

**2022 IEEE/ACM 1st  
International Conference on AI  
Engineering - Software  
Engineering for AI (CAIN 2022)**

**Pittsburgh, Pennsylvania, USA  
16-17 May 2022**



IEEE Catalog Number: CFP22AK2-POD  
ISBN: 978-1-6654-5206-9

**Copyright © 2022, Association for Computing Machinery (ACM)  
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:	CFP22AK2-POD
ISBN (Print-On-Demand):	978-1-6654-5206-9
ISBN (Online):	978-1-4503-9275-4

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

CURRAN ASSOCIATES INC.  
**proceedings**  
.com

# 2022 IEEE/ACM 1st International Conference on AI Engineering – Software Engineering for AI (CAIN) **CAIN 2022**

## Table of Contents

Message from the CAIN 2022 Conference Chairs .....	ix
CAIN 2022 Organizing Committee .....	xi
CAIN 2022 Program Committee .....	xiii

### Session 1: Quality Assurance

What is Software Quality for AI Engineers? Towards a Thinning of the Fog .....	1
<i>Valentina Golendukhina (University of Innsbruck, Austria), Valentina Lenarduzzi (University of Oulu, Finland), and Michael Felderer (University of Innsbruck, Austria)</i>	
Exploring ML Testing in Practice - Lessons Learned from an Interactive Rapid Review with Axis Communications .....	10
<i>Qunying Song (Lund University, Lund, Sweden), Markus Borg (RISE Research Institutes of Sweden and Lund University, Lund, Sweden), Emelie Engström (Lund University, Lund, Sweden), Håkan Ardö (Axis Communications, Lund, Sweden), and Sergio Rico (Lund University, Lund, Sweden)</i>	
Quality Assurance of Generative Dialog Models in an Evolving Conversational Agent Used for Swedish Language Practice .....	22
<i>Markus Borg (RISE Research Institutes of Sweden, Sweden), Johan Bengtsson (Lund University, Sweden), Harald Österling (Lund University, Sweden), Alexander Hagelborn (NordAxon AB, Sweden), Isabella Gagner (NordAxon AB, Sweden), and Piotr Tomaszewski (RISE Research Institutes of Sweden, Sweden)</i>	

### Session 2: Posters

MLOps: Five Steps to Guide its Effective Implementation .....	33
<i>Beatriz M.A. Matsui (Federal University of ABC - UFABC, Brazil) and Denise H. Goya (Federal University of ABC - UFABC, Brazil)</i>	
Towards a Methodological Framework for Production-Ready AI-Based Software Components .....	35
<i>Markus Haug (University of Stuttgart, Germany) and Justus Bogner (University of Stuttgart, Germany)</i>	

Preliminary Insights to Enable Automation of the Software Development Process in Software StartUps: An Investigation Study from the use of Artificial Intelligence and Machine Learning .....	37
<i>Olimar Borges (PUCRS University, Brazil), Valentina Lenarduzzi (University of Oulu, Finland), and Rafael Prikladnicki (PUCRS University, Brazil)</i>	
Identification of Out-of-Distribution Cases of CNN using Class-Based Surprise Adequacy .....	39
<i>Mira Marhaba (Polytechnique Montreal, Canada), Ettore Merlo (Polytechnique Montreal, Canada), Foutse Khomh (Polytechnique Montreal, Canada), and Giuliano Antoniol (Polytechnique Montreal, Canada)</i>	
Robust Active Learning: Sample-Efficient Training of Robust Deep Learning Models .....	41
<i>Yuejun Guo (University of Luxembourg), Qiang Hu (University of Luxembourg), Maxime Cordy (University of Luxembourg), Mike Papadakis (University of Luxembourg), and Yves Le Traon (University of Luxembourg)</i>	
Structural Causal Models as Boundary Objects in AI System Development .....	43
<i>Hans-Martin Heyn (University of Gothenburg, Sweden) and Eric Knauss (University of Gothenburg, Sweden)</i>	
TopSelect: A Topology-Based Feature Selection Method for Industrial Machine Learning .....	46
<i>Hadil Abukwaik (ABB Corporate Research Center Germany), Lefter Sula (ABB Corporate Research Center Germany), and Pablo Rodriguez (ABB Corporate Research Center Germany)</i>	
Pynblint: A Static Analyzer for Python Jupyter Notebooks .....	48
<i>Luigi Quaranta (University of Bari, Italy), Fabio Calefato (University of Bari, Italy), and Filippo Lanubile (University of Bari, Italy)</i>	
Traceable Business-to-Safety Analysis Framework for Safety-Critical Machine Learning Systems .....	50
<i>Jati H. Husen (Waseda University, Japan), Hironori Washizaki (Waseda University, Japan), Hnin Thandar Tun (Waseda University, Japan), Nobukazu Yoshioka (Waseda University, Japan), Yoshiaki Fukazawa (Waseda University, Japan), and Hironori Takeuchi (Musashi University, Japan)</i>	
A New Approach for Machine Learning Security Risk Assessment - Work in Progress .....	52
<i>Jun Yajima (Fujitsu Limited, Japan), Maki Inui (Fujitsu Limited, Japan), Takanori Oikawa (Fujitsu Limited, Japan), Fumiyoshi Kasahara (Fujitsu Limited, Japan), Ikuya Morikawa (Fujitsu Limited, Japan), and Nobukazu Yoshioka (Waseda University, Japan)</i>	

## Session 3: Training & Learning

An Empirical Evaluation of Flow Based Programming in the Machine Learning Deployment Context .....	54
<i>Andrei Paleyes (University of Cambridge, United Kingdom), Christian Cabrera (University of Cambridge, United Kingdom), and Neil D. Lawrence (University of Cambridge, United Kingdom)</i>	

