

**2025 IEEE/ACM 4th
International Conference on AI
Engineering - Software
Engineering for AI (CAIN 2025)**

**Ottawa, Ontario, Canada
27-28 April 2025**



**IEEE Catalog Number: CFP25AK2-POD
ISBN: 979-8-3315-0220-1**

**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:	CFP25AK2-POD
ISBN (Print-On-Demand):	979-8-3315-0220-1
ISBN (Online):	979-8-3315-0219-5

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/ACM 4th International Conference on AI Engineering – Software Engineering for AI (CAIN) **CAIN 2025**

Table of Contents

Message from the Conference Chairs	x
Organizing Committee	xii
Program Committee	xiii
Steering Committee	xvi

Research Papers

Rule-Based Assessment of Reinforcement Learning Practices using Large Language Models	1
<i>Evangelos Ntontos (University of Vienna, Austria), Stephen John Warnett (University of Vienna, Austria), and Uwe Zdun (University of Vienna, Austria)</i>	
Designing and Implementing LLM Guardrails Components in Production Environments	12
<i>Mateus Devino (IBM Research, Brazil), Evaline Ju (IBM Research, Brazil), and Paulo Marques Caldeira (IBM Research, Brazil)</i>	
Themes of Building LLM-Based Applications for Production: A Practitioner's View	18
<i>Alina Mailach (Leipzig University), Sebastian Simon (Leipzig University), Johannes Dorn (Leipzig University), and Norbert Siegmund (Leipzig University)</i>	
InsightAI: Root Cause Analysis in Large Log Files with Private Data using Large Language Model	31
<i>Maryam Ekhlesi (Polytechnique Montreal, Canada), Anurag Prakash (Ciena, Canada), Maxime Lamothe (Polytechnique Montreal, Canada), and Michel Dagenais (Polytechnique Montreal, Canada)</i>	
Developing Multi-Agent LLM Applications through Continuous Human-LLM Co-Programming ...	42
<i>Hui Song (SINTEF Digital, Norway), Arda Goknil (SINTEF Digital, Norway), Xiaojun Jiang (Oslo University Hospital, Norway), Espen Melum (Oslo University Hospital, Norway), Hyunwhan Joe (Seoul National University, Korea), Caterina Gazzotti (University of Modena, Italy), Valerio Frascolla (Intel Deutschland GmbH, Germany), Adela Nedisan Videsjorden (SINTEF Digital, Norway), and Phu Nguyen (SINTEF Digital, Norway)</i>	

How Do Model Export Formats Impact the Development of ML-Enabled Systems? A Case Study on Model Integration	48
<i>Shreyas Kumar Parida (Vrije Universiteit Amsterdam, The Netherlands), Ilias Gerostathopoulos (Vrije Universiteit Amsterdam, The Netherlands), and Justus Bogner (Vrije Universiteit Amsterdam, The Netherlands)</i>	
RAGProbe: Breaking RAG Pipelines with Evaluation Scenarios	60
<i>Shangeetha Sivasothy (Deakin University, Australia), Scott Barnett (Deakin University, Australia), Stefanus Kurniawan (Deakin University, Australia), Zafaryab Rasool (Deakin University, Australia), and Rajesh Vasa (Deakin University, Australia)</i>	
On-Device or Remote? On the Energy Efficiency of Fetching LLM-Generated Content	72
<i>Vince Nguyen (Vrije Universiteit Amsterdam, The Netherlands), Vidya Dhopate (Vrije Universiteit Amsterdam, The Netherlands), Hieu Huynh (Vrije Universiteit Amsterdam, The Netherlands), Hiba Bouhlal (Vrije Universiteit Amsterdam, The Netherlands), Anusha Annengala (Vrije Universiteit Amsterdam, The Netherlands), Gian Luca Scoccia (Gran Sasso Science Institute, Italy), Matias Martinez (Universitat Politècnica de Catalunya, Spain), Vincenzo Stoico (Vrije Universiteit Amsterdam, The Netherlands), and Ivano Malavolta (Vrije Universiteit Amsterdam, The Netherlands)</i>	
LoCoML: A Framework for Real-World ML Inference Pipelines	83
<i>Kritin Maddireddy (IIIT Hyderabad, India), Santhosh Kotekal Methukula (IIIT Hyderabad, India), Chandrasekar Sridhar (IIIT Hyderabad, India), and Karthik Vaidhyanathan (IIIT Hyderabad, India)</i>	
Towards Continuous Experiment-Driven MLOps	89
<i>Keerthiga Rajenthiram (Vrije Universiteit Amsterdam, The Netherlands), Milad Abdullah (Charles University, Czech Republic), Ilias Gerostathopoulos (Vrije Universiteit Amsterdam, The Netherlands), Petr Hnětýnka (Charles University, Czech Republic), Tomáš Bureš (Charles University, Czech Republic), Gerard Pons (Universitat Politècnica de Catalunya, Spain), Besim Bilalli (Universitat Politècnica de Catalunya, Spain), and Anna Queralt (Universitat Politècnica de Catalunya, Spain)</i>	
ImageBiTe: A Framework for Evaluating Representational Harms in Text-to-Image Models	95
<i>Sergio Morales (Universitat Oberta de Catalunya, Spain), Robert Clarisó (Universitat Oberta de Catalunya, Spain), and Jordi Cabot (University of Luxembourg, Luxembourg)</i>	
Approach Towards Semi-Automated Certification of Low Criticality ML-Enabled Airborne Applications	107
<i>Chandrasekar Sridhar (IIIT Hyderabad, India), Vyakhya Gupta (IIIT Hyderabad, India), Prakhar Jain (IIIT Hyderabad, India), and Karthik Vaidhyanathan (IIIT Hyderabad, India)</i>	
From Hazard Identification to Controller Design: Proactive and LLM-Supported Safety Engineering for ML-Powered Systems	113
<i>Yining Hong (Carnegie Mellon University), Christopher S. Timperley (Carnegie Mellon University), and Christian Kästner (Carnegie Mellon University)</i>	

