafi Jodat (University of Ottawa, Canada)
Luciano Baresi (Politecnico di Milano, Italy)
Lwin Khin Shar (Singapore Management University, Singapore)
Weiyi Shang (University of Waterloo, Canada)
Ridwan Salihin Shariffdeen (SonarSource, Singapore)
Abhishek Kumar Singh (Meta, USA)
Andrea Stocco (Technical University of Munich, fortiss, Germany)
Tim Storer (University of Glasgow, UK)
Uraz Cengiz Turker (University of Lancaster, UK)
Franz Wotawa (Technische Universität Graz, Austria)

ASE 2026 NIER: Program Committee

Elena Sherman (Boise State University, USA), Co-chair
Kathryn Stolee (North Carolina State University, USA), Co-chair

Iftekhar Ahmed (University of California at Irvine, USA)
Mohamed Almukhtar (University of Michigan-Flint, USA)
Eman Abdullah AlOmar (Stevens Institute of Technology, USA)
Jon Ayerdi (Mondragon University, Spain)
Roberto Casadei (University of Bologna, Italy)
Maxime Cauz (University of Namur, Belgium)
Ferruccio Damiani (University of Turin, Italy)
Martina De Sanctis (Gran Sasso Science Institute, Italy)
Sebastian Elbaum (University of Virginia, USA)
Sen Fang (North Carolina State University, USA)
Kevin Feichtinger (Karlsruhe Institute of Technology, Germany)
Noah Frontzkowski (Hamburg University of Technology, Germany)
Aayush Garg (Luxembourg Institute of Science and Technology, Luxembourg)
Paul Gazzillo (University of Central Florida, USA)
Giammaria Giordano (University of Salerno, Italy)
Sandra Greiner (University of Southern Denmark, Denmark)
Esther Guerra (Universidad Autónoma de Madrid, Spain)

Soneya Binta Hossain (University of Texas at Dallas, USA)
Emanuele Iannone (Hamburg University of Technology, Germany)
Einar Broch Johnsen (University of Oslo, Norway)
Eduard Kamburjan (IT University of Copenhagen, Denmark)
Katja Kevic (Microsoft, UK)
Wing Lam (George Mason University, USA)
Jiawei Li (University of California, Irvine, USA)
Ze Shi (Zane) Li (University of Oklahoma, USA)
Shandler Mason (Norfolk State University, USA)
Kristof Meixner (TU Wien, Austria)
Neelofar Neelofar (Royal Melbourne Institute of Technology, Australia)
Tien N. Nguyen (University of Texas at Dallas, USA)
Vesna Nowack (Imperial College London, UK)
Moses Openja (Polytechnique Montreal, Canada)
Giovanni Quattrocchi (University of Milan, Italy)
Chaiyong Ragkhitwetsagul (Mahidol University, Thailand)
José Miguel Rojas (University of Sheffield, UK)
Mohammed Sayagh (ETS Montreal, University of Quebec, Canada)
Chia-Yi Su (York University, Canada)
Yutian Tang (University of Glasgow, UK)
Rajesh Vasa (Deakin University, Australia)
Song Wang (York University, Canada)
Xiaoyin Wang (University of Texas at San Antonio, USA)
Shiyi Wei (University of Texas at Dallas, USA)
Aashish Yadavally (University of Central Florida, USA)
Zhen Yang (Shandong University, China)
Zhou Yang (University of Alberta, Canada and Alberta Machine Intelligence Institute, Canada)
Tingting Yu (University of Connecticut, USA)
Fiorella Zampetti (University of Sannio, Italy)
Cen Zhang (Microsoft, USA)
Hongyu Zhang (Chongqing University, China)
Ying Zhang (Wake Forest University, USA)
Yanjie Zhao (Huazhong University of Science and Technology, China)

ASE 2026 Student Research Competition: Program Committee

Andrea Stocco (Technical University of Munich, fortiss, Germany), Co-chair
Dominik Winterer (University of Manchester, UK), Co-chair

Gabin An (Korea University, South Korea)
Norman Becker (CISPA Helmholtz Center for Information Security, Germany)
Masudul Hasan Masud Bhuiyan (CISPA Helmholtz Center for Information Security, Germany)

Matteo Biagiola (USI Lugano, Switzerland and University of St. Gallen, Switzerland)
Islem Bouzenia (CISPA Helmholtz Center for Information Security, Germany)
Jürgen Cito (TU Wien, Austria)
Marcelo d'Amorim (North Carolina State University, USA)
Aryaz Eghbali (CISPA Helmholtz Center for Information Security, Germany)
Khashayar Etemadi (ETH Zurich, Switzerland)
Alessio Gambi (Austrian Institute of Technology, Austria)
Bernd Gruner (German Aerospace Center, Germany)
Emanuela Guglielmi (University of Molise, Italy)
Huimin Hu (CISPA - Helmholtz Center for Information Security, Germany)
Shan Huang (East China Normal University, China)
Nargiz Humbatova (Università della Svizzera italiana, Switzerland)
Gunel Jahangirova (King's College London, UK)
Jiayi Jiang (East China Normal University, China)
Cong Li (ETH Zurich, Switzerland)
Jing Liu (Max Planck Institute for Security and Privacy, Germany)
Junming Liu
Rafael Menezes (University of Manchester, UK)
Matteo Paltenghi (University of Stuttgart, Germany)
Vincenzo Riccio (University of Udine, Italy)
Chengyu Zhang (Loughborough University, UK)

ASE 2026 Student Research Competition: Judges

Ruijie Meng (CISPA Helmholtz Center for Information Security, Germany)
Shanto Rahman (University of Texas at Austin, USA)
Banani Roy (University of Saskatchewan, Canada)
Shengcheng Yu (Technical University of Munich, Germany)

ASE 2026 Journal First: Program Committee

Michael Schwarz (National University of Singapore, Singapore), Co-chair
Vesal Vojdani (University of Tartu, Estonia), Co-chair

Gabin An (Korea University, South Korea)
Maxime Cordy (University of Luxembourg, Luxembourg)
Coen De Roover (Vrije Universiteit Brussel, Belgium)
Xiaoning Du (Monash University, Australia)
Chengwei Liu (Nankai University, China)
Michelle Thalakottur (Northeastern University, USA)

Alexi Turcotte (CISPA, Germany)
Yakun Zhang (Harbin Institute of Technology (Shenzhen), China)

ASE 2026 Artifact Evaluation: Committee

Jürgen Cito (TU Wien, Austria), Co-chair
Ben Hermann (University of Stuttgart, Germany), Co-chair

Paola Accioly (Federal University of Pernambuco, Brazil)
Elijah Kayode Adejumo (George Mason University, USA)
Meah Tahmeed Ahmed (University of Texas at Dallas, USA)
Henri Aïdasso (École de technologie supérieure, Canada)
Khadijah Al Safwan (King Fahd University of Petroleum and Minerals, Saudi Arabia)
Jessy Ayala (University of California, Irvine, USA)
Zeerak Babar (Google)
Doehyun Baek (CISPA Helmholtz Center for Information Security, Germany)
Yubo Bai (University of California, Davis, USA)
Mahi Begoug (Concordia University, Canada)
Vlad Birsan (University of Texas at Dallas, USA)
Arifa Islam Champa (Idaho State University, USA)
Manas Chaudhari (Meta)
Songqiang Chen (Hong Kong University of Science and Technology, China)
Zhiyuan Chen (Rochester Institute of Technology, USA)
Yi-Hung Chou (University of California, Irvine, USA)
Lázaro Costa (University of Porto, Portugal and HASLab/INESC, Portugal)
Ruben Cuevas Menendez (Amazon)
Leuson Da Silva (Polytechnique Montreal, Canada)
Aidan Dakhama (University of Edinburgh, UK)
Arjya Das (UC Davis, USA)
Arnav Das (India Internet Foundation, India)
Kanak Das (University of California, Riverside, USA)
Vishnu Asutosh Dasu (Pennsylvania State University, USA)
Nadeeshan De Silva (William & Mary, USA)
Higor Amario de Souza (University of São Paulo, Brazil)
Antonio Della Porta (University of Salerno, Italy)
Mouna Dhaouadi (University of Montreal, Canada)
Hridya Dhulipala (University of Texas at Dallas, USA)
Liming Dong (CSIRO's Data61, Australia)
Simantika Bhattacharjee Dristi (University of Virginia, USA)
Hang Du (University of California at Irvine, USA)
Mohamed Elwakil (U.S. Coast Guard Academy, USA)
Syeda Tasnim Fabiha (University of Southern California, USA)

Moshood A. Fakorede (Louisiana State University, USA)
Gianluca Filippone (Gran Sasso Science Institute, Italy)
Samuel W. Flint (Dakota State University, USA)
Radu Găceanu (Babes-Bolyai University, Cluj-Napoca, Romania)
Mustafa Ghani (Hasso Plattner Institute, University of Potsdam, Germany)
Sangharatna Godboley (NIT Warangal, India)
Wenbo Guo (Nanyang Technological University, Singapore)
Sachin Gupta (eBay)
Md. Asif Haider (University of California, Irvine, USA)
Maryum Hamdani (University of Eastern Finland, Finland)
Rakib Ul Haque (University of Texas San Antonio, USA)
Ruida Hu (Harbin Institute of Technology, Shenzhen, China)
Muhammad Imran (Università degli Studi dell'Aquila, Italy)
Gerardo Iuliano (University of Salerno, Italy)
Aaron Jeyaraj (Fluidstack)
Shuyao Jiang (Chinese University of Hong Kong, Hong Kong SAR China)
Bihui Jin (University of Waterloo, Canada)
Anemone Kampkötter (University of Stuttgart, Germany)
Shravya Kanchi (Virginia Tech, USA)
Marcus Kessel (University of Mannheim, Germany)
Tooba Khan (University of Southern California, USA)
Afrina Khatun (University of California Irvine, USA)
Mugdha Khedkar (Heinz Nixdorf Institute, Paderborn University, Germany)
Sehoon Kim (Ulsan National Institute of Science and Technology, South Korea)
Moritz Kliemke (LMU Munich, Germany)
Sriteja Kummita (Paderborn University, Germany)
Cristian Leo (Amazon)
Fengjie Li (Tianjin University, China)
Ningke Li (National University of Singapore, Singapore)
Yuchen Ling (Nanjing University, China)
Wei Liu (Concordia University, Montreal, Canada)
Xin Liu (National University of Defense Technology, Changsha, China)
Ye Liu (Beijing Institute of Technology, China)
Yixuan Liu (Nanyang Technological University, Singapore)
Zichen Liu (Arizona State University, USA)
Jie Ma (Beihang University, China and Zhongguancun Laboratory, China)
Zhongkui Ma (University of Queensland, Australia)
Ardi Madadi (Max Planck Institute for Security and Privacy, Germany)
Tarek Mahmud (Texas A&M University-Kingsville, USA)
Suneet Malhotra (Independent Researcher, USA)
Amir Mashmool (University of Bremen, Germany)
Agon Memeti (University of Tetova, North Macedonia)
Miao Miao (University of Texas at Dallas, USA)

Nusrat Jahan Mozumder (University of Virginia, USA)
Palaniappan Muthuraman (Heinz Nixdorf Institute at Paderborn University, Germany)
Steven Ngo (University of California, Irvine, USA)
Tuan Ngo (University of Southern California, USA)
Daniel Ogenrwot (University of Nevada Las Vegas, USA)
Milos Ojdanic (Goodyear Innovation Center)
Uju Okoye (Louisiana State University, USA)
Rafael Oliveira (USP)
Kunal Suresh Pai (University of California, Los Angeles, USA)
Ranindya Paramitha (North Carolina State University, USA)
Jiaqian Peng (Institute of Information Engineering, Chinese Academy of Sciences, China)
Sara Pettinari (Gran Sasso Science Institute, Italy)
Jyoti Prakash (University of Southern Denmark, Denmark)
Yuxin Qiu (University of California at Riverside, USA)
Lili Quan (Singapore Management University, Singapore)
Md Fazle Rabbi (Idaho State University, USA)
Mohammad Rafieian (University of Texas at Dallas, USA)
Ravishka Rathnasuriya (University of Texas at Dallas, USA)
Gilberto Recupito (Vrije Universiteit Brussel, Belgium)
Joy Saha (Computer Science, University of Virginia, USA)
Sudipa Saha (University of South Florida, USA)
Rohith Kumar Shanmuganathan (Carl von Ossietzky Universität Oldenburg, Germany)
Shuai Shao (University of Connecticut, USA)
Reshabh K Sharma (University of Washington, USA)
Rithik Sharma (University of California Santa Cruz, USA)
Iti Shree (King's College London, UK)
Shweta (National Institute of Technology, Calicut, India)
Florian Sihler (Ulm University, Germany)
Sagor Sikdar (Penn State University, USA)
Vipin Singh (Google, USA)
Soaibuzzaman (Bauhaus-University Weimar, Germany)
Shrikant Sonparote (Lowe's Companies, USA)
Majd Soud (University of Doha for Science and Technology, Qatar and University of Iceland, Iceland)
Chia-Yi Su (York University, Canada)
Tasfia Tasnim (University of Texas at Dallas, USA)
Aditya Thimmaiah (University of Texas at Austin, USA)
Yu-Jye Tung (Pennsylvania State University, USA)
Ashwin Prasad Shivarpatna Venkatesh
Jian Wang (Nanyang Technological University, China)
Ke Wang (Central South University, China)
Zifan Xie (Chongqing University, China)
Shangzhi Xu (University of New South Wales, Australia)
Drishti Yadav (University of Luxembourg, Luxembourg)

Yanming Yang (Wuhan University, China)
Jiaqiang Yao (National University of Defense Technology, China)
Srinivas Kalyan Yellanki (Expedia Group)
Jiongchi Yu (Nanyang Technological University, Singapore)
Zehong Yu (Tsinghua University, China)
José Antonio Zamudio Amaya (CISPA Helmholtz Center for Information Security, Germany)
Gennaro Zanfardino (University of L'Aquila, Italy)
Samvid Avinash Zare (Amazon Web Services)
Ruihan Zhang (Singapore Management University, Singapore)
Weixing Zhang (Karlsruhe Institute of Technology, Germany)
Zejun Zhang (Nanjing University, China)
Huan Zhao (National University of Singapore, Singapore)
Zitong Zhou (University of California, Los Angeles, USA)
Shihao Zhu (Institute of Software, Chinese Academy of Sciences, China)

ASE Steering Committee

Bernd Fischer (Stellenbosch University, South Africa), Chair
David Lo (Singapore Management University, Singapore), Vice-chair

Tegawendé F. Bissyandé (University of Luxembourg, Luxembourg)
Marcel Böhme (CISPA Helmholtz Center for Information Security, Germany)
Tevfik Bultan (University of California, Santa Barbara, USA)
Marsha Chechik (University of Toronto, Canada)
Myra Cohen (Iowa State University, USA)
Massimiliano Di Penta (University of Sannio, Italy)
Gordon Fraser (University of Passau, Germany)
Lars Grunske (Humboldt-Universität zu Berlin, Germany)
Dan Hao (Peking University, China)
Jacques Klein (University of Luxembourg, Luxembourg)
Michael R. Lowry (NASA Ames Research Center, USA)
Shahar Maoz (Tel Aviv University, Israel)
Darko Marinov (University of Illinois Urbana-Champaign, USA)
Tim Menzies (North Carolina State University, USA)
Tien Nguyen (University of Texas at Dallas, USA)
Denys Poshyvanyk (William & Mary, USA)
Baishakhi Ray (Columbia University, USA)
Julia Rubin (University of British Columbia, Canada)
Federica Sarro (University College London, UK)
Shin Yoo (KAIST, South Korea)

Lingming Zhang (University of Illinois Urbana-Champaign, USA)
Minghui Zhou (Peking University, China)
Thomas Zimmermann (University of California, Irvine, USA)

Sponsors and Supporters of ASE 2026

Sponsors





Supporters

Platinum





Gold





In-Kind



Contents

Frontmatter

Welcome from the Chairs . . . iii
ASE 2026 Organization . . . vii
Sponsors . . . xxx

Keynotes

Trust the Delta: Rethinking Software Validation in the AI Era (Keynote)
Michael Pradel — *CISPA Helmholtz Center for Information Security, Germany* . . . 1
Beyond the Hype: What AI Really Did to Software Engineering (Keynote)
Joost Noppen — *British Telecommunications, UK; King's College London, UK* . . . 2
Lost in Automated Specifications: A Time and Space Odyssey in System Testing (Keynote)
Myra B. Cohen — *Iowa State University, USA* . . . 3

Main Research Papers

SpecFSM: Extracting and Repairing Finite State Machines from Protocol Specification Documents
Xiangdong Li, DaWei Huang, Jingjing Guan, and Hui Li — *Beijing University of Posts and Telecommunications, Beijing, China* . . . 4
Trace-Driven Automated PoC Synthesis for On-Chain Attacks
Xing Su, Hao Wu, Hanzhong Liang, Yunlin Jiang, Yuxi Cheng, Yating Liu, Fengyuan Xu, and Sheng Zhong — *Nanjing University, Nanjing, China* . . . 16
Context-Aware Trust Verification for Identity-Based Software Signing
Chinenye Okafor, James C. Davis, and Santiago Torres-Arias — *Purdue University, USA* . . . 29
A11YRepair: Bridging Web Accessibility Barriers via Knowledge-Enhanced Divide-and-Conquer Repair
Kai Huang, Ling Zhu, Jian Zhang, Xiaofei Xie, and Chunyang Chen — *TU Munich, Germany; Nanyang Technological University, Singapore; Singapore Management University, Singapore* . . . 42
GALA: Graph-Augmented LLM Agents for Root Cause Analysis and Incident Response in Microservices
Yifang Tian, Yaming Liu, Zichun Chong, Zihang Huang, Yiran Li, and Hans-Arno Jacobsen — *University of Toronto, Canada* . . . 55
IntOAgent: A Practical Framework for Automated Detection and Triggering of Integer Overflow Vulnerabilities
Yuelin Wang, Jiongchi Yu, Xiaofei Xie, Yaohui Sun, Tianyu Shi, Yanbang Sun, Qiang Hu, and Junjie Wang — *Tianjin University, China; Singapore Management University, Singapore* . . . 68
LLM-Based Agents for Identifying Bug-Introducing Commits
Niklas Risse and Marcel Böhme — *MPI-SP, Germany; CISPA Helmholtz Center for Information Security, Germany* . . . 81
ARMOR: A Robust Self-Supervised Framework for Root Cause Analysis in Microservices under Missing Modality
Wenzhuo Qian, Hailiang Zhao, Ziqi Wang, Zhipeng Gao, Jiayi Chen, Zhiwei Ling, and Shuiguang Deng — *Zhejiang University, China* 93
Beyond Text Matching: Towards Reference-Free Evaluation for Human-Oriented Binary Reverse Engineering
Xiuwei Shang, Li Hu, Xiao Jiang, Jieke Shi, Junda He, Zhou Yang, Shaoyin Cheng, Guoqiang Chen, Weiming Zhang, and David Lo — *University of Science and Technology of China, China; Singapore Management University, Singapore; University of Alberta, Canada; Anhui Province Key Laboratory of Digital Security, Hefei, China* . . . 106
Source-Free Detection and Impact Analysis of Compiler Optimization Problems in Mobile Applications
Han Hu, Xiaoheng Xie, Bo Sun, Jian Gu, Gang Fan, and Li Li — *Independent Researcher, Hong Kong; Huawei, China; Monash University, Australia; Beihang University, China* . . . 119
Quantum Squeeziness: An Information Theoretical Metric for Quantum Software Testability
Avner Bensoussan, Hector Menendez, and Mohammad Reza Mousavi — *King's College London, UK* . . . 132
Spec2V: Planning-Guided Specification-to-Verilog Generation Framework with Verification-Driven Debugging
Xuming Liu, Chi Zhang, Jinglong Xu, Shan Zhou, and Jinbo Wang — *Technology and Engineering Center for Space Utilization at Chinese Academy of Sciences, China; University of Chinese Academy of Sciences, China* . . . 145
Characterizing the Landscape of Open-Source Satellite Software
Jinfeng Wen, Qi Liang, Yuehan Sun, Federica Sarro, Ao Zhou, Xuanzhe Liu, and Shangguang Wang — *Beijing University of Posts and Telecommunications, China; University College London, UK; Peking University, China* . . . 158

Bridging the Gap between Intent and Impact: An Empirical Study of GPU Optimizations in Deep Learning Frameworks
Siqi Wang, Fanjing Luo, Xing Hu, Xinyu Wang, and Xin Xia — *Zhejiang University, China; Hangzhou High-Tech Zone (Binjiang) Institute of Blockchain and Data Security, China* . . . 171

CIPIHunter: Detecting Configuration-Induced Prediction Instability in Deep Learning Frameworks
Yanzhou Mu, Shuo Meng, Mijung Kim, Xiang Chen, Chunrong Fang, Zhenyu Chen, and Juan Zhai — *UNIST, Republic of Korea; Nanjing University, China; Nantong University, China; University of Massachusetts at Amherst, USA* . . . 184

TasmScan: Continuation-Aware Taint Analysis for TVM Bytecode with Savelist Abstraction
Yixuan Liu, Yin Wu, and Yi Li — *Nanyang Technological University, Singapore; Xi'an Jiaotong University, China* . . . 197

Quick Bug Detection through Black-Box Checking: A Systematic Evaluation
Bram Pellen, María Belén Rodríguez, Frits Vaandrager, and Petra van den Bos — *Radboud University, Netherlands; University of Twente, Netherlands* . . . 209

A Scalable Rule-Based Deep Reinforcement Learning Framework for the Next Release Problem
Hemanth Gudaparthi, Nan Niu, Prem Swaroopaanda Ramalingam, and Laxman Chowdary Kakara — *Governors State University, USA; University of North Florida, USA* . . . 222

DynCA: An Effective Algorithm for 3-Wise Combinatorial Interaction Testing in Highly Configurable Systems
Ruizhi Shi, Shuangyu Lyu, Renyu Yang, Wei Wu, Chanjuan Liu, Chunming Hu, and Chuan Luo — *Beihang University, China; Central South University, China; Dalian University of Technology, China* . . . 234

Effective and Efficient Context Retrieval via Partial Dependency Graph for Repository-Level Code Generation
Zhongxin Liu, Zhonghao Jiang, Zhifan Ye, Haoye Wang, Jiakun Liu, and Xiaoxue Ren — *Zhejiang University, China; Hangzhou City University, China; Harbin Institute of Technology, China* . . . 247

When Ambiguity Meets Atypicality: Dual-Perspective Test Input Prioritization for DNNs
Haoran Li, Shihai Wang, Bin Liu, Jialuo Chen, Wenjing Zhu, Yu Liu, Tengfei Shi, and Shudi Guo — *Beihang University, China; Zhejiang University, China* . . . 260

Not In My Git Yard: Catching Backdoors at Commit and Release Time
Dimitri Kokkonis, Michaël Marcozzi, and Stefano Zacchiroli — *Université Paris-Saclay, France; CEA List, France; Télécom Paris, France; Institut Polytechnique de Paris, France* . . . 273

EffiHolmes: Differential Profiling-Guided Repository Level Time Inefficiency Fix Localization
Haowen Yang, Yun Peng, and Zishuo Ding — *Hong Kong University of Science and Technology, Guangzhou, China; Chinese University of Hong Kong, Hong Kong* . . . 285

Maximal Format-Free Data Repair
Zijian Luo, Xi Wu, Hong Jin Kang, Alan Fekete, and Rahul Gopinath — *University of Sydney, Australia* . . . 298

AgentExecutor: Partial Code Execution via Agentic Context Generation
Junkai Chen, Chengran Yang, Xing Hu, Zhenhao Li, Xin Xia, and David Lo — *Singapore Management University, Singapore; Zhejiang University, China; York University, Canada* . . . 310

Automated Lemma Discovery in Agentic Program Verification
Huan Zhao, Haoxin Tu, Zhengyao Liu, Martin C. Rinard, and Abhik Roychoudhury — *National University of Singapore, Singapore; Massachusetts Institute of Technology, USA* . . . 323

Regression Accumulation in Multi-turn LLM Programming Conversations
Yonghui (Andie) Huang, Lin Ma, Amjed Tahir, Qian Zhang, Liwen Xiao, and Lysa Xiao — *Massey University, New Zealand; Alibaba Research Center for Complexity Sciences, China; Hangzhou Normal University, China; University of Otago, New Zealand* . . . 336

TheBotCompany: Self-Organizing Multi-agent Systems for Continuous Software Development
Wenhan Lyu, Yue Xiao, Yixuan Zhang, and Yifan Sun — *William & Mary, USA* . . . 349

When Comments Mislead: An Empirical Study of Outdated Comments in LLM-Based Code Completion
Zhikun Xia, Yuan Huang, Xiangping Chen, and Zibin Zheng — *Sun Yat-sen University, China* . . . 362

Property-Based Testing of Wafer Handling Robots via Model-Based Planning and Fuzzing
Ruiyang Xu, Jingjing Liang, Guoyue Zheng, Geguang Pu, and Ting Su — *East China Normal University, China* . . . 374

ATLAS: Agentic Taxonomy of LArge-Scale Software Ecosystems
Junyi Lu, Mengyao Lyu, Jiahui Wu, Lei Yu, Chengwei Liu, Fengjun Zhang, Li Yang, Chun Zuo, and Yang Liu — *Institute of Software at Chinese Academy of Sciences, China; University of Chinese Academy of Sciences, China; Nanyang Technological University, Singapore; Nankai University, China; Sinosoft, China* . . . 386

SimServing: Native-Execution Simulation for Evolution-Resilient LLM Serving Configuration Tuning
Yujie Huang, Jiazhen Gu, Zhihan Jiang, and Michael R. Lyu — *Chinese University of Hong Kong, Hong Kong* . . . 398

LLM Agents Can See Code Repositories
Dongjian Ma, Silin Chen, Yufei Yang, Yuling Shi, Yanfu Yan, and Xiaodong Gu — *Shanghai Jiao Tong University, China; Xi'an Jiaotong University, China; Zhejiang University, China* . . . 411

Are Reasoning Traces Faithful in LLM-Based Code Vulnerability Detection?
Jialai Wang, Juncheng Lu, Songtao Yang, Bingyi He, Zeyu Gao, Chao Zhang, and Shuai Wang — *National University of Singapore, Singapore; Tsinghua University, China; Zhongguancun Laboratory, China; Huazhong University of Science and Technology, China; Hong Kong University of Science and Technology, Hong Kong* 424

CtxFuzz: Context-Aware Directed Fuzzing via Runtime Fuzzer-Agent Cooperation
Jiho Kim, Dongkwan Kim, HyungSeok Han, Youngjoon Kim, Soyeon Park, Sangwoo Ji, Joshua Wang, Dae R. Jeong, and Taesoo Kim — *Georgia Institute of Technology, USA; Microsoft, USA; Independent Researcher, USA; Samsung Electronics, Republic of Korea; Seoul National University, Republic of Korea* 437

SustainSR: Interpretable OSS Sustainability Prediction via Symbolic Regression
Yang Deng, Mustafa Misir, Jinfu Chen, and Jifeng Xuan — *Wuhan University, China; Duke Kunshan University, China* 449

Stop When It Matters: Detectability-Guided Microbenchmarking for Performance Regression Testing
Zongxiong Chen, Derui Zhu, Jinfu Chen, Kundi Yao, Alexander Pretschner, Weiyi Shang, Manfred Hauswirth, and Sonja Schimmler — *Fraunhofer FOKUS, Germany; Rochester Institute of Technology, USA; Wuhan University, China; Ontario Tech University, Canada; TU Munich, Germany; fortiss, Germany; University of Waterloo, Canada; Technische Universität Berlin, Germany* 461

Evolution-Aware Regression Test Prioritization of ML-Enabled Systems using Gradient-Based Behavior Vectors
Eunho Cho, Donghwan Shin, and In-Young Ko — *KAIST, Republic of Korea; University of Sheffield, UK* 474

Multi-level Code Optimization via Mixture of Prompts
Yun Peng, Jun Wan, Jiakun Liu, Shuzheng Gao, David Lo, and Xiaoxue Ren — *Fudan University, China; Zhejiang University, China; Harbin Institute of Technology, China; Chinese University of Hong Kong, Hong Kong; Singapore Management University, Singapore; Hangzhou High-Tech Zone (Binjiang) Institute of Blockchain and Data Security, China* 486

Evaluating the Maintainability of ViewModel Tests using Multi-modal DSLs: An Industrial Mixed-Methods Study
Mario Fuksa, Sandro Speth, and Steffen Becker — *University of Stuttgart, Germany; TU Munich, Germany* 499

