2024 IEEE/ACM 3rd International Conference on AI Engineering - Software Engineering for AI (CAIN 2024)

Lisbon, Portugal
14-15 April 2024

IEEE Catalog Number: CFP24AK2-POD
ISBN: 979-8-3503-5219-1

Copyright © 2024, 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: CFP24AK2-POD
ISBN (Print-On-Demand): 979-8-3503-5219-1
ISBN (Online): 979-8-4007-0591-5

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

2024 IEEE/ACM 3rd International Conference on AI Engineering – Software Engineering for AI (CAIN)

CAIN 2024

Table of Contents

Message from the CAIN 2024 Conference Chairs ..... **xi**
Organizing Committee ..... **xiii**
Program Committee ..... **xiv**
Steering Committee ..... **xvi**

Architecting, Designing, Managing, and Modeling AI-Enabled Systems

A Taxonomy of Foundation Model based Systems through the Lens of Software Architecture 1
Qinghua Lu (Data61, CSIRO), Liming Zhu (Data61, CSIRO), Xiwei Xu (Data61, CSIRO), Yue Liu (Data61, CSIRO), Zhenchang Xing (Data61, CSIRO), and Jon Whittle (Data61, CSIRO)

Investigating the Impact of SOLID Design Principles on Machine Learning Code Understanding... 7
Raphael Cabral (PUC-Rio, Brazil), Marcos Kalinowski (PUC-Rio, Brazil), Maria Teresa Baldassarre (University of Bari, Italy), Hugo Villamizar (PUC-Rio, Brazil), Tatiana Escovedo (PUC-Rio, Brazil), and Hélio Lopes (PUC-Rio, Brazil)

Identifying architectural design decisions for achieving green ML serving 18
Francisco Durán (Universitat Politècnica de Catalunya, Spain), Silverio Martínez-Fernández (Universitat Politècnica de Catalunya, Spain), Matias Martinez (Universitat Politècnica de Catalunya, Spain), and Patricia Lago (Vrije Universiteit Amsterdam, Netherlands)

Modeling Resilience of Collaborative AI Systems 24
Diaeddin Rimawi (Free University of Bozen-Bolzano), Antonio Liotta (Free University of Bozen-Bolzano), Marco Todescato (Fraunhofer Italia), and Barbara Russo (Free University of Bozen-Bolzano)

An Exploratory Study of V-Model in Building ML-Enabled Software: A Systems Engineering Perspective 30
Jie Wu (Independent)

Engineering Challenges in Industrial AI 41
Martin Hollender (ABB, Germany), Chaojun Xu (ABB, Germany), and Ruomu Tan (ABB, Germany)

Data Engineering and Management for AI-Enabled Systems

What About the Data? A Mapping Study on Data Engineering for AI Systems 43
Petra Heck (Fontys University of Applied Sciences)

Unmasking Data Secrets: An Empirical Investigation into Data Smells and Their Impact on Data Quality 53
Gilberto Recupito (University of Salerno), Raimondo Rapacciuolo (University of Salerno), Dario Di Nucci (University of Salerno), and Fabio Palomba (University of Salerno)

An Exploratory Study of Dataset and Model Management in Open Source Machine Learning Applications 64
Tajkia Rahman Toma (University of Alberta, Canada) and Cor-Paul Bezemer (University of Alberta, Canada)

DVC in Open Source ML-development: The Action and the Reaction 75
Lorena Barreto Simedo Pacheco (Concordia University, Canada), Musfiqur Rahman (Concordia University, Canada), Fazle Rabbi (Concordia University, Canada), Pouya Fathollahzadeh (Queen's University, Canada), Ahmad Abdellatif (University of Calgary, Canada), Emad Shihab (Concordia University, Canada), Tse-Hsun Chen (Concordia University, Canada), Jinqiu Yang (Concordia University, Canada), and Ying Zou (Queen's University, Canada)

Generative AI Engineering

Developer Experiences with a Contextualized AI Coding Assistant: Usability, Expectations, and Outcomes 81
Gustavo Pinto (Zup Innovation & UFPA), Cleidson de Souza (UFPA), Thayssa Rocha (Zup Innovation & UFPA), Igor Steinmacher (Nothern Arizona University), Alberto de Souza (Zup Innovation), and Edward Monteiro (StackSpot)

Privacy and Copyright Protection in Generative AI: A Lifecycle Perspective 92
Dawen Zhang (CSIRO's Data61, Australia), Boming Xia (CSIRO's Data61, Australia), Yue Liu (CSIRO's Data61, Australia), Xiwei Xu (CSIRO's Data61, Australia), Thong Hoang (CSIRO's Data61, Australia), Zhenchang Xing (CSIRO's Data61, Australia), Mark Staples (CSIRO's Data61, Australia), Qinghua Lu (CSIRO's Data61, Australia), and Liming Zhu (CSIRO's Data61, Australia)

Innovating Translation: Lessons Learned from BWX Generative Language Engine 98
Vanilson Buregio (Bureau Works, USA), Iverson Pereira (Bureau Works, USA), and Henrique Cabral (Bureau Works, USA)

