

**2025 2nd IEEE/ACM  
International Conference on  
AI-powered Software  
(AIware 2025)**

**Seoul, South Korea  
19-20 November 2025**



**IEEE Catalog Number: CFP255BO-POD  
ISBN: 979-8-3315-8270-8**

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

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

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

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

IEEE Catalog Number:	CFP255BO-POD
ISBN (Print-On-Demand):	979-8-3315-8270-8
ISBN (Online):	979-8-3315-8269-2

**Additional Copies of This Publication Are Available From:**

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

CURRAN ASSOCIATES INC.  
**proceedings**  
.com

**2025 2nd IEEE/ACM  
International Conference on  
AI-powered Software  
(AIware)  
AIware 2025**

**Table of Contents**

Message from the Chairs .....	x
Organizing Committee .....	xii
Program Committee .....	xiii

**Main Papers**

CHASE: LLM Agents for Dissecting Malicious PyPI Packages .....	1
<i>Takaaki Toda (Waseda University, Japan) and Tatsuya Mori (Waseda University, Japan)</i>	
Understanding the Characteristics of LLM-Generated Property-Based Tests in Exploring Edge Cases .....	11
<i>Hidetake Tanaka (Nara Institute of Science and Technology, Japan), Haruto Tanaka (Nara Institute of Science and Technology, Japan), Kazumasa Shimari (Nara Institute of Science and Technology, Japan), and Kenichi Matsumoto (Nara Institute of Science and Technology, Japan)</i>	
Neuro-Symbolic Compliance: Integrating LLMs and SMT Solvers for Automated Financial Legal Analysis .....	19
<i>Yung-Shen Hsia (National ChengChi University, Taiwan), Fang Yu (National ChengChi University, Taiwan), and Jie-Hong Roland Jiang (National Taiwan University, Taiwan)</i>	
Understanding LLM-Driven Test Oracle Generation .....	29
<i>Adam Bodicoat (University of Auckland, New Zealand), Gunel Jahangirova (King's College London, United Kingdom), and Valerio Terragni (University of Auckland, New Zealand)</i>	

Turning Manual Tasks into Actions: Assessing the Effectiveness of Gemini-Generated Selenium Tests .....	40
<i>Myron David Lucena Campos Peixoto (Computing Institute - Federal University of Alagoas, Brazil), Balduino Fonseca (Computing Institute - Federal University of Alagoas, Brazil), Davy de Medeiros Baia (Computing Institute - Federal University of Alagoas, Brazil), Kevin Lira (North Carolina State University, USA), Márcio Ribeiro (Computing Institute - Federal University of Alagoas, Brazil), Wesley K. G. Assunção (North Carolina State University, USA), Nathalia Nascimento (Penn State University, USA), and Paulo Alencar (University of Waterloo, Canada)</i>	
How Does Quantization Impact Privacy Risk on LLMs for Code? .....	50
<i>Md Nazmul Haque (North Carolina State University, United States), Hua Yang (North Carolina State University, United States), Zhou Yang (University of Alberta, Canada), and Bowen Xu (North Carolina State University, United States)</i>	
Envisioning Future Interactive Web Development: Editing Webpage with Natural Language .....	61
<i>Truong Hai Dang (Singapore Management University, Singapore), Jingyu Xiao (The Chinese University of Hong Kong, Hong Kong), and Yintong Huo (Singapore Management University, Singapore)</i>	
Human to Document, AI to Code: Comparing GenAI for Notebook Competitions .....	67
<i>Tasha Setteuwong (Nara Institute of Science and Technology, Japan), Youmei Fan (Nara Institute of Science and Technology, Japan), Raula Gaikovina Kula (The University of Osaka, Japan), and Kenichi Matsumoto (Nara Institute of Science and Technology, Japan)</i>	
Examining the Usage of Generative AI Models in Student Learning Activities for Software Programming .....	76
<i>Rufeng Chen (McGill University, Canada), Shuaishuai Jiang (McGill University, Canada), Jiyun Shen (Carnegie Mellon University, USA), AJung Moon (McGill University, Canada), and Lili Wei (McGill University, Canada)</i>	
Beyond Code Explanations: A Ray of Hope for Cross-Language Vulnerability Repair .....	87
<i>Kevin Lira (North Carolina State University, USA), Balduino Fonseca (Federal University of Alagoas, Brazil), Wesley K. G. Assunção (North Carolina State University, USA), Davy Baía (Federal University of Alagoas, Brazil), and Márcio Ribeiro (Federal University of Alagoas, Brazil)</i>	
Software Testing with Large Language Models: An Interview Study with Practitioners .....	96
<i>Deolinda Santana (CESAR School, Brazil), Cleyton Magalhaes (Universidade Federal Rural de Pernambuco, Brazil), and Ronnie de Souza Santos (University of Calgary, Canada)</i>	
Model-Assisted and Human-Guided: Perceptions and Practices of Software Professionals Using LLMs for Coding .....	105
<i>Italo Santos (University of Hawai'i at Mānoa, USA), Cleyton Magalhaes (UFRPE, Brazil), and Ronnie de Souza Santos (University of Calgary, Canada)</i>	