Sound and Efficient Statistical Model Checking for Probabilities and Bounded Rewards
Hao Bu, Lin Huang, Tao Wei, and Jingyi Wang — *Ant Group, China; Zhejiang University, China* 511

Tracing the Invisible: Semantic Message Flow Discovery via Data Contracts in Real-World Distributed Systems
Youlong Chen, Jie Lu, Mingtao Huang, Chenghang Shi, Yongheng Huang, Haofeng Li, Dong Liu, Mengna Ma, Yong Liu, Qinfen Hao, and Lian Li — *Institute of Computing Technology at Chinese Academy of Sciences, China; University of Chinese Academy of Sciences, China; Zhongguancun Laboratory, China; Qi An Xin Technology Group, China* 523

How Your Credentials Are Leaked by LLM Agent Skills: An Empirical Study
Zhihao Chen, Ying Zhang, Yi Liu, Gelei Deng, Yuekang Li, Yanjun Zhang, Jianting Ning, Leo Yu Zhang, Lei Ma, and Zhiqiang Li — *Griffith University, Australia; Wake Forest University, USA; Nanyang Technological University, Singapore; University of New South Wales, Australia; Zhejiang Sci-Tech University, China; University of Tokyo, Japan; University of Alberta, Canada; Independent Researcher, China* 536

WebCQ: Cooperative Multi-agent Deep Reinforcement Learning for Scalable Web GUI Testing
Yujia Fan, Sinan Wang, Zebang Fei, Yao Qin, Huaxuan Li, and Yepang Liu — *Southern University of Science and Technology, China* . 549

MalSkills: Detecting Malicious Skills in the Agentic Supply Chain via Neuro-symbolic Reasoning
Shenao Wang, Junjie He, Yanjie Zhao, Yayi Wang, Kan Yu, and Haoyu Wang — *Huazhong University of Science and Technology, China; Ant Group, China* 562

Less Is More: Tuning Configurable Systems with Imperfect Fidelity
Yulong Ye, Miqing Li, and Tao Chen — *University of Birmingham, UK* 575

Programming by Chat: A Large-Scale Behavioral Analysis of 11,579 Real-World AI-Assisted IDE Sessions
Ningzhi Tang, Chaoran Chen, Zihan Fang, Gelei Xu, Maria Dhakal, Yiyu Shi, Collin McMillan, Yu Huang, and Toby Jia-Jun Li — *University of Notre Dame, USA; Vanderbilt University, USA* 588

Understanding Bugs in Modern Agentic Frameworks: A Study of Symptoms, Root Causes, and Triggering Conditions
Xiaowen Zhang, Hannuo Zhang, and Shin Hwei Tan — *Concordia University, Canada* 601

Neuro-symbolic Requirements Elicitation: Utilizing Formal Verification Counterexamples as Contextual Prompts
Okba Tibermacine and Chouki Tibermacine — *National School of Artificial Intelligence, Sidi Abdallah, Algeria; University of Southern Brittany, France* 613

SmartPoC: Validating Smart Contract Vulnerability Findings through PoC Generation and Execution
Longfei Chen, Ruibin Yan, Taiyu Wong, Yiyang Chen, Jialai Wang, and Chao Zhang — *Tsinghua University, China; Peking University, China; National University of Singapore, Singapore* 624

Every Scenario Matters: Scenario-Guided White-Box Testing for Static Code Checkers Powered by LLMs
Jun Liu, Liqiang Ji, Jiwei Yan, Linjie Pan, Jinsong Ren, Jun Yan, and Jian Zhang — *Institute of Software at Chinese Academy of Sciences, China; University of Chinese Academy of Sciences, China; Beijing Key Laboratory of Intelligent Software Engineering, China; Huawei Technologies, China* 637

Adaptive Proof Refinement with LLM-Guided Strategy Selection
Minghai Lu, Zhe Zhou, Danning Xie, Songlin Jia, Benjamin Delaware, and Tianyi Zhang — *Purdue University, USA; Meta, USA* . . . 649

Assessing the Cross-Version Applicability of Java Library Vulnerability Exploits
Zirui Chen, Qi Zhan, Jiayuan Zhou, Xing Hu, Xin Xia, and Xiaohu Yang — *Zhejiang University, China; Queen's University, Canada; Hangzhou High-Tech Zone (Binjiang) Institute of Blockchain and Data Security, China* . . . 662

TDRepro: A Neurosymbolic Approach to Reproducing Timing-Dependent Flaky Test Failures
Shanto Rahman, Talank Baral, August Shi, and Wing Lam — *University of Texas at Austin, USA; George Mason University, USA* . . . 675

Think Before You Code: Dual Reasoning for the NLSafety–Utility Trade-Off in LLM Code Generation
Honghao Tan, Haibo Wang, and Shin Hwei Tan — *Concordia University, Canada* . . . 687

Fragmented Markers, Mixed Results: A Systematic Review of AI Coding Assistants and Developer Productivity
Annemarie Wittig, Alina Mailach, and Norbert Siegmund — *Leipzig University, Germany* . . . 700

How Reasoning Shapes Social Bias in LLM-Generated Code?
Weifeng Sun, Jieke Shi, Zhou Yang, Yuchen Chen, Hongyan Li, Meng Yan, and David Lo — *Singapore Management University, Singapore; University of Alberta, Canada; Alberta Machine Intelligence Institute, Canada; Nanjing University, China; Chongqing University, China* 713

From Runnable Code to Shippable Applications: Test-Driven Development for Full-Stack Web Application Generation
Yuxuan Wan, Tingshuo Liang, Jiakai Xu, Jingyu Xiao, Yintong Huo, and Michael R. Lyu — *Chinese University of Hong Kong, Hong Kong; Columbia University in the City of New York, USA; Singapore Management University, Singapore* . . . 726

Planner-Fuzz: Overcoming Diagnostic Saturation in UAV Planners via Boundary-Oriented Fuzzing
Rongbo Chen, Chengcheng Zhao, Peng Cheng, Jiming Chen, and Taegyu Kim — *Zhejiang University, China; Pennsylvania State University, USA* . . . 739

Automated Assertion Generation and Regression Testing for Machine Learning Notebooks
Yingao (Elaine) Yao, Vedant Nimje, Varun Viswanath, and Saikat Dutta — *Cornell University, USA; Veermata Jijabai Technological Institute, India; Dwarkadas J Sanghvi College of Engineering, India* . . . 752

Learning the Lexical Structure of Black-Box Systems by Parsing Systematic String Edit Mutations
Moeketsi Raselimo, Lars Grunske, and Bernd Fischer — *Humboldt University of Berlin, Germany; Stellenbosch University, South Africa* 764

KQFuzz: Knowledge-Guided Fuzzing for Quantum Libraries via Large Language Models
Fuyuan Xia, Qixin Zhang, Chenhao Ying, Haojin Zhu, Shuai Wang, Yuan Luo, Pingchuan Ma, and Yuxuan Du — *Shanghai Jiao Tong University, China; Nanyang Technological University, Singapore; Hong Kong University of Science and Technology, Hong Kong; Zhejiang University of Technology, China* . . . 777

What Breaks When LLMs Code? Characterizing Operational Safety Failures of Agentic Code Assistants
Alif Al Hasan and Sumon Biswas — *Case Western Reserve University, USA* . . . 790

When Compression Becomes an Attack Surface: Black-Box Attacks on Prompt-Compressed LLM Agents
Zesen Liu, Zhixiang Zhang, Yuchong Xie, and Dongdong She — *Hong Kong University of Science and Technology, China; Fudan University, China* . . . 803

ExplainBench: Evaluating Code Explanations from Agents
Zhiyuan Pan, Sungmin Kang, Imam Nur Bani Yusuf, and Abhik Roychoudhury — *Zhejiang University, China; National University of Singapore, Singapore* . . . 815

Finding Typing-Related Bugs in JVM Implementations
Yanzhou Lu, Junjie Chen, Yingquan Zhao, Tianchang Gao, and Zan Wang — *Tianjin University, China* . . . 828

Compressing Code Context for LLM-Based Issue Resolution
Haoxiang Jia, Earl T. Barr, and Sergey Mechtaev — *Peking University, China; University College London, UK* . . . 841

When Dependencies Depend: A Large-Scale Study on Conditional Dependencies and Their Security Implications in C/C++
Yifeng Di, Juewei Zhang, and Tianyi Zhang — *Purdue University, USA; Carnegie Mellon University, USA* . . . 854

Towards Fully Automated Medical Imaging Code Generation via Validation-Based Context Engineering
Zixiao Zhao, Jing Sun, Zhe Hou, Cheng-Hao Cai, Qian Liu, Mengze Li, Zijian Zhang, and Jin Song Dong — *University of Auckland, New Zealand; Griffith University, Australia; Monash University, Suzhou, China; Beijing Institute of Technology, China; National University of Singapore, Singapore* . . . 867

From Plan to Action: How Well Do Agents Follow the Plan?
Shuyang Liu, Saman Dehghan, Jatin Ganhotra, Martin Hirzel, and Reyhaneh Jabbarvand — *University of Illinois at Urbana-Champaign, USA; IBM, USA* . . . 880

Needle in the Repo: Diagnosing Maintainability Failures in AI-Generated Repository Edits
Haichao Zhu, Qian Zhang, Jiyuan Wang, Zhaorui Yang, and Yuxin Qiu — *Qualcomm, USA; University of California at Riverside, USA; Tulane University, USA* . . . 892

An Empirical Study of Data Access Practice in Android AR Apps to Understand User Privacy Risks
Sabbir Hussain Meraj, Long Trac, Xiaoyin Wang, Xusheng Xiao, and Wei Wang — *University of Texas at San Antonio, USA; Arizona State University, USA* . . . 904

Clarity Is Not Assumed: Understanding LLM-Based Code Generation under Ambiguous Requirements
Di Yang, Xinou Xie, Xiuwen Yang, Ming Hu, Yihao Huang, Yueling Zhang, Weikai Miao, Ting Su, Chengcheng Wan, and Geguang Pu — *East China Normal University, China; Shanghai Innovation Institute, China* . . . 916

Conflict Extraction in Probabilistic Datalog Analyses
Siyu Chen, Chungha Sung, Xuyang Li, and Jingbo Wang — *Purdue University, USA; Amazon Web Services, USA* . . . 929

EMERGE: Checking Implementation Equivalence for LLMs via Rule Synthesis
Qi Zhan, Xing Hu, Xin Xia, and Shanping Li — *Zhejiang University, China; Hangzhou High-Tech Zone (Binjiang) Institute of Blockchain and Data Security, China* . . . 942

SpecTrum: Specification-Guided Differential Fuzzing for Ethereum Consensus Clients
Seokhun Jeong, Gyeongmin Dan, Sukyoung Ryu, and Sungjae Hwang — *KAIST, Republic of Korea; Sungkyunkwan University, Republic of Korea* . . . 955

Where Does Balance Break? Boundary Discovery for Game Balance Testing under a Finite Simulation Budget
Hiroki Mukai, Yusaku Kato, Norihiro Yoshida, Erina Makihara, and Katsuro Inoue — *Ritsumeikan University, Japan* . . . 967

On the Effects of Customized Configurations of Static Code Analysis Tools: A Prospective Cohort Study of SonarQube Cloud
Sabato Nocera, Sira Vegas, and Giuseppe Scanniello — *University of Salerno, Italy; Polytechnic University of Madrid, Spain* . . . 980

RefineAct: Automatic Runtime Verification of LLM Agent Actions
Fraol Batole, Foutse Khomh, and Hridesh Rajan — *Tulane University, USA; Polytechnique Montréal, Canada* . . . 993

Verifier-in-the-Loop LLM Solving of Heap Constraints for Concolic Execution
Shaoran Xia, Leyi Cheng, Caihua Dong, Dongdong She, Bo Wang, Xin Peng, and Zhen Dong — *Fudan University, China; Soochow University, China; Hong Kong University of Science and Technology, China; Beijing Jiaotong University, China* . . . 1005

RealAIGC: Towards More Realistic Evaluation of AI-Generated Code Detection in Real-World Code Repository
Wentao Chen, Yufei Chen, Huiqun Yu, Guisheng Fan, Kaiwen Zhi, Peite Guo, and Jianyuan Yu — *East China University of Science and Technology, China; Shanghai Engineering Research Center of Smart Energy, China* . . . 1018

Hybrid Fuzzing in Input Generation Logic
Siwei Wei, Shihao Zhu, Ruijie Meng, and Yan Cai — *Institute of Software at Chinese Academy of Sciences, China; University of Chinese Academy of Sciences, China; National University of Singapore, Singapore* . . . 1031

Understanding the Code and Architectural Smells Generated by LLMs and Autonomous Agents
Yuecai Zhu, Nikolaos Tsantalis, and Peter C. Rigby — *Concordia University, Canada* . . . 1043

Toward Inclusive Programming Support for Novice Programmers with Dyslexia: Insights from a Comparative Study
Zihan Fang, Janice Chung, Ruijia Chen, Marcia Barnes, Yuhang Zhao, and Yu Huang — *Vanderbilt University, USA; University of Wisconsin-Madison, USA* . . . 1054

AEG-Edit: Bridging API Evolution and Code Semantics for Accurate API Usage in Code Generation
Yitong Lin, Haoye Tian, Dehai Zhao, Chao Ni, Xiaohu Yang, and Xiaoxue Ren — *Zhejiang University, China; Hangzhou High-Tech Zone (Binjiang) Institute of Blockchain and Data Security, China; Aalto University, Finland* . . . 1067

ReqEvolve: User-Oriented Software Self-Evolution through Automatic Requirement Interpretation
Md Asif Iqbal Fahim and Alessio Ferrari — *University College Dublin, Ireland; National Research Council, Italy* . . . 1080

Testing IFrame Rendering in Web Browsers
Dahyeon Park, Yuseok Jeon, Seongil Wi, and Mijung Kim — *UNIST, Republic of Korea; Korea University, Republic of Korea* . . . 1093

Real-World Perturbation Testing of Autonomous Driving Systems
Stefano Carlo Lambertenghi, Matthias Weil, and Andrea Stocco — *TU Munich, Germany; fortiss, Germany* . . . 1106

SwiftGram: Dynamic Lexical Forking for Precise and Efficient Grammar-Constrained Decoding
Shuyi Ling, Liangyi Kang, Jie Liu, Shuai Wang, Zhirou Ma, and Dan Ye — *Institute of Software at Chinese Academy of Sciences, China* 1119

GraphLedger: Repository-Level Vulnerability Detection via Graph-Based Compression and Assumption Validation
Thi-Hong-Cuc Le, Hoang-Quoc-Bao Hua, and Xuan-Bach Le — *Ho Chi Minh City University of Technology - Vietnam National University Ho Chi Minh City, Vietnam* . . . 1132

Escaping the Self-Repair Trap: Improving Test Oracle Generation via Dual-Context Awareness
Kefan Li, Hongyue Yu, and Yuan Yuan — *Beihang University, China* . . . 1144

PSearch: Search-Based Patch Generation in the Era of LLM-Based Automated Program Repair
Haichuan Hu, Ye Shang, Weifeng Sun, and Quanjun Zhang — *Nanjing University of Science and Technology, China; Nanjing University, China; Singapore Management University, Singapore* . . . 1157

Same Scrutiny, More Time: Eye Tracking Insights into Reviewing LLM-Labelled Code
Ranim Khojah, Francisco Gomes de Oliveira Neto, Mazen Mohamad, Julian Frattini, and Philipp Leitner — *Chalmers University of Technology, Sweden; University of Gothenburg, Sweden; RISE Research Institutes of Sweden, Sweden* . . . 1170

Breaking Customized LLMs for Coding: Automated Red Teaming for Instruction Backdoor Attacks
Yuchen Chen, Wei Cheng, Yuan Xiao, Weisong Sun, Chunrong Fang, Yang Liu, Zhenyu Chen, and Baowen Xu — *Nanjing University, China; Nanjing University of Aeronautics and Astronautics, China; Nanyang Technological University, Singapore* 1182

On the Reliability of Code Comprehension Proxies
Erfan Arvan, Nadeeshan De Silva, Oscar Chaparro, and Martin Kellogg — *New Jersey Institute of Technology, USA; William & Mary, USA* 1195

Discovering Performance Archetypes: Critical-Path-Aware Pattern Analysis and Regression Detection
Kaveh Shahedi, Heng Li, Maxime Lamothe, and Foutse Khomh — *Polytechnique Montréal, Canada* 1208

Barista: Synthesizing Typestate Specifications with LLM Agents
Catarina Gamboa, Paulo Canelas, Ricardo Costa, Márcio Caetano, Jonathan Aldrich, and Alcides Fonseca — *Carnegie Mellon University, USA; University of Lisbon, Portugal* 1221

A Longitudinal Study of Android Apps Signing Key Protection
Mark Huasong Meng, Qing Zhang, Weirao Lu, and Chunyang Chen — *University College Dublin, Ireland; JD Group, China; TU Munich, Germany* 1234

Clarity: From Formal System Designs to Verified RL Controllers
Austin O'Quinn, Conor Snedeker, Max Taylor, Lance Joneckis, and Christopher Stewart — *Ohio State University, USA; Boise State University, USA; Idaho National Lab, USA* 1246

Towards Verifiable Fairness Testing: Auditable Coverage over Data-Supported Input Domains
Yingqian Guo, Wentian Zhao, and Tian Song — *Beijing Institute of Technology, China* 1258

PL4SA: Optimized Partial Library Selection for Efficient Static Analysis
Guohao Feng, Yifei Lu, and Minxue Pan — *Nanjing University, China* 1270

VeriSBOM: Secure and Verifiable SBOM Sharing via Zero-Knowledge Proofs
Gianpietro Castiglione, Shahriar Ebrahimi, and Narges Khakpour — *Newcastle University, UK* 1283

Bifrost: Empowering Pretrained Language Model with Fallibility Representation for Log-Based Fault Diagnosis
Minghua He, Tong Jia, Lingzhe Zhang, Chiming Duan, Xinlong Zhao, Leyi Pan, Cheng Wang, Kangjin Wang, Yinghao Yu, Liping Zhang, Yifan Wu, and Ying Li — *Peking University, China; Tsinghua University, China; Alibaba Group, China* 1295

All or Nothing: How Library Type Annotations Affect Client-Side Errors in Python
Eric Asare, Luca Di Grazia, and Sarah Nadi — *New York University Abu Dhabi, United Arab Emirates; University of St. Gallen, Switzerland* 1308

Tensor-Based Batch Fuzzing with Adaptive Perturbation Scaling for Deep Neural Networks
Guanqin Zhang and Yulei Sui — *University of New South Wales, Australia* 1320

ChangeWiki: Repository-Level Change Summarization Augmented with Architecture Semantics
Zijun Wang, Wuxia Jin, Jiale Li, Zhanhong Mou, and Ting Liu — *Xi'an Jiaotong University, China* 1333

Refine2Diff: Detecting Protocol Specification–Implementation Inconsistencies via Specification-Driven Code Refinement
Yuekun Wang, Lili Quan, and Xiaofei Xie — *Singapore Management University, Singapore* 1346

Free Is More: Empowering Software Agents via Issue-Free Trajectory Learning and Entropy-Aware RLVR Training
Xin-Cheng Wen, Binbin Chen, Haoxuan Lan, Hang Yu, Peng Di, and Cuiyun Gao — *Ant Group, China; Zhejiang University, China; University of New South Wales, Australia; Chinese University of Hong Kong, China* 1358

How Quantum Bugs Live and Die: A Lifecycle-Based Empirical Study of Bugs in Quantum Software
Yasai Shi, Xiangxin Meng, Xiangjie Huang, Jian Zhang, Tianyu Wo, and Xu Wang — *Beihang University, China* 1371

Automated Table Reproduction via Code Generation
Doehyun Baek and Michael Pradel — *CISPA Helmholtz Center for Information Security, Germany* 1384

MicroAgent: Context-Augmented Multi-agent Framework for Automatic Microservice Decomposition
Zishan Su, Junjie Huang, Shiwen Shan, Xingyan Chen, Hui Zeng, Yuxin Su, Yanlin Wang, and Michael R. Lyu — *Chinese University of Hong Kong, China; Sun Yat-sen University, China* 1396

Optimization-Free Repair of Code LLMs: A Closed-Form Fisher-Guided Bayesian Approach
Zhiyu Duan, Chong Bian, Yizhuo Zhang, and Shunkun Yang — *Beihang University, China; China Mobile, China* 1409

Pattern over Pixels: Measuring Pattern Completion Bias in Multimodal Code Generation
Khai-Nguyen Nguyen, Oscar Chaparro, and Antonio Mastropaolo — *William & Mary, USA* 1421

GPU-Accelerated Belief Propagation for Program Analysis
Haoyu Feng and Xin Zhang — *Peking University, China* 1434

Inferring the Shape of Data Frames in R Programs using Abstract Interpretation
Oliver Gerstl, Florian Sihler, and Matthias Tichy — *Ulm University, Germany* 1446

Translation Tag Team: Formal Rules and LLMs Translate More Macros Together Than Apart
Brent Pappas, Joseph Zalusky, Zachary Burkett, and Paul Gazzillo — *University of Central Florida, USA* 1457

ProbGuard: Proactive Runtime Monitoring for LLM Agent Safety via Probabilistic Prediction
Haoyu Wang, Christopher M. Poskitt, Jiali Wei, and Jun Sun — *Singapore Management University, Singapore; Xi'an Jiaotong University, China* . . . 1470

Tangent: An Empirical Study of Testing Practices for LLM-Based Agent Applications
Rangeet Pan, Tyler Stennett, Divya Sankar, Bridget McGinn, Alessandro Orso, Raju Pavuluri, Saurabh Sinha, and Maja Vukovic — *IBM Research, USA; Georgia Institute of Technology, USA; University of Georgia, USA* . . . 1483

A Unified Model for Cross-Domain Clone Detection via Model Merging
Palash R. Roy, Banani Roy, Kevin A. Schneider, and Chanchal K. Roy — *University of Saskatchewan, Canada* . . . 1496

Don't Retest Everything: Auditable Regression Scoping for Safety-Critical Systems
Francesco Basciani, Daniele Masti, Patrizio Pelliccione, Alberto Petrucci, Franco Raimondi, Emerson Sales, and Paolo Serri — *Gran Sasso Science Institute, Italy; Gran Sasso Tech, Italy; Thales Alenia Space Italia, Italy* . . . 1509

Recovering Revisions of Pull Requests with Altered History
Gengyi Sun, Georges Aaron Randrianaina, Tao Xiao, Yasutaka Kamei, and Shane McIntosh — *University of Waterloo, Canada; Kyushu University, Japan* . . . 1521

Making Theft Useless: Adulteration-Based Protection of Proprietary Knowledge Graphs in GraphRAG Systems
Weijie Wang, Peizhuo Lv, Yan Wang, Weiqing Huang, Rujie Dai, Guokun Xu, Qiujian Lv, Hangcheng Liu, Hao Hu, and Jiaheng Zhang — *Institute of Information Engineering at Chinese Academy of Sciences, China; University of Chinese Academy of Sciences, China; Nanyang Technological University, Singapore; Beijing University of Technology, China; Henan Key Laboratory of Information Security, China; National University of Singapore, Singapore* . . . 1534

Decompiling the Web: Automated Semantic Recovery of Post-compilation Abstraction Leaks
Shriyans Sudhi and Yinxi Liu — *Rochester Institute of Technology, USA* . . . 1547

AutoSQL: Extracting SQL Templates from Imperative ORM Code in Large-Scale Repositories
Junsong Pu, Yichen Li, Zhuangbin Chen, Zhihan Jiang, and Zibin Zheng — *Sun Yat-sen University, Zhuhai, China; Chinese University of Hong Kong, Hong Kong* . . . 1560

SIMD-Accelerated Sparse Bit-Vectors for Pointer Analysis
Zhaoyang Tan, Peisen Yao, and Kui Ren — *Zhejiang University, China* . . . 1573

DesignFix: Context-Augmented Repair of Design Rule Violations
Mainul Hossain and Thomas D. LaToza — *George Mason University, USA* . . . 1586

LaMAR: Latent Multi-agent Collaboration via KV-Cache Communication for Automated Program Repair
Rui Jiao, Yue Zhang, Zhexuan Feng, and Jinku Li — *Xidian University, China; Shandong University, China* . . . 1599

Enhancing Trace-Based Root Cause Analysis for Microservice Systems via Code Change Understanding
Min Zhang, Chenxi Zhang, Senyu Xie, Shihong Chen, Lei Wu, and Xin Peng — *Xidian University, China; Fudan University, China; McDonald's, China* . . . 1612

Assessing Uncertainty in Performance Modeling: Conformal Prediction vs. Bayesian Regression
Stefan Jahns, Johannes Dorn, Max Weber, Sven Apel, and Norbert Siegmund — *Leipzig University, Germany; Saarland University, Germany; ScaDS.AI, Germany* . . . 1625

RACE-Bench: A Reasoning-Augmented Benchmark for Repository-Level Code Agents on Feature Addition
Shuhan Liu, Zhiyi Zhao, Xing Hu, Kui Liu, Xiaohu Yang, and Xin Xia — *Zhejiang University, China; Independent Researcher, China; Hangzhou High-Tech Zone (Binjiang) Institute of Blockchain and Data Security, China* . . . 1638

How Effective Are NPM Malicious Package Detectors? A Large-Scale Empirical Study
Wenbo Guo, Zhongwen Chen, Zhengzi Xu, Chengwei Liu, Ming Kang, Shiwen Song, Chengyue Liu, Yijia Xu, Weisong Sun, and Yang Liu — *Nanyang Technological University, Singapore; Sichuan University, China; Imperial Global Singapore, Singapore; Nankai University, China; Singapore Management University, Singapore* . . . 1651

Localizing Data Annotation Faults in Semantic Segmentation Systems
Pengyi Yang, Teng Long, Zhiwu Xu, and Linhai Jing — *China University of Geosciences (Beijing), China; Shenzhen University, China* 1664

"Impossible to Hide Secret ...": Uncovering Security and Privacy Issues in LLM-Native IDEs
Mostafijur Rahman Akhond, Md Afif Al Mamun, Gias Uddin, and Song Wang — *York University, Canada; University of Calgary, Canada* 1677

Smart Contract Synthesis via Multi-modal Specifications
Tanglin Chen, Haoxian Chen, Huilin Xiang, and Yuepeng Wang — *ShanghaiTech University, China; Simon Fraser University, Canada* 1690

BDiff: Block-Aware and Accurate Text-Based Code Differencing
Yao Lu, Wanwei Liu, Tanghaoran Zhang, Kang Yang, Yang Zhang, Wenyu Xu, Longfei Sun, Xinjun Mao, Shuzheng Gao, and Michael R. Lyu — *National University of Defense Technology, China; Chinese University of Hong Kong, China* . . . 1702

PCodeTrans: Translate Decompiled Pseudocode to Trace-Level Equivalent Source Code
Yuxin Cui, Zeyu Gao, Siliang Qin, Yuanda Wang, Jiaming Wang, Xize Luo, Peng Zhou, and Chao Zhang — *Tsinghua University, China; Institute of Information Engineering at Chinese Academy of Sciences, China; Peking University, China; Hunan University, China* . . . 1715

AlphaCu: A Transformation-Driven Synthesis Framework for LLM-Based GPU Kernel Generation
Yangjie Zhou, Xing Ma, Zihan Liu, Qing Wang, Weihao Cui, Wu Sun, Hanjing Wang, Binhang Qi, Jingwen Leng, Yun Lin, Minyi Guo, and Jin Song Dong — *National University of Singapore, Singapore; Shanghai Jiao Tong University, China; Shanghai Qi Zhi Institute, China; Nanjing University, China; Harvard University, USA* 1728

How Do LLMs Read Bug Reports? An Empirical Study of Attention in LLMs for Automated Program Repair
Ramtin Ehsani, Irene Manotas, Saurabh Pujar, Luca Buratti, and Preetha Chatterjee — *Drexel University, USA; IBM Research, USA* . . 1741

Mind the Gap: Do Widely-Used SAST Tools Really Cover the Language-Specific Top CWEs?
Zhilin Li, Yiran Nie, Xianglong Qi, Jiwei Yan, and Jun Yan — *Institute of Software at Chinese Academy of Sciences, China; University of Chinese Academy of Sciences, China; Hangzhou Institute for Advanced Study at University of Chinese Academy of Sciences, Hangzhou, China; Beijing University of Technology, China; Beijing Key Laboratory of Intelligent Software Engineering, China* 1754