Towards a Responsible AI Metrics Catalogue: A Collection of Metrics for AI Accountability 100
Boming Xia (CSIRO's Data61 & University of New South Wales), Qinghua Lu (CSIRO's Data61 & University of New South Wales), Liming Zhu (CSIRO's Data61 & University of New South Wales), Sung Une Lee (CSIRO's Data61), Yue Liu (CSIRO's Data61 & University of New South Wales), and Zhenchang Xing (CSIRO's Data61 & Australian National University)

Energy-Aware AI Engineering

Green Runner: A tool for efficient deep learning component selection 112
Jai Kannan (Applied Artificial Intelligence Institute), Scott Barnett (Applied Artificial Intelligence Institute), Luís Cruz (Delft University of Technology), Anj Simmons (Applied Artificial Intelligence Institute), and Taylan Selvi (Applied Artificial Intelligence Institute)

Engineering Carbon Emission-aware Machine Learning Pipelines 118
Erik Johannes Husom (SINTEF), Sagar Sen (SINTEF), and Arda Goknil (SINTEF)

The Impact of Knowledge Distillation on the Energy Consumption and Runtime Efficiency of NLP Models 129
Ye Yuan (Vrije Universiteit Amsterdam), Jingzhi Zhang (Vrije Universiteit Amsterdam), Zongyao Zhang (Vrije Universiteit Amsterdam), Kaiwei Chen (Vrije Universiteit Amsterdam), Jiacheng Shi (Vrije Universiteit Amsterdam), Vincenzo Stoico (Vrije Universiteit Amsterdam), and Ivano Malavolta (Vrije Universiteit Amsterdam)

Green AI: A Preliminary Empirical Study on Energy Consumption in DL Models Across Different Runtime Infrastructures 134
Negar Alizadeh (Utrecht University) and Fernando Castor (University of Twente)

LLMs and Testing

A Combinatorial Approach to Hyperparameter Optimization 140
Krishna Khadka (University of Texas at Arlington), Jaganmohan Chandrasekaran (National Security Institute, Virginia Tech), Yu Lei (University of Texas at Arlington), Raghu N. Kacker (Information Technology Laboratory, National Institute of Standards and Technology), and D. Richard Kuhn (Information Technology Laboratory, National Institute of Standards and Technology)

Mutation-based Consistency Testing for Evaluating the Code Understanding Capability of LLMs 150
Ziyu Li (University of Sheffield) and Donghwan Shin (University of Sheffield)

LLMs for Test Input Generation for Semantic Applications 160
Zafaryab Rasool (Deakin University), Scott Barnett (Deakin University), David Willie (Deakin University), Stefanus Kurniawan (Deakin University), Sherwin Balugo (Deakin University), Srikanth Thudumu (Deakin University), and Mohamed Abdelrazek (Deakin University)

(Why) Is My Prompt Getting Worse? Rethinking Regression Testing for Evolving LLM APIs 166
Wanqin Ma (The Hong Kong University of Science and Technology), Chenyang Yang (Carnegie Mellon University), and Christian Kästner (Carnegie Mellon University)

Welcome Your New AI Teammate: On Safety Analysis by Leashing Large Language Models 172
Ali Nouri (Volvo cars), Beatriz Cabrero-Daniel (University of Gothenburg), Fredrik Törner (Volvo Cars), Håkan Sivencrona (Zenseact), and Christian Berger (University of Gothenburg)

ML-On-Rails: Safeguarding Machine Learning Models in Software Systems – A Case Study 178
Hala Abdelkader (Deakin University, Australia), Mohamed Abdelrazek (Deakin University, Australia), Scott Barnett (Deakin University, Australia), Jean-Guy Schneider (Monash University, Australia), Priya Rani (RMIT University, Australia), and Rajesh Vasa (Deakin University, Australia)

System Qualities

Approach for Argumenting Safety on Basis of an Operational Design Domain 184
Gereon Weiss (Fraunhofer Institute for Cognitive Systems IKS, Germany), Marc Zeller (Siemens AG, Germany), Hannes Schoenhaar (Siemens AG, Germany), Christian Drabek (Fraunhofer Institute for Cognitive Systems IKS, Germany), and Andreas Kreutz (Fraunhofer Institute for Cognitive Systems IKS, Germany)

Seven Failure Points When Engineering a Retrieval Augmented Generation System 194
Scott Barnett (Deakin University), Stefanus Kurniawan (Deakin University), Srikanth Thudumu (Deakin University), Zach Brannelly (Deakin University), and Mohamed Abdelrazek (Deakin University)

POLARIS: A framework to guide the development of Trustworthy AI systems 200
Maria Teresa Baldassarre (University of Bari "A. Moro"), Domenico Gigante (Ser&Practices Srl), Marcos Kalinowski (Pontifical Catholic University of Rio de Janeiro (PUC-Rio)), and Azzurra Ragone (University of Bari "A. Moro")

Worst-Case Convergence Time of ML Algorithms via Extreme Value Theory 211
Saeid Tizpaz-Niari (University of Texas at El Paso) and Sriram Sankaranarayanan (University of Colorado Boulder)

