# 2025 IEEE International Conference on Artificial Intelligence Testing (AITest 2025)

## Tucson, Arizona, USA
## 21-24 July 2025

IEEE Catalog Number: CFP25S64-POD
ISBN: 979-8-3315-8914-1

**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: | CFP25S64-POD |
| ISBN (Print-On-Demand): | 979-8-3315-8914-1 |
| ISBN (Online): | 979-8-3315-8913-4 |
| 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

CURRAN ASSOCIATES INC.
proceedings.com

# 2025 IEEE International Conference on Artificial Intelligence Testing (AITest)

# AITest 2025

## Table of Contents

**Message from the General Chairs** ........ ix

**Message from the PC Chairs** ........ x

**Committees** ........ xi

**Panel: Future AI and Intelligent Applications** ........ xiii

## Invited Papers

State-of-the-Art of Ontology-Based Testing ........ 1
*Tobias Kulmburg (Graz University of Technology, Austria), Nayab Shahzad (Graz University of Technology, Austria), and Franz Wotawa (Graz University of Technology, Austria)*

A Framework for Safe AI: Data Governance and Ecosystem Structure ........ 11
*Wei-Tek Tsai (Arizona State University, USA; Research Center of Social Security and Enterprise Finance University of International Business and Economics, China) and Li Zhang (Beihang University, China)*

## Session 1 - Testing of AI applications

Single-Valued Risk Estimation in Classification for Dependent Data ........ 24
*Alexander Günther (University of Kaiserslautern-Landau (RPTU), Germany), Sebastian Vollmer (Deutsches Forschungszentrum f¨ur K¨unstliche Intelligenz (DFKI), Germany; University of Kaiserslautern-Landau (RPTU), Germany), and Peter Liggesmeyer (Fraunhofer-Institut f¨ur Experimentelles Software Engineering (IESE), Germany; University of Kaiserslautern-Landau (RPTU), Germany)*

The Power of Patterns in Detecting News Articles Written by AI ........ 34
*Lilou Sicard-Noel (San Jose State University, USA), Eric Wu (San Jose State University, USA), and Ching Seh Mike Wu (San Jose State University, USA)*

Rethinking AI Safety and Ethics: A State-of-the-Art Multi-Task Model for Bias and Toxicity Detection via Task-Specific Supervision and Data-Centric Fine-Tuning ........ 42
*Adya Jha (EY GDS Consulting, India), Ankush Verma (EY GDS Consulting, India), and Tanusree De (EY GDS Consulting, India)*

Verification and Validation of LLM-RAG for Industrial Automation ............ 50
*Ziran Min (Siemens Corporation, Foundational Technologies, USA) and Christof J. Budnik (Siemens Corporation, Foundational Technologies, USA)*

## Session 2 - Attacks and defense in ML

Adversarial AI Training to Mitigate Cyber Attacks ............ 54
*Pranav Gangwani (Florida International University, USA), Jayesh Soni (Florida International University, USA), Abiel Almonte (Florida International University, USA), Alexander Perez-Pons (Florida International University, USA), and Himanshu Upadhyay (Florida International University, USA)*

Online Detection of Adversarial Examples by Activation Profile Inspection ............ 62
*Sofiane Bessaï (Numalis, France), Matthieu Martel (Université de Perpignan Via Domitita, France; Numalis, France), and Arnault Ioualalen (Numalis, France)*

Surveying the RAG Attack Surface and Defenses: Protecting Sensitive Company Data ............ 69
*Lynn Vonderhaar (Embry-Riddle Aeronautical University, USA), Daniel Machado (Embry-Riddle Aeronautical University, USA), and Omar Ochoa (Embry-Riddle Aeronautical University, USA)*

Formal Verification of Synthetic Image Datasets using Large Language Models ............ 77
*Alexandra Davidoff (Embry-Riddle Aeronautical University, Florida), Lynn Vonderhaar (Embry-Riddle Aeronautical University, Florida), Timothy Elvira (Embry-Riddle Aeronautical University, Florida), and Omar Ochoa (Embry-Riddle Aeronautical University, Florida)*

## Panel 1

Beyond Accuracy: Engineering Trustworthy AI Systems in Production ............ 85
*Satya Venkata Anusha Dwivedula (Morningstar Inc., USA)*

## ONLINE Session 1

Hybrid LSTM-Transformer Model for Stock Market Prediction: A Deep Learning Approach ............ 87
*Naga Sathwik Reddy Gona (University of North Carolina, Charlotte), Direesh Reddy Aunugu (Independent Researcher, USA), Vijayalaxmi Methuku (Independent Researcher, USA), Manan Agrawal (University of North Carolina, Charlotte), and Praveen Kumar Myakala (Independent Researcher, USA)*

CODEX: Testing Machine Learning with the Coverage of Data Explorer Tool ............ 94
*Erin Lanus (National Security Institute Virginia Tech, USA), Brian Lee (National Security Institute, Virginia Tech, USA), Dylan Steberg (Department of Statistics, Virginia Tech, USA), Jaganmohan Chandrasekaran (Sanghani Center for Artificial Intelligence and Data Analytics, Virginia Tech, USA), and Laura Freeman (National Security Institute, Virginia Tech, USA)*