Checkpointing and Deterministic Training for Deep Learning .....	65
<i>Xiangzhe Xu (Purdue University, USA), Hongyu Liu (Huawei Galois Lab, China, Purdue University, USA), Guanhong Tao (Purdue University, USA), Zhou Xuan (Purdue University, USA), and Xiangyu Zhang (Purdue University, USA)</i>	
Influence-Driven Data Poisoning in Graph-Based Semi-Supervised Classifiers .....	77
<i>Adriano Franci (SnT, Université de Luxembourg), Maxime Cordy (SnT, Université de Luxembourg), Martin Gubri (SnT, Université de Luxembourg), Michail Papadakis (SnT, Université de Luxembourg), and Yves Le Traon (SnT, Université de Luxembourg)</i>	
Engineering a Platform for Reinforcement Learning Workloads .....	88
<i>Ali Kanso (Microsoft, USA) and Kinshuman Patra (Microsoft, USA)</i>	
Method Cards for Prescriptive Machine-Learning Transparency .....	90
<i>David Adkins (Meta AI, USA), Bilal Alsallakh (Meta AI, USA), Adeel Cheema (Meta AI, USA), Narine Kokhlikyan (Meta AI, USA), Emily McReynolds (Meta AI, USA), Pushkar Mishra (Meta AI, UK), Chavez Procope (Meta AI, USA), Jeremy Sawruk (Meta AI, USA), Erin Wang (Meta AI, USA), and Polina Zvyagina (Meta AI, USA)</i>	

## Session 4: Keynote - Christopher Re

## Session 5: AI Engineering Practices

Towards a Roadmap on Software Engineering for Responsible AI .....	101
<i>Qinghua Lu (CSIRO, Australia), Liming Zhu (CSIRO, Australia), Xiwei Xu (CSIRO, Australia), Jon Whittle (CSIRO, Australia), and Zhenchang Xing (CSIRO, Australia)</i>	
AI Governance in the System Development Life Cycle: Insights on Responsible Machine Learning Engineering .....	113
<i>Samuli Laato (University of Turku, Finland), Teemu Birkstedt (University of Turku, Finland), Matti Mäntymäki (University of Turku, Finland), Matti Minkinen (University of Turku, Finland), and Tommi Mikkonen (University of Jyväskylä, Finland)</i>	
The Goldilocks Framework: Towards Selecting the Optimal Approach to Conducting AI Projects... 124	
<i>Rimma Dzhusupova (Mcdermott, The Netherlands), Jan Bosch (Chalmers University of Technology, Sweden), and Helena Holmström Olsson (Malmö University, Sweden)</i>	
What Is an AI Engineer? An Empirical Analysis of Job Ads in The Netherlands .....	136
<i>Marcel Meesters (Fontys University of Applied Science, The Netherlands), Petra Heck (Fontys University of Applied Science, The Netherlands), and Alexander Serebrenik (Eindhoven University of Technology, The Netherlands)</i>	
Data is About Detail - An Empirical Investigation for Software Systems with NLP at Core .....	145
<i>Anmol Singhal (TCS Research, India), Preethu Rose Anish (TCS Research, India), Pratik Sonar (TCS Research, India), and Smita S Ghaisas (TCS Research, India)</i>	

## Session 6: AI Models & Pipelines

Practical Insights of Repairing Model Problems on Image Classification .....	157
<i>Akihito Yoshii (Fujitsu Limited, Japan), Susumu Tokumoto (Fujitsu Limited, Japan), and Fuyuki Ishikawa (National Institute of Informatics, Japan)</i>	
UDAVA: An Unsupervised Learning Pipeline for Sensor Data Validation in Manufacturing .....	159
<i>Erik Johannes Husom (SINTEF Digital, Norway), Simeon Tverdal (SINTEF Digital, Norway), Arda Goknil (SINTEF Digital, Norway), and Sagar Sen (SINTEF Digital, Norway)</i>	
Black-Box Models for Non-Functional Properties of AI Software Systems .....	170
<i>Daniel Friesel (Universität Osnabrück, Germany) and Olaf Spinczyk (Universität Osnabrück, Germany)</i>	
Improving Generalizability of ML-Enabled Software Through Domain Specification .....	181
<i>Hamed Barzamini (Northern Illinois University, USA), Mona Rahimi (Northern Illinois University, USA), Murtuza Shahzad (Northern Illinois University, USA), and Hamed Alhoori (Northern Illinois University, USA)</i>	
Data Sovereignty for AI Pipelines: Lessons Learned from an Industrial Project at Mondragon Corporation .....	193
<i>Marcel Altendeitering (Fraunhofer ISST, Germany), Julia Pampus (Fraunhofer ISST, Germany), Felix Larrinaga (Mondragon Unibertsitatea, Spain), Jon Legaristi (Mondragon Unibertsitatea, Spain), and Falk Howar (TU Dortmund University, Germany)</i>	

## Session 7: Smells

Data Smells in Public Datasets .....	205
<i>Arumoy Shome (Delft University of Technology, Netherlands), Luis Cruz (Delft University of Technology, Netherlands), and Arie van Deursen (Delft University of Technology, Netherlands)</i>	
Code Smells for Machine Learning Applications .....	217
<i>Haiyin Zhang (ING Group, Netherlands), Luis Cruz (Delft University of Technology, Netherlands), and Arie van Deursen (Delft University of Technology, Netherlands)</i>	
Data Smells: Categories, Causes and Consequences, and Detection of Suspicious Data in AI-Based Systems .....	229
<i>Harald Foidl (University of Innsbruck, Austria), Michael Felderer (University of Innsbruck, Austria), and Rudolf Ramler (Software Competence Center Hagenberg GmbH, Austria)</i>	

## Session 8: Keynote - Saleema Amershi

Author Index .....	241
--------------------	-----