Smart Brain: Semantic Anomaly Detection for Operational Time Series in Large Scale Service Systems
Hang Cui, Zexin Wang, Jingjing Li, Juncheng Hu, Haotian Si, Cenjie Hu, Quan Zhou, Yongchang Hu, Lei Han, Dan Pei, Changhua Pei, and Gaogang Xie — *University of Chinese Academy of Sciences, China; Computer Network Information Center at Chinese Academy of Sciences, China; Jilin University, China; Independent Researcher, China; Shenyang Institute of Automation at Chinese Academy of Sciences, China; Beijing Academy of Science and Technology, China; Huawei Technologies, China; Tsinghua University, China* 1767

AgentChaos: Chaos Engineering for Agent Systems via Programmatic Fault Injection
Gou Tan, Zhensu Sun, Jieke Shi, Ting Zhang, Zilong He, Qingfu Wu, Shuai Liang, Weifeng Sun, Junda He, Pengfei Chen, Chuanfu Zhang, Lwin Khin Shar, and David Lo — *Sun Yat-sen University, Guangzhou, China; Singapore Management University, Singapore; Monash University, Australia; China Unicom Software Research Institute, Beijing, China* 1780

Precise and Efficient Static Data Race Detection
Pei Wang, Zhihang Sun, and Fei He — *Tsinghua University, China; Key Laboratory for Information System Security, China; Beijing National Research Center for Information Science and Technology, China* 1793

Best-Effort GR(1) Synthesis: Exploiting Environmental Cooperation under Unrealizability
Sirui Liu, Yating Zhang, Chi Hu, and Wei Dong — *National University of Defense Technology, China; China Academy of Engineering Physics, China* 1806

TELLER: Non-intrusive Cross-Layer Root-Cause Analysis for LLM Inference
Ruilin Xu, Junyi Li, Pengfei Chen, and Zongxuan Xie — *Sun Yat-sen University, China* 1819

Predicate-Guided Synthesis of Input Generators via Gentoo
Vasudev Vikram and Rohan Padhye — *Carnegie Mellon University, USA* 1831

DepUpgrade: Automating Dependency Upgrade through State-Path Exploration
Xiangxi Ma, Yifan An, Wentong Tian, Xuanqi Wang, Qingao Dong, Xiang Gao, and Hailong Sun — *Beihang University, China* . . . 1844

PWCC: Probability-Weighted Combinatorial Coverage for Risk-Informed Scenario Testing of Autonomous Systems
Borhaneddine Hamadou, Chokri Mraidha, and Nikolaos Matragkas — *Université Paris-Saclay, France; CEA List, France* 1857

Lost in Translation, Found in Summary: Summary-Driven Supervision for Cross-Language Code Retrieval
Zhijie Liu, Yisi Liu, Linxi Jiang, Yutian Tang, Lili Xiao, and Zhipeng Xue — *Independent Researcher, China; Donghua University, China; University of Glasgow, UK; Zhejiang University, China* 1870

ExeCRE: Execution-Consistency Guided Reliability Estimation for Self-Correcting Code Generation
Yiru Dong, Richong Zhang, Fanshuang Kong, and Si Chen — *Beihang University, China* 1883

DAGGER: Dynamic Agentic Graph-Guided Efficient Reasoner for Container Image Dependency Optimization
Shikhar Vashistha and Neetesh Kumar — *IIT Roorkee, India* 1896

RamFuzz: LLM-Guided Greybox Fuzzing for Spatial Memory Corruption via Valid Range Violation
Shangzhi Xu, Wei Song, Yuekang Li, Nan Sun, Muhammad Ejaz Ahmed, Willy Susilo, Benjamin Turnbull, Xiao Cheng, and Siqi Ma — *University of New South Wales, Australia; CSIRO's Data61, Australia; University of Wollongong, Australia; Macquarie University, Australia* 1908

To Ban or Not to Ban? How Open Source Projects Govern GenAI Contributions
Wenhao Yang, Runzhi He, and Minghui Zhou — *Peking University, China* 1921

GraphDroid: Asynchronous LLM-Based Mobile App GUI Testing via History-Aware Exploration and Hybrid Intent Fulfillment
Xiaolei Li, Jialun Cao, Zhijian Hou, Yuzhi Zhao, Yepang Liu, and Shing-Chi Cheung — *Hong Kong University of Science and Technology, China; Southern University of Science and Technology, China; Guangzhou HKUST Fok Ying Tung Research Institute, China; City University of Hong Kong, China* 1934

VARIES: Verification Harness Synthesis and Efficient Scheduling for Unsoundness Detection in Rust Libraries
Huan Li, Xing Hu, Xin Xia, and Xinyu Wang — *Zhejiang University, China; Hangzhou High-Tech Zone (Binjiang) Institute of Blockchain and Data Security, China* 1947

Prism: A Multi-solution Reasoning and Synthesis Framework for Repository-Level Issue Resolution
Yibo Wang, Guotian Wang, Zhihao Peng, Ying Wang, Chang Xu, and Zhiliang Zhu — *Northeastern University, China; Nanjing University, China* 1960

RepoProbe: Benchmarking Architecture-Aware Repository Comprehension with Checklists
Yuexi Yang, Alyssa Wu, Ji Luo, Richeng Xuan, Zhichao Hu, Yuhong Liu, and Zhen Qin — *Zhejiang University, China; Tencent, China* 1973

RFC2TLA+: Extracting and Verifying Formal Models from RFC Documents using Continuous LLM Feedback
Guozhen Ding, Kexin Li, Ilya Grishchenko, and David Lie — *University of Toronto, Canada* 1986

It Takes Two to Tango: Accelerating Code Generation via Cache-Inspired Hierarchical Speculative Decoding
Yinggang Qiu, Shangwen Wang, Yihao Qin, Mingyang Geng, and Dezun Dong — *National University of Defense Technology, China; Information Support Force Engineering University, China* 1998

One Is Not Enough: The Untold Story of Multiple Security Patches for One Vulnerability
Fangyuan Zhang, Lyuye Zhang, Lingling Fan, Chengwei Liu, Yinan Li, Liang Huang, Yang Liu, Zheli Liu, and Sen Chen — *Nankai University, China; Nanyang Technological University, Singapore; Qi An Xin Technology Group, China* 2011

Not as Sweet by Another Name: An Empirical Study of Format Robustness in LLM Document Workflows
Xiaoyu Zhang, Xianyun Cheng, Tianlin Li, Yuwei Zheng, Yue Yang, and Yang Liu — *Nanyang Technological University, Singapore; National University of Singapore, Singapore; Beihang University, China; Xi'an University of Architecture and Technology, China* 2024

RealisticTritonBench: A Benchmark for Triton-Kernel Generation in Real-World AI Frameworks
Jinjun Huang, Zhongzhen Wen, Tongtong Xu, Meng Yan, Xin Xia, and Zhongxin Liu — *Zhejiang University, China; Nanjing University, China; Huawei, China; Chongqing University, China* 2037

From Guessing to Seeing: Enhancing LLM-Based Program Repair via Trace-Guided Multi-strategy Debate
Jiaqing Wu, Tong Wu, Manqing Zhang, Yunwei Dong, and Bo Shen — *Northwestern Polytechnical University, China* 2050

When Extensions Lie: Large-Scale Study of Privacy Policy Compliance in Browser Extensions
Zilun Wang, Zeyang Zhuang, Mingxue Zhang, Wei Meng, and Michael R. Lyu — *Chinese University of Hong Kong, China; Zhejiang University, China; Hangzhou High-Tech Zone (Binjiang) Institute of Blockchain and Data Security, China* 2063

MultiKernelBench: A Multi-platform Benchmark for Kernel Generation
Zhongzhen Wen, Yinghui Zhang, Zhong Li, Zhongxin Liu, Linna Xie, and Tian Zhang — *Nanjing University, China; Zhejiang University, China* 2076

Post-quantum Cryptography in the Wild: Assessing the Readiness of Open-Source Ecosystems
Tongxin Yuan, Zhanpeng Liu, Jiashuo Liang, Zhuosheng Zhang, Gongshen Liu, Yang Yu, and Guancheng Li — *Shanghai Jiao Tong University, China; Tencent Xuanwu Lab, China* 2089

CodeFault: Predicting Fault Risks from Code Changes via Multi-modal Learning
Yifan Xiao, Shijie Li, and Yu Huang — *Peking University, China; China Southern Power Grid, China* 2101

ReLog: Execution-Aware Logging with Runtime Feedback for LLM-Oriented Debugging
Xin Wang, Yang Feng, Xiaoqian Jiao, Yang Zhang, Zhenhao Li, and Zishuo Ding — *Hong Kong University of Science and Technology, Guangzhou, China; Hebei University of Science and Technology, China; York University, Canada* 2114

Documentation vs. Code Patterns: What Drives LLM-Based Exception Oracle Generation?
Soneya Binta Hossain, Matthew B. Dwyer, and Tasfia Tasnim — *University of Texas at Dallas, USA; University of Virginia, USA* . . . 2127

Harnessing Uncertainty in Code Language Models: Lessons from Vulnerability Detection
Haodong Li, Xiao Cheng, Xudong Wang, Zhihao Guo, and Haoyu Wang — *Huazhong University of Science and Technology, China; Macquarie University, Australia; Zhongguancun Institute of Artificial Intelligence, China; University of Technology Sydney, Australia* . 2139

Breaking the Isolation: Coordinated Multi-agent Fuzzing for Smart Contracts with Multi-dimensional Objective Learning
Ruichao Liang, Ruochen Cao, Jing Chen, Kun He, Yebo Feng, Yue Xue, and Cong Wu — *Wuhan University, China; Nanyang Technological University, Singapore; MetaTrust Labs, Singapore* 2152

HyTri: Hybrid Triage of CI Failures via LLM Guided Semantic Reasoning and Change Attribution
Lior Broide, Argaman Mordoch, and Roni Stern — *Ben-Gurion University of the Negev, Israel* 2165

Latent Reuse in Agent Skills: Multi-modal Clone Detection at Ecosystem Scale
Jiaying Zhu, Lyuye Zhang, Wenbo Guo, and Yang Liu — *Nanyang Technological University, Singapore; Nankai University, China* . . 2177

Specification-Guided Synthesis of Deadlock-Free Communication Protocol Refinements with Large Language Models
Yang Li, Ping Hou, and Nobuko Yoshida — *University of Oxford, UK* 2190

Execution-as-Configuration: Security Smells in Model Configuration Artifacts
Mohammed Latif Siddiq, Prince Noah Johnson, and Joanna C. S. Santos — *University of Notre Dame, USA* 2203

Evidence-Grounded Logic Vulnerability Detection for Smart Contracts via Runtime Context Reasoning
Guang Gong, Yucong Jiang, Rong Jian, Xiaotao Feng, Zhihan Wei, Qi Li, and Ke Xu — *Tsinghua University, China; 360 Vulnerability Research Institute, China* 2216

Fully Automating Template Polyhedral Analysis by Leveraging LLMs
Renjie Huang, Liqian Chen, Hongfei Fu, Banghu Yin, Dengping Wei, and Ji Wang — *National University of Defense Technology, China; Shanghai University of Finance and Economics, China* 2229

When Does AI Actually Help in Incident Response? Identifying Good First Messages in Cloud Service Incidents
Minghua Ma, Rujia Wang, Chetan Bansal, Saravan Rajmohan, Yingnong Dang, and Hongyu Zhang — *Microsoft, USA; Chongqing University, China* 2242

We Must Have Missed This Comment: Detecting and Repairing Stale Function References in Linux Kernel Comments
Kexin Sun, Yunbo Lyu, Xutong Ma, Hongyu Kuang, Ratnadira Widyasari, He Zhang, Xiaoxing Ma, Julia Lawall, and David Lo — *Nanjing University, China; Singapore Management University, Singapore; Inria, France* 2254

LARA: Static ReDoS Detection for Regular Expressions with Lookahead via Matching Automata
Mengxi Wang, Chengyao Peng, Weihao Su, Guiyi He, Rongchen Li, Hong Huang, and Haiming Chen — *Institute of Software at Chinese Academy of Sciences, China; University of Chinese Academy of Sciences, China* 2267

ConfFuzz: Parameter-Aware Greybox Fuzzing for Configurable Cloud Systems
Shuai Wang, Hao Wang, Darko Marinov, Tianyin Xu, and Yongle Zhang — *University of Illinois at Urbana-Champaign, USA; University of California at Berkeley, USA; Purdue University, USA* 2280

Synthesizing File-Level Data for Unit Test Generation with Chain-of-Thoughts via Self-Debugging
Ziyue Hua, Tianyu Chen, Yeyun Gong, Shuai Lu, Peng Cheng, Qinglin Zhu, Yibo He, Yingjie Fu, Wenpin Jiao, Wei Yang, and Tao Xie — *Peking University, China; Microsoft Research, China; King's College London, UK; Fudan University, China* 2293

FAiLGen: Calibrating LLM for Failing Test Generation with Monte Carlo Tree Search
Tsz-On Li, Xiaochuan Yan, Dehui Yi, Yuhang Jiang, Kexin Li, Congying Xu, Ying Wang, and Shing-Chi Cheung — *Hong Kong University of Science and Technology, Hong Kong; Guangzhou HKUST Fok Ying Tung Research Institute, China; Northeastern University, China* . 2306

Feature-Focused Test Generation
José Antonio Zamudio Amaya, Gaetano Sapia, Alexi Turcotte, David Benavides, Marcel Böhme, and Andreas Zeller — *CISPA Helmholtz Center for Information Security, Germany; MPI-SP, Germany; Universidad de Sevilla, Spain* 2319

SecChain: Operationalizing Security Guidelines for Multi-agent Smart Contract Generation
Dianxiang Sun, Liming Nie, and Yang Liu — *Nanyang Technological University, Singapore; Shenzhen Technology University, China* . . 2332

SafeDep: Detecting State Mutation Unsoundness in Rust via Dependency Slicing and LLM Reasoning
Zihao Rao, Yilin Chen, Chengjun Chen, and Hui Xu — *Fudan University, China; Ant Group, China* 2345

A Selective Undo/Redo Algorithm for Dependency-Constrained Operation Histories Based on Change Propagation
Xiaoxiong Fan, Hou Tam, Hanxi Zhu, and Song-Hai Zhang — *Tsinghua University, China* 2358

WiDepFuzz: Finding Wi-Fi Protocol Bugs in Mobile Hotspot via Field Dependency Model
Yuanliang Chen, Fuchen Ma, Taotao Gu, and Yu Jiang — *Renmin University of China, China; Tsinghua University, China* 2370

CTForge: Automatically Generating Test Suites for Software Configuration
Yuanliang Zhang, Zhizheng Zheng, Shanshan Li, Zhouyang Jia, Chaopeng Luo, Liqian Chen, Zhenbang Chen, Ji Wang, and Xiangke Liao — *National University of Defense Technology, China* 2383

Panorama: Unveiling Latent Dependencies for Microservice Autoscaling via Meta-learning
Zhuangbin Chen, Hongjiang Feng, Yang Liu, Juzheng Zheng, Xiaoyu He, and Zibin Zheng — *Sun Yat-sen University, China* 2396

ReCodeAgent: A Multi-agent Workflow for Language-Agnostic Translation and Validation of Large-Scale Repositories
Ali Reza Ibrahimzada, Brandon Paulsen, Daniel Kroening, and Reyhaneh Jabbarvand — *University of Illinois at Urbana-Champaign, USA; Amazon, USA* 2409

Aligning User Requirement with Large Language Model in Code Generation
Jia Li, Ruiqi Bai, Yangkang Luo, Yiran Zhang, Wentao Yang, Zeyu Sun, Tiankuo Zhao, Dongming Jin, Lei Li, and Zhi Jin — *Wuhan University, China; Nanyang Technological University, Singapore; Institute of Software at Chinese Academy of Sciences, China; Peking University, China; University of Hong Kong, Hong Kong* 2422

QDoctor: Probe-Based Testing for Quantum Programs
Qichen Wang, Shangzhou Xia, Xiaoyu Guo, and Jianjun Zhao — *Kyushu University, Japan* 2435

Doc2Feat-Bench: Evaluating Documentation-Driven Feature Addition
Zhonghao Jiang, Le Deng, Jialun Cao, Michael Pradel, and Zhongxin Liu — *Zhejiang University, China; Hong Kong University of Science and Technology, Hong Kong; CISPA Helmholtz Center for Information Security, Germany* 2447

RegoMender: A Multi-agent Framework for Automated Rego Policy Repair
Ruijie Zhang, Huaxiao Yin, Li Hu, Fanqi Kong, Haozhe Liang, Shaoyin Cheng, Kun Zhang, and Bibo Tu — *Institute of Information Engineering at Chinese Academy of Sciences, China; University of Chinese Academy of Sciences, China; University of Science and Technology of China, China; Peking University, China; Anhui Province Key Laboratory of Digital Security, China* 2460

Defensive Capability Analysis for JavaScript Libraries
Wenyuan Xu and Anders Møller — *Aarhus University, Denmark* 2473

RAVEN: Measuring Detection Instability in Anti-virus Engines under Semantics-Preserving Perturbations
Younghoon Ban, Beomjin Jin, Doowon Kim, Hyoungshick Kim, and Haehyun Cho — *Soongsil University, Republic of Korea; Ajou University, Republic of Korea; University of Tennessee, Knoxville, USA; Sungkyunkwan University, Republic of Korea* 2486

CiL-Agent: Compiler-in-the-Loop LLM Agent for Memory-Safe C Program Repair
Yang Shen, Yating Zhang, Jiaqiang Yao, Minghao Liu, and Wei Dong — *National University of Defense Technology, China* 2499

Safe Multi-site Binary Instrumentation of Running x86-64 Programs
Jihun Baek, Sanghoon Han, and Hyungon Moon — *UNIST, Republic of Korea* 2512

Mobile Bug Reproduction via Global State Reprioritization and LLM-Guided Trajectory Exploration
Dingbang Wang, Sidong Feng, William G. J. Halfond, and Tingting Yu — *University of Connecticut, USA; Chinese University of Hong Kong, China; University of Southern California, USA* 2524

IncSFS: Incremental Full-Sparse Flow-Sensitive Pointer Analysis for C/C++
Kunlin Liu, Zhenbang Chen, Piyi Zu, Yide Du, and Ji Wang — *National University of Defense Technology, China* 2537

ExcavatorFuzz: LLM-Guided Path-Difficulty-Driven Directed Greybox Fuzzing
Kaiyu Xie, Siyuan Li, Jiaqian Peng, Yun Yang, Yaowen Zheng, Kai Cheng, Jiayu Zhang, Shizhao Tian, Yubo Li, and Hongsong Zhu — *Institute of Information Engineering at Chinese Academy of Sciences, Beijing, China; Shandong University, China* 2550

GANADI: Uncovering C/C++ OSS Reuse Genealogies via Pivotal Function-Based Clustering to Enhance Supply Chain Security
Dongyeon Kim, Seunghoon Woo, and Heejo Lee — *Korea University, Republic of Korea* 2562

OrthCoder: Balancing Security and Utility in Secure Code Generation via Orthogonal Adaptation
Menghan Tian, Youliang Yuan, Zhiqing Zhong, and Pinjia He — *Chinese University of Hong Kong, Shenzhen, China* 2574

Zero-Knowledge VMs: Security and Usability in Practice
Jinan Jiang, Jinzhao Chu, Pengyu Xue, Xinghao Peng, Haoran Qin, and Xiapu Luo — *Hong Kong Polytechnic University, Hong Kong* 2587

Formal Verification-Integrated Reinforcement Learning for Automated ORM Schema Synthesis
Md Rashedul Hasan and Hamid Bagheri — *University of Nebraska-Lincoln, USA* 2600

Repository-Aware Metamorphic Relation Generation for Augmented Reality Applications using Large Language Models
Dibyendu Brinto Bose, Jiawei Qin, and Chris Brown — *Virginia Tech, USA* 2613

Learning to Triage Vulnerability Reports from Program Analysis: An Empirical Study in Node.js
Ronghao Ni, Aidan Z.H. Yang, Min-Chien Hsu, Nuno Sabino, Limin Jia, Ruben Martins, Darion Cassel, and Kevin Cheang — *Carnegie Mellon University, USA; Athanor AI, USA; University of California at Berkeley, USA* 2626

Code Transformation Rule Synthesis using LLMs: Potential and Limits
Axel Allain, Aymeric Blot, Djamel Eddine Khelladi, and Mathieu Acher — *INSA, France; University of Rennes, France; Inria, France; CNRS, France; IRISA, France* 2639

A Low-Cost Human-in-the-Loop Investigation of Toxicity on GitHub at Scale
Rahat Rizvi Rahman, Mia Mohammad Imran, and Kostadin Damevski — *Virginia Commonwealth University, USA; Missouri University of Science and Technology, USA* 2652

Generation of Web Apps with Agentic IDEs: An Empirical Assessment
Manuel Marceca, Maria Teresa Rossi, and Leonardo Mariani — *University of Milano-Bicocca, Italy; Gran Sasso Science Institute, Italy* 2665

Looks Good, but Does It Run? A Large-Scale Empirical Study of the Executability of Example Code in Model Cards
Simeng Yao, Jialin Zhao, Yang Zhang, Tun Li, Tao Wang, Changrong Xie, Zezhou Tang, and Yiwen Wu — *National University of Defense Technology, China* 2677

From Thread to Pool: Modeling and Static Detection of Thread Pool Misuses in Java Programs
Baoquan Cui, Yucheng Qiu, Rui Han, Hang Gao, and Jian Zhang — *Institute of Software at Chinese Academy of Sciences, Beijing, China; Hangzhou Institute for Advanced Study at University of Chinese Academy of Sciences, Hangzhou, China; University of Chinese Academy of Sciences, China* 2690

HERO: Hypothesis-Centered Root-Cause Analysis for Microservice Incidents
Jiewei Lyu, Junquan Yi, Shu Liang, Pengfei Chen, and Long Pan — *Sun Yat-sen University, China; Tencent, China* 2703

Gray-Box Fuzzing in Local Space
Martin Jonáš, Jan Strejček, and Marek Trtík — *Masaryk University, Czech Republic* 2716

Revisiting LLMs on New Feature Implementation in Real-World Software Development Practices
Jia Li, Tingxuan Huang, Tiankuo Zhao, Yiming Wang, Yiran Zhang, Anshun Hu, Kechi Zhang, Lecheng Wang, Zinan Sheng, Kui Liu, Ge Li, and Zhi Jin — *Wuhan University, China; Nanyang Technological University, Singapore; Peking University, China; Huawei Technologies, China* 2728

Self-Discovering Security Oracles: Meta-learning Vulnerability Detection Strategies through Adversarial Self-Play
Md Rashedul Hasan and Hamid Bagheri — *University of Nebraska-Lincoln, USA* 2741

S2CAgent: A Self-Learning Agent Knowledge Graph Framework for Software Supply Chain Analysis
Youfei Huang, Zhong Li, Minxue Pan, and Xuandong Li — *Nanjing University, China* 2754

ArkEval: Benchmarking and Evaluating Automated Code Repair for ArkTS
Bang Xie, Senjian Zhang, Zhiyuan Peng, Wei Chen, Xin Yin, Chenhao Ying, and Yuan Luo — *Shanghai Jiao Tong University, China; Zhejiang University, China* 2767

From Vibe to Spec: An Empirical Study to Demystify Specifications in Code Generation
Ruirui He, Hanqing Lu, Lezhi Ma, Kai Ye, Qiong Wu, Shangqing Liu, Lei Bu, and Xuandong Li — *Nanjing University, China* 2780

Instruction Alignment for Binary Code Representation Learning
Huaijin Wang and Shuai Wang — *Shandong University, China; Hong Kong University of Science and Technology, Hong Kong* 2793

Formula Completion Suggestions for Alloy Models with Selection Guidance
Mohammad Nurullah Patwary and Allison Sullivan — *University of Texas at Arlington, USA* 2806

COBug: A Bug Localization Pipeline for COBOL: Does It Work? An Exploratory Study
Kavyasri Gajula, Phanindra Kodela, and Sridhar Chimalakonda — *Indian Institute of Technology Tirupati, India* 2819

Accurate and Low-Cost Residual Risk Assessment via Sampled Profiling and Structure-Aware Coverage Amplification
Seongmin Lee, Işıl Özgü, Marcel Böhme, and Miryung Kim — *University of California at Los Angeles, USA; MPI-SP - CISPA Helmholtz Center for Information Security, Germany, Germany* 2831

SPSYN: Synthesizing DeFi Price-Manipulation Exploits via Semantic Recovery and State-Guided Search
Bosi Zhang, Ningyu He, Guangdong Bai, and Haoyu Wang — *Huazhong University of Science and Technology, China; Hong Kong Polytechnic University, Hong Kong; City University of Hong Kong, Hong Kong* 2843

To Think or Not to Think: Evaluating LLM Reasoning and Agents in Vulnerability Detection
Hua Beng Tan, Ratnadira Widyasari, Merve Astekin, Arda Goknil, Hasan Sözer, Yan Naing Tun, Erik Johannes Husom, Lwin Khin Shar, and Lingxiao Jiang — *Singapore Management University, Singapore; SINTEF, Norway; Ozyegin University, Türkiye* 2856

Demystifying Solana Bots: From GitHub Blueprints to On-Chain Fingerprints
Xiaoye Zheng, Yujing Chen, Minghao Wu, David Lo, Difan Xie, Daoyuan Wu, Xiaohu Yang, and Zhiyuan Wan — *Zhejiang University, China; Singapore Management University, Singapore; Hangzhou High-Tech Zone (Binjiang) Institute of Blockchain and Data Security, China; Lingnan University, Hong Kong* 2869

CLEAR: Causal Context-Based Agentic Reasoning for Vulnerability Detection
Sungju Yun, Sijune Hwang, Yeonjoon Lee, Kyungtae Kang, and Sungbin Park — *Hanyang University, Republic of Korea* 2882

Large Language Model Enabled Symbolic Execution for Automated Functional Analysis
Meixi Liu, Zhenbang Chen, Xudong Wang, Ziran He, Jiachen Gu, Minghui Chen, Lei Wang, and Wei Dong — *National University of Defense Technology, China* 2895

Structure-Guided Semantic Deobfuscation of Android Identifiers
Mingyang Chen, Zhentao Xie, Yaqi Gao, Shuang Li, and Wenrui Diao — *Shandong University, China* 2908

Orbis: Guiding Symbolic Execution Techniques to Maximize Option-Related Branch Coverage
Minjong Kim, Sungjae Hwang, and Sooyoung Cha — *Sungkyunkwan University, Republic of Korea* 2921

MultiFixer: A Coordinator-Proposer Based Multi-agent Framework for Fixing Multi-hunk Bugs
Haichuan Hu, Chunrong Fang, Ye Shang, Jiawei Liu, Weifeng Sun, Guoqing Xie, Chenxing Zhong, and Quanjun Zhang — *Nanjing University of Science and Technology, China; Nanjing University, China; Singapore Management University, Singapore* 2934

Mining Tactics for Automated Theorem Proving
Jian Fang, Yixun Yao, and Yingfei Xiong — *Ministry of Education, China; Peking University, China; Zhongguancun Laboratory, China* 2947

Leveraging System-Level Observations to Inform Bayesian Learning of Model Parameters for Quantitative Verification
Simos Gerasimou and Xingyu Zhao — *Cyprus University of Technology, Cyprus; Wuhan University, China; University of Warwick, UK* 2960

Generating Loop Invariants for Deductive Verification using Iterator Profiles
Keisuke Nishimura, Jean-Pierre Lozi, and Julia Lawall — *Inria, France* 2973

RustGTC: Constraint-Guided Test Generation for Rust APIs via Comment Analysis and Code Context
Yichi Zhang, Jinhao Huang, and Yang Feng — *Peking University, China; Institute of Software at Chinese Academy of Sciences, China; Nanjing University, China* 2985

HyBinMAS: LLM-Powered Hybrid Analysis for Function Name Recovery of Stripped Binaries
Chenxi Wang, Lei Xu, Xincheng He, Gaozhan Yu, and Ye Wang — *Nanjing University, Nanjing, China; Jiangnan University, China; Nanjing University, China; Southeast University, China* 2997

Beyond Syntactic Differencing: Semantic-Aware Scheduling for Cross-Configuration Incremental Builds
Xiaolong Dai, Shanshan Li, Jun Lyu, and Xintao Niu — *Nanjing University, China* 3010

Probe to Generate: Program Variant-Guided Test Augmentation for Repository-Level Repair Benchmarks
Chenglin Li, Yisen Xu, Zehao Wang, Shin Hwei Tan, and Tse-Hsun (Peter) Chen — *Concordia University, Canada* 3023