Automated Extract Method Refactoring with Open-Source LLMs: A Comparative Study .....	113
<i>Sivajeet Chand (Technical University of Munich, Germany), Melih Kilic (Technical University of Munich, Germany), Roland Würsching (Technical University of Munich, Germany), Sushant Kumar Pandey (University of Groningen, The Netherlands), and Alexander Pretschner (Technical University of Munich, Germany)</i>	
CFCEval: Evaluating Security Aspects in Code Generated by Large Language Models .....	123
<i>Cheng Cheng (Concordia University, Canada) and Jinqiu Yang (Concordia University, Canada)</i>	
Are We Aligned? A Preliminary Investigation of the Alignment of Responsible AI Values between LLMs and Human Judgment .....	133
<i>Asma Yamani (King Fahd University of Petroleum and Minerals, Saudi Arabia), Malak Baslyman (King Fahd University of Petroleum and Minerals, Saudi Arabia), and Moataz Ahmed (King Fahd University of Petroleum and Minerals, Saudi Arabia)</i>	
A Vision for Value-Aligned AI-Driven Systems .....	143
<i>Humphrey O. Obie (Swinburne University of Technology, Australia)</i>	
Security in the Wild: An Empirical Analysis of LLM-Powered Applications and Local Inference Frameworks .....	149
<i>Julia Gomez-Rangel (Texas A&amp;M University-Corpus Christi, USA), Young Lee (Texas A&amp;M University-San Antonio, USA), and Bozhen Liu (Texas A&amp;M University-Corpus Christi, USA)</i>	
Judge the Votes: A System to Classify Bug Reports and Give Suggestions .....	160
<i>Emre Dinç (Bilkent University, Türkiye) and Eray Tüzün (Bilkent University, Türkiye)</i>	
The Future of Generative AI in Software Engineering: A Vision from Industry and Academia in the European GENIUS Project .....	170
<i>Robin Gröpler (ifak - Institut für Automation und Kommunikation e. V., Germany), Steffen Klepke (Siemens AG, Germany), Jack Johns (BT Group PLC, UK), Andreas Dreschinski (Akkodis Germany Solutions GmbH, Germany), Klaus Schmid (Universität Hildesheim, Germany), Benedikt Dornauer (Universität Innsbruck, Austria), Eray Tüzün (Bilkent University, Türkiye), Joost Noppen (BT Group PLC, UK), Mohammad Reza Mousavi (King's College London, UK), Yongjian Tang (Siemens AG, Germany), Johannes Viehmann (Fraunhofer FOKUS, Germany), Selin Şirin Aslangül (Beko, Türkiye), Beum Seuk Lee (BT Group PLC, UK), Adam Ziolkowski (BT Group PLC, UK), and Eric Zie (GoCodeGreen, UK)</i>	
PromptExp: Multi-Granularity Prompt Explanation of Large Language Models .....	182
<i>Ximing Dong (Huawei, Canada), Shaowei Wang (University of Manitoba), Dayi Lin (Huawei, Canada), Gopi Krishnan Rajbahadur (Huawei, Canada), and Ahmed E. Hassan (Queen's University)</i>	
Combining Reasoning Optimized LLMs and SMT Solvers for Automated Loop Invariant Synthesis ....	192
<i>Varun Bharti (IIIT Delhi, India), Shashwat Jha (IIIT Delhi, India), Dhruv Kumar (BITS Pilani, India), and Pankaj Jalote (IIIT Delhi, India)</i>	

