2026 IEEE International Conference on Artificial Intelligence Testing (AITest 2026)

Fukuoka, Japan
27-30 July 2026

IEEE Catalog Number: CFP26S64-POD
ISBN: 979-8-3195-0563-7

IEEE
Advancing Technology for Humanity

Copyright © 2026, IEEE

All Rights Reserved

Copyright and Reprint Permissions:

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

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

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

IEEE Catalog Number:	CFP26S64-POD
ISBN (Print-On-Demand):	979-8-3195-0563-7
ISBN (Online):	979-8-3195-0562-0
ISSN:	2835-3552

Additional Copies of This Publication Are Available From:

Curran Associates, Inc
57 Morehouse Lane
Red Hook, NY 12571 USA

Phone:	(845) 758-0400
Fax:	(845) 758-2633
E-mail:	curran@proceedings.com
Web:	www.proceedings.com



TABLE OF CONTENTS

AutoQABench: A Three-Level UX Benchmark for Automated Evaluation of Open-Ended LLM Responses 1
Tianyue Wang, Fei Yuan, ChiaJu Miao, Muralidharan Seshadri Kaniyanoor, Kazuo Matsumoto

LLM-Based Automatic Metamorphic Test Case Generation for LLM Fairness Testing 9
Francisca Isabella Manope, Adam Rafif Faqih, Daffa Hilmy Fadhlurrohman, Jati H. Husen Rosa Reska Riskiana

Dual-Track Red Teaming and Evaluation for Traditional Chinese Localized LLMs 17
Ren Wei Lu, Wan Ping Chen, Fang Yu

Do AI Agents Exhibit Greed in Shared Resource Environments? 27
Landy Rakotoarison, Fanamby T. Randriamahenintsoa

Bridging the Interpretation Gap in Accessibility Testing: Empathetic and Legal-Aware Bug Report Generation via Large Language Models 31
Ryoya Koyama, Zhiyao Wang, Devi Karolita, Jialong Li, Kenji Tei

AI4SE for CI/CD: Explainable Code Smell Risk Analysis 40
João Domingues, Daniela Duarte, António Sousa, Nuno Pombo

A RAG approach for assisting testers in database query testing 46
Raquel Blanco, María José Suárez-Cabal, Javier Tuya, Claudio De La Riva

A Dynamic Evaluation Approach to Repository-Level Code Generation Via LLM 50
Fuze Kuang, Dongmei Liu, Yongli Wang, Honghua Huan, Mingzhe Zhao, Longyang Mi, XiaoHong Zheng

LLM-Based Detection and Test Generation for Command Injection Vulnerabilities in Python 58
Yuxuan Wang, Jingshu Chen, Qingyang Wang

Improving LLM-Based Unit Test Generation Through Root-Cause-Driven Prompt Design 66
Mizuki Yamada, Masahiko Kato, Juichi Takahashi

VectorSec: A Web-Based AI Security Scanner for Systematic Evaluation of LLM Vulnerabilities 74
Muhammad Mudassar Yamin

Evaluating DL Test Adequacy Metrics: A Correlation Study with DeepCrime's Mutation Score 78
Tien Duc Nguyen, Gonzalo Oberreuter, Siegfried Steckenbiller, Gordon Fraser, Patrick Tkalcic

Evaluating the Efficacy and Diversity of DNN Test Input Prioritisation Techniques 86
Gianmarco De Vita, Nargiz Humbatova, Paolo Tonella

Formal Trajectory Analysis for Testing Agentic AI in Stateful Environments 96
Ahmed Twabi, Yepeng Ding, Tohru Kondo

TE-TCP-Net: Parallel Transformer-Encoders for Test Case Prioritization 104
Osama Saad, Mohamed Abdelkarim, Reem ElAdawi

Robustness Verification of Recurrent Neural Networks with Abstraction Refinement 110
Li-Jen Lin, Chih-Duo Hong

Ontology-based Testing of Reinforcement Learning Applications 118
Nayab Shahzad, Franz Wotawa

Beyond Bypass: Measuring Vulnerability Regression in LLM-Generated Security Patches 126
Yasmeen Rawajfih, Adiba Mahmud, Ross Arnold

Behavioral Regression Testing for Continuously Updated Machine Learning Models 134
Vinil Pasupuleti, Srinivasateja Songa, Nagesh Gulkotwar, Shyalendar Reddy Allala, Shrey Tyagi

An Empirical Study of Transformer Representations for Fine-Grained Biological Operation Localization 138
Tang Phu Thien Nhan, Tomohiro Shibata

Improving Robustness of Semantic Segmentation for Autonomous Driving: A Case Study 146
Minggu Wang, Zhiyong Zhou, Fuyuan Zhang, Jianjun Zhao, Vinh Truong Duy Truong

Connecting Uncertainty Quantification to Safety Engineering: Uncertainty Gate Framework for Safe ML Decision Making 154
Mateus Molina, Alexander Günther, Peter Liggesmeyer

Passive Testing of AI-Enhanced Financial Systems via Protocol Traffic Analysis and Property-Based Validation 160
Alena Bulda, Daria Degtiarenko, Alexey Itkin, Iosif Itkin

A C-SSRS-Based Risk-Aware Evaluation Framework for Self-Harm Safety in LLMs 166
Lingling Zhang, Fen Zhao, Xin Yang

Metamorphic Testing of Multi-Agent LLM Systems: A Trace-Based Behavioral Oracle Framework 175
Gopalakrishnan Marimuthu

Direct Transfer Learning for Cross-Project Test Case Prioritization under Cold-Start Conditions 183
Michael Johnson, Khaled Slhoub

All Smoke, No Alarm: Oracle Signals in Agent-Authored Test Code 187
Dipayan Banik, Kowshik Chowdhury, Shazibul Islam Shamim

Bandit-Controlled Semi-Ensemble Learning for Accuracy-Energy Trade-off Under Real-Time Edge Conditions 191
Lifeng Ni, Lixin Tao

Conversational AI for Safety-Critical Virtual and Extended Reality: Intelligent Agents Across Health Informatics, Navigation, and Urban Sensing 197
Johannas Katikala, Kanishk Sharma, Tharun Swaminathan Ravi Kumar, Sharad Sharma

A Survey on Intelligent Robot Testing and Automation 201
Karthik Rengarajan, Siddharth Pokuri, Jerry Gao

LongSumEval: Question-Answering Based Evaluation and Feedback-Driven Refinement for Long Document Summarization 212
Huyen Nguyen, Haoxuan Zhang, Yang Zhang, Junhua Ding, Haihua Chen

MASTEST: A LLM-Based Multi-Agent System For Testing RESTful APIs 225
Xiaoke Han, Hong Zhu

Multi-Agent Change Impact Analysis and Test Optimization for AI-Enabled Software Systems 239
Nariman Mani, Amr S. Abdelfattah, Shakthi Weerasinghe, Xiaozhou Li, Tomas Cerny