Safeguarding LLM Agents from Misalignment through Provenance Analysis
Yining She, Yiliang Liang, and Eunsuk Kang — *Carnegie Mellon University, USA* 3036

PropGen: Automated Property Generation for Property-Based Testing of Mobile Apps
Yiheng Xiong, Shiwen Song, Bo Ma, Ting Su, and Xiaofei Xie — *Singapore Management University, Singapore; East China Normal University, China* 3049

From Function to Repository: Towards Understanding and an Agentic Approach for Repo-Level Directed Input Generation
Haoyu Wang, Ming Wen, Haoran Yan, Jiayi Huang, Xuanhua Shi, and Hai Jin — *Huazhong University of Science and Technology, China*3062

A Comprehensive Evaluation of Code Language Models for Security Patch Detection
Nils Loose, Joseph Bienhüls, Kristoffer Hempel, Felix Mächtle, and Thomas Eisenbarth — *University of Lübeck, Germany* 3075

Exact-IaC: Execution-Driven Exact Knowledge Injection for Reliable IaC Synthesis
Hua Ge and Minxue Pan — *Nanjing University, China* 3088

TrackTest: Trajectory Generation and Prioritization for Testing Multiple Object Tracking System
Xinyu Gao, Haoxin Chen, Shuoxiao Zhang, Enyi Tang, Hengrui Xu, Haoxiang Tian, Minghui Wei, and An Guo — *Nanjing University, China; Nanyang Technological University, Singapore; Hong Kong Polytechnic University, China* 3101

PELSE: Asymmetric-Evidence-Guided MCU Peripheral Model Synthesis
Guohao Wu, Chunlin Wang, Rui Yao, Feng Hong, Hongliang Liang, and Qiuping Yi — *Beijing University of Posts and Telecommunications, China; Fudan University, China; Beijing Institute of Computer Technology and Applications, China* 3114

Benchmarking Automated Security Patch Backporting: How Far Are We?
Jincheng Yang, Yulong Fu, Chengwei Liu, Lyuye Zhang, Fangyuan Zhang, Bingyang Ren, Yang Liu, and Hui Li — *Xidian University, China; Nankai University, China; Nanyang Technological University, Singapore* 3127

When Knowledge Changes: Metamorphic Testing of RAG Systems with Mutations
Jinhan Kim, Samuele Pasini, and Paolo Tonella — *USI Lugano, Switzerland* 3140

Testing Static Analyzers via Semantic-Preserving Mutators Learned from Real-World Refactoring Practice
Meilin Li, Kaixuan Li, Zifan Xie, Shiyu Qiu, Ming Wen, Maolin Sun, and Hongyu Zhang — *Chongqing University, China; Nanyang Technological University, Singapore; Huazhong University of Science and Technology, China; Nanjing University, China* 3152

ReSEDA: Automatic and Precise Program-Level ReDoS Triage via Interprocedural Taint Analysis and Agentic Exploit Generation
Rongchen Li, Weihao Su, Chengyao Peng, Hong Huang, Haiming Chen, and Guiyi He — *Institute of Software at Chinese Academy of Sciences, China; University of Chinese Academy of Sciences, China* 3164

A Few Pages of Markdown: Committed AI Configuration and Lower Quality Cost after Coding-Agent Adoption
Yegor Denisov-Blanch, Shyam Agarwal, Pavel Azaletskiy, Hao He, Rylan Schaeffer, Brando Miranda, Bogdan Vasilescu, and Sanmi Koyejo — *Stanford University, USA; Carnegie Mellon University, USA; Grid Dynamics, USA* 3176

Perception, Reasoning, and Action: A ReAct-Based Vulnerability Detection Framework with Repository-Level Reasoning
Xin-Cheng Wen, Hanyue Luo, Cuiyun Gao, Hao Zhu, Yang Xiao, and Ge Li — *Harbin Institute of Technology, Shenzhen, China; Peking University, China; Institute of Information Engineering at Chinese Academy of Sciences, China; University of Chinese Academy of Sciences, China* 3188

TDiFf: Detecting Bugs in DataFrame Systems via Transferred DBMS Test Cases
Jiaxin Hu, Shaowei Chen, and Rongxin Wu — *Xiamen University, China; Fujian Agriculture and Forestry University, China* 3201

MiniPoly: Automatic Extraction of Efficient JavaScript Polyfill from Language Specification
Jungwoong Kim, Youngmin Cho, and Jihyeok Park — *Korea University, Republic of Korea* 3214

IcFuzz: Fuzzing Isaac Sim with Semantic Stage Guidance and Multi-level Mutation
Zhixiang Chen, Zhuangbin Chen, Ruoxi Jia, Zeqin Liao, Wei Li, Jinyang Liu, and Zibin Zheng — *Sun Yat-sen University, China; Nanyang Technological University, Singapore; Chinese University of Hong Kong, Hong Kong* 3226

How Well Do LLMs Generate Taxonomies in the SE Domain? A Multi-perspective Evaluation Framework
Sota Nakashima, Yuta Ishimoto, Masanari Kondo, Tao Xiao, and Yasutaka Kamei — *Kyushu University, Japan; University of Osaka, Japan* 3239

Implicit, Yet Impactful: Understanding Hidden Dependencies in Java Projects
Lyuye Zhang, Chengwei Liu, Fangyuan Zhang, Yiran Zhang, Yuan Zhou, and Yang Liu — *Nankai University, China; Nanyang Technological University, Singapore; Zhejiang Sci-Tech University, China* 3252

Piece by Piece: Automating Combination Interaction GUI Testing via Planning and Dual Memory
Zhe Liu, Oujin Wang, Junjie Wang, Chunyang Chen, Mengzhuo Chen, Boyu Wu, Yuekai Huang, and Qing Wang — *Institute of Software at Chinese Academy of Sciences, China; Tsinghua University, China; TU Munich, Germany; University of Chinese Academy of Sciences, Beijing, China* 3265

The CodeInverter Suite: Structure- and Data-Aware Binary Decompilation with Efficient LLMs
Peipei Liu, Jian Sun, Rongkang Sun, Li Chen, Zhaoteng Yan, Xiaoling Zhang, Dawei Wang, Dapeng Sun, Peizheng Zhang, and Dan Li — *Zhongguancun Laboratory, China; Tsinghua University, China* 3278

Evaluating Inference-Time Defenses against Package Hallucination in LLM-Generated Code
Alberick Euraste Djire, Iyiola E. Olatunji, Melissa Tessa, Earl T. Barr, Jacques Klein, and Tegawendé F. Bissyandé — *University of Luxembourg, Luxembourg; AI4D (CITADEL), Ouagadougou, Burkina Faso; University College London, UK* 3290

Post-hoc Attention Steering of Large Language Models for Robust Code Understanding under Obfuscation
Xiaokai Rong, Aashish Yadavally, and Tien N. Nguyen — *University of Texas at Dallas, USA; University of Central Florida, USA* . . . 3303

PAIChecker: Uncovering and Checking PR-Issue Misalignment in SWE-Bench-Like Benchmarks
Manyi Wang, Junjielong Xu, and Pinjia He — *Chinese University of Hong Kong, Shenzhen, China* 3315

Can Large Language Models Decompose User Stories into Tasks? Exploring the Role of Prompting Strategies and Models
Delina Ly, Fatma Başak Aydemir, and Fabiano Dalpiaz — *VX Company, Netherlands; Utrecht University, Netherlands* 3328

Certified Program Synthesis with a Multi-modal Verifier
Yueyang Feng, Dipesh Kafle, Vladimir Gladshtein, Vitaly Kurin, George Pîrlea, Qiyuan Zhao, Peter Müller, and Ilya Sergey — *National University of Singapore, Singapore; Neapolis University Pafos, Cyprus; ETH Zurich, Switzerland* 3341

Tools and Datasets

Graphectory Viewer: A Tool for Process-Centric Analysis of Agentic Software Trajectories
Charlie Jyu, Shuyang Liu, and Reyhaneh Jabbarvand — *University of Illinois at Urbana-Champaign, USA* 3354

VidSumEval: A Web Platform and Benchmark for Evaluating AI-Generated Programming Video Summaries
Osama AlGhamdi, Ameer AlGhamdi, Faisal Alorabi, Mohammad D. Alahmadi, Abdulmalik Alzahrani, and Abdullah Almazroui — *University of Jeddah, Saudi Arabia* 3359

BESSER-NN: Visual Neural Network Design and Automated Code Generation with the BESSER Low-Code Platform
Nadia Daoudi, Armen Sulejmani, and Jordi Cabot — *Luxembourg Institute of Science and Technology, Luxembourg; University of Luxembourg, Luxembourg* 3364

TraceGate: A Framework for Policy-Controlled Failure Evidence in LLM-Assisted Program Repair
Nicolas Schuler, Vincenzo Scotti, and Raffaela Mirandola — *KIT, Germany* 3369

SAF: An AI-Agent-Ready and Browser-Accessible Static Analysis Framework for LLVM IR
Yuekang Li, Wei Li, Wei Song, Yi Liu, and Gelei Deng — *University of New South Wales, Australia; Griffith University, Australia; Nanyang Technological University, Singapore* 3373

TLCrafter: Peeling the Layers of Natural Language to Reveal Temporal Logic Specifications
Cheng Wen, Wenjun Ke, Xiao Liang, Zhi Ma, Xiaoyong Xue, Shengchao Qin, and Cong Tian — *Xidian University, China; Northwest University, China; Peking University, China* 3377

VidTutorAssistant: Automating Responses to Programming Tutorial Questions
Ahmad Tayeb, Sonia Haiduc, and Mohammad D. Alahmadi — *King Abdulaziz University, Saudi Arabia; Florida State University, USA; University of Jeddah, Saudi Arabia* 3382

ProofPulse: Interactive Proof Coverage Analysis for Dafny
Álvaro F. Silva, Ruben Martins, and Alexandra Mendes — *INESC TEC, Portugal; Universidade do Porto, Portugal; Carnegie Mellon University, USA* 3387

DT-Drive: A Tool for Deterministic Replay-Based Testing and Debugging of Autonomous Driving Systems
Sanjeetha Pennada, Matthew Leach, Olek Osikowicz, and Donghwan Shin — *University of Sheffield, UK* 3392

FairLint-DL: An IDE-Native Tool for Fairness Debugging of Deep Learning Software
Archit Rathod and Saeid Tizpaz-Niari — *University of Illinois at Chicago, USA* 3397

VerusSeek: Retrieval-Augmented LLM-Based Proof Synthesis for Rust Programs
Yuchen Zhang, Cheng Wen, Jialun Cao, Dugang Liu, Zhiwu Xu, Yuwei Liu, and Shengchao Qin — *Xidian University, China; Hong Kong University of Science and Technology, Hong Kong; Shenzhen University, China; Institute of Software at Chinese Academy of Sciences, China* 3402

xDECAF: An Extensible Data Flow Diagram Analysis Framework for Information Security
Benjamin Arp, Felix Schwickerath, Alexander Vogt, Tom Hüller, Nils Niehues, and Nicolas Boltz — *KIT, Germany* 3407

AISmellBench: A Reusable Benchmark of AI Code Smells in ML and LLM-Based Systems
Brahim Mahmoudi, Naouel Moha, Quentin Stiévenart, and Florent Avellaneda — *École de Technologie Supérieure de Montréal, Canada; Université du Québec à Montréal, Canada* 3411

CppPerf: An Automated Pipeline and Dataset for Performance-Improving C++ Commits
Tommy Ho, Khashayar Etemadi, and Zhendong Su — *ETH Zurich, Switzerland* 3415

FlightRecorder: Tracing Fine-Grained Development with AI Assistants
Tobias Dick, Benedict Bliem, Kallistos Weis, Alisa Welter, and Sven Apel — *Saarland University, Germany* 3420

ThreatCraft: Automated Attack Scenario Generation via Hybrid Rule-Based and LLM-Driven Reasoning
Dohee Kang, Jiwon Kwak, Geunwoo Baek, and Seungjoo Kim — *Korea University, Republic of Korea* 3425

QUTest: A Native Testing Framework for Quantum Programs
José Campos — *University of Porto, Porto, Portugal; University of Lisbon, Lisbon, Portugal* 3430

DafnyRAG: A Retrieval-Augmented and Verification-Driven Tool for Automated Repair of Dafny Code
Yanzhen Liu, Xu Lu, Bin Yu, Haoxiang Zhang, Chu Chen, and Meng Wang — *Xidian University, China; Qufu Normal University, China; Hebei University, China* 3435

SmellCC: A Tool for Automated Code Smells Remediation
Xiaoting Zhang, Yujie Zhang, Zhipeng Gao, Xing Hu, and Xin Xia — *Zhejiang University, China; Hangzhou High-Tech Zone (Binjiang) Institute of Blockchain and Data Security, China* 3440

UIBenchKit: A Unified Toolkit for Design-to-Code Model Evaluation
Chinh T. Le, Trevor Yee Siang Ong, Jingyu Xiao, Yuxuan Wan, and Yintong Huo — *Singapore Management University, Singapore; Chinese University of Hong Kong, Hong Kong* . . . 3445

CAS2UML: A Handwritten Sketch-to-PlantUML Dataset for Class and Activity Diagrams
Simon Scholz and Mersedeh Sadeghi — *University of Cologne, Germany* . . . 3450

The EVerest Dataset for Secure Software Engineering
Sophie Corallo, Debora Grupp, Dominik Fuchß, Jan Keim, Frederik Reiche, Tobias Hey, and Anne Koziolek — *KIT, Germany* . . . 3454

AIGen: Automating AI Bill of Materials Generation through Hybrid MLOps Integration
Federica Pepe, Daniele Bifolco, Costantino Martignetti, Aureliano D'Amici, Fabiano Izzo, Damian Andrew Tamburri, and Massimiliano Di Penta — *University of Sannio, Italy; Smart Shaped, Italy* . . . 3459

TypeScript Repository Indexing for Code Agent Retrieval
Junsong Pu, Yichen Li, and Zhuangbin Chen — *Sun Yat-sen University, Zhuhai, China; Chinese University of Hong Kong, Hong Kong; Sun Yat-sen University, China* . . . 3464

multiplex: A Modular LLM-Based Mutation Framework
Megan Maton, Gregory M. Kapfhammer, and Phil McMinn — *University of Sheffield, UK; Allegheny College, USA* . . . 3468

ScratchLog+: Live Analytics of Learners' Exercise Progress
Benedikt Fein and Gordon Fraser — *University of Passau, Germany* . . . 3473

NL2VBench: A Benchmark for End-to-End Natural-Language to Verifiable Rust Generation
Lianhe Hu, Cheng Wen, Dugang Liu, Zhiwu Xu, Zhuohua Li, Jie Su, Bin Yu, and Shengchao Qin — *Xidian University, China; Shenzhen University, China; Chinese University of Hong Kong, Hong Kong* . . . 3478

A Smart Assistant for Debugging
Steven P. Reiss and Mohamed Salem — *Brown University, USA; University of Pécs, Hungary* . . . 3483

The ARDoCo Tool Landscape: REST API, TraceView, and TraceViz for Architecture Traceability
Jan Keim, Dominik Fuchß, Sophie Corallo, Tobias Hey, Julian Winter, and Kevin Feichtinger — *KIT, Germany* . . . 3488

FallibleUser: An Imperfect User Agent for Realistic Evaluation of Interactive AI Systems
Yaotian Yang, Zhi Jin, Dongming Jin, and Xiaohong Chen — *Beijing Forestry University, China; Peking University, China; East China Normal University, China* . . . 3492

SplitCA: An Effective Tool for Generating 4-wise Covering Arrays for Large-Scale Highly Configurable Systems
Shuangyu Lyu, Ruizhi Shi, Yuechen Duan, Chunming Hu, and Chuan Luo — *Beihang University, China* . . . 3497

MIMOS-Tools: An Integrated Toolchain for Model-Based Design of Embedded Systems
Wang Yi, Chengzi Huang, Behnam Khodabandeloo, and Duc Anh Nguyen — *Uppsala University, Sweden* . . . 3502

MASSIV: A Hybrid RAG for Automated Defeater Extraction in Autonomous Driving Safety Cases
Tihomir Rohlinger, Daniel Ratiu, and Stefan Wagner — *University of Stuttgart, Germany; Cariad, Germany; TU Munich, Germany* . . . 3507

PITMuS: A Tool for Automated Bug Dataset Generation via Source-Level Mutant Reconstruction
Tasfia Tasnim and Soneya Binta Hossain — *University of Texas at Dallas, USA* . . . 3512

StarFuse: LLM Guided Multimodal Interactive Theorem Proving
Sharvil Athaley, Samyak Singhania, Khushboo Agrawal, and Subhajit Roy — *IIT Kanpur, India* . . . 3517

ACSLBench: A Verified C/ACSL Corpus and Benchmark with Compositional Call Chains for Formal Specification Synthesis
Junjie Hu, Cheng Wen, Bin Yu, Jialun Cao, Dugang Liu, Weidi Sun, Haokun Li, Shengchao Qin, and Cong Tian — *Xidian University, China; Hong Kong University of Science and Technology, Hong Kong; Shenzhen University, China; Peking University, China* . . . 3522

DSpec2Test: Specification-Driven Test Generation in Dafny
Sofia Vieira Pinto, Álvaro F. Silva, João Pascoal Faria, and Alexandra Mendes — *INESC TEC, Portugal; Universidade do Porto, Portugal* 3527

MergeSE: Post-Hoc Model Merging for Software Engineering Tasks without Retraining
Palash R. Roy, Banani Roy, Kevin A. Schneider, and Chanchal K. Roy — *University of Saskatchewan, Canada* . . . 3532

AIRBORNE: Augmented fIne-grained softwaRe Bill Of mateRials geNerator for stack ovErflow
Sabato Nocera, Massimiliano Di Penta, Simone Romano, and Giuseppe Scanniello — *University of Salerno, Italy; University of Sannio, Italy* . . . 3537

LintCFG: DSL-Driven Linter Configuration Generation for Coding Standards
Yi Tang, Zejun Zhang, Zhenchang Xing, Minxue Pan, Tian Zhang, and Xuandong Li — *Nanjing University, China; CSIRO's Data61, Australia* . . . 3542

InsightQL: A Unified Code Database for Fuzz Blocker Analysis
Wentao Gao, Renata Borovica-Gajic, Sang Kil Cha, Michael Fu, Tian Qiu, and Van-Thuan Pham — *University of Melbourne, Australia; KAIST, Republic of Korea* . . . 3547

AgentSysFilter: A Multi-agent Analysis Approach for Scenario-Aware Syscall Limitation
Dongyang Zhan, Xingchen Yan, Zhaofeng Yu, Yangfan Guo, Kai Tan, and Langtong Zhang — *Harbin Institute of Technology, China; Macau University of Science and Technology, China* . . . 3552

AutoFLC: An LLM-Based Flowchart Generation Tool for Legacy Code Repository
Yixing Luo, Longxing Yang, Xiaofeng Li, Bin Gu, and Zhi Jin — *Beijing Institute of Control Engineering, China; Beijing Key Laboratory of Trustworthy Code Large Language Models, Beijing, China; Peking University, China; Wuhan University, China* 3557

UCT: A Versatile Code Translation Framework Based on a Unified Intermediate Representation
Wangcai Li, Dongyan Wang, Zehong Yu, Yifan Cheng, Ning Ge, Chunming Hu, and Zhuo Su — *Beihang University, China; Renmin University of China, China; Tsinghua University, China; University of Electronic Science and Technology of China, China* 3562

Introducing Safe LLM-Agent Execution via Typed Tool Mediation and World-Effect Evaluation
Jun Zhou and Tao Jia — *Chongqing Normal University, China; Chongqing Electric Power College, China; Southwest University, China* . 3567

ReproBreak: A Dataset of Reproducible Web Locator Breaks
Thiago Santos de Moura, Leon Adamietz, Samra Mehboob, and Yannic Noller — *Ruhr University Bochum, Germany* 3572

Cam2Sim: Neural Scenario Reconstruction for Closed-Loop Autonomous Driving Simulation
Davide Jannussi, Stefano Carlo Lambertenghi, Constantin Carste, and Andrea Stocco — *Politecnico di Torino, Italy; TU Munich, Germany; fortiss, Germany* 3576

EnergyTrackr: A Modular Energy Regressions Detection Tool
François Bechet, Jérôme Maquoi, Luís Cruz, Benoît Vanderose, and Xavier Devroey — *University of Namur, Belgium; TU Delft, Netherlands* 3581

FlakeFighters: A Pytest Plugin to Detect Flaky Test Failures
Michael Foster, Owain Parry, Michael Hilton, Gregory M. Kapfhammer, and Phil McMinn — *University of Sheffield, UK; University of Edinburgh, UK; Carnegie Mellon University, USA; Allegheny College, USA* 3586

UltraInstinctVR: A Model-Based Tool for Automated System Testing of VR Applications
Gerry Longfils, Maxime Cauz, Arnaud Blouin, and Xavier Devroey — *University of Namur, Belgium; INSA Rennes, France* 3591

Marv: An Interactive Tool for Visualising and Reviewing the Results of Mutation Analysis
Daniel Wells, Zalán Lévai, and Phil McMinn — *University of Sheffield, UK* 3596

GoLiSA: An Advanced Static Analysis Tool for Detecting Potential Risks and Vulnerabilities in Hyperledger Fabric Smart Contracts
Luca Olivieri, Luca Negrini, Vincenzo Arceri, Pietro Ferrara, Fausto Spoto, and Agostino Cortesi — *Ca' Foscari University of Venice, Italy; University of Parma, Italy; University of Verona, Italy* 3601

Baseliner: A Plugin-Oriented, Hardware-Agnostic GPU Benchmarking Library
Côme Eyraud and Maxime Lamothe — *Polytechnique Montréal, Canada* 3606

FeatX: Editing Software by Editing Features for Repository-Level Code Evolution
Xutian Li, Yifeng Zhu, Xianlin Zhao, Yanzhen Zou, Lu Zhang, and Bing Xie — *Peking University, China* 3611

IoTutorMine: A Tool for Mining Hardware Bills of Materials from IoT Tutorial Videos
Abdullah A. Alahmadi, Ahmed O. Bahaj, and Mohammad D. Alahmadi — *University of Jeddah, Saudi Arabia* 3616

E-CoDrive: A Co-Simulation Framework for Testing Energy-Critical Driving Scenarios
Manfredi Napolitano, Alessandra Somma, Alessio Gambi, Andrea Stocco, and Nicola Mazzocca — *University of Naples Federico II, Italy; Austrian Institute of Technology, Austria; Italian Institute of AI for Industry, Italy; TU Munich, Germany; fortiss, Germany* 3621

JavaVulBench: A Java Vulnerability Benchmark with Realistic Splits, a Unified Multi-backend Harness, and a Leakage-Aware Evaluation Mode
Norbert Sándor Szolnoki and Gábor Antal — *University of Szeged, Hungary; FrontEndART, Hungary* 3626

OpenStack-RCA-Bench: A Reproducible IaaS Root Cause Analysis Dataset
Igor Bogomolov and Oleg Borisenko — *Trusted AI Research Center, RAS, Russian Federation; System Solutions, Russian Federation* . . 3631

Cochise: A Reference Harness for Autonomous Penetration Testing
Andreas Happe and Jürgen Cito — *TU Wien, Austria* 3635

Agent-HAnS: Maintaining Feature and Code Traceability in Agentic Development
Johan Martinson, Jacob Bengtsson, Moltas Hultin, Kevin Hermann, and Thorsten Berger — *Ruhr University Bochum, Germany; Centiro Solutions, Sweden; Chalmers University of Technology, Sweden; University of Gothenburg and Chalmers University of Technology, Sweden* 3640

TORI: A Tool for Oracle Quality Assessment
Facundo Molina, Nazareno Aguirre, and Alessandra Gorla — *Complutense University of Madrid, Spain; University of Rio Cuarto, Argentina; CONICET, Argentina; Guangdong Technion-Israel Institute of Technology, China; IMDEA Software Institute, Spain* 3645

RELY-SE: A Dashboard for Measuring Operational Reliability of AI Coding Agents
Akilesh P, Shivadharshan S, Sridhar Chimalakonda, Vibhu Saujanya Sharma, and Vikrant Kaulgud — *IIT Tirupati, India; Accenture Labs, India* 3650

AIALIB: A Threat Library for AI-Generated Code
Ali Aljaberi, Jian Kang, and Youcheng Sun — *MBZUAI, United Arab Emirates* 3655

Industry Showcases

DragonCrawl: A Generative, Intent-Based Framework for Scalable Mobile End-to-End Testing
Sowjanya Puligadda, Mengdie Zhang, Ali Zamani, Dhruva Dixith Kurra, Eric Chen, and Juan Marcano — *Uber Technologies, USA* . 3660

Bridging Stakeholder and Product Requirements: An Empirical Study of Requirement Engineering in the Automotive Industry
Zixu Wang, Shengcheng Yu, Zhenchang Xing, Tobias Wenzel, and Chunyang Chen — *TU Munich, Germany; CSIRO's Data61, Australia; Infineon Technologies, Germany* 3672

MicroWeaver: A Constraint-Guided Structure–Semantic Aware Approach to Automated Microservice Decomposition
Huitian Xue, Kairui Zhang, Chenxi Zhang, Zheng Li, Chuan Luo, Yilong Yang, Tianyu Wo, Chunming Hu, Yanliang Li, Yang Zhou, Penghao Zhang, Tiezi Zhang, and Renyu Yang — *Beihang University, China; Xidian University, China; Queen's University Belfast, UK; Independent Researcher, China; Kuaishou, China* 3684

When Elo Lies: Hidden Biases in Codeforces-Based Evaluation of Large Language Models in Practice
Shenyu Zheng, Ximing Dong, Xiaoshuang Liu, Chun Yong Chong, Gustavo Oliva, Dayi Lin, Boyuan Chen, Shaowei Wang, and Ahmed E. Hassan — *Huawei Canada, Canada; Huawei, China; Monash University, Malaysia; Queen's University, Canada; University of Manitoba, Canada* 3697

Maintaining Control States for Initialization-Aware GRPO in Industrial LLM Workflows
Chen Liu, Zhenzhao Yuan, Fengyuan Fan, Limei Che, Qiao Qian, Qingyuan Liang, Zeyu Sun, and Lu Zhang — *Peking University, China; Tencent, China; Institute of Software at Chinese Academy of Sciences, China* 3703

FPSieve: LLM-Powered Cascade for Filtering False Positives in OS Migration Testing
Jie Dong, Ying Fu, Chao Hu, Yuhang Chen, Qi Xiao, and Heyuan Shi — *Central South University, China; Tsinghua University, China; Hunan Congmao Technology, China* 3716

Disassembly-Augmented Root Cause Analysis for C/C++ Segmentation Faults in SAP HANA
Jihoon Jung, Chansong You, Jingun Hong, Thomas Bach, and Gabin An — *SAP Labs Korea, Republic of Korea; SAP, Germany; Korea University, Republic of Korea* 3729

Characterizing and Mitigating Productivity Loss in Large-Scale Model Training: An Empirical Study
Dinghao Xue, Yanxiang Chen, Junhong Liu, Yepeng Zhang, Guangsen Ni, Menghao Zhang, Tianyu Wo, Zheng Zheng, Chunming Hu, Jin Ouyang, and Renyu Yang — *Beihang University, China; Kuaishou, China; Unaffiliated, China* 3741

DBRepro: Automated Database Synthesis via a Hybrid Constraint-Solving Approach for Reproducing Slow Queries
Zhaoyang Zhang, Shuang Liu, Dengfeng Xu, Wei Lu, Jianquan Leng, Sheng Du, and Xiaoyong Du — *Renmin University of China, China; China Electronics Technology Kingbase (Beijing) Technologies, China* 3754

WiseSword: An Automated Computation Graph Partitioning Tool for Industrial DLRM Deployment on Ascend NPUs
Ke Cheng, Yuhan Zhang, Bo Pan, Jiawei Zhu, Qidong Zhang, Yuting Yan, Sheng Zhang, Yibo Jin, and Zhengyong Zhang — *Nanjing University, China; Huawei Technologies, China* 3767

Airdrop Scams in Practice: An Industry-Oriented Taxonomy, Benchmark, and Tool Audit
Ningyu He, Xiaohui Hu, and Haoyu Wang — *Hong Kong Polytechnic University, Hong Kong; Huazhong University of Science and Technology, China* 3779

