

2024 IEEE International Conference on Artificial Intelligence Testing (AITest 2024)

**Shanghai, China
15-18 July 2024**



**IEEE Catalog Number: CFP24S64-POD
ISBN: 979-8-3503-6506-1**

**Copyright © 2024 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:	CFP24S64-POD
ISBN (Print-On-Demand):	979-8-3503-6506-1
ISBN (Online):	979-8-3503-6505-4
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

2024 IEEE International Conference on Artificial Intelligence Testing (AITest) **AITest 2024**

Table of Contents

Message from the IEEE CISOSE 2024 General Co-Chairs	viii
Message from the IEEE AITest 2024 General Chairs	x
Message from the IEEE AITest 2024 PC Chairs	xi
IEEE AITest 2024 Organizing Committee	xii
IEEE AITest 2024 Program Committee	xiii
Keynote Speakers	xv
Tutorials	xxi
Plenary Panel - GenAI, Where Are You Going?	xxvi
Plenary Panel - Sustainable AI: Pursuing Computational Efficiency and Environmental Harmony	xxviii
Plenary Panel - Ethical, Societal, and Policy Implications of AI	xxx
Sponsors	xxxi

Session 1: Testing of AI Applications

Outline of an Independent Systematic Blackbox Test for ML-based Systems	1
<i>Hans-Werner Wiesbrock (ITPower Solutions GmbH) and Jürgen Grossmann (Fraunhofer FOKUS)</i>	
On a Systematic Test of ML-Based Systems: Experiments on Test Statistics	11
<i>Nicolas Grube (ITPower Solutions GmbH), Mozdeh Massah (ITPower Solutions GmbH, Germany), Michael Tebbe (ITPower Solutions GmbH, Germany), Paul Wancura (ITPower Solutions GmbH, Germany), Hans-Werner Wiesbrock (ITPower Solutions GmbH, Germany), Jürgen Grossmann (Fraunhofer-Institut für Offene Kommunikationssysteme FOKUS, Germany), and Sami Kharmah (Fraunhofer-Institut für Offene Kommunikationssysteme FOKUS, Germany)</i>	
Computer Vision Intelligence Test Modeling and Generation: A Case Study on Smart OCR	21
<i>Jing Shu (San Jose State University, USA), Bing-Jiun Miu (San Jose State University, USA), Eugene Chang (ALPSTouchStone, Inc., USA), Jerry Gao (San Jose State University, USA), and Jun Liu (San Jose State University, USA)</i>	

Session 2: Security Verification of AI Models

MEH-FEST-NA: An Ensemble Defense System Against Adversarial Attacks in Speaker Verification Systems	29
<i>Zesheng Chen (Purdue University Fort Wayne, USA), Jack Li (Purdue University Fort Wayne, USA), and Chao Chen (Purdue University Fort Wayne, USA)</i>	
Quantifying the Vulnerability of Anomaly Detection Implementations to Nondeterminism-based Attacks	37
<i>Muayed Ahmed (New Jersey Institute of Technology, USA) and Iulian Neamtii (New Jersey Institute of Technology, USA)</i>	
FaultLines - Evaluating the Efficacy of Open-Source Large Language Models for Fault Detection in Cyber-Physical Systems	47
<i>Herbert Mühlburger (Institute of Software Technology, Graz University of Technology, Austria) and Franz Wotawa (Institute of Software Technology, Graz University of Technology, Austria)</i>	

Session 3: Use of ML Techniques for Software Testing - I

ScenEval: A Benchmark for Scenario-Based Evaluation of Code Generation	55
<i>Debalina Ghosh Paul (Oxford Brookes University, UK), Hong Zhu (Oxford Brookes University, UK), and Ian Bayley (Oxford Brookes University, UK)</i>	
Proxima: A Proxy Model-Based Approach to Influence Analysis	64
<i>Sunny Shree (The University of Texas at Arlington, USA), Yu Lei (The University of Texas at Arlington, USA), Raghu N Kacker (National Institute of Standards and Technology, USA), and D. Richard Kuhn (National Institute of Standards and Technology, USA)</i>	
Utilizing Genetic Algorithms for Generating Critical Scenarios for Testing Autonomous Driving Functions	73
<i>Florian Klück (AVL List GmbH, Austria), Daniel Sumann (Graz University of Technology, Austria), and Franz Wotawa (Graz University of Technology, Austria)</i>	

Session 4: Use of ML techniques for software testing - II

SpecNLP: A Pre-trained Model Enhanced with Spectrum Profile for Bug Localization	81
<i>Mahdi Farzandway (University of Tehran, Iran) and Fatemeh Ghassemi (University of Tehran, Tehran Institute for Advanced Studies, Khatam University, Iran)</i>	

Session 5: The First IEEE International Workshop on Testing and Evaluation of Large Language Models (TELLMe 2024)

Benchmarks and Metrics for Evaluations of Code Generation: A Critical Review	87
<i>Debalina Ghosh Paul (Oxford Brookes University, UK), Hong Zhu (Oxford Brookes University, UK), and Ian Bayley (Oxford Brookes University, UK)</i>	

Leveraging Large Language Models for Python Unit Test	95
<i>Jiri Medlen (LIGS University) and Emese Bari (LIGS University)</i>	
Black-box L1 and L2 Adversarial Attack Based on Genetic Algorithm	101
<i>Jiyuan Sun (Kyushu University), Haibo Yu (Kyushu Sangyo University), and Jianjun Zhao (Kyushu University)</i>	

Invited Papers

User Centric Evaluation of Code Generation Tools	109
<i>Tanha Miah (Oxford Brookes University, UK) and Hong Zhu (Oxford Brookes University, UK)</i>	
A Survey on Data Quality Dimensions and Tools for Machine Learning	120
<i>Yuhan Zhou (University of North Texas), Fengjiao Tu (University of North Texas), Kewei Sha (University of North Texas), Junhua Ding (University of North Texas), and Haihua Chen (University of North Texas)</i>	
AI Test Modeling and Analysis for Intelligent Chatbot Mobile App - A Case Study on Wysa	132
<i>Jerry Gao (San Jose State University, USA; ALPSTouchStone, Inc., USA.), Prerna Garsole (San Jose State University, USA.), Radhika Agarwal (ALPSTouchStone, Inc., USA.), and Shengqiang Liu (Computer Network Information Center, Chinese Academy of Sciences, Tiangong University, China)</i>	
Evaluation of Question-Answering Based Text Summarization using LLM	142
<i>Junhua Ding (University of North Texas), Huyen Nguyen (University of North Texas), and Haihua Chen (University of North Texas)</i>	

Fast Abstracts

Real-Time Anomaly Detection in Time Series Using Transformer-Like Architecture	150
<i>Meixian Zhang (Baidu Group), Xue Shi (Baidu Group), and Jiaxin Huang (Baidu Group)</i>	
Test Case Prioritization for Regression Testing Using Machine Learning	152
<i>Priyanka Sawant (University of New Hampshire)</i>	

Author Index	155
---------------------------	------------