

# **37th International Conference on Software Engineering & Knowledge Engineering (SEKE2025)**

Pompeii, Italy and Online  
29 September - 4 October 2025

ISBN: 979-8-3313-2907-5

**Printed from e-media with permission by:**

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



**Some format issues inherent in the e-media version may also appear in this print version.**

Copyright© (2025) by KSI Research Inc.  
All rights reserved.

Printed by Curran Associates, Inc. (2026)

For permission requests, please contact KSI Research Inc.  
at the address below.

KSI Research Inc.  
156 Park Square Lane  
Pittsburgh, PA 15238  
USA

Phone: 412-606-5022

[office@ksiresearch.org](mailto:office@ksiresearch.org)

**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  
Email: [curran@proceedings.com](mailto:curran@proceedings.com)  
Web: [www.proceedings.com](http://www.proceedings.com)

# Table of Contents

## Session 1: Software Engineering and Knowledge Engineering

**011** *David Parsons* **AI Literacy Development for Software Engineering Education: Mapping AI from the SWEBOK to the SAIL Framework** 1

**022** *Sergio Di Meglio, Aniello Somma, Luigi Libero Lucio Starace, Fabio Scippacercola, Giancarlo Sperl and Sergio Di Martino* **Large Language Models in the Travel Domain: An Industrial Experience** 7

**044** *Loredana Caruccio, Grazia Margarella and Giuseppe Polese* **GitHub Issue Mining to enhance Developer Profiling: A pipeline based on NLP tasks (S)** 13

**085** *Wenya Wang, Wang Gao, Zhenghe Wang, Sicong Cao, Xingwei Lin and Dawu Gu* **AePPollo: Automated Exploit Generation for Prototype Pollution Vulnerabilities in Node.js Application** 19

## Session 2 Knowledge Engineering and Software Engineering

**075** *Christos Konstantopoulos and Kostas Tsihclas* **A Framework for Explanations in Weighted Networks: The Community Detection Case** 25

**100** *Huan Jin, Yingkai Yuan, Peng Ye, Hongyan Wan, Zhiyuan Zou and Bangchao Wang* **How to enhance requirements-to-code traceability? From the perspective of project artifacts** 31

**107** *Bingyang Wei, Robin Chataut and Lin Deng* **Project Pulse: Enhancing Peer Evaluation and Team Accountability in Senior Design Projects** 37

**083** *Khalid Hasan and Jamil Saquer* **Beyond Architectures: Evaluating the Role of Contextual Embeddings in Detecting Bipolar Disorder on Social Media** 43

## Session SCA: Social Computing and Applications

- 093** *Xinzhi Wang, Zhennan Li, Yiming Zhang, Jiayan Qian and Yinghua Ma* **Substructure Mining Based on Fact Reflection** 50
- 058** *Niannian Zhao and Xiao Wei* **StepRAG: Step-by-Step Evidence Retrieval and Aggregation for Complex QA (S)** 56
- 095** *Minglong Zhang, Weidong Liu and Kai Wang* **Patent Concept Standardization Prediction Model** 60
- 131** *Yicun Liu, Xinyu Zheng and Xianguo Zhang* **MDC-GCN: Multi-Directional Contextual Graph Convolution Network for Aspect Sentiment Triplet Extraction** 66
- 132** *Pengfei Liang, Lirong Chen and Dongsong Zhang* **Detecting Machine-Generated Text in the Wild via a Hybrid Semantic-Statistical Model** 72
- 129** *Tianqi Fu, Lirong Chen, Qiang Li and Yao Wang.* **ABSTRE: A Multi-Feature Fusion Approach for Mobile App Review Classification** 78

## Session DMM: Data Modeling and Mining

- 049** *Claudia Cavallaro and Vincenzo Cutello* **An efficient heuristic algorithm to compute minimal and stable weighted feedback arc sets (S)** 84
- 086** *Tao Wu, Yi Yang, Chunming Yang, Bo Li and Hui Zhang* **Individual Fairness in K-means Clustering: A Density-Aware Similarity Approach (S)** 88
- 080** *Claudia Cavallaro* **From Structure to Suspicion: Cross-Network Detection of Fraudulent Behavior in Real-World Signed Graphs (S)** 92

### Session 3: Software Testing and Analysis

071 *Elsa Paulson and Renee Bryce*. Enhancing Automated Test Efficiency: A Hybrid Machine Learning Approach for Test Suite Reduction in Gray-Box Web Application Testing (S) 102

108 *Asrar Qassem, Renee Bryce and Khalid Alkhalidi*. A Survey on Visual GUI Testing for Automatic Test Case Generation: Tools, AI Techniques, and Emerging Trends (S) 108