ABACI: Automated Downstream Bug Resolution via Semantic-Aware Bisection and Convergent Backporting
Runzhe Wang, Yuanhang Zhou, Qingming Su, Yong Yang, Xiaohai Shi, and Tao Ma — *Alibaba Cloud Computing, China* 3792

KRCA: An Efficient Root Cause Analysis System in Hyper-scale Microservice Systems via Agentic AI
Jiamin Jiang, Jingfei Feng, Yu Luo, Qingliang Zhang, Yongqian Sun, Wenwei Gu, Shenglin Zhang, Tianyu Cui, Yao Wu, Jielong Huang, Nan Qi, and Dan Pei — *Nankai University, China; Kuaishou Technology, China; Tsinghua University, China* 3804

Understanding the Performance-Effectiveness Trade-Offs of AddressSanitizer in Real-World Programs: An Empirical Study
Jiaxin Xi, Cheng Wen, Zhiwei Lin, Yuandao Cai, Jie Su, Bin Yu, Zhuohua Li, and Shengchao Qin — *Xidian University, China; Hong Kong University of Science and Technology, China* 3816

Debug2Fix: Can Interactive Debugging Help Coding Agents Fix More Bugs?
Spandan Garg and Yufan Huang — *Microsoft Corporation, USA* 3829

ActionNex: A Production-Grade System for Next-Best-Action Recommendation in Cloud Outage Management
Zhenfeng Lin, Haoji Hu, Ming Hao, Xuchao Zhang, Yudi He, Oleg Kulygin, Hatay Tuna, Chetan Bansal, Ze Li, Murali Chintalapati, Sheila Jiang, Salman Zafar, Angie Anderson, and Ryan Zhang — *Microsoft, USA; Microsoft, UK* 3840

Expanding Guard Reach: Online Learning-Driven Incremental Kernel Panic Diagnosis
Yicheng Sui, Xiaotian Wang, Hailin Zhang, Shenglin Zhang, Binhong Jiang, Heng Zheng, Yongqian Sun, Yuzhi Zhang, Yang Zhang, Xin Wu, Liangyan Peng, Linlin Han, Feng Wang, Xiaozhou Liu, and Yu Zhang — *Nankai University, China; ByteDance, China* . . . 3852

An Empirical Study of Security Risks in the Android Push Notification Ecosystem
Shilong Hu, Zikan Dong, Chao Wang, Tianming Liu, and Haoyu Wang — *Huazhong University of Science and Technology, China* . . 3865

eARCO: Efficient Automated Root Cause Analysis with Prompt Optimization
Drishti Goel, Raghav Magazine, Supriyo Ghosh, Akshay Nambi, Prathamesh Deshpande, Xuchao Zhang, Chetan Bansal, and Saravan Rajmohan — *University of Illinois at Urbana-Champaign, USA; Microsoft Research, India; Inception, United Arab Emirates; Microsoft, India; Microsoft Research, USA; Microsoft, USA; Microsoft 365, USA* 3877

Prompt Coach: An Empirical Evaluation of an Agentic Tutor for Learning Prompt Engineering in Software Development
Rohit Mehra, Kapil Singi, Vikrant Kaulgud, Vibhu Saujanya Sharma, Swapnajeet Gon Choudhury, Swati Sharma, Adam P. Burden, and Majd Sakr — *Accenture Labs, India; Accenture, India; Accenture, USA* 3889

OpsAgent: An Evolving Multi-agent System for Incident Management in Microservices
Yu Luo, Jiamin Jiang, Jingfei Feng, Lei Tao, Qingliang Zhang, Xidao Wen, Yongqian Sun, Shenglin Zhang, Tong Liu, Wenjie Zhang, and Dan Pei — *Nankai University, China; Alibaba Cloud, China; Lenovo, China; Tsinghua University, China* 3896

Towards Reliable Alarm Flood Reduction in Data Center via Grammar-Guided Verifiable LLM Reasoning
Jiacheng Liu, Guocong Lin, Jichao Yan, Huiwen Zheng, Hongli Ma, Changhao Zhou, Zhiyi Yu, Yuanchun Li, Yunxin Liu, Ding Li, Tao Xie, and Yao Guo — *Peking University, China; Tsinghua University, China; Beijing University of Chemical Technology, China; GDS Holdings, China* 3908

Test Suite Generation for Simulink Models: An Experience with Reactis
Andrea Bombarda, Nunzio Marco Bisceglia, Marco Saini, Federico Formica, Angelo Gargantini, and Claudio Menghi — *University of Bergamo, Italy; Gran Sasso Science Institute, Italy; McMaster University, Canada* 3920

Beyond Single-Level PGO: Understanding and Harnessing Cross-Level Optimization Effects in Industrial Systems
Dong Liu, Yifan Cheng, Yunfeng Cui, Ting Chen, Jiachi Chen, Zhihong Xue, Chong Hu, and Shengyu Cheng — *ZTE Corporation, China; University of Electronic Science and Technology of China, China; Zhejiang University, China* 3933

DCAS: Decoupling CLI Agent Scaffolding to Internalize Planning across Scaffolds
Kishanthan Thangarajah, Boyuan Chen, and Ahmed E. Hassan — *Huawei Canada, Canada; Queen's University, Canada* 3945

CAPMix: Robust KPI Anomaly Detection for AIOps in Noisy and Dynamic Environments
Xudong Mou, Rui Wang, Tiejun Wang, Zexin Wu, Fangda Guo, Jie Sun, Shiru Chen, Penghao Zhang, Tiezi Zhang, Tianyu Wo, Hao Peng, Chunming Hu, Xudong Liu, and Renyu Yang — *Beihang University, China; Chinese Academy of Sciences, China; Inspur, China; Kuaishou, China* 3957

ControlsDSL: A Language for Verifiable Cloud Configuration Controls
Miyriee Alair, Shrutarshi Basu, Sean Cai, Loris D'Antoni, Julio Delgado Jr., Antonio Filieri, Andrew Gacek, Anand Goyal, Sebastiaan Joosten, Varad Panch, Sorawee Porncharoenwase, Vishwanath Raman, Niloofar Razavi, Neha Rungta, Mike Schliep, Preethi Sekaran, and Chase Johnson — *Amazon Web Services, USA; University of Minnesota, USA* 3969

Why Is This on My Bill? Documentation-Grounded Rule Learning for Itemized Bill Explanation
John Konstantinides, Jason Titus Lefever, Muhammed Numair Mansur, Martin Schaef, and Willem Visser — *Amazon Web Services, USA; Drexel University, USA; Amazon Web Services, Germany* 3981

Requirements-Augmented Generation for Trustworthy Acceptance Testing of LLM-Based Software
Fanyu Wang, Chetan Arora, Zhenping Xie, Yonghui Liu, Kla Tantithamthavorn, Aldeida Aleti, and Siwei Jiang — *Monash University, Australia; Jiangnan University, China; Australian National University, Australia; Yiwuyishi Intelligent Technology (Nantong), China; Tianjin University, China* 3993

Debugging the Debuggers: Failure-Anchored Structured Recovery for Software Engineering Agents
Chenyu Zhao, Shenglin Zhang, Yihang Lin, Wenwei Gu, Zhimin Chen, Yongqian Sun, Dan Pei, Chetan Bansal, Saravan Rajmohan, and Minghua Ma — *Nankai University, China; Tsinghua University, China; Microsoft, USA; Microsoft 365, USA* 4006

Automating Low-Risk Code Review at Meta: RADAR, Risk Calibration, and Review Efficiency
Chris Adams, Arjun Singh Banga, Parveen Bansal, Souvik Bhattacharya, Rujin Cao, Pedro Canahuati, Nate Cook, Brian Ellis, Prabhakar Goyal, Gurinder Grewal, Tianyu He, Matt Labunka, Alex Manners, David Molnar, Ging Cee Ng, Vishal Parekh, Jiefu Pei, Frederic Sagnes, James Saindon, Will Shackleton, Sid Sidhu, Gursharan Singh, Karthik Chengayan Sridhar, Matt Steiner, Pratibha Udmalpet, Sean Xia, Stacey Yan, Audris Mockus, Peter C. Rigby, and Nachiappan Nagappan — *Meta, USA; Meta, UK; University of Tennessee, USA; Concordia University, Canada* 4018

Understanding Real-World Ordering-Related Concurrency Bugs under Weak-Memory and Cross-Architecture Settings
Shaohao Wang, Cheng Wen, Yuandao Cai, Shengchao Qin, Jie Su, Mengda He, Xiaoxue Ma, and Cong Tian — *Xidian University, China; Hong Kong University of Science and Technology, China; Teesside University, UK; City University of Hong Kong, Hong Kong* 4030

CADscanner: Automating Malware Analysis for Compiled AutoCAD Scripts in Engineering Environments
Luxing Yin, Chang Sun, Chenghui Liang, Chu Wang, Shufu Zhi, Tianshi Zheng, Ziang Liu, Xin Liu, and Qingguo Zhou — *Lanzhou University, China; Sangfor Technologies, China* 4042

GraphQLer: Enhancing GraphQL Security with Context-Aware API Testing
Omar Tsai, Jianing Li, Tsz Tung Cheung, Lejing Huang, Hao Zhu, Jianrui Xiao, Iman Sharafaldin, and Mohammad A. Tayebi — *Simon Fraser University, Canada; Forward Security, Canada* 4054

Top-Down Multi-agent LLM System with Runtime Data for MISRA C Violation Detection
Ahcheong Lee, Yeongbin Kang, Sechang Jang, Heechan Yang, and Moonzoo Kim — *KAIST, Republic of Korea; VplusLab, Republic of Korea* 4066

Fuzz4DB: A Practice of LLM-Agent-Guided Fuzzing for Database Feature-Level Delta Testing
Chunling Qin, Yong Hu, Xiao Zhang, Jinchuan Chen, Fangtao Gu, Baoxun Wang, Yuxing Chen, Anqun Pan, and Lixiong Zheng — *Renmin University of China, China; Tencent, China* 4078

Kani: A Model Checker for Rust
Rémi Delmas, Zyad Hassan, Qinheping Hu, Rahul Kumar, Felipe R. Monteiro, Thanh Nguyen, Adrián Palacios, Celina Val, Michael Tautschnig, Justus Adam, Daniel Schwartz-Narbonne, and Carolyn Zech — *Amazon Web Services, USA; Amazon Web Services, Canada; Queen Mary University, UK; Brown University, USA; Datadog, USA; Massachusetts Institute of Technology, USA* 4091

Closing the Gap: Automated Discovery of Secure Dockerfile Reference Standards via Semantic Clustering in Enterprise Inner Source
Jessica Hösl, Benedikt Hofmann, and Patrick Stöckle — *TU Munich, Germany; Siemens, Germany* 4104

OpsLens: Runbook-Free Root Cause Analysis via Multi-agent Collaboration and Tool Management
Haiyu Huang, Jiewei Lyu, Gou Tan, Pengfei Chen, Zhihan Jiang, Guangba Yu, Penghui Mi, Fuqiang Wang, Jian Han, Chongkang Tan, and Michael R. Lyu — *Sun Yat-sen University, China; Chinese University of Hong Kong, China; Huawei, China; Independent Researcher, China* 4116

Li-WaC: A Lightweight WebAssembly Container for Independently Upgradable IoT Applications in Xiaomi Vela
Bin Liao, Canfeng Zheng, Shiwen Shan, Qi Huang, Junjie Dong, Xiang Xiao, Yaoyao Gu, and Yuxin Su — *Beijing Xiaomi Mobile Software, China; Sun Yat-sen University, China* 4128

KVerus: Scalable and Resilient Formal Verification Proof Generation for Rust Code
Yuwei Liu, Xinyi Wan, Yanhao Wang, Minghua Wang, Lin Huang, and Tao Wei — *Ant Group, China; Independent Researcher, China* 4139

Beyond Benchmarks: A Case Study of LLM-Generated Verus Specification Failures on Asterinas Vostd
Dugang Liu, Ding Wang, Chao Jiang, Cheng Wen, Zhiwu Xu, Zhong Ming, Yuwei Liu, Xinyi Wan, and Lin Huang — *Shenzhen University, China; Ant Group, China* 4152

DepWareTrans: Dependency-Aware Incremental Repository Migration across Co-executable Languages
Sivajeet Chand, Alexander Pretschner, Steve Haupt, Derui Zhu, and Sushant Kumar Pandey — *TU Munich, Germany; andrena objects, Munich, Germany; University of Groningen, Netherlands* 4164

Search-Based Testing of Vision Language Models for In-Car Scene Understanding
Lev Sorokin, Chen Yang, Ken E. Friedl, and Andrea Stocco — *TU Munich, Germany; BMW Group, Germany; fortiss, Germany* 4175

Doc2Test: An Industrial Multi-agent Framework for Document-Driven End-to-End Selenium Test Generation
Andrea Lops, Antonio Ferrara, Fedelucio Narducci, Azzurra Ragone, and Michelantonio Trizio — *Polytechnic University of Bari, Italy; Wideverse, Italy; University of Bari, Italy* 4186

Customizing an LLM for Enterprise Software Engineering
Aditya Kini, Satish Chandra, Milad Hashemi, Saksham Thakur, Aditya Pandey, Vincent Nguyen, Marc Brockschmidt, Franjo Ivančić, Danny Tarlow, Parthasarathy Ranganathan, Petros Maniatis, Ahmed Omran, Zaheer Abbas, Anita Gergely, Martin Sevenich, Gufeng Zhang, Amy Hua, and Alexander Frömmgen — *Google, USA; Meta, USA; Google DeepMind, USA; Google DeepMind, UK; Google DeepMind, Canada; Google, Germany* 4199

From Security Findings to Verified Fixes: Automated Repair of Infrastructure-as-Code Misconfigurations
Ying Li, Tamjid Al Rahat, Yingjun Lyu, Omer Tripp, and George Argyros — *University of California at Los Angeles, USA; Amazon Web Services, USA* 4210

RiskScope: Pre-audit Triage for Incentive and Control Risks in Smart Contracts
Yu Pan, Qi Zhang, Yuannan Yang, and Dan She — *CertiK, USA* 4222

Catching Developers in the Flow: Low-Latency Agentic Program Repair at Google Scale
Celal Ziftci, Spencer Greene, Ray Liu, Livio Dalloro, and Lorenzo Dini — *Google, USA* 4229

IAM Policy Autopilot: Static Analysis for Policy Generation from Application Code
Antonio Filieri, Luke Kennedy, Kevin Luo, Matt Luttrell, Adrián Palacios, Akash Panda, Neha Rungta, Matthias Schlaipfer, Karan Jit Singh, Anvesh Tanuku, Nick Winans, Diana Yin, and Weiben Zhang — *Amazon Web Services, USA; Amazon Web Services, Germany* 4242

Vulnerability Tracking using Normalized Scope+Offset
Julian Thome, Hua Yan, Lucas Charles, Craig Smith, and Jason Leasure — *GitLab, Luxembourg; GitLab, Australia; GitLab, USA; Independent Researcher, USA* 4254

DAL: Decoupled Aspect-Based Layering of I/O Processing for Automotive Software
Muhammet Uslu, Atif Imran Penkar, Robert Rasche, Olaf Spinczy, Marc Fessler, and Michael Unterreiner — *Cariad - Volkswagen Group, Germany; TU Dortmund, Germany; Tensor embedded, Germany; Osnabrück University, Germany; Porsche, Germany* 4260

A Longitudinal Analysis of Software Development Evolution in a Financial Organization
Wentao Chen, Zehong Zhang, Huiqun Yu, Guisheng Fan, Huan Wang, and Kaiwen Zhi — *East China University of Science and Technology, China; Orient Securities Company Limited, China* 4266

Manifesting Implicit Dependencies in Requirements Task: An Automated Method Integrating Code Awareness and Change Analysis
Hui Ma, Jiajun Yang, Yan Chen, Pingjia Liao, Baoyu Zhao, and Daoping Jiang — *Shenzhen Poweroak Newener, China* 4279

Deployment Risk Assessment using Diff-Aware Features: A Case Study at Prime Video
Mayur Premkumar Kurup, Hyunjae Suh, Swathi Vaidyanathan, Pranesh Vyas, Srinidhi Madabhushi, and Yegor Silyutin — *Amazon.com, USA; University of California at Irvine, USA* 4286

MERA: A Multi-expert LLM Framework for Severity Assessment of User Tickets in Mobile OS Beta Testing
Yongqian Sun, Yimin Zuo, Qingliang Zhang, Bowen Hao, Wenwei Gu, Jiaqi Luan, Sixu Zhou, Shenglin Zhang, and Dan Pei — *Nankai University, China; Tsinghua University, China* 4292

REAP: Automatic Curation of Coding Agent Benchmarks from Interactive Production Usage
Smriti Jha, Matteo Paltenghi, Chandra Maddila, Vijayaraghavan Murali, Shubham Ugare, and Satish Chandra — *Meta Platforms, USA* 4304

If You Build It, Will They Come? Lessons from 20 Years of Industrial Use of a Code Transformation DSL
Johan Fabry and Vadim Zaytsev — *Raincode Labs, Belgium; University of Twente, Netherlands* 4315

LLM-Assisted Joint Ticket and Log Analysis for Incident Triage in Intelligent and Connected Vehicles
Ruowei Fu, Shenglin Zhang, Wenwei Gu, Weiguo Li, Yongqian Sun, and Dan Pei — *Nankai University, China; Huawei, China; Tsinghua University, China* 4328

No Isolated Evolution, No Blind Search: Memory-Guided Prompt Co-evolution for Industrial Multi-agent Vulnerability Detection
Jie Xu, Zihan Wu, Yun Peng, Chun Yong Chong, Linghan Meng, Kui Liu, and Xiaohua Jia — *City University of Hong Kong, Hong Kong; Chinese University of Hong Kong, Hong Kong; Monash University, Malaysia; Huawei Technologies, China* 4340

AgenticSCR: An Autonomous Agentic Secure Code Review for Immature Vulnerabilities Detection
Wachiraphan Charoenwet, Kla Tantithamthavorn, Patanamon Thongtanunam, Hong Yi Lin, Minwoo Jeong, and Ming Wu — *University of Melbourne, Australia; Monash University, Australia; Atlassian, USA* 4352

AlarmClaw: Context-Enriched Alarm Management with Category-/Severity-Aware Incident Graphs
Siyu Yu, Meizhen Li, Jiacheng Yang, Yifan Wu, Ning Wang, Zhaoxing Chen, Ming Liu, Xinchi Ren, Xincheng Ren, Chan Li, Tong Jia, Xiang Zhang, and Ying Li — *Peking University, China; ByteDance, China* 4364

EvoLLM: High-Coverage, Readable, and Token-Efficient Unit Test Generation for Industrial Java Codebases
Seokhyeon Moon, Myungho Lee, Jihun Kim, Yoon-Chan Jhi, Jihoon Cho, and Hakjoo Oh — *Samsung SDS, Republic of Korea; Korea University, Republic of Korea* 4376

InfraAPR: An Agent-Based Repair Framework for Cloud Infrastructure-as-Code Programs
Quan Chen, Zhao Liu, Chunfang Li, Yifan Zhang, Ning Hu, and Qing Liao — *Harbin Institute of Technology, Shenzhen, China; Pengcheng Laboratory, China; Hunan University, China; Nanchang University, China* 4389

From Backlog Items to Security Guidance: Towards Continuous Security Compliance
Ignacio García Núñez, Florian Angermeir, and Fabiola Moyón Constante — *Technical University of Munich, Germany; fortiss, Germany; Blekinge Institute of Technology, Sweden; Siemens Technology, Germany* 4402

LiteScope: Revealing the Security Paradox of Android Lite Apps in Resource-Constrained Ecosystems
Yuanxiang Shi, Liangcai Su, Kai Ye, Xi Xiao, and Chenxiong Qian — *University of Hong Kong, Hong Kong; Tsinghua University, Shenzhen, China* 4413

Mutation Testing of Simulink Cyber-Physical System Models: Challenges and Solutions in Practice
Murat Kavak, Onur Kilincceker, Serge Demeyer, Kevin Vandenbroucke, and Halim Abdurrahman Ceylan — *University of Antwerp, Belgium; Flanders Make, Belgium; Duco, Belgium; Ege University, Türkiye* 4425

FlowChecker: One Command to Model Check Control-Flow Rules in Industrial Java
Qiuye Wang, Liqing Cao, Wei Hu, and Dengcheng He — *Alibaba Cloud Computing, China* 4432

Kozuchi Agent: A Language-Agnostic Open-Weight Agent for Software Repair
Mehdi Bahrami, Kosaku Kimura, Satoshi Munakata, Satoshi Nakashima, Yu Ishikawa, Kosuke Maeda, Nao Soma, Kenichi Kobayashi, Keisuke Miyazaki, Keizo Kato, Shigeki Fukuta, Tatsuo Kumano, Nobutaka Imamura, Kevin Musgrave, Shahbaz Abdul Khader, Kwun Ho Ngan, Joe Townsend, Fayas Asharindavida, Matthieu Parizy, Akira Sakai, Yuma Ichikawa, Yang Zhao, Michiaki Takizawa, Taku Fukui, Hiroki Ohtsuji, Wei-Peng Chen, and Hiromichi Kobashi — *Fujitsu Research of America, USA; Fujitsu Research, Japan; Fujitsu Research of Europe, UK; Fujitsu Research & Development, China* 4443

Continuous On-Commit Certification Testing of Railway Control Software: An Experience Report
Jacopo Maltagliati, Salvatore De Simone, and Giovanni Denaro — *University of Milano-Bicocca, Italy; Rete Ferroviaria Italiana, Italy* 4456

IP-Preserving Enterprise System Modernization via LLM-Driven Specification Extraction and Multi-target Code Generation
Andreas Biesdorf, Eckhard Seibert, Nicole Wengatz, Bart Carpels, and Tom Schoonbaert — *Trier University of Applied Sciences, Germany; Siemens, Germany; Euroports, Belgium* 4463

Towards Reliable C-to-Rust Translation with Rule-Guided Reasoning and Reinforcement Learning
Feng Luo, Jiachen Liu, Cuiyun Gao, Jia Feng, and Kui Liu — *Harbin Institute of Technology, Shenzhen, China; Huawei Technologies, China* 4469

Continuum: Automated Construction and Retrieval of Software Decision Knowledge Graphs from Developer-AI Conversations
Mohammad Ali Shehral, Karthik Ravi, Nikhil Trivedi, and Akram Bayat — *Northeastern University, USA* 4481

Command Coding: A Methodological Framework for AI-Accelerated Software Development
Thomas S. Villani and Martin Kellogg — *New Jersey Innovation Institute, USA; New Jersey Institute of Technology, USA* 4493

Model-Guided Property-Based Testing of WeChat Pay at Billion-User Scale
Xiangchen Shen, Yiting Wang, Ting Su, Jingjing Liang, Jingling Sun, Xixian Liang, Haiying Sun, Xinjie Xu, Haochuan Lu, Yuetang Deng, Pengcheng Wang, Geguang Pu, Zhendong Su, and John Hughes — *East China Normal University, China; University of Electronic Science and Technology of China, China; Tencent, China; ETH Zurich, Switzerland; Chalmers University of Technology, Sweden* 4504

New Ideas and Emerging Results

Beyond the First Crash: Interaction-Aware Multi-fault Fuzzing
Omar I. Al-Bataineh — *Gran Sasso Science Institute, Italy* 4517

Phantom Fixes: Fault Interaction Ambiguity in Multi-fault Concurrent Programs
Omar I. Al-Bataineh — *Gran Sasso Science Institute, Italy* 4522

Empirical Computation: Prompting versus Programming
Eric Tang, Jing Liu, and Marcel Böhme — *Carnegie Mellon University, USA; MPI-SP, Germany; CISPA Helmholtz Center for Information Security, Germany* 4527

End-to-End Regulatory Compliance Evolution Pipeline: From Legal Requirements to Developer Tests
Carolin Brandt and Sallam Abualhaija — *Delft University of Technology, Netherlands; University of Luxembourg, Luxembourg* 4532

Towards Diagnostic Explainability in Workflow Verification via Shapley Attribution
Radosław Klimek and Jakub Błażowski — *AGH University of Krakow, Poland* 4538

PR2Trigger: From Concurrency-Fix PR Diffs to Executable Triggering Tests for Go
Risheng Wei, Yuheng Shen, Yiming Ye, Yuchen Yan, Yihao Peng, Hao Xi, Yue Chen, Hai Wan, and Xibin Zhao — *Tsinghua University, China* 4543

ECLAIR: A Causally-Grounded AI Framework for Scientific Discovery in Empirical Software Engineering
Alejandro Velasco, Daniel Rodriguez-Cardenas, Dipin Khati, David N. Palacio, and Denys Poshyvanyk — *William & Mary, USA; Microsoft, USA* 4548

Let's Chat about Design: A Transformation-Oriented Viewpoint for LLM-Native Software
Víctor A. Braberman and Flavia Bonomo-Braberman — *University of Buenos Aires, Argentina; CONICET, Argentina* 4554

Agentic Code Review in the Terminal: A Trajectory-Level Analysis of Behavior, Cost, and Human-Alignment
Wachiraphan Charoenwet, Kla Tantithamthavorn, Patanamon Thongtanunam, Hong Yi Lin, Minwoo Jeong, and Ming Wu — *University of Melbourne, Australia; Monash University, Australia; Atlassian, USA* 4559

Is Three the Magic Number? An Empirical Evaluation of LLM-Based Repair Loops
Tobias Kiecker, Eik Reichmann, Hosung Kang, Gabin An, and Lars Grunske — *Humboldt-Universität zu Berlin, Germany; Korea University, Republic of Korea* 4564

BanditFuzz: A Synergistic Differential Fuzzing Framework for Compilers
Donghui Li, Yingying Li, Jinlong Xu, Bo Zhao, Nan Wang, Jinyang Yao, and Lili Liu — *Information Engineering University, China; Henan Normal University, China* 4569

Can Formal Specifications Be Synthesized from Tests Alone?
Tianhai Liu, Maximilian Müller, Tobias Hey, Vitus Lüntzel, Muhammad Minhas, Anne Koziolek, and Bernhard Beckert — *KIT, Germany* 4575

SkillMOO: Multi-objective Optimization of Agent Skills for Software Engineering
Jingzhi Gong, Ruizhen Gu, Zhiwei Fei, Yazhuo Cao, Lukas Twist, Alina Geiger, Shuo Han, Dominik Sobania, Federica Sarro, and Jie M. Zhang — *King's College London, UK; Queen's University Belfast, UK; Nanjing University, China; Johannes Gutenberg University Mainz, Germany; University College London, UK; paluno, Germany; University of Duisburg-Essen, Germany* 4580

Specifying the Delegated-Autonomy Boundary: Requirements Engineering for Agentic AI
Chetan Arora, Andreas Vogelsang, and Abbi Sharma — *Monash University, Australia; University of Duisburg-Essen, Germany* 4585

FaultForge: A Business-Semantics-Driven Cross-Layer Fault Injection Framework for Microservice
Yongxin Yang, Yang Yang, Anni Jiang, Mingyuan Yang, and Peng Yu — *Beijing University of Posts and Telecommunications, Beijing, China* 4590

The Ground Is Shifting: A Reflection on the Foundations of Software Measurement
Thomas Bock, Audris Mockus, and Bogdan Vasilescu — *Carnegie Mellon University, USA; University of Tennessee, USA* 4595

Do Not Copy/Paste: Soft Barriers for Copying in AI-Assisted Programming
Iyiola E. Olatunji, Alberick Euraste Djire, Jacques Klein, and Tegawendé F. Bissyandé — *University of Luxembourg, Luxembourg* . . 4601

Distributional Program Analysis: Treating Large Language Models as Program Samplers
Phat T. Tran-Truong and Xuan-Bach Le — *Ho Chi Minh City University of Technology - Vietnam National University Ho Chi Minh City, Vietnam* 4606

UIXPOSE: Screen-Conditioned Runtime Discrepancy Analysis for Security-Aware Android GUI Exploration
Amirmohammad Pasdar, Van-Thuan Pham, and Toby Murray — *University of Melbourne, Australia* 4611

Agents That Teach: Towards Designing Incidental Learning Back into AI-Assisted Software Development
Rohit Mehra, Samdyuti Suri, Prithviraj K. Tagadinamani, Kapil Singi, Vikrant Kaulgud, and Adam P. Burden — *Accenture Labs, India; Accenture, USA* 4616

Towards Multilingual AI-Assisted Software Engineering
Shivadharshan S, Anirudh Arrepu, and Sridhar Chimalakonda — *Indian Institute of Technology Tirupati, India* 4621