Fixing Performance Bugs Through LLM Explanations ........ 102
*Suryansh Singh Sijwali (Pennsylvania State University, USA), Angela Marie Colom (Pennsylvania State University, USA), Anbi Guo (Pennsylvania State University, USA), and Suman Saha (Pennsylvania State University, USA)*

Evaluation of Ethical Decision Making in Large Language Models Across Classical Moral Frameworks ........ 110
*Bishal Thapa (Texas State University, USA), Alden Duarte-Vasquez (California State University, Dominguez Hills, USA), Sahar Hooshmand (California State University, Dominguez Hills, USA), and Heena Rathore (Texas State University, USA)*

Smart Test Plan Generation Using Large Language Models for Utility Systems: A Multi-Component Framework ........ 118
*Sidhartha Shukla (Amazon, USA), Prasad Yacham (Amazon, USA), Shamnad Mohamed Shaffi (Amazon Web Services, USA), Jinal Mehta (Amazon, USA), Priyamvada Govil (Amazon, USA), and Gaurav Mittal (California State University, Fullerton, USA)*

## Session 3 - Use of ML  techniques for software testing

Enhancing Adaptive Test Healing with Graph Neural Networks for Dependency-Aware Decision Making ........ 126
*Nariman Mani (Nutrosal Inc., Canada) and Salma Attaranasl (Nutrosal Inc., Canada)*

Ontology-Based Testing Revisited ........ 134
*Tobias Kulmburg (Graz University of Technology, Austria) and Franz Wotawa (Graz University of Technology, Austria)*

Prioritizing Metamorphic Relations via Execution Profile Dissimilarity for Improved Fault Detection ........ 142
*Madhusudan  Srinivasan (East Carolina University, USA) and Upulee Kanewala (University of North Florida, USA)*

A Multi-Layer Framework for AI-Driven Quality Control in Large-Scale Data Production ........ 152
*Behrouz Banitalebi (Morningstar Inc., Canada) and Satya Venkata Anusha Dwivedula (Morningstar Inc., USA)*

## ONLINE Session 2

REVISE: A Framework for Paragraph-Level Misinformation Correction in Large Language Models...... 160
*Shivangi Tripathi (Texas State University, USA), Teancy Jennifer (Texas State University, USA), Henry Griffith (San Antonio College, USA), and Heena Rathore (Texas State University, USA)*

Are Large Language Models Capable of Deep Relational Reasoning? Insights from DeepSeek-R1 and Benchmark Comparisons ........ 168
*Chi Chiu So ( The Hong Kong Polytechnic University, China), Yueyue Sun (Beijing Institute of Technology, China), Jun-Min Wang (Beijing Institute of Technology, China), Siu Pang Yung (The University of Hong Kong, China), Anthony Wai Keung Loh (The Hong Kong Polytechnic University, China), and Chun Pong Chau (The Hong Kong Polytechnic University, China)*

Assessing the Ethical Alignment of Large Language Models Using Deep Embedding Models ........ 178
*Jacob Silva (San Antonio College, TX), Henry Griffith (San Antonio College, TX), and Heena Rathore (Texas State University, TX)*

## Session 5 - Testing LLM-based AI applications

GenFair: Systematic Test Generation for Fairness Fault Detection in Large Language Models ........ 182
*Madhusudan Srinivasan (East Carolina University, USA) and Jubril Abdel (East Carolina University, USA)*

Efficient Fairness Testing in Large Language Models: Prioritizing Metamorphic Relations for Bias Detection ........ 191
*Madhusudan Srinivasan (East Carolina University, USA), Suavis Giramata (East Carolina University, USA), Venkat Naidu Gudivada (East Carolina University, USA), and Upulee Kanewala (University of North Florida, USA)*

The Impact of Imputation Timing on Model Performance Estimation ........ 201
*Ben Mitchell (Swarthmore College, USA) and Shikha Shrestha (Swarthmore College, USA)*

Mutant-Driven Test Generation for Ethereum Smart Contracts via LLMs ........ 209
*Morena Barboni (University of Camerino, Italy), Filippo Lampa (University of Camerino, Italy), Andrea Morichetta (University of Camerino, Italy), Andrea Polini (University of Camerino, Italy), and Edward Zulkoski (Quantstamp, USA)*

## Session 6 - Performance evaluation of machine learning models

Decoding 3D Geometry: Deep Networks for Mesh Classification ........ 217
*Aravind Shankara Narayanan (Georgia Institute of Technology, USA), Ekansh Gupta (Georgia Institute of Technology, USA), Junaid Syed (Georgia Institute of Technology, USA), and Matan Carmon (Georgia Institute of Technology, USA)*

A Systematic Evaluation of AI-Generated Security Patches: Analyzing and Exploiting AI-Generated Fixes ........ 224
*Adiba Mahmud (Tuskegee University, USA), Yasmeen Rawajfih (Tuskegee University, USA), and Fan Wu (Tuskegee University, USA)*

The Convergence of UX Design and AI: A Scientometric Exploration of Research Fronts ........ 232
*Serene Wang (University of North Texas, USA) and Xin Wang (University of North Texas, USA)*

**Author Index** ........ **237**