114 *Imran Pinjari, Vaishnavi Kattula and Michael Shin*. Bridging Static and Dynamic Analysis using IBCs (S) 114

043 *Chendong Lin, Leyi Cheng, Jiayi Wu and Xiaofang Zhang*. SGTTest: A Semantics-Guided Framework for LLM-Driven Unit Test Generation 118

053 *Yiwei Li, Jiaxin Liu, Yanfeng Hu, Chen Liu, Yating Zhang, Liangze Yin and Wei Dong*. Beyond Test Cases: Multi-Agent Collaboration for Detecting Errors in Full-Score Code Implementations 124

050 *Seth Wolfgang, Lan Lin and Fengguang Song*. Automated Statistical Testing and Certification of a Reliable Model-Coupling Server for Scientific Computing 130

### Session 4: Formal Methods and Protocol Verification

010 *Xunhao Wang and Wen Bai*. Improving Local Search for Maximum Satisfiability by the Feasible Solution Constraint (S) 137

020 *Canh Minh Do and Kazuhiro Ogata*. Formal Specification and Model Checking of the BB84 Protocol in Maude (S) 141

<b>042</b>	<i>Huiying Liu, Huibiao Zhu and Sini Chen.</i>	<b>Formal Modeling and Verification of AMQP Protocol Using Multiparty Session Types (S)</b>	147
<b>092</b>	<i>Di Wu, Wenting Dong and Huibiao Zhu.</i>	<b>Formalization and Verification of XMPP Communication Mechanism Using CSP (S)</b>	153
<b>023</b>	<i>Trong Binh Hoang, Duong Dinh Tran, Canh Minh Do and Kazuhiro Ogata.</i>	<b>Formal Specification and Analysis of Post-quantum OpenPGP Protocol in CafeOBJ</b>	159
<b>048</b>	<i>Han Bao, Chenhui Wang and Huibiao Zhu.</i>	<b>Verifying the Communication and Security Mechanisms of WAMP</b>	165
<b>Session 5: Machine Learning for Software Quality and Reliability</b>			
<b>122</b>	<i>Qinxuan Shi, Zhanglong Yang and Sicong Shao.</i>	<b>DualATLog: Dual-mode Anomaly Transformer-based Log Anomaly Detection Framework for Securing Software Systems (S)</b>	171
<b>018</b>	<i>Naizhu Jin, Zhong Li, Tian Zhang and Qingkai Zeng.</i>	<b>GUARD: Dual-Agent based Backdoor Defense on Chain-of-Thought in Neural Code Generation</b>	175
<b>025</b>	<i>Yike Liu and Jinfeng Liu.</i>	<b>Detecting Pure Functions in JavaScript Using Machine Learning</b>	181
<b>101</b>	<i>Chenyi Zhou, Ran Mo and Weijun Hu.</i>	<b>Predicting Defective Code Clones in Autonomous Driving Software</b>	187
<b>126</b>	<i>Zhanglong Yang, Qinxuan Shi and Sicong Shao.</i>	<b>Energy-efficient Anomaly Detection for Securing Water Treatment and Distribution Systems</b>	193
<b>098</b>	<i>Changrong Xie, Yang Zhang, Xinjun Mao and Huiping Zhou.</i>	<b>Decoding Serverless Security: Exploring Developer Challenges and Solutions from Stack Overflow</b>	199

## Session 6: LLM-Based Content Generation and Reasoning

- 038** *Hiroyuki Nakagawa and Shinichi Honiden. LLM-based Adaptive Requirements Elicitation for Innovative Systems (S)* 205
- 084** *Tong Li, Xinyu Hu and Gang Xiao. LLM-Augmented KBQA: Robust Reasoning over Error-Prone Programs (S)* 211
- 123** *To Kim Bao Pham, Ahmad Alshomar, Sam Supakkul, Tom Hill and Lawrence Chung. Generating Synthetic Nonfunctional Requirements using Large Language Models for Training Machine Learning Classifiers (S)* 217
- 012** *Longfei Sun, Yao Lu, Xinjun Mao, Tanghaoran Zhang, Zhang Zhang and Huiping Zhou. ConflictLens: an LLM-Based Method for Detecting Semantic Merge Conflicts* 223
- 019** *Yifei Liu, Junyi Lu, Li Yang, Yang Wang, Zheng Zhang, Xiaoxiao Ma, Jiajia Ma, Fengjun Zhang and Chun Zuo. EXE-Reviewer: Towards EXplainable and Effective Review Comments Generation* 229
- 102** *Aneesa Bashir and Rong Peng. GraphFusion: A Hybrid Semantic 璽砂倪 Symbolic Retrieval Approach to NL2Cypher Translation for Domain-Specific KGQA using LLM* 235