Software Engineering Meets the EU AI Act: Artifacts, Patterns, and Future Directions
Sridhar Chimalakonda and Upendra Karimi — *Indian Institute of Technology Tirupati, India* 4626

K-COBAgent: A Knowledge-Driven Multi-agent Framework for COBOL Modernization
Kavyasri Gajula, Phanindra Kodela, and Sridhar Chimalakonda — *Indian Institute of Technology Tirupati, India* 4631

SRC

PCLA-Bridge: The Reusability of CARLA-Trained Autonomous Driving Agents beyond the CARLA Simulator
Masoud Jamshidiyan Tehrani — *Università della Svizzera italiana, Switzerland* 4636

Where There's One Patch, There's Another: Taxonomy-Driven Discovery of Missing Security Patches
Fangyuan Zhang — *Nankai University, China* 4639

Adaptive Parallel Testing for Floating-Point Branch Coverage
Yunxiang Wu — *Nanjing University, China* 4642

An Approach toward Automated Generation of Contextual Privacy Policies for Mobile Applications
Zhen Tao — *TU Munich, Germany* 4645

Making Vulnerability Proof-of-Concepts Trustworthy and Usable: From Empirical Diagnosis to Operational Validation
Wenjing Dang — *Tianjin University, China* 4648

Correct According to Whom? Auditing What LLM-Generated Tests Actually Check
Sarah Dagger — *University of British Columbia, Canada* 4651

Danger Is a Counterfactual: Toward a Falsifiable Continuous Test Oracle for Autonomous Driving Systems
Borhaneddine Hamadou — *Université Paris-Saclay, France; CEA List, France* 4654

Code Detectors Have a Half-Life: Obsolescence and Metric Illusions in LLM-Generated Code Detection
Alberick Euraste Djire — *University of Luxembourg, Luxembourg* 4657

Which Camera Perturbation Should We Test First? Offline-Guided Prioritization of On-Vehicle ADS Tests
Matthias Weil — *TU Munich, Germany* 4660

Beyond Outcomes: Toward Process-Aware Software Agents
Shuyang Liu — *University of Illinois at Urbana-Champaign, USA* 4663

Order-Aware State Elimination for Algebraic Program Analysis
Boxuan Zheng — *Zhejiang University, China* 4666

Toward Automatically Detecting Cross-Engine Asymptotic Complexity Divergence in JavaScript
Minseok Choe — *Korea University, Republic of Korea* 4669

Toward Authentic and Scalable Evaluation of Terminal Agents in Real-World Environments
Zhaoyang Chu — *University College London, UK* 4672

WiseSpec: Requirements-Driven Agents for Code Generation
Zhao Tian — *Tianjin University, China* 4675

Multi-modal Boundary Testing of Vision–Language Models
Ruben Kaiser — *TU Munich, Germany* 4678

Reconstructing Executable Programs from Informal Bug Reports
Ragib Shahariar Ayon — *Texas State University, USA* 4681

Measuring the Cost of Prompt Structure in LLM Code Generation
Yuetian Mao — *TU Munich, Germany* 4684

From Accessibility to Reachability: LLM-Driven Contrastive Accessibility Testing for Android Apps
Shen Hu — *TU Munich, Germany* 4687

From Automated Program Repair to Automated Agent Repair: Why Failed Trajectories Matter
Xiangxin Zhao — *University College London, UK* 4690

MACO-AI: Multi-agent Code Optimization via Annotations and Intrinsics
Junyu Liu and Zongyan Xian — *Zhejiang University, China* 4693

Secrets without Signatures: An LLM Agent for Logic-Defined Exposures in Recovered JavaScript
Shriyans Sudhi — *Rochester Institute of Technology, USA* 4696

Author Index 4699

Test Suite Generation for Simulink Models: An Experience with Reactis

Andrea Bombarda
andrea.bombarda@unibg.it
University of Bergamo
Bergamo, Italy

Nunzio Marco Bisceglia
nunziomarco.bisceglia@gssi.it
Gran Sasso Science Institute
L'Aquila, Italy

Marco Saini
m.saini1@studenti.unibg.it
University of Bergamo
Bergamo, Italy

Federico Formica
formicaf@mcmaster.ca
McMaster University
Hamilton, ON, Canada

Angelo Gargantini
angelo.gargantini@unibg.it
University of Bergamo
Bergamo, Italy

Claudio Menghi
claudio.menghi@unibg.it
University of Bergamo
Bergamo, Italy
McMaster University
Hamilton, ON, Canada

Abstract

Testing Simulink® models is a widely recognized, relevant, and challenging problem for automated software engineering. It is relevant to industry practitioners since Simulink® is a development and simulation language widely used by the Cyber-Physical System (CPS) industry. It is challenging since CPS are usually complex, entail many components, and failures are not easily identifiable without automated support. This paper reports our experience in applying Reactis®, an industrial tool for testing Simulink® models, to the ten Cyber-Physical Challenges Simulink® models from Lockheed Martin. We considered 48 requirements from these challenge models. We assess the effectiveness of Reactis® in identifying failure-revealing test cases and its efficiency. We compare our results and those from an academic testing tool (HC), and two model checking tools (QVtrace and CoCoSim). Our results show that Reactis® could return a failure-revealing test case for a requirement that was considered correct. They also show that the efficiency of Reactis® is comparable to that of existing tools.

CCS Concepts

• **Software and its engineering → Software testing and debugging**.

Keywords

Model Checking, Model Testing, Simulink® Model, Reactis

ACM Reference Format:
Andrea Bombarda, Nunzio Marco Bisceglia, Marco Saini, Federico Formica, Angelo Gargantini, and Claudio Menghi. 2026. Test Suite Generation for Simulink Models: An Experience with Reactis. In *Proceedings of the 41st IEEE/ACM International Conference on Automated Software Engineering (ASE '26), October 12–16, 2026, Munich, Germany*. ACM, New York, NY, USA, 13 pages. https://doi.org/10.1145/3832783.3834483



This work is licensed under a Creative Commons Attribution 4.0 International License.
ASE '26, Munich, Germany
© 2026 Copyright held by the owner/author(s).
ACM ISBN 979-8-4007-2882-2/2026/10
https://doi.org/10.1145/3832783.3834483

1 Introduction

Cyber-Physical Systems (CPS) development requires engineers to model the behavior of the software and its environment [22, 44, 62, 74]. The model of the environment describes the behavior of the physical entities where the software is deployed. For example, it specifies how physical objects move and react to external stimuli. The software model describes how the software components sense and modify their environment. For example, it can describe how robots respond to the presence of humans in their surroundings.

Simulink® [46] is the standard development and simulation language for CPS design. We consider Simulink® since it is the prevalent modeling language for many CPS domains, such as automotive [49, 78]. It is used by more than 60% of CPS engineers [4, 76].

Considering the safety-critical nature of CPSs, failures must be avoided. Testing is a standard practice used to search for failures within Simulink® models. There are many industrial (e.g., Simulink® Design Verifier [45], Reactis® [65], QVtrace [64][1]) and academic tools (e.g., ARIsTEO [55], ATheNA [31], `FalCAuN` [71], FlexiFal [42], ForeSee [75], and Ψ-TaLiRo [70]) that support the verification of Simulink® models. There are also mature academic tools ready for industrial applications (e.g., S-TaLiRo [30] and Breach [23]). For academic tools, competitions aim to establish the benefits and limitations of different solutions. For example, the falsification category of the annual International Competition on Verifying Continuous and Hybrid Systems (ARCH-COMP) focuses on testing tools that support Simulink® models and outlines results in yearly reports (e.g., [25–27, 40, 53]). These competitions provide relevant and valuable information on the progress made by the research community. However, they usually do not provide information related to commercial tools. Additionally, their objective is not to report experiences of applying testing solutions for Simulink® models in practice and identify practical challenges. Dedicated experience papers focus instead on industrial testing tools. For example, a recent work [59] assessed the usefulness of the industrial model checker QVtrace by comparing its effectiveness and that of an academic implementation (HC) of a hill-climbing-based search-based approach. The results show that model checking and model testing are complementary.

[1] QVtrace is no longer available.

To document how industrial software engineering testing solutions behave on Simulink® models, we report on our experience in applying Reactis® in practice. Reactis® is a commercial tool [18, 65] for testing, simulating, and validating Simulink® models and C code. It has been extensively used to test models from different domains, such as the automotive [17, 20, 38, 58, 68] and aerospace/defense [18] domains. We analyzed Reactis® by considering the Lockheed Martin Cyber-Physical challenge models [24, 51, 59]. These models were used to assess many tools for requirement elicitation and analysis (e.g., FRET [36] and Theano [54, 73]), synthesis of assumptions (e.g., EPIcuRus [34, 35]), and many others. They were also used as a benchmark for comparing testing solutions (e.g., [59] and [14]). In this work, we compared Reactis® with three baseline solutions: HC, QVTrace, and CoCoSim [14]. HC and CoCoSim are products from the research community; QVTrace is an industrial tool. We selected these tools because they are standard frameworks for verifying Simulink® models that were previously considered by industrial experience reports [50, 59]. Additionally, they were assessed over the Lockheed Martin Cyber-Physical challenge models. To assess the usefulness of Reactis®, we compare its effectiveness with that of our baselines. We analyzed 48 requirements from our challenge models, and we compared the effectiveness and efficiency of Reactis® and our baselines.

Our results show that Reactis® reported a failure-revealing test case for 100% (25 out of 25) of the cases in which other existing testing tools (HC) returned a failure-revealing test case. Remarkably, Reactis® could return a failure-revealing test case for a requirement that was considered correct, i.e., always satisfied, by two model checkers (QVTrace and CoCoSim). Our results also show that the efficiency of Reactis® is comparable to that of our baselines. We reflect on our experience in applying Reactis® in practice, and highlight practical challenges and lessons learned.

This paper is organized as follows. Section 2 introduces the Reactis® testing tool. Section 3 evaluates Reactis® on the Cyber-Physical Challenges Simulink® models from Lockheed Martin and compares its results with our baseline tools. Section 4 reflects on our results and discusses their practical implications. Section 5 presents related work. Section 6 concludes our work and outlines directions for future research.

2 The Reactis® Testing Tool

Reactis® [18, 65] is an industrial tool for testing, simulation, and validation of Simulink® models and C code. Figure 1a shows the Reactis® graphical interface. The right side of the graphical interface shows the Simulink® model loaded for analysis. For example, the right side of Figure 1a shows the Neural Network (NN) model, a model from our benchmark that will be discussed in Section 3. The left side of the interface provides a model browser that enables navigating model components. The menu bar and a toolbar provide the *play* button (to start the execution of a test suite) and the *step* button that supports simulation and trace navigation.

Reactis® allows engineers to perform three tasks: *Requirements specification*, *automatic test generation*, and *interactive simulation*.

Requirements specification. Reactis® enables engineers to add requirements within their Simulink® models. The bolt symbol within the Simulink® model indicates the presence of requirements. For example, the bolt symbols from Figure 1a indicate the presence of requirements Req 1, Req 2, and Req 3. Blue arrows indicate the inclusion of an input or output signal within the requirement, as shown in Figure 1d. Reactis® enables engineers to model requirements as assertions (i.e., conditions that must always hold). Assertions are c-like Boolean expressions. They can contain mathematical ("+", "−", "∗", "/"), relational (">", "<", ">=", "<=", "==", and "! ="), Boolean ("!", "− >", "||", "&&") operators, and conditional evaluations (if-then-else). Assertions also enable engineers to use mathematical functions (e.g., *pow*, *abs*, *exp*), vectors, or values from the previous simulation steps (*pre*). For example, the assertion $z < 1.1$ for the NN model requires the value of the z variable to be always lower than 1.1.

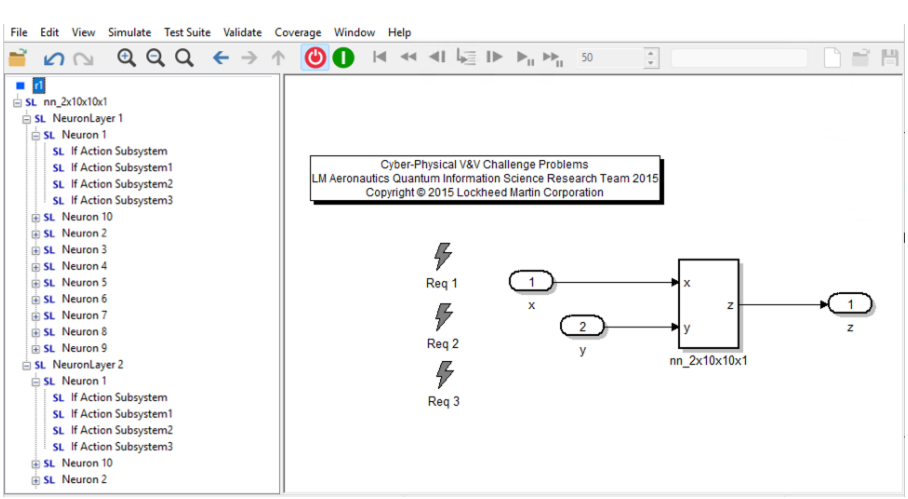
Automatic test generation. Reactis® generates test cases to detect requirements violations driven by coverage goals. Reactis® assesses how a test suite exercises the model under test and reports whether modes, transitions, blocks, or decision points have been exercised. Reactis® exploits *guided simulation* [19] to generate test suites, i.e., an automated exploration of the model behavior that iteratively steers the input values toward yet-uncovered coverage targets. The generation is driven by uncovered targets and the selected coverage metrics and includes a pseudo-random component and an SMT-based component [18]. Specifically, given a coverage criterion, guided-simulation iteratively generates inputs by generating a portion of the input, tracking which coverage targets have been hit and which have not, and using this information to generate the next portion of the input. To generate the next portion of the input (a.k.a. input selection), Reactis® can be configured to use a constraint solver (supported by the SMT technology, e.g., Microsoft Z3 [21]) or the Monte Carlo method [57] (supported by several heuristics). The constraint solver solution generates inputs that are guaranteed to extend coverage. However, in practical scenarios, constructing the constraints can be infeasible, constraints may fall outside the theories supported by the solver, or computing the solution might be too computationally expensive. Unlike the constraint solver, the Monte Carlo method generates inputs very quickly and supports any model, but does not ensure that coverage grows across iterations.

Figure 1b shows the configuration window of the test generator. Engineers can configure a Reactis® tester run by setting the duration of test generation, i.e., the number of tests and steps (*Run for*), the coverage criteria to be achieved (*Coverage*), and the assertions to be considered. The Reactis® Validator searches for executions that violate these assertions.

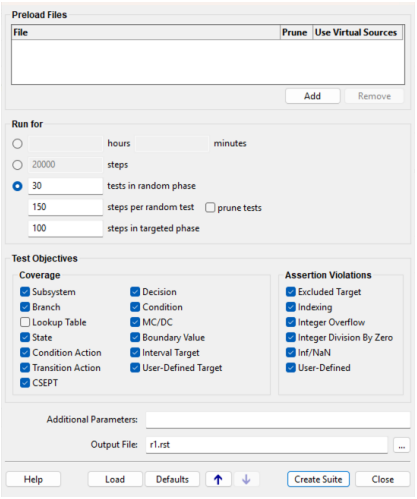
For each assertion, Reactis® returns one of the following results [66]:

(1) *Violated*, if a violation of the assertion is found;
(2) *Not violated*, if no violation has been detected; and
(3) *Impossible to violate*, if Reactis® has statically determined that the assertion cannot be violated.

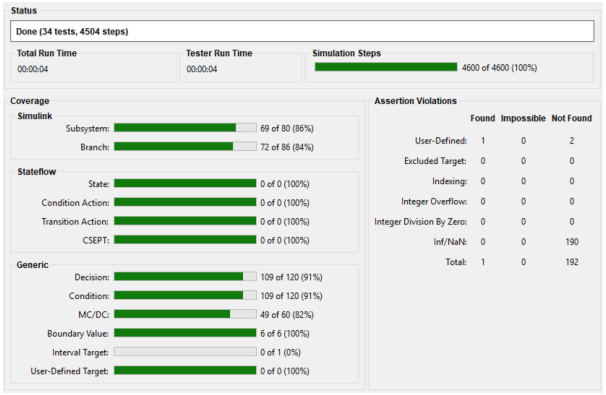
We use the symbols "✗", "?", and "⊥" to respectively indicate the "*Violated*", "*Not violated*", and "*Impossible to violate*" verdicts. Reactis® is not a model checker: if it does not find any violation for an assertion (?), it does not guarantee that the requirement is satisfied. The "*Impossible to violate*" verdict is typically returned when the requirements can not be violated by the Simulink® model under



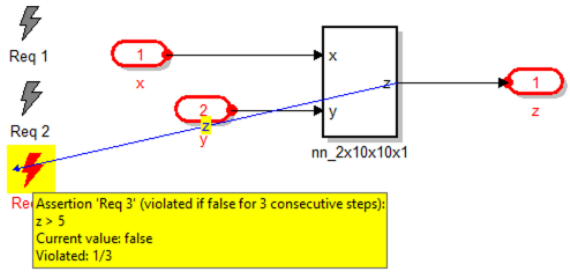
(a) Graphical interface.



(b) Configuration Window.



(c) Test generation report.



(d) Example of requirement violation.

Figure 1: Graphical interface, test case configuration window, test generation report, and requirement violation from Reactis®.

test, e.g., because the requirements are tautologies based on input value domains.

If an assertion violation is found, Reactis® returns a failure-revealing test case (i.e., an input sequence) that shows the requirement violation. For example, Figure 1c shows the report of the Reactis® test generation process, where the achieved coverage levels are reported (*Coverage*), together with *Assertion Violations*. Figure 1d reports the visual representation of an assertion violation: The corresponding requirement is highlighted in yellow, and the description of a violation is shown in a yellow box.

Interactive simulation. Reactis® enables engineers to re-execute failure-revealing test cases through simulation. Specifically, Reactis® enables engineers to execute the model step-by-step. Engineers can observe how the system's state changes in response to changes in the input signals.

3 Evaluation

Our evaluation considers two research questions:

- **Effectiveness (RQ1)**. How *effective* is Reactis® in generating failure-revealing test cases? (Section 3.4)
- **Efficiency (RQ2)**. How *efficient* is Reactis® in generating failure-revealing test cases? (Section 3.5)

We describe the baseline techniques (Section 3.1), our benchmark of Simulink® models (Section 3.2), and the methodology (Section 3.3) we used to assess the effectiveness and efficiency of Reactis®.

The material to replicate our experiments is available online [11].

3.1 Baselines

We compare Reactis® with three baselines: HC, QVTrace, and CoCoSim. We selected HC and QVTrace because they were considered in a previous study [59] that compared the effectiveness of model checking and model testing for verifying Simulink® models. We selected CoCoSim since it supports the analysis of discrete Simulink® dataflow models. For all three baselines, we rely on the results reported in prior work. We reproduced the corresponding experiments (when the baseline tool is still available) only when necessary to clarify or resolve potential ambiguities in our results.

Table 1: The Ten Lockheed Martin Cyber-Physical Challenges.

ID	Name	Description
TSM	Triplex Signal Monitor	A redundancy management system preventing errors from propagating in an airborne application.
FSM	Finite State Machine	An autopilot system interacting with an independent sensor to ensure safe operations near hazards.
TUI	Tustin Integrator	A flight software utility for signal integration with non-linear elements.
REG	Control Loop Regulators	A regulator architecture for feedback control applications.
NLG	Nonlinear Guidance	A guidance system for an aerial vehicle to follow a moving target while respecting safety distances.
NN	Neural Network	A neural network predictor with two hidden layers and a feedforward architecture.
EB	Effector Blender	A method to calculate optimal effector configurations for vehicles.
AP	6DoF Autopilot	A full six-degree-of-freedom simulation of an airplane with autopilot capabilities.
SWIM	System-Wide Monitor	Monitors airspeed to warn operators when evasive maneuvers are unachievable.
EUL	Euler Transformation	Creates a rotation matrix for inertial frame rotations in Euclidean space.

HC (Hill Climbing). It is a testing framework based on local search. We considered the implementation from Matinnejad et al., [48] since it is already applied to our benchmark models, and it was used as a baseline for comparing the effectiveness of model checking and model testing on the provided models. HC relies on a robustness metric used as a fitness function and on hill climbing for test case generation. It returns two verdicts:

(1) *Violation found* indicating that the requirement does not hold on the model; and
(2) *No violation found* indicating that the tool could not return a failure-revealing test case. If HC does not return a failure-revealing test case, it does not ensure that the requirement is satisfied.

We use the symbols "✗" and "?" to indicate these verdicts.

QVtrace [64]. It is an SMT-based model checker for Simulink® models developed by QRA Corp [63].[2] To verify whether a model satisfies a requirement, QVtrace generates a logical formula and uses an SMT-solver to prove its satisfiability. Depending on the results returned by the SMT-solver, QVtrace can return four verdicts:

(1) *No violation exists*, indicating that the requirement is satisfied;
(2) *No violation exists for* $0 \leq k \leq k_{max}$. The model satisfies the requirement in the time interval $[0, k_{max}]$. However, there is no guarantee that a violation will not occur after k_{max};
(3) *Violations found* indicating that the requirement does not hold on the model; and
(4) *Inconclusive* indicating that QVtrace can not check the validity of the requirement due to scalability and incompatibility issues.

We use "✓", "≡", "✗", "?" to indicate each of these verdicts.

CoCoSim [14]. It is an open source framework to verify Simulink® models. It supports discrete-time Simulink® constructs and compiles them into LUSTRE [15], a data-flow language for programming systems that interact with their environments. Currently, it supports around 100 Simulink® blocks typically used by developers. The LUSTRE model is then verified using an SMT-based (e.g., Kind2 [16]) model checker, which verifies the validity of the specified properties. Considering the Kind2 results, CoCoSim produces five verdicts:

(1) *Valid*. If the model satisfies the property;
(2) *Invalid*. If the model violates the property;
(3) *Inconclusive/Undecided*. If the model checker can not confirm the validity or detect that the property is not satisfied;
(4) *Timeout*. If the time budget allocated for the analysis expires; and
(5) *Unsupported*. If the property can not be verified due to unsupported Simulink® blocks.

We use the symbols "✓", "✗", "?", "T", and "U" to respectively indicate the *Valid*, *Invalid*, *Inconclusive/Undecided*, *Timeout*, and *Unsupported* verdicts.

3.2 Benchmark

Our benchmark consists of the ten Lockheed Martin Cyber-Physical challenge models. These Simulink® models were originally proposed by Chris Elliott [24], considered by Nejati et al., [59], and later formalized, analyzed, and explained by Mavridou et al., [51]. These models are related to typical issues encountered in vehicle management and flight control systems. They were used to assess two tools from NASA: the Formal Requirements Elicitation Tool (FRET) [36] that enables to formalize requirements in the FRETISH semi-formal language, and CoCoSim [14]. They were also used to assess a tool (Theano [54, 73]) to analyze Simulink® Requirements Tables [47, 54], a tool that synthesizes CPS assumptions (EPIcuRus [34, 35]), and recently considered to assess the verification of Over-the-Air updates [10].

Table 1 describes each Lockheed Martin Cyber-Physical challenge model. Each model is associated with a set of requirements expressed in English. We consider the formalization of these requirements in QCT, the language for requirements specification supported by QVtrace. We made this choice because QVtrace is no longer available. Therefore, we can not perform any additional experiments with this tool. Our choice ensures the presence of results for all our baselines. Specifically, the replication package from Nejati et al., [59] provides four requirements for TSM (i.e., TSM-001, TSM-002, TSM-003a, TSM-004), 13 requirements for FSM (i.e., FSM-001, FSM-002, ..., FSM-013), five requirements for TUI (i.e., TUI-001, TUI-002, TUI-003, TUI-004a, TUI-004b), ten requirements for REG (i.e., REG-001, REG-002, ..., REG-010), two requirements for NLG

[2] QVtrace is no longer supported and available.

(i.e., NLG-006, NLG-007), two requirements for NN (i.e., NN-001, NN-002), three requirements for EB (i.e., EB-001, EB-002, and EB-004), no requirements for AP, two requirements for SWIM (i.e., SWIM-001, SWIM-002), seven requirements for EUL (i.e., EUL-001, EUL-002, EUL-003, EUL-004, EUL-006, EUL-007, EUL-008). Therefore, we consider nine of the Lockheed Martin challenge models (no requirement is provided for AP) and 48 (4+13+5+10+2+2+3+2+7) requirements in total.

3.3 Methodology

To assess the effectiveness and efficiency of Reactis®, we considered the nine models and 48 requirements from our benchmark. We formalized the 48 requirements using the support provided by Reactis® assertions. We run Reactis®. We recorded the verdict returned by Reactis® (i.e., "*Violated*", "*Not violated*", or "*Impossible to violate*") and the time required to produce these verdicts. We considered the default configuration parameters for Reactis® (30 tests, 150 steps per random test, with 150 steps in the targeted phase — see Figure 1b) and the assumptions on input signals specified by our baseline testing tool (HC).

To assess the effectiveness of Reactis®, we considered each requirement. We extracted the verdicts returned by HC ("*Violation found*" and "*No violation found*"), and QVTrace ("*No violation exists*", "*No violation exists for* $0 \leq k \leq k_{max}$.", "*Violations found*", and "*Inconclusive*") from the replication package provided by Nejati et al., [59], and the verdicts returned by CoCoSim (*Valid*, *Invalid*, *Inconclusive/Undecided*, *Timeout*, and *Unsupported*) from the documentation provided as supplementary material by Bourbouh et al., [14]. We compare the results of Reactis® and our baselines.

To assess the efficiency of Reactis®, we considered each requirement. We extracted the time required by QVTrace to produce its verdicts from the replication package. QVTrace runs several solvers (i.e., Z3 [67], SLDV [45], CVC [5]). We considered the time required by Z3. Z3 was chosen since it is an industry-strength awarded tool to check the satisfiability of logic-based formulae [29, 69]. Information on the time required by HC to generate the failure-revealing test cases was not distributed within the replication package. Since HC was configured to run 150 iterations to generate the failure-revealing test cases across these iterations, we multiplied 150 by the time required for the model to perform a single simulation since in each iteration HC executes the model once, and the time required by the search is negligible compare to the model execution time. We emphasize that this is an approximation: HC implements an early-stopping mechanism, in the case failure-revealing test cases were found before the 150th iteration. Finally, we extracted the time required by CoCoSim to analyze these requirements from a recent report [50]. Note that our objective in RQ2 is not to produce a detailed comparison of the efficiency of Reactis® and our baselines, but to have a ballpark estimation regarding the presence of (significant) differences in their efficiency.

Our experiments were executed on an Intel® Core™ i7-13700K CPU, 16 cores, 50 MB SmartCache, 32 GB RAM.

3.4 Effectiveness (RQ1)

Table 2 reports our results. The rows of the table report the results obtained for the different requirements. The columns refer to different tools. To analyze the effectiveness of Reactis®, we proceeded as follows:

- *Step 1.* We compare the results obtained by our two baseline model checkers (QVTrace and CoCoSim). Since model checkers can not only find counterexamples but also prove that a requirement is satisfied, they can be used to define the baseline verdicts expected by Reactis®.
- *Step 2.* We compare the results from HC and the baseline model checking verdicts. This step enables us to define a baseline expectation for a testing tool.
- *Step 3.* We compare the results from Reactis® and HC.

We detail each of these steps in the following.

Step 1 (QVTrace vs CoCoSim). We checked whether the verdicts returned by QVTrace and CoCoSim coincide. The two model checkers confirm the satisfaction or violation of 26 out of 48 (54.2%) requirements: 13 out of 48 (27.1%) requirements are satisfied, and 13 out of 48 (27.1%) requirements are violated. For 2 out of 48 (4.2%) requirements, the CoCoSim result was not available. For 6 out of 48 (12.5%) requirements, both QVTrace and CoCoSim could neither confirm the satisfaction nor find a failure-revealing test case. For 14 out of 48 (29.2%) requirements (marked with a yellow background), the verdicts returned by QVTrace and CoCoSim do not coincide. For TSM-001, QVTrace confirms that *No violation exists for* $0 \leq k \leq k_{max}$, while CoCoSim confirms that the requirement is satisfied. Despite the differences, these results are consistent with CoCoSim proving that the property does hold regardless of the bound selected for the analysis. By looking at the formalization of the requirements considered by QVTrace and CoCoSim, we realized that CoCoSim considers a broader requirement compared to QVTrace. For this reason, CoCoSim can provide a stronger guarantee for the property satisfaction. For REG-001, REG-002, REG-004, and REG-005, QVTrace returned a failure-revealing test case (REG-001, REG-002, REG-004) and confirms that *No violation exists for* $0 \leq k \leq k_{max}$ (REG-005), while CoCoSim returned an *Inconclusive/Undecided* (?) verdict for all these requirements. For REG-003, QVTrace and CoCoSim returned contradictory results: QVTrace found a failure-revealing test case and CoCoSim confirmed that the requirement is satisfied. We verified, and the two model checkers used the same formalization of the requirements. We conjecture that QVTrace and CoCoSim used a slightly different model for the regulator benchmark. Indeed, although the models used by QVTrace are publicly available in the replication package accompanying the referenced paper [59], we were unable to access the replication package associated with the CoCoSim technical report [50]. For NLG-007, QVTrace returned a failure-revealing test case, while CoCoSim returned an *Inconclusive/Undecided* verdict. By inspecting the models, we identified an explanation as a textual comment within the Simulink® model (likely from the Lockheed Martin engineers) confirming that this requirement is violated. For SWIM-002, we identified an error (likely a typo) in the result reported by Nejati et al., [59]. While the authors reported that the requirement is satisfied, the logs reported in the replication

Table 2: Verdicts (V.) produced by the different tools and the time (Time - seconds) required to produce them.