Bringing Machine Learning Models Beyond the Experimental Stage with Explainable AI	119
<i>Niels With Mikkelsen (Jyske Bank A/S, Denmark), Lasse Steen Pedersen (Jyske Bank A/S, Denmark), Mansoor Hussain (Jyske Bank A/S, Denmark), Victor Foged (Deloitte, Denmark), and Ekkart Kindler (Technical University of Denmark, Denmark)</i>	
Challenges in AI Projects for Machinery and Plant Engineering	130
<i>Richard Nordsieck (XITASO IT & Software Solutions GmbH, Germany), Jan-Philipp Steghöfer (XITASO IT & Software Solutions GmbH, Germany), and Manish Bhandari (XITASO IT & Software Solutions GmbH, Germany)</i>	
Towards a Domain-Specific Modeling Language for Streamlined Change Management in AI Systems Development	132
<i>Razan Abualsaud (L'Institut de Recherche en Informatique de Toulouse (IRIT))</i>	
An AI-Driven Requirements Engineering Framework Tailored for Evaluating AI-Based Software ..	138
<i>Hamed Barzamini (Northern Illinois University), Fatemeh Nazaritiji (Northern Illinois University), Annalise Brockmann (Northern Illinois University), Hasan Ferdowsi (Northern Illinois University), and Mona Rahimi (Northern Illinois University)</i>	
MLScnt: A Tool for Anti-Pattern Detection in ML Projects	150
<i>Karthik Shivashankar (University of Oslo, Norway) and Antonio Martini (University of Oslo, Norway)</i>	
Debugging and Runtime Analysis of Neural Networks with VLMs (A Case Study)	161
<i>Boyue Caroline Hu (University of Toronto, Canada), Divya Gopinath (NASA Ames, USA), Corina S. Păsăreanu (Carnegie Mellon Univ., USA), Nina Narodytska (VMware Research, USA), Ravi Mangal (Colorado State University, USA), and Susmit Jha (SRI International, USA)</i>	
Investigating Issues that Lead to Code Technical Debt in Machine Learning Systems	173
<i>Rodrigo Ximenes (PUC-Rio, Brazil), Antonio Pedro Santos Alves (PUC-Rio, Brazil), Tatiana Escovedo (PUC-Rio, Brazil), Rodrigo Spinola (Virginia Commonwealth University (VCU), USA), and Marcos Kalinowski (PUC-Rio, Brazil)</i>	
Addressing Quality Challenges in Deep Learning: The Role of MLOps and Domain Knowledge	184
<i>Santiago del Rey (Universitat Politècnica de Catalunya (BarcelonaTech)), Adrià Medina (Universitat Politècnica de Catalunya (BarcelonaTech)), Xavier Franch (Universitat Politècnica de Catalunya (BarcelonaTech)), and Silverio Martínez-Fernández (Universitat Politècnica de Catalunya (BarcelonaTech))</i>	
DDPT: Diffusion-Driven Prompt Tuning for Large Language Model Code Generation	190
<i>Jinyang Li (University of Adelaide, Australia), Sangwon Hyun (University of Adelaide, Australia), and M. Ali Babar (University of Adelaide, Australia)</i>	

Engineering LLM Powered Multi-Agent Framework for Autonomous CloudOps	201
<i>Kannan Parthasarathy (MontyCloud Inc), Karthik Vaidhyanathan (IIIT Hyderabad, India), Rudra Dhar (IIIT Hyderabad, India), Venkat Krishnamachari (MontyCloud Inc), Adyansh Kakran (IIIT Hyderabad, India), Sreemae Akshathala (IIIT Hyderabad, India), Shrikara Arun (IIIT Hyderabad, India), Amey Karan (IIIT Hyderabad, India), Basil Muhammed (MontyCloud Inc), Sumant Dubey (MontyCloud Inc), and Mohan Veerubhotla (MontyCloud Inc)</i>	
Generating and Verifying Synthetic Datasets with Requirements Engineering	212
<i>Lynn Vonderhaar (Embry-Riddle Aeronautical University, USA), Timothy Elvira (Embry-Riddle Aeronautical University, USA), and Omar Ochoa (Embry-Riddle Aeronautical University, USA)</i>	
LLM-Based Safety Case Generation for Baidu Apollo: Are We There Yet?	222
<i>Oluwafemi Odu (York University, Canada), Alvine Boaye Belle (York University, Canada), and Song Wang (York University, Canada)</i>	
SqPal - Text to SQL GenAI Tool for PayPal	234
<i>Dan Liyanage (PayPal, USA), Mahshid Moha (PayPal, USA), and Sandy Suresh (PayPal, USA)</i>	