SEALGuard: Safeguarding the Multilingual Conversations in Southeast Asian Languages for AI-Powered Software .....	197
<i>Wenliang Shan (Monash University, Australia), Michael Fu (The University of Melbourne, Australia), Rui Yang (Monash University, Australia), and Chakkrit Tantithamthavorn (Monash University, Australia)</i>	
On the Promises and Challenges of AI-Powered XR Glasses as Embodied Software .....	207
<i>Ruizhen Gu (The University of Sheffield, UK), Jingqiong Zhang (The University of Sheffield, UK), José Miguel Rojas (The University of Sheffield, UK), and Donghwan Shin (The University of Sheffield, UK)</i>	
HPCAgentTester: A Multi-Agent LLM Approach for Enhanced HPC Unit Test Generation .....	213
<i>Rabimba Karanjai (University Of Houston), Lei Xu (Kent State University), and Weidong Shi (University Of Houston)</i>	
Securing Smart Contract Languages with a Unified Agentic Framework for Vulnerability Repair in Solidity and Move .....	223
<i>Rabimba Karanjai (University Of Houston), Lei Xu (Kent State University), and Weidong Shi (University Of Houston)</i>	
Generative AI and Empirical Software Engineering: A Paradigm Shift .....	233
<i>Christoph Treude (Singapore Management University, Singapore) and Margaret-Anne Storey (University of Victoria, Canada)</i>	

## Benchmark & Dataset

Benchmarking Web API Integration Code Generation .....	240
<i>Daniel Maninger (Technische Universität Darmstadt, Germany), Leon Chemnitz (Pariton AI, Germany), Amir Molzam Sharifloo (Technische Universität Darmstadt, Germany), Jannis Brugger (Technische Universität Darmstadt, Germany), and Mira Mezini (Technische Universität Darmstadt, Germany)</i>	
Where Do LLMs Still Struggle? An In-Depth Analysis of Code Generation Benchmarks .....	249
<i>Amir Molzam Sharifloo ( Technische Universität Darmstadt, Germany), Maedeh Heydari (Technische Universität Darmstadt, Germany), Parsa Kazerooni (Technische Universität Darmstadt, Germany), Daniel Maninger (Technische Universität Darmstadt, Germany), and Mira Mezini (Technische Universität Darmstadt, Germany)</i>	
From Search to Reasoning: A Five-Level RAG Capability Framework for Enterprise Data .....	254
<i>Gurbinder Gill (Corvic, Inc., USA), Ritvik Gupta (Corvic, Inc., USA), Denis Lusson (Corvic, Inc., USA), Anand Chandrashekar (Corvic, Inc., USA), and Donald Nguyen (Corvic, Inc., USA)</i>	
A Tool for Benchmarking Large Language Models' Robustness in Assessing the Realism of Driving Scenarios .....	263
<i>Jiahui Wu (Simula Research Laboratory, Norway; University of Oslo, Norway), Chengjie Lu (Simula Research Laboratory, Norway; University of Oslo, Norway), Aitor Arrieta (Mondragon University, Spain), and Shaukat Ali (Simula Research Laboratory, USA)</i>	

Secure Code Generation at Scale with Reflexion .....	268
<i>Arup Datta (University of North Texas, USA), Ahmed Aljohani</i> <i>(University of North Texas, USA), and Hyunsook Do (University of North</i> <i>Texas, USA)</i>	
SWE-Sharp-Bench: A Reproducible Benchmark for C# Software Engineering Tasks .....	277
<i>Sanket Mhatre (Microsoft, India), Yasharth Bajpai (Microsoft, India),</i> <i>Sumit Gulwani (Microsoft, USA), Emerson Murphy-Hill (Microsoft, USA),</i> <i>and Gustavo Soares (Microsoft, USA)</i>	
Assertion-Aware Test Code Summarization with Large Language Models .....	281
<i>Anamul Haque Mollah (University of North Texas, USA), Ahmed Aljohani</i> <i>(University of North Texas, USA), and Hyunsook Do (University of North</i> <i>Texas, USA)</i>	
<b>Author Index .....</b>	<b>291</b>