Req.	Reactis®		HC		QVtrace		CoCoSim	
	V.	Time	V.	Time	V.	Time	V.	Time
TSM-001	NA (✗)	(2.00)	✗	19.28	≡	20.3	✓	0.72
TSM-002	✗	2.00	?	19.28	✓	0.91	✓	0.72
TSM-003	?	2.00	?	19.28	✓	0.86	✓	0.72
TSM-004	✗	2.00	✗	19.28	✗	0.88	✗	0.43
FSM-001	✗	17.00	✗	1.37	✗	0.10	✗	0.25
FSM-002	?	2.00	?	1.37	✓	0.58	✓	0.25
FSM-003	✗	2.00	✗	1.37	✗	0.08	✗	0.19
FSM-004	✗	2.00	✗	1.37	✗	0.58	✗	1.19
FSM-005	?	2.00	?	1.37	✓	0.61	✓	0.25
FSM-006	?	2.00	?	1.37	✓	0.59	✓	0.25
FSM-007	✗	2.00	✗	1.37	✗	0.12	✗	0.13
FSM-008	✗	2.00	✗	1.37	✗	0.08	✗	0.16
FSM-009	?	2.00	?	1.37	✓	0.56	✓	0.25
FSM-010	?	2.00	?	1.37	✓	0.60	✓	0.13
FSM-011	✗	2.00	✗	1.37	✗	0.12	✗	0.11
FSM-012	?	2.00	?	1.37	✓	0.60	✓	0.13
FSM-013	?	2.00	?	1.37	✓	0.58	✓	0.13
TUI-001	⊥	11.00	?	16.15	✓	0.20	✓	0.08
TUI-002	⊥	2.00	?	16.15	✓	0.19	✓	0.06
TUI-003	✗	4.00	✗	16.15	✗	0.04	✗	0.09
TUI-004a	✗	12.00	✗	16.15	✗	1.49	NA	NA
TUI-004b	✗	4.00	✗	16.15	?	1121.02	NA	NA
REG-001	? (✗)	1.00 (59.0)	✗	14.22	✗	23.90	?	7200.00
REG-002	✗	1.00	✗	14.22	✗	17.20	?	7200.00
REG-003	? (✗)	1.00 (59.0)	✗	14.22	✗	28.20	✓	10.04
REG-004	? (✗)	1.00 (59.0)	✗	14.22	✗	21.30	?	7200
REG-005	? (✗)	1.00 (59.0)	✗	14.22	≡	46.9	?	7200
REG-006	✗	1.00	✗	14.22	✗	0.08	✗	6.00
REG-007	✗	1.00	✗	14.22	✗	0.07	✗	6.00
REG-008	✗	1.00	✗	14.22	✗	0.09	✗	6.00
REG-009	✗	1.00	✗	14.22	✗	0.06	✗	6.00
REG-010	? (✗)	1.00 (59.0)	✗	14.22	✗	0.09	✗	6.00
NLG-006	✗	6.00	✗	22.22	?	1071.10	?	7200
NLG-007	✗	3.00	✗	22.22	✗	0.12	?	7200
NN-001	?	5.00	?	22.65	?	1958.90	?	7200
NN-002	?	4.00	?	22.65	?	847.14	?	7200
EB-001	NA (✗)	(2.00)	✗	8.05	?	9475.40	?	7200
EB-002	NA (?)	(2.00)	?	8.05	?	4371.95	?	7200
EB-004	NA (✗)	(2.00)	✗	8.05	?	37.87	?	7200
SWIM-001	?	1.00	?	11.56	✓	0.22	✓	2.48
SWIM-002	✗	1.00	?	11.56	✓ (✗)	0.14	✗	1.93
EUL-001	✗	1.00	?	9.03	✓	0.06	✗	0.87
EUL-002	?	1.00	?	9.03	✓	0.07	✓	3.07
EUL-003	✗	1.00	?	9.03	✓	0.06	✗	0.14
EUL-004	?	1.00	?	9.03	✓	0.03	✗	0.18
EUL-006	✗	1.00	?	9.03	✓	0.07	✗	2.44
EUL-007	✗	1.00	?	9.03	✓	0.06	✗	0.19
EUL-008	✗	1.00	?	9.03	✓	0.05	✗	0.15

The verdicts associated with the different tools are defined as follows:
Reactis®: *Violated* (✗), *Not violated* (?), *Impossible to violate* (⊥), and *Not Available* (*NA*).
HC: *Violation found* (✗), *No violation found* (?).
QVtrace: *No violation exists* (✓), *No violation exists for* $0 \leq k \leq k_{max}$ (≡), *Violations found* (✗), *Inconclusive* (?).
CoCoSim: *Valid* (✓), *Invalid* (✗), *Inconclusive/Undecided* (?), *Timeout* ("T"), *Unsupported* ("U").
NA: Result Not Available.

package confirm that QVtrace found a violation. For this reason, Table 2 reports the correct result within parentheses. For EUL-001, EUL-003, EUL-004, EUL-006, EUL-007, and EUL-008, QVtrace confirmed the satisfaction of the requirements while CoCoSim found a counterexample. Unlike QVtrace, the approximation of the sine (sin) and cosine (cos) trigonometric functions used by CoCoSim leads to an incorrect verdict.

Based on these observations, there are no cases in which CoCoSim provides additional information compared to QVtrace; Therefore, we consider the verdicts of QVtrace as baseline results. According to the results reported by QVtrace, at least 20 out of 48 (41.7%) requirements are violated, and 19 out of 48 (39.6%) requirements are satisfied. For 9 out of 48 (18.7%) requirements, QVtrace returned an inconclusive verdict or a bounded satisfaction.

Therefore, a testing tool can identify requirements violations for at least 20 and at most 29 requirements.

Step 2 (HC vs QVtrace). We assessed whether the verdicts returned by HC and QVtrace coincide. For 52.1% (25 out of 48) of the requirements, HC returned a failure-revealing test case. For 19 of these 25 requirements, HC returned a failure-revealing test case, and this verdict is confirmed by the result returned by QVtrace. For the remaining 6 requirements (marked with a green background), HC returned a failure-revealing test case, while QVtrace could not find a violation and returned an "Inconclusive" (?) verdict. For TUI-004b, NLG-006, EB-001, and EB-004, HC could find a requirement violation, while QVtrace was inconclusive. For TSM-001, and REG-005 HC could find a requirement violation, while QVtrace returned that *No violation exists for* $0 \leq k \leq k_{max}$ (TSM-001 and REG-005). For 1 out of 48 (2.1%) requirements (i.e., SWIM-002), HC returned no failure-revealing test case while QVtrace returned a counterexample. For the remaining 22 out of 48 (45.8%) requirements, HC returned no failure-revealing test case, and QVtrace either confirmed that the requirement is satisfied (19 reqs.), or it returned an inconclusive verdict (3 reqs.).

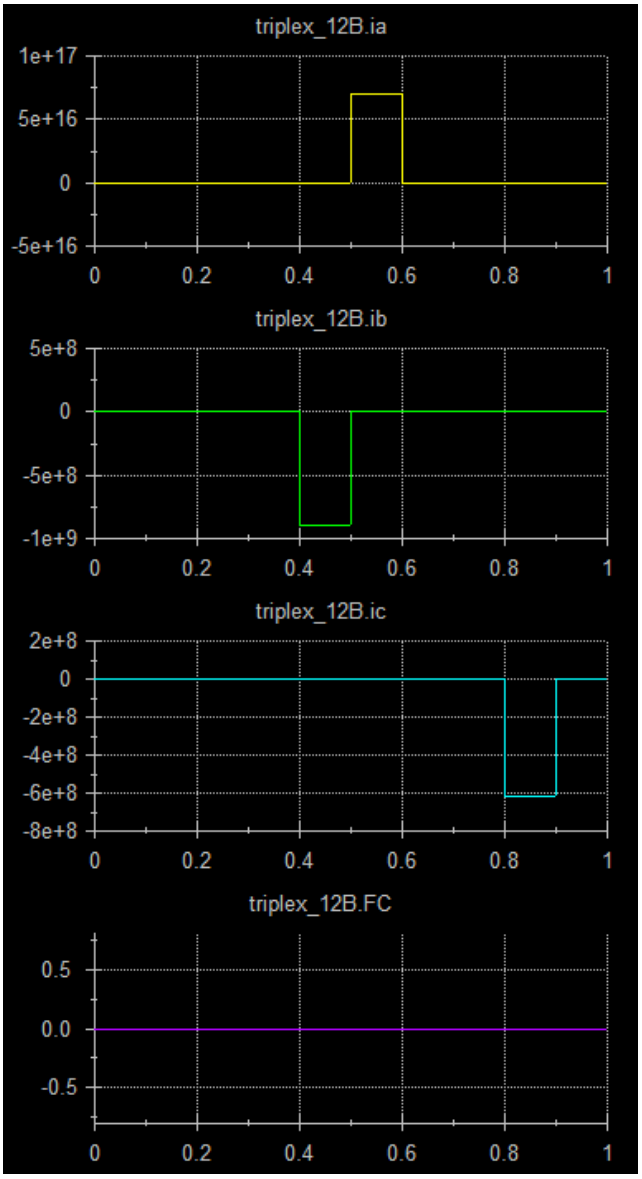
Therefore, our baseline testing tool returned a failure-revealing test case for 25 requirements (it identified a violation for TSM-001, but not detect on SWIM-002).

Step 3 (Reactis® vs HC). We compare the verdicts returned by Reactis® and HC.

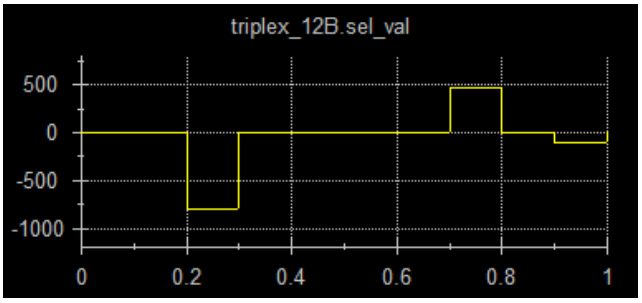
For 62.5% (30 out of 48) of the requirements, the verdicts (*Violated* - ✗, *Not violated* - ?) from Reactis® matched the one returned by HC. For 35.4% (17 out of 48) of the requirements, both Reactis® and HC returned a failure-revealing test cases. For 27.1% (13 out of 48) of the requirements, neither Reactis® nor HC returned a failure-revealing test case. Therefore, for these requirements, Reactis® was as effective as HC. We discuss the remaining 37.5% (18 out of 48) of the requirements (identified with a blue background in Table 2) below.

For 4.2% (2 out of 48) of the requirements (TSM-002, SWIM-002), Reactis® could return a failure-revealing test case that HC could not identify. For TSM-002, we analyzed the failure-revealing test cases returned by Reactis®. We report input and output signals for such a failure-revealing test case in Figure 2, with the failure occurring at $t = 1$. The test confirms that when *FC* is set to 0, in the presence of a value for *ib* ($-5 \cdot 10^{-324}$) comprised between *ia* (0.0) and *ic* (−1.0), as in Figure 2a, the TSM model instead of returning the value of *ib*, returns the value of *ia*, as shown in Figure 2b. This result is surprising considering that HC could not return any failure-revealing test cases for this example, and both QVTrace and CoCoSim confirmed the model correctness. Unfortunately, we could not replicate this experiment for HC since the authors uploaded all their scripts except for this model, which was likely forgotten. We could also not replicate the experiments for QVTrace because it is no longer supported. However, we managed to install CoCoSim and replicate their experiment. We found that the results reported for CoCoSim in the original work [50] were correct and that the model checker could verify the TSM-002 requirement. However, we exported the test case generated by Reactis® and executed it in Simulink®, confirming that it was indeed failure-revealing. We speculate that this result is due to an assumption made by the CoCoSim model checker (e.g., an assumption on the type of the solver or input value ranges) that is not verified when the model is executed. For SWIM-002, while HC could not identify a failure-revealing test case, QVTrace and CoCoSim confirm that the requirement is violated. We emphasize that the same input ranges were considered for all tools in the analysis of TSM-002. Thus, although the counterexample generated by Reactis includes extreme numerical values that may not be physically plausible in practice, these values are admissible under the formal assumptions encoded in the benchmark. For this reason, the counterexample is meaningful from a verification and testing perspective: it exposes a violation that exists within the specified input domain, even if further domain-specific assumptions would be required to determine whether the scenario can occur in real operation. Under the same formal input ranges, the violation was in principle potentially observable by all tools.

For 2.1% (1 out of 48) of the requirements (TSM-001), we could not express the requirements in Reactis® with c-like assertions. Indeed, the requirement is an implication forcing a condition to hold depending on the satisfaction of a pre-condition. However, the precondition is a conjunction of three predicates. Each predicate



(a) Inputs (*ia*, *ib*, and *ic*) and flag *FC* for the TSM model.



(b) Output (*sel_val*) of the TSM model.

Figure 2: Violation for the requirement TSM-002 with Reactis®.

specifies a condition that should hold for the entire simulation. Reactis® does not enable engineers to add temporal conditions on specific predicates. However, discussions with Reactis® engineers revealed that assertions can be modeled using a diagram-based approach: an external Simulink® subsystem can be created for each assertion and connected to the system under test as a diagram assertion. We adopted such approach and we could generate a failure revealing test case (see the results in Table 2 between brackets). This result is compliant with what obtained with HC and it is remarkable, as the requirement was considered valid by CoCoSim and valid up to a bound by QVtrace.

For 10.4% (5 out of 48) of the requirements (REG-001, REG-003, REG-004, REG-005, REG-010), unlike HC, Reactis® did not return any failure-revealing test case. Requirements REG-001, REG-003, REG-004, REG-005 specify that a specific measure (e.g., the acceleration) should not exceed a desired threshold (e.g., 50) for more than 100 seconds. We inspected the results of our experiments, the models, and the configuration we used to run our experiments. Our findings are as follows. First, HC considered a slightly different requirement than QVtrace and CoCoSim, requiring the *absolute value* of the measure to not exceed the desired threshold (e.g., 50) for more than 100 seconds. Second, the input ranges specified by the scripts from the replication package of HC do not match the ones considered by QVtrace and CoCoSim and do not enable HC to return a failure-revealing test case. We conjecture that these values were used by the authors of HC to perform some sanity check on their procedure after obtaining the results from Table 2. Therefore, we re-run our experiments by considering the input ranges from QVtrace and CoCoSim. Considering these values, Reactis® could effectively return a failure-revealing test case. Unlike the other requirements, REG-010 constraints change in angular roll acceleration to be lower than a specific threshold. Reactis® did not find a failure-revealing test case within 30 tests with 150 steps per random test, and 150 steps in the targeted phase. However, increasing the number of tests to 150, with 1000 steps per random test and 1000 steps in the targeted phase, enabled it to find the violation.

For 4.2% (2 out of 48) of the requirements (TUI-001, TUI-002), Reactis® could prove that the requirements are impossible to violate. Indeed, based on the bounds for the input signals, Reactis® can prove that TUI-001 and TUI-002 are tautologies. This verdict is consistent with the results returned by HC, QVtrace, and CoCoSim.

For 6.2% (3 out of 48) of the requirements (EB-001, EB-002, and EB-004), Reactis® returned the "*Matrix inversion not supported*" error. Therefore, to verify the EB model, we initially replaced the Matrix inversion Simulink® block with a set of multiplication/sum/-division blocks. After implementing this change, we observed that simulating EB with Reactis® required a significant amount of time due to the large input space, which consists of matrices. We conjectured that when Reactis® uses SMT solvers to generate test cases, each element of a matrix is encoded as a separate SMT variable, thereby significantly enlarging the input space while looking for unreachable targets (i.e., assertions which can never be covered regardless of input values). This prevented Reactis® from reaching a verdict. Our conjecture was confirmed by Reactis® engineers, who are working to fix this issue and suggested a way to make the EB model compatible. In particular, we modified the Simulink® matrix inversion block with a MATLAB function block implementing the matrix inversion operation. With this modification, we could generate test suites for the three requirements (the results are reported in Table 2 between brackets). The obtained results are consistent with HC and, interestingly, Reactis® was able to find failure-revealing test cases for two requirements (EB-001 and EB-004) for which both QVtrace and CoCoSim were inconclusive.

For 10.5% (5 out of 48) of the requirements (EUL-001, EUL-003, EUL-006, EUL-007, and EUL-008), unlike HC, Reactis® found a failure-revealing test case. This outcome differs from the results obtained with QVtrace, but is consistent with those produced by CoCoSim. To further investigate this discrepancy, we manually inspected the EUL benchmark and reformulated the affected requirements by replacing strict equality checks with comparisons that include a small tolerance. Under this modified specification, Reactis® was no longer able to generate failure-revealing test cases for any of the five requirements. Based on this observation, we conjecture that, similarly to QVtrace, Reactis® may internally use an approximate implementation of the sine (sin) and cosine (cos) functions, which could explain the difference in the results.

Therefore, Reactis® reported a failure-revealing test case for 100% (25 out of 25) of the cases in which other testing tools returned a failure-revealing test case. Unlike other testing tools, Reactis® could return a failure-revealing test case for SWIM-002 and could return a failure-revealing test case for a requirement (TSM-002) that was considered correct.

RQ1 - Effectiveness

Reactis® returned a reported-revealing test case for 100% (25 out of 25) of the cases in which other testing tools returned a failure-revealing test case. Reactis® could return a failure-revealing test case for a requirement that was considered correct.

3.5 Efficiency (RQ2)

The boxplots from Figure 3 report the time required by Reactis®, HC, QVtrace, and CoCoSim to analyze the requirements from our benchmark. Diamonds depict the average. We remark that our objective is not to perform a detailed comparison of the efficiency of Reactis® and our baseline tools (as we could not repeat the experiments for QVtrace), but rather to get a ballpark estimation regarding their efficiency. Indeed, since for the three baselines (HC, QVtrace, and CoCoSim) we did not replicate their experiments, their times are obtained with a different hardware setup.

Reactis®, HC, QVtrace, and CoCoSim required on average 7.92 *s* (*min*=1.00 *s*, *max*=59.00 *s*, *StdDev*=16.90 *s*, *Mdn*=2.00 *s*), 10.79 *s* (*min*= 1.37 *s*, *max*=22.65 *s*, *StdDev*=6.98 *s*, *Mdn*=11.56 *s*), 396.92 *s* (*min*= 0.03 *s*, *max*=9475.40 *s*, *StdDev*=1517.98 *s*, *Mdn*=0.57 *s*), and 1722.99 *s* (*min*= 0.06 *s*, *max*=7200.00 *s*, *StdDev*=3104.40 *s*, *Mdn*=0.79 *s*) to verify our requirements.[3] Therefore, Reactis® is on average 4911.62% and 21654.92% faster than QVtrace and CoCoSim, and 36.24% faster than HC. These results confirm that the efficiency of Reactis® is comparable to that of existing testing frameworks from the research literature (HC). As expected, it is much more efficient than

[3]For Reactis®, the average was computed by also considering the additional experiments reported within parentheses.

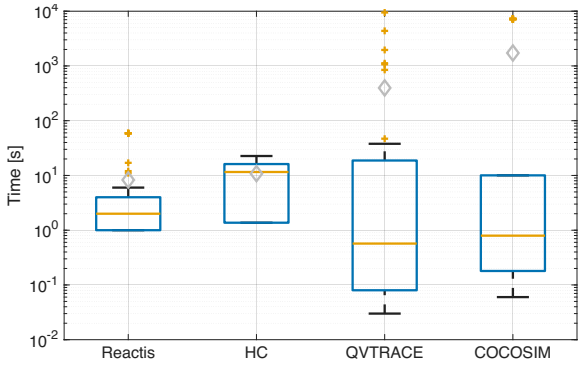


Figure 3: Time Required by Reactis®, HC, QVtrace, and CoCoSim to analyze the requirements from our benchmark.

model-checking tools, since these tools can also prove the satisfaction of requirements (which requires additional computational time).

RQ2 - Efficiency

The efficiency of Reactis® is comparable to that of existing testing tools from the research literature.

4 Discussion and Threats to Validity

We first discuss our results and present the lessons we learned. Then, we present threats to validity.

Lesson L1. The results from RQ1 demonstrate that Reactis® could find failure-revealing test cases that other testing tools could not find. Specifically, for TSM-002, none of the existing alternatives (HC, QVtrace, and CoCoSim) could detect the failure-revealing test case. Despite it being a single case, this result demonstrates the benefit of the usage of Reactis®. Reactis® could find a failure-revealing test case for TSM-001 that was considered valid by CoCoSim and up to a bound by QVtrace. In practical applications, failures can have catastrophic consequences and also imply significant economic losses. Within these contexts, being able to identify failures (with the support of different tools) is pivotal. Even being able to identify one new failure justifies the use of an additional tool. We also remark that our evaluation is based on 48 requirements. We do expect to have many more of these cases in practice.

Lesson L2. Unlike academic testing tools (HC), Reactis® provides support to detect the presence of requirements that are impossible to violate. Our results demonstrate the benefits of this functionality: This feature could prove that the TUI model can not violate requirements (TUI-001, TUI-002).

Lesson L3. Unlike HC, Reactis® could not analyze requirements (EB-001, EB-002, and EB-004) since it does not support the "*Matrix inversion*" Simulink® block. Indeed, Reactis® does not support all the Simulink® blocks. Although in this specific case we could replace the "*Matrix inversion*" with a set of equivalent Simulink® blocks, this may not be generally possible. Therefore, we learned that the capability of Reactis® to analyze the model depends on its blocks.

Lesson L4. The results from RQ2 demonstrate that for a given verification objective (testing or model-checking), the performance of existing solutions is comparable. Although for our benchmark models Reactis® performed better than HC, we did not notice significant differences in the performance of Reactis® and HC: Both can provide results (i.e., failure-revealing test cases) within minutes. Similarly, the performance of QVtrace and CoCoSim from their publications is comparable. This result confirms that the performance of technologically similar solutions is comparable.

Lesson L5. Our results enabled us to understand the synergy between academic and industrial solutions. First, despite HC being a preliminary implementation of a research idea, it could analyze all the requirements and provide insights that we found useful for the analysis of the results from Reactis®. Therefore, there might be some concepts, or additional support for Simulink® blocks (e.g., the matrix inversion for the EB model) that could be integrated within Reactis® to support additional requirements. Second, for the other requirements, the results for Reactis® are generally as good as (in some cases even better) those from the research literature. These results demonstrate that existing industrial testing tools for Simulink® models are up to speed with the research progress (and in some cases even provide better results). Our results demonstrate that industrial and academic communities should join forces, and the interchange of ideas can help obtain results with practical impact.

Lesson L6. Reactis® also offers graphical support, debugging facilities, and many other functionalities HC does not provide. Indeed, HC is a preliminary implementation of a research idea. We exploited and used the Reactis® functionalities during our experimentation, and we learned their usefulness. We understood that these functionalities may be pivotal for industrial practitioners who may also exploit dedicated client support channels that academic solutions (like HC) do not provide. Therefore, Reactis® is a mature industrial tool.

Lesson L7. Comparing the results of four different tools was complicated. First, despite the authors generally organizing good replication packages, we found a few missing files and configuration setups that were not extensively described. Second, QVtrace is no longer supported. Therefore, we could not replicate any of the experiments. Third, CoCoSim's last commit was nine years ago. Over the last nine years, many of the tools used by CoCoSim have evolved. We had to contact the developers of CoCoSim (who were responsive and helpful) to help us install this tool for analyzing the requirement SWIM-002. From this experience, we learned the importance of replicating packages and properly documenting the experiments and their results. Our study benefited from extensive reports and documentation from these packages.

Lesson L8. The models and requirements considered for testing strongly affect the results of the verification activities. The results for requirements TSM-001, REG-001, REG-002, REG-004, REG-005, EUL-001, EUL-003, EUL-006, EUL-007, and EUL-008 are affected by slight differences in the models or their requirements, affecting the testing and model-checking verdicts. Therefore, we learned the importance of appropriately using models and requirements versioning to support experiment replication.

Lesson L9. We learned that the input ranges used for generating the test cases significantly influence the verification and validation verdicts. For REG-001, REG-003, REG-004, and REG-005, the input ranges specified by the scripts from the replication package of HC do not match the ones considered by QVtrace and CoCoSim and do not enable HC to return a failure-revealing test case. This finding

is confirmed by recent works [34, 35] that advocate the need for approaches that automatically synthesize assumptions, i.e., conditions for which a property holds. Assumptions are also considered a primary element to perform the safety analysis of over-the-air updates for CPS [10]. We learned that a precise documentation of software assumptions is needed for a detailed and precise comparison of testing solutions for Simulink® models.

Lesson L10. In our study, we compared two model testing tools (Reactis® and HC) and two model checking tools (QVtrace and CoCoSim). Our results show how to use model-testing tools as an instrument to test model-checkers, and viceversa. They also show how to use the model-testing tools as a baseline for assessing other model testers. Specifically, comparing the verdicts returned by different solutions provides an instrument to analyze the soundness of existing solutions. Within this setting, we remark that being Reactis® an industrial tool, it offers much more reliable results compared to academic testing tools and solutions. Indeed, Reactis® has a more stable implementation, it continuously receive support and updates, it is ISO 26262 certified, and it has the *Interactive simulation* mode to analyze the test cases and the source of the requirement violation. The sum of all these factors makes it more reliable than HC, which is no longer maintained and provides only a 'prototype'-level implementation.

Lesson L11. The effort for performing this study was significant. We had to analyze documents and publications related to three different baselines, reconcile differences in the model and requirements considered by different solutions. In many cases, we had to analyze the replication packages, reverse engineer scripts, and results contained within the logs (e.g., SWIM-002) to perform our comparison. We learned the value of the activity performed by international competitions (e.g., the ARCH Competition [25–27, 40, 41, 53, 53], CT-Competition [12, 13, 43], or Test-Comp [6–9]) that provide an environment and facilities to compare existing testing solutions. These competitions define a standard way of configuring testing tools and provide results about the comparison of existing testing solutions.

Lesson L12. Reactis® could natively analyze 44 out of 48 requirements, i.e., the results of four requirements are tagged as not available ("NA") in Table 2. For three requirements (EB-001, EB-002, and EB-003), the problem was a model construct that was not supported. For only one requirement TSM-001, the problem was a limitation of the expressiveness of the Reactis® specification language. However, in this case, through interaction with Reactis® engineers, we discovered that assertions can be expressed using a diagram-based approach, which could support the specification of TSM-001. For the other benchmark requirements, expressing them using the requirements specification language from Reactis® was relatively straightforward.

Threats to Validity. The selection of the models of our benchmark could threaten the *external validity* of our results. Considering other benchmark models could lead to different results. However, the facts that the Lockheed Martin Cyber-Physical challenge models contain models from different domains and that they were largely used within the research literature to assess different solutions (e.g., [10, 34–36, 51, 54, 59, 73]) mitigate this threat. Indeed, authors of these works would have likely documented the limitations of the generalizability of the results obtained using models. The comparison with baseline tools may be affected by differences in requirement formalisms and execution environments: although we carefully reused the available formalizations and results from prior studies, different tools encode requirements using different specification languages and were run on different hardware.