Is Your Anomaly Detector Ready for Change? Adapting AIOps Solutions to the Real World 222
Lorena Poenaru-Olaru (Delft University of Technology), Natalia Karpova (Delft University of Technology), Luis Cruz (Delft University of Technology), Jan S. Rellermeyer (Leibniz University Hannover), and Arie van Deursen (Delft University of Technology)

Novel Contract-based Runtime Explainability Framework for End-to-End Ensemble Machine Learning Serving 234
Minh-Tri Nguyen (Aalto University, Finland), Hong-Linh Truong (Aalto University, Finland), and Tram Truong-Huu (Singapore Institute of Technology, Singapore)

Trustworthy AI: Industry-Guided Tooling of the Methods 245
Zakaria Chihani (Université Paris-Saclay, CEA, List, France)

Doctoral Symposium

Continuous Quality Assurance and ML Pipelines under the AI Act 247
Matthias Wagner (Lund University, Sweden)

Component-based Approach to Software Engineering of Machine Learning-enabled Systems . 250
Vladislav Indykov (University of Gothenburg | Chalmers)

Optimizing Data Analytics Workflows through User-driven Experimentation 253
Keerthiga Rajenthiram (Vrije Universiteit Amsterdam, The Netherlands)

Software Design Decisions for Greener Machine Learning-based Systems 256
Santiago del Rey (Universitat Politècnica de Catalunya)

Energy-Efficient Development of ML-Enabled Systems: A Data-Centric Approach 259
Rafiullah Omar (UNIVAQ)

Threat Modeling of ML-intensive Systems: Research Proposal 264
Felix Viktor Jedrzejewski (Blekinge Institute of Technology)

Posters

A Domain Specific Language for the Specification of Risk-oriented Object Detection Requirements 267
Junji Hashimoto (GREE, Inc.) and Nobukazu Yoshioka (Waseda University)

AI Security Continuum: Concept and Challenges 269
Hironori Washizaki (Waseda University) and Nobukazu Yoshioka (Waseda University)

A Roadmap for Enriching Jupyter Notebooks Documentation with Kaggle Data 271
Mojtaba Mostafavi Ghahfarokhi (Sharif University of Technology, Iran), Hamed Jahantigh (Sharif University of Technology, Iran), Alireza Asadi (Sharif University of Technology, Iran), Sepehr Kianiangolafshani (Sharif University of Technology, Iran), Ashkan Khademian (Sharif University of Technology, Iran), and Abbas Heydarnoori (Sharif University of Technology, Iran)

Automating Patch Set Generation from Code Reviews Using Large Language Models 273
Md Tajmilur Rahman (Gannon University, USA), Rahul Singh (Gannon University, USA), and Mir Yousuf Sultan (Gannon University, USA)

Data Selection Driven by Item Difficulty: On Investigating Data Efficient Practice for Hyperparameter Search 275
Gustavo Rodrigues dos Reis (NAVER LABS Europe/Univ. Grenoble Alpes), Adrian Mos (NAVER LABS Europe), Mario Cortes-Cornax (Univ. Grenoble Alpes, CNRS, Inria, Grenoble INP, LIG), and Cyril Labbé (Univ. Grenoble Alpes, CNRS, Inria, Grenoble INP, LIG)

Beyond Syntax: Unleashing the Power of Computational Notebooks Code Metrics in Documentation Generation 278
Mojtaba Mostafavi Ghahfarokhi (Sharif University of Technology, Iran), Ashkan Khademian (Sharif University of Technology, Iran), Sepehr Kianiangolafshani (Sharif University of Technology, Iran), Alireza Asadi (Sharif University of Technology, Iran), Hamed Jahantigh (Sharif University of Technology, Iran), and Abbas Heydarnoori (Sharif University of Technology, Iran)

Can causality accelerate experimentation in software systems? 280
Andrei Paleyes (University of Cambridge, UK), Han-Bo Li (University of Cambridge, UK), and Neil D. Lawrence (University of Cambridge, UK)

Custom Developer GPT for Ethical AI Solutions 282
Lauren Olson (VU Amsterdam)

Evaluation of The Generality of Multi-view Modeling Framework for ML Systems 284
Jati H. Husen (Waseda University), Jomphon Runpakprakun (Waseda University), Sun Chang (Waseda University), Hironori Washizaki (Waseda University), Hnin Thandar Tun (Waseda University), Nobukazu Yoshioka (Waseda University), and Yoshiaki Fukazawa (Waseda University)

Prompt Smells: An Omen for Undesirable Generative AI Outputs 286
Krishna Ronanki (University of Gothenburg), Beatriz Cabrero-Daniel (University of Gothenburg), and Christian Berger (University of Gothenburg)

Taxonomy of Generative AI Applications for Risk Assessment 288
Hiroshi Tanaka (Fujitsu Limited, Japan), Masaru Ide (Fujitsu Limited, Japan), Jun Yajima (Fujitsu Limited, Japan), Sachiko Onodera (Fujitsu Limited, Japan), Kazuki Munakata (Fujitsu Limited, Japan), and Nobukazu Yoshioka (Waseda University, Japan)

Author Index **291**