Doctoral Symposium

Optimizing Data Analytics Workflows through User-Driven Experimentation: Progress and Updates	236
<i>Keerthiga Rajenthiram (Vrije Universiteit Amsterdam, The Netherlands)</i>	
CoCo Challenges in ML Engineering Teams: How to Collaboratively Build ML-Enabled Systems ..	241
<i>Aidin Azamnouri (Technical University of Munich)</i>	
Towards a Privacy-by-Design Framework for ML-Enabled Systems	244
<i>Yorick Sens (Ruhr University Bochum, Germany)</i>	
Towards an Adoption Framework to Foster Trust in AI-Assisted Software Engineering	247
<i>Marvin Muñoz Barón (Technical University of Munich, Germany)</i>	
Assessing and Enhancing the Robustness of LLM-Based Multi-Agent Systems Through Chaos Engineering	250
<i>Joshua Owotogbe (Tilburg University, Netherlands)</i>	
Designing ML-Enabled Software Systems with ML Model Composition: A Green AI Perspective ..	253
<i>Rumbidzai Chitakunye (Vrije Universiteit Amsterdam, Netherlands)</i>	
A Metrics-Oriented Architectural Model to Characterize Complexity on Machine Learning-Enabled Systems	256
<i>Renato Cordeiro Ferreira (University of São Paulo, Brazil)</i>	
Identification and Optimization of Redundant Code using Large Language Models	261
<i>Shamse Tasnim Cynthia (University of Saskatchewan, Canada)</i>	
Systematic Testing of Security-Related Vulnerabilities in LLM-Based Applications	264
<i>Hasan Kaplan (Tilburg University, Netherlands)</i>	
Model-Based Verification for AI-Enabled Cyber-Physical Systems through Guided Falsification of Temporal Logic Properties	267
<i>Hadiza Umar Yusuf (University of Michigan-Dearborn, USA)</i>	

A Holistic Framework for Evolving AI-Based Systems	270
<i>Merel Veracx (Fontys UAS, The Netherlands)</i>	

Posters

All You Need is an AI Platform: A Proposal for a Complete Reference Architecture	273
<i>Benjamin Weigell (University of Augsburg, Germany), Fabian Stieler (University of Augsburg, Germany), and Bernhard Bauer (University of Augsburg, Germany)</i>	
Safeguarding LLM-Applications: Specify or Train?	275
<i>Hala Abdelkader (Deakin University, Australia), Mohamed Abdelrazek (Deakin University, Australia), Sankhya Singh (Deakin University, Australia), Irini Logothetis (Deakin University, Australia), Priya Rani (RMIT University, Australia), Rajesh Vasa (Deakin University, Australia), and Jean-Guy Schneider (Monash University, Australia)</i>	
Navigating the Shift: Architectural Transformations and Emerging Verification Demands in AI-Enabled Cyber-Physical Systems	277
<i>Hadiza Umar Yusuf (University of Michigan-Dearborn, USA) and Khoulood Gaaloul (University of Michigan-Dearborn, USA)</i>	
Task Decomposition and RAG as Design Patterns for LLM-Based Systems	279
<i>Orlando Marquez Ayala (ServiceNow)</i>	
Finding Trojan Triggers in Code LLMs: An Occlusion-Based Human-in-the-Loop Approach	281
<i>Aftab Hussain (Texas A&M University, USA), Md Rafiqul Islam Rabin (University of Houston, USA), Toufique Ahmed (University of California, USA), Mohammad Amin Alipour (University of Houston, USA), Bowen Xu (North Carolina State University, USA), and Stephen Huang (University of Houston, USA)</i>	
Evaluating Reinforcement Learning Safety and Trustworthiness in Cyber-Physical Systems	283
<i>Katherine Dearstyne (University of Notre Dame), Pedro Alarcon Granadeno (University of Notre Dame), Theodore Chambers (University of Notre Dame), and Jane Cleland-Huang (University of Notre Dame)</i>	
Random Perturbation Attack on LLMs for Code Generation	285
<i>Qiulu Peng (Carnegie Mellon University), Chi Zhang (Carnegie Mellon University), Ravi Mangal (Colorado State University), Corina Pasareanu (Carnegie Mellon University), and Limin Jia (Carnegie Mellon University)</i>	
Author Index	289