## Session 7: Code Development, Optimization, and Tools

- 026** *Leyi Cheng, Jiayi Wu, Hao Zhu and Xiaofang Zhang. Towards the Impact of Code Readability on Vulnerability Detection: An Empirical Study* 241
- 037** *Shuo He, Hongbao Liu, Tang Tao and Jiawei Jiang. A Privacy Preserving Cross-Domain Label Propagation Algorithm* 247
- 076** *Haolin Pan, Yin Zhang, Mingjie Xing and Yanjun Wu. Navigating the SIMD Optimization Maze: A Reinforcement Learning Approach to Library and Compiler Co-Optimization* 253

**070** Yang Liu, Shihao Gao, Hanghang Cao, Hongbin Zhang, Mingjie Xing and Yanjun Wu. **MLProf: A Multi-Level Runtime Performance Profiling Framework for MLIR Operations** 259

**067** Khalid Alkhalidi, Asrar A. Qassem and Stephanie Ludi. **Kencode: A Novel Methodology for Voice-Based Programming** 265

**069** Yuyang Wang, Yutong Ye, Yingbo Zhou, Qi Wen, Xiang Lian, Xian Wei and Mingsong Chen. **MuseCNN: Embedding-Guided Polyphonic Music Accompaniment Generation** 271

## **Session 8: Machine Learning for Prediction and Recommendation**

**051** Yunpeng Han, Yugen Du, Xia Chen, Yingwei Luo, Guoxing Tang, Benchi Ma and Jingjing Huang. **MFAE: Multi-Feature-Aware Expert Modeling for Web Service QoS Prediction (S)** 277

**064** Xu Zhiqianru, Zeng Cheng, Wang Aoyu and Zeng Chao. **Dynamic Uncertainty-Aware Fusion for Few-Shot Multimodal Learning (S)** 281

**016** Jiyun Li, Teng Li, Chen Qian and Jiabao Zhao. **FedDMC: Dual-Model Dynamic Interaction with Multi-Stage Correction for Noisy Federated Learning** 285

**039** Zhongqin Bi, Meiyun Xiang and Weina Zhang. **Diffusion Model-based Intent Contrastive Learning for Sequential Recommendation** 291

**087** Hao Yang and Meijuan Yin. **Contrastive Learning for Entity Matching with Similar Sample Pairs** 297

## **Session 9: Domain-Specific and Applied LLM Research**

**036** Mahmoud Abusaqer and Jamil Saquer. **Parameter-Efficient Hate Speech Detection: A Comprehensive Evaluation of LoRA-Adapted LLMs Across 18 Architectures** 304

**088** Turki Albalawi and Renee Bryce. **Evaluating Large Language Models for Extracting Usability Issues from User Bug Reports** 312

**104** *Haoying Jin and Song Yang. When LLMs Meet Multi-Category Complex Classification: A Comparative Study* 318

**111** *Sarra Laaouar and Nacim Yanes. CLARA: A Context-Aware Drug Recommendation Framework Using Multi-Level Integration of Large Language Models and Graph Neural Networks* 324

**125** *Aryan K C and Unnati Shah. Multi-Level Software Requirements Classification: Assessing BERT Variants and their Hybrid Models* 330

## **Session 10: Systems Modeling, Theory, and Learning Frameworks**

**054** *Wenzheng Yang, Yongxin Zhao, Qiang Wang, Yongjian Li and Yudan Tian. Learning Behavior Trees for Automated Guided Vehicles via Genetic and Reinforcement Methods* 336

**057** *Xiangyu Li, Yihao Zhang and Meng Sun. Component Composition in MedTiny: Multi-Level Constructs and Operational Semantics* 342

**072** *Tushar Agrawal, Gursimran Walia and Vaibhav Anu. Insights from an Industry Survey on Software Design Errors* 348

**075** *Christos Konstantopoulos and Kostas Tsihlias. A Framework for Explanations in Weighted Networks: The Community Detection Case* 354

**077** *Liu Mao, Ning Li, Zechen Meng, Xia Xie, Mei Yu, Ruiguo Yu and Mankun Zhao. Enhancing Knowledge Tracing with Inspiration of Psychological Learning Theories* 360

**109** *Jiang Xiong, Jing Liu and Mingxing Liu. Stochastic Z<sub>stack</sub> lus: Stochastic Hybrid Modeling and Verification for Nuclear I&C System* 366

**Notes: (S) is a short paper**