The values selected for the Reactis® configuration parameters (30 tests, 150 steps for both random test and the targeted phase) could threaten the *internal validity* of our results. We mitigated this threat by considering the default values for these parameters. Since Reactis® includes pseudo-random components during test generation, different runs may produce slightly different tests and execution times. This could threaten the *internal validity* of our results. However, our goal was not to perform a statistically rigorous efficiency benchmark, but rather to obtain a ballpark comparison with existing tools, which were also taken from prior studies and used different hardware setups. Therefore, although multiple runs would provide a more precise estimate of runtime variability, we do not expect them to affect the main conclusion of RQ2.

Notice that there are no ground truth results for the Lockheed Martin Cyber-Physical challenge models. In the absence of these results, we compared the results from Reactis® and those from our baseline techniques. Therefore, although unlikely, it is possible that all tools (while agreeing on the verdict) could be reporting an incorrect result. A relevant field to research concerns evaluating the reliability of the verdicts.

The failure-revealing test cases produced by Reactis® that we manually inspected (using the Reactis® interactive simulation engine) lead to a violation of the requirement: We did not experience any false alarms. However, the behavior of the same test case may differ when executed by a different simulation engine or when the same simulation engine is executed with different configuration options.

5 Related Work

Many search-based testing solutions have been proposed in the research literature. The interested reader can consult existing surveys on the topic (e.g., [1, 37, 52, 61]). Many of them target Simulink® models. For example, tools such as ARIsTEO [55], Athena [31], Breach [23], FalCAuN [71], falsify [72], FalStar [28], ForeSee [75], FReaK [3], Moonlight [60], NNFal [42], S-TaLiRo [2], Ψ-TaLiRo [70], and zlscheck [77] are testing solutions for Simulink® models that consider requirements expressed as logic-based formulae. These tools participated in the ARCH competition [25–27, 40, 41, 53]. The interested reader can refer to the report from this competition for information about their effectiveness. HECATE [32, 33] is a testing solution that supports requirements expressed as Test Assessment or Requirements Tables blocks. These blocks are already included within Simulink® and do not require engineers to learn new (logic-based) languages. A recent study [39] proposed a formal framework that enables the automated generation of tests from requirements written in a structured natural language. While this work is related to those considering requirements expressed as logic-based formulae and was applied to the ten Lockheed Martin challenge models, the authors did not perform a comparison with these frameworks and did not relate their results to those obtained

by HC, QVtrace, and CoCoSim. In our work, we compared the effectiveness of Reactis® with three baseline solutions. These tools have been selected because they are standard frameworks for verifying Simulink® models and were previously considered by industrial experience reports [50, 59].

Many reseach works considered the Lockheed Martin challenge models [10, 24, 34–36, 51, 54, 56, 59, 73]. However, they do not focus on the testing problem.

6 Conclusion

We report on our experience in applying Reactis®. Reactis® is an industrial tool for testing Simulink® models. We used this tool to test the ten Cyber-Physical Challenges Simulink® models from Lockheed Martin. We considered 48 requirements from these challenge models. We assessed the effectiveness and efficiency of Reactis® in identifying failure-revealing test cases by comparing it with an academic testing tool (HC) and two model checking tools (QVtrace and CoCoSim).

Our results show that Reactis® reported a failure-revealing test case for 100% (25 out of 25) of the requirements that our baseline testing tool could violate. Additionally, Reactis® could return a failure-revealing test case for one requirement that was considered correct (i.e., the model checkers QVtrace and CoCoSim proved it correct). Our results also demonstrate that the efficiency of Reactis® is comparable to that of our baselines. We presented twelve lessons we learned from our experience and discussed threats to the validity of our results.

Acknowledgment

We thank Dr. Pierre-Loïc Garoche and Dr. Andreas Katis for their guidance and instructions in the installation of CoCoSim, and Michael Marzella for the help in inspecting the formalization of the requirements in CoCoSim. Finally, we thank Dr. Steve Sims from Reactive Systems, Inc. for the insightful feedback and the productive discussion on our results.

Data Availability Statement

A complete replication package is available at [11].

References

[1] Saswat Anand, Edmund K. Burke, Tsong Yueh Chen, John Clark, Myra B. Cohen, Wolfgang Grieskamp, Mark Harman, Mary Jean Harrold, Phil McMinn, Antonia Bertolino, J. Jenny Li, and Hong Zhu. 2013. An orchestrated survey of methodologies for automated software test case generation. *Journal of Systems and Software* 86, 8 (2013), 1978–2001. doi:10.1016/j.jss.2013.02.061

[2] Yashwanth Annpureddy, Che Liu, Georgios Fainekos, and Sriram Sankaranarayanan. 2011. S-taliro: a tool for temporal logic falsification for hybrid systems. In *Proceedings of the 17th International Conference on Tools and Algorithms for the Construction and Analysis of Systems: Part of the Joint European Conferences on Theory and Practice of Software* (Saarbrücken, Germany) *(TACAS'11/ETAPS'11)*. Springer-Verlag, Berlin, Heidelberg, 254–257. doi:10.1007/978-3-642-19835-9_21

[3] Stanley Bak, Sergiy Bogomolov, Abdelrahman Hekal, Niklas Kochdumper, Ethan Lew, Andrew Mata, and Amir Rahmati. 2024. Falsification using Reachability of Surrogate Koopman Models. In *Proceedings of the 27th ACM International Conference on Hybrid Systems: Computation and Control* (Hong Kong SAR, China) *(HSCC '24)*. Association for Computing Machinery, New York, NY, USA, Article 27, 13 pages. doi:10.1145/3641513.3650141

[4] Luciano Baresi, Marcio Delamaro, and Paulo Nardi. 2017. Test oracles for simulink-like models. *Automated Software Engg.* 24, 2 (June 2017), 369–391. doi:10.1007/s10515-016-0207-9

[5] Clark W. Barrett, Christopher L. Conway, Morgan Deters, Liana Hadarean, Dejan Jovanovic, Tim King, Andrew Reynolds, and Cesare Tinelli. 2011. CVC4. In *Computer Aided Verification - 23rd International Conference, CAV 2011, Snowbird, UT, USA, July 14-20, 2011. Proceedings (Lecture Notes in Computer Science, Vol. 6806)*, Ganesh Gopalakrishnan and Shaz Qadeer (Eds.). Springer, Berlin, Heidelberg, 171–177. doi:10.1007/978-3-642-22110-1_14

[6] Dirk Beyer. 2019. *International Competition on Software Testing (Test-Comp)*. Springer International Publishing, Cham, 167–175. doi:10.1007/978-3-030-17502-3_11

[7] Dirk Beyer. 2020. *Second Competition on Software Testing: Test-Comp 2020*. Springer-Verlag, Berlin, Heidelberg, 505–519. doi:10.1007/978-3-030-45234-6_25

[8] Dirk Beyer. 2021. *Status Report on Software Testing: Test-Comp 2021*. Springer International Publishing, Cham, 341–357. doi:10.1007/978-3-030-71500-7_17

[9] Dirk Beyer. 2022. *Advances in Automatic Software Testing: Test-Comp 2022*. Springer-Verlag, Berlin, Heidelberg, 321–335. doi:10.1007/978-3-030-99429-7_18

[10] Nunzio Marco Bisceglia, Aurora Francesca Zanenga, Mehrnoosh Askarpour, Sahar Kokaly, S. Ramesh, Marsha Chechik, and Claudio Menghi. 2026. Safety Analysis of Over-the-Air Updates for CPS: A Contract-Driven Approach. *IEEE Transactions on Software Engineering* 52, 3 (2026), 889–907. doi:10.1109/TSE.2026.3654819

[11] Andrea Bombarda, Nunzio Marco Bisceglia, Marco Saini, Federico Formica, Angelo Gargantini, and Claudio Menghi. 2026. *Replication Package*. https://doi.org/10.5281/zenodo.19604809

[12] Andrea Bombarda, Edoardo Crippa, and Angelo Gargantini. 2021. An environment for benchmarking combinatorial test suite generators. In *2021 IEEE International Conference on Software Testing, Verification and Validation Workshops (ICSTW)* (Porto de Galinhas, Brazil). IEEE, 48–56. doi:10.1109/ICSTW52544.2021.00021

[13] Andrea Bombarda and Angelo Gargantini. 2024. Design, implementation, and validation of a benchmark generator for combinatorial interaction testing tools. *Journal of Systems and Software* 209 (2024), 111920. doi:10.1016/j.jss.2023.111920

[14] Guillaume Bourbouh, Hamza Band Brat and Pierre-Loic Garoche. 2020. CoCoSim: an automated analysis framework for Simulink/Stateflow. In *Model Based Space Systems and Software Engineering-European Space Agency Workshop (MBSE)*. European Space Agency (ESA), Virtual Event. https://indico.esa.int/event/329/

[15] P. Caspi, D. Pilaud, N. Halbwachs, and J. A. Plaice. 1987. LUSTRE: a declarative language for real-time programming. In *Proceedings of the 14th ACM SIGACT-SIGPLAN Symposium on Principles of Programming Languages* (Munich, West Germany) *(POPL '87)*. Association for Computing Machinery, New York, NY, USA, 178–188. doi:10.1145/41625.41641

[16] Adrien Champion, Alain Mebsout, Christoph Sticksel, and Cesare Tinelli. 2016. The Kind 2 Model Checker. In *Computer Aided Verification*, Swarat Chaudhuri and Azadeh Farzan (Eds.). Springer International Publishing, Cham, 510–517. doi:10.1007/978-3-319-41540-6_29

[17] Ying Chen and Brandon Jones. 2007. Software Testing Strategies for Model-Based Chassis Control Systems. In *SAE Technical Paper Series (ANNUAL, Vol. 1)*. SAE International. doi:10.4271/2007-01-0505

[18] Rance Cleaveland, David Hansel, Steve Sims, and Scott A. Smolka. 2025. Two Decades of Industrializing Formal Verification: The Reactis Story. In *Model Checking Software*, Thomas Neele and Anton Wijs (Eds.). Springer Nature Switzerland, Cham, 87–105. doi:10.1007/978-3-031-66149-5_5

[19] Rance Cleaveland, Steve T Sims, and David Hansel. 2010. System and method for automatic test-case generation for software. US Patent 7644398B2. https://patents.google.com/patent/US7644398B2/en

[20] Rance Cleaveland, Scott A. Smolka, and Steven T. Sims. 2008. An Instrumentation-Based Approach to Controller Model Validation. In *Model-Driven Development of Reliable Automotive Services*. Springer, 84–97. doi:10.1007/978-3-540-70930-5_6

[21] Leonardo de Moura and Nikolaj Bjørner. 2008. *Z3: An Efficient SMT Solver*. Springer Berlin Heidelberg, Berlin, Heidelberg, 337–340. doi:10.1007/978-3-540-78800-3_24

[22] Patricia Derler, Edward A. Lee, and Alberto Sangiovanni Vincentelli. 2012. Modeling Cyber–Physical Systems. *Proc. IEEE* 100, 1 (2012), 13–28. doi:10.1109/JPROC.2011.2160929

[23] Alexandre Donzé. 2010. Breach, A Toolbox for Verification and Parameter Synthesis of Hybrid Systems. In *International Conference on Computer Aided Verification (CAV)*. Springer, 167–170. doi:10.1007/978-3-642-14295-6_17

[24] Chris Elliott. 2015. On example models and challenges ahead for the evaluation of complex cyber-physical systems with state of the art formal methods V&V, Lockheed Martin Skunk Works. In *Safe & Secure Systems and Software Symposium (S5)*. 9–11. http://www.mys5.org/index.php/proceedings/agenda-2015

[25] Gidon Ernst, Paolo Arcaini, Ismail Bennani, Aniruddh Chandratre, Alexandre Donzé, Georgios Fainekos, Goran Frehse, Khouloud Gaaloul, Jun Inoue, Tanmay Khandait, Logan Mathesen, Claudio Menghi, Giulia Pedrielli, Marc Pouzet, Masaki Waga, Shakiba Yaghoubi, Yoriyuki Yamagata, and Zhenya Zhang. 2021. ARCH-COMP Category Report: Falsification with Validation of Results. In *Workshop on Applied Verification of Continuous and Hybrid Systems*, Vol. 80. EasyChair, 133–152. doi:10.29007/xwl1

[26] Gidon Ernst, Paolo Arcaini, Ismail Bennani, Alexandre Donze, Georgios Fainekos, Goran Frehse, Logan Mathesen, Claudio Menghi, Giulia Pedrielli, Marc Pouzet, Shakiba Yaghoubi, Yoriyuki Yamagata, and Zhenya Zhang. 2020. ARCH-COMP

2020 Category Report: Falsification. In *ARCH20. 7th International Workshop on Applied Verification of Continuous and Hybrid Systems (ARCH20) (EPiC Series in Computing, Vol. 74)*, Goran Frehse and Matthias Althoff (Eds.). EasyChair, 140–152. doi:10.29007/trr1

[27] Gidon Ernst, Paolo Arcaini, Georgios Fainekos, Federico Formica, Jun Inoue, Tanmay Khandait, Mohammad Mahdi Mahboob, Claudio Menghi, Giulia Pedrielli, Masaki Waga, et al. 2022. Arch-comp 2022 category report: Falsification with ubounded resources. In *International Workshop on Applied Verification of Continuous and Hybrid Systems (ARCH). EPiC Series in Computing*, Vol. 90. EasyChair, 204–221. doi:10.29007/fhnk

[28] Gidon Ernst, Sean Sedwards, Zhenya Zhang, and Ichiro Hasuo. 2021. Falsification of Hybrid Systems Using Adaptive Probabilistic Search. *ACM Trans. Model. Comput. Simul.* 31, 3, Article 18 (2021), 22 pages. doi:10.1145/3459605

[29] ETAPS. 2018. The 2018 Test of Time Award. https://www.etaps.org/2018/test-of-time-award.html.

[30] Georgios E. Fainekos, Sriram Sankaranarayanan, Koichi Ueda, and Hakan Yazarel. 2012. Verification of Automotive Control Applications Using S-TaLiRo. In *American Control Conference (ACC)*. IEEE, 3567–3572. doi:10.1109/ACC.2012.6315384

[31] Federico Formica, Tony Fan, and Claudio Menghi. 2023. Search-based software testing driven by automatically generated and manually defined fitness functions. *ACM Transactions on Software Engineering and Methodology* 33, 2 (2023), 1–37. doi:10.1145/3624745

[32] Federico Formica, Tony Fan, Akshay Rajhans, Vera Pantelic, Mark Lawford, and Claudio Menghi. 2024. Simulation-Based Testing of Simulink Models With Test Sequence and Test Assessment Blocks. *IEEE Transactions on Software Engineering* 50, 2 (2024), 239–257. doi:10.1109/TSE.2023.3343753

[33] Federico Formica, Chris George, Shayda Rahmatyan, Vera Pantelic, Mark Lawford, Angelo Gargantini, and Claudio Menghi. 2025. Search-based Testing of Simulink Models with Requirements Tables. https://arxiv.org/abs/2501.05412

[34] Khouloud Gaaloul, Claudio Menghi, Shiva Nejati, Lionel C. Briand, and Yago Isasi Parache. 2022. Combining Genetic Programming and Model Checking to Generate Environment Assumptions. *IEEE Transactions on Software Engineering* 48, 9 (2022), 3664–3685. doi:10.1109/TSE.2021.3101818

[35] Khouloud Gaaloul, Claudio Menghi, Shiva Nejati, Lionel C. Briand, and David Wolfe. 2020. Mining assumptions for software components using machine learning. In *Proceedings of the 28th ACM Joint Meeting on European Software Engineering Conference and Symposium on the Foundations of Software Engineering* (Virtual Event, USA) *(ESEC/FSE 2020)*. Association for Computing Machinery, New York, NY, USA, 159–171. doi:10.1145/3368089.3409737

[36] Dimitra Giannakopoulou, Anastasia Mavridou, Julian Rhein, Thomas Pressburger, Johann Schumann, and Nija Shi. 2020. Formal requirements elicitation with FRET. In *International Working Conference on Requirements Engineering: Foundation for Software Quality (REFSQ)*. CEUR Workshop Proceedings. https://ceur-ws.org/Vol-2584/PT-paper4.pdf

[37] Mark Harman, S. Afshin Mansouri, and Yuanyuan Zhang. 2012. Search-based software engineering: Trends, techniques and applications. *Comput. Surveys* 45, 1, Article 11 (Dec. 2012), 61 pages. doi:10.1145/2379776.2379787

[38] Aditya Kanade, Rajeev Alur, Franjo Ivančić, S. Ramesh, Sriram Sankaranarayanan, and K. C. Shashidhar. 2009. Generating and Analyzing Symbolic Traces of Simulink/Stateflow Models. In *Computer Aided Verification*, Ahmed Bouajjani and Oded Maler (Eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 430–445. doi:10.1007/978-3-642-02658-4_33

[39] Andreas Katis, Anastasia Mavridou, and Tom Pressburger. 2025. A Streamlined, Formal Approach to Requirements-Based Testing. In *NASA Formal Methods*, Aaron Dutle, Laura Humphrey, and Laura Titolo (Eds.). Springer Nature Switzerland, Cham, 159–179. doi:10.1007/978-3-031-93706-4_10

[40] Tanmay Khandait, Federico Formica, Paolo Arcaini, Surdeep Chotaliya, Georgios Fainekos, Abdelrahman Hekal, Atanu Kundu, Ethan Lew, Michele Loreti, Claudio Menghi, et al. 2024. Arch-comp 2024 category report: Falsification. In *International Workshop on Applied Verification for Continuous and Hybrid Systems*. EasyChair, EasyChair, 122–144. doi:10.29007/hgfv

[41] Tanmay Khandait, Deyun Lyu, Paolo Arcaini, Georgios Fainekos, Federico Formica, Sauvik Gon, Abdelrahman Hekal, Atanu Kundu, Claudio Menghi, Giulia Pedrielli, et al. 2025. ARCH-COMP 2025 Category Report: Falsification. In *International Workshop on Applied Verification for Continuous and Hybrid Systems*, Vol. 108. EasyChair, 169–189. doi:10.29007/dgnn

[42] Atanu Kundu, Sauvik Gon, and Rajarshi Ray. 2026. Data-Driven Falsification of Cyber-Physical Systems. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems* 45, 4 (2026), 1989–2002. doi:10.1109/TCAD.2025.3608632

[43] Manuel Leithner, Andrea Bombarda, Michael Wagner, Angelo Gargantini, and Dimitris E. Simos. 2024. State of the CArt: evaluating covering array generators at scale. *Int. J. Softw. Tools Technol. Transf.* 26, 3 (May 2024), 301–326. doi:10.1007/s10009-024-00745-2

[44] Bo Liu, Yuan-rui Zhang, Xue-lian Cao, Yu Liu, Bin Gu, and Tie-xin Wang. 2020. A survey of model-driven techniques and tools for cyber-physical systems. *Frontiers of Information Technology & Electronic Engineering* 21, 11 (2020), 1567–1590. doi:10.1631/FITEE.2000311

[45] MathWorks. 2020. *Simulink Design Verifier*. Retrieved August 2020 from https://www.mathworks.com/products/simulink-design-verifier.html

[46] MathWorks. 2022. Simulink®. Release R2022b. https://www.mathworks.com/products/simulink.html

[47] MathWorks. 2023. *Use a Requirements Table Block to Create Formal Requirements*. https://www.mathworks.com/help/slrequirements/ug/use-requirements-table-block.html

[48] Reza Matinnejad, Shiva Nejati, Lionel Briand, Thomas Bruckmann, and Claude Poull. 2015. Search-based automated testing of continuous controllers: Framework, tool support, and case studies. *Information and Software Technology* 57 (2015), 705–722. doi:10.1016/J.INFSOF.2014.05.007

[49] Reza Matinnejad, Shiva Nejati, Lionel C Briand, and Thomas Bruckmann. 2016. Automated test suite generation for time-continuous Simulink models. In *International Conference on Software Engineering* (Austin, Texas). Association for Computing Machinery, New York, NY, USA, 595–606. doi:10.1145/2884781.2884797

[50] Anastasia Mavridou, Hamza Bourbouh, Pierre-Loïc Garoche, and Mohammad Hejase. 2019. *Evaluation of the FRET and CoCoSim tools on the ten Lockheed Martin cyber-physical challenge problems*. Technical Report. NASA. 84 pages.

[51] Anastasia Mavridou, Hamza Bourbouh, Dimitra Giannakopoulou, Thomas Pressburger, Mohammad Hejase, Pierre-Loic Garoche, and Johann Schumann. 2020. The Ten Lockheed Martin Cyber-Physical Challenges: Formalized, Analyzed, and Explained. In *2020 IEEE 28th International Requirements Engineering Conference (RE)*. IEEE, 300–310. doi:10.1109/RE48521.2020.00040

[52] Phil McMinn. 2011. Search-Based Software Testing: Past, Present and Future. In *IEEE International Conference on Software Testing, Verification and Validation Workshops*. IEEE, 153–163. doi:10.1109/ICSTW.2011.100

[53] Claudio Menghi, Paolo Arcaini, Walstan Baptista, Gidon Ernst, Georgios Fainekos, Federico Formica, Sauvik Gon, Tanmay Khandait, Atanu Kundu, Giulia Pedrielli, et al. 2023. ARCH-COMP 2023 category report: Falsification. In *International Workshop on Applied Verification of Continuous and Hybrid Systems*. EasyChair, EasyChair, 151–169. doi:10.29007/6nqs

[54] Claudio Menghi, Eugene Balai, Darren Valovcin, Christoph Sticksel, and Akshay Rajhans. 2025. Completeness and Consistency of Tabular Requirements: An SMT-Based Verification Approach. *IEEE Transactions on Software Engineering* 51, 2 (2025), 595–620. doi:10.1109/TSE.2025.3530820

[55] Claudio Menghi, Shiva Nejati, Lionel Briand, and Yago Isasi Parache. 2020. Approximation-refinement testing of compute-intensive cyber-physical models: an approach based on system identification. In *Proceedings of the ACM/IEEE 42nd International Conference on Software Engineering* (Seoul, South Korea) *(ICSE '20)*. Association for Computing Machinery, New York, NY, USA, 372–384. doi:10.1145/3377811.3380370

[56] Claudio Menghi, Shiva Nejati, Khouloud Gaaloul, and Lionel C. Briand. 2019. Generating automated and online test oracles for Simulink models with continuous and uncertain behaviors. In *Proceedings of the 2019 27th ACM Joint Meeting on European Software Engineering Conference and Symposium on the Foundations of Software Engineering* (Tallinn, Estonia) *(ESEC/FSE 2019)*. Association for Computing Machinery, New York, NY, USA, 27–38. doi:10.1145/3338906.3338920

[57] Nicholas Metropolis and Stanislaw Ulam. 1949. The Monte Carlo method. *J. Amer. Statist. Assoc.* 44, 247 (1949), 335–341. doi:10.2307/2280232

[58] Swarup Mohalik, Ambar A. Gadkari, Anand Yeolekar, K.C. Shashidhar, and S. Ramesh. 2014. Automatic test case generation from Simulink/Stateflow models using model checking. *Software Testing, Verification and Reliability* 24, 2 (2014), 155–180. doi:10.1002/stvr.1489

[59] Shiva Nejati, Khouloud Gaaloul, Claudio Menghi, Lionel C. Briand, Stephen Foster, and David Wolfe. 2019. Evaluating model testing and model checking for finding requirements violations in Simulink models. In *Proceedings of the 2019 27th ACM Joint Meeting on European Software Engineering Conference and Symposium on the Foundations of Software Engineering* (Tallinn, Estonia) *(ESEC/FSE 2019)*. Association for Computing Machinery, New York, NY, USA, 1015–1025. doi:10.1145/3338906.3340444

[60] Laura Nenzi, Ezio Bartocci, Luca Bortolussi, Simone Silvetti, and Michele Loreti. 2023. MoonLight: a lightweight tool for monitoring spatio-temporal properties. *International Journal on Software Tools for Technology Transfer* 25, 4 (2023), 503–517. doi:10.1007/s10009-023-00710-5

[61] Annibale Panichella, Fitsum Meshesha Kifetew, and Paolo Tonella. 2018. Automated Test Case Generation as a Many-Objective Optimisation Problem with Dynamic Selection of the Targets. *IEEE Transactions on Software Engineering* 44, 2 (2018), 122–158. doi:10.1109/TSE.2017.2663435

[62] Andr Platzer. 2018. *Logical Foundations of Cyber-Physical Systems* (first ed.). Vol. 662. Springer, Cham. doi:10.1007/978-3-319-63588-0

[63] QRACorp. 2020. *QRACorp*. Retrieved August 2020 from QVtrace.https://qracorp.com/

[64] QRACorp. 2020. *QVtrace*. Retrieved August 2020 from QVtrace.https://qracorp.com/qvtrace/

[65] Reactive Systems Inc. 2026. *Automated Testing & Verification for Simulink & C Code*. https://reactive-systems.com/

[66] Reactive Systems Inc. 2026. *How Assertions Impact Testing*. https://reactive-systems.com/doc/reactis-for-simulink/reactis-assertions.html#how-assertions-impact-testing

[67] Microsoft Research. 2020. *Z3*. Microsoft Research. https://github.com/Z3Prover/z3

[68] Manoranjan Satpathy, Anand Yeolekar, and S. Ramesh. 2008. Randomized directed testing (REDIRECT) for Simulink/Stateflow models. In *Proceedings of the 8th ACM International Conference on Embedded Software* (Atlanta, GA, USA) *(EMSOFT '08)*. Association for Computing Machinery, New York, NY, USA, 217–226. doi:10.1145/1450058.1450088

[69] ACM SIGPLAN. 2020. Programming Languages Software Award. http://www.sigplan.org/Awards/Software/.

[70] Quinn Thibeault, Jacob Anderson, Aniruddh Chandratre, Giulia Pedrielli, and Georgios Fainekos. 2021. PSY-TaLiRo: A Python Toolbox for Search-Based Test Generation for Cyber-Physical Systems. In *Formal Methods for Industrial Critical Systems*. Springer, 223–231. doi:10.1007/978-3-030-85248-1_15

[71] Masaki Waga. 2020. Falsification of cyber-physical systems with robustness-guided black-box checking. In *Proceedings of the 23rd International Conference on Hybrid Systems: Computation and Control* (Sydney, New South Wales, Australia) *(HSCC '20)*. Association for Computing Machinery, New York, NY, USA, Article 11, 13 pages. doi:10.1145/3365365.3382193

[72] Yoriyuki Yamagata, Shuang Liu, Takumi Akazaki, Yihai Duan, and Jianye Hao. 2021. Falsification of Cyber-Physical Systems Using Deep Reinforcement Learning. *IEEE Transactions on Software Engineering* 47, 12 (2021), 2823–2840. doi:10.1109/TSE.2020.2969178

[73] Aurora Francesca Zanenga, Nunzio Marco Bisceglia, Benedetta Ippoliti, Andrea Bombarda, Angelo Gargantini, Akshay Rajhans, and Claudio Menghi. 2025. Theano: A Tool for Verifying the Consistency and Completeness in Tabular Requirements. In *Proceedings of the 33rd ACM International Conference on the Foundations of Software Engineering* (Clarion Hotel Trondheim, Trondheim, Norway) *(FSE Companion '25)*. Association for Computing Machinery, New York, NY, USA, 1148–1152. doi:10.1145/3696630.3728599

[74] Jing Zeng, Laurence T. Yang, Man Lin, Huansheng Ning, and Jianhua Ma. 2020. A survey: Cyber-physical-social systems and their system-level design methodology. *Future Generation Computer Systems* 105 (2020), 1028–1042. doi:10.1016/j.future.2016.06.034

[75] Zhenya Zhang, Deyun Lyu, Paolo Arcaini, Lei Ma, Ichiro Hasuo, and Jianjun Zhao. 2021. Effective Hybrid System Falsification Using Monte Carlo Tree Search Guided by QB-Robustness. In *Computer Aided Verification*, Alexandra Silva and K. Rustan M. Leino (Eds.). Springer International Publishing, Cham, 595–618. doi:10.1007/978-3-030-81685-8_29

[76] Xi Zheng, Christine Julien, Miryung Kim, and Sarfraz Khurshid. 2015. Perceptions on the state of the art in verification and validation in cyber-physical systems. *IEEE Systems Journal* 11, 4 (2015), 2614–2627. doi:10.1109/JSYST.2015.2496293

[77] zlscheck 2024 [Online]. *zlscheck*. Retrieved August 2024 from https://github.com/ismailbennani/zlscheck

[78] Paolo Zuliani, André Platzer, and Edmund M Clarke. 2013. Bayesian statistical model checking with application to Stateflow/Simulink verification. *Formal Methods in System Design* 43, 2 (2013), 338–367. doi:10.1007/s10703-013-0195-3