The Workshop of Observability and Explainability of AI Systems

Yan Liu
Department of Electrical and Computer Engineering
Concordia University
Montréal, Québec, Canada
yan.liu@concordia.ca

Abdelwahab Hamou-Lhadj
Department of Electrical and Computer Engineering
Concordia University
Montréal, Québec, Canada
wahab.hamou-lhadj@concordia.ca

***Abstract*—This workshop aims to bring together researchers, professionals, and academics to discuss and advance the state of the art and best practices in observability and explainability of AI systems. An explainable AI system allows users and stakeholders to understand the automated decision-making process for ensuring trust, and compliance with regulations and ethical standards. Observability, which involves the ability to measure, monitor, and understand the internal states and behaviors of AI systems. The combination of observability and explainability contributes to form the data centric base for responsible AI development. This workshop invites contributions that explore the intersection of observability and explainability in AI, focusing on methods, frameworks, and practices that enhance our understanding and development of AI-driven systems.**

***Index Terms*—AI systems, observability, explainability, system design, software development**

I. Workshop Relevance

The rapid advancements in complex neural networks enables the ability of AI system processing and integrating different modalities of data. Inevitably, the adoption of AI systems grows in a wide domain of the society. Meanwhile, the constant growth of neural networks' scale and complexity also leads to high barrier of cost to develop the AI systems from the original training. The practices are adopting cloud-based AI services or adopting open source AI models to deploy as services. Retrieving domain related data and augmenting the prompts for foundation models become data driven in different scales. Within such a context, observability provides the trace of data usage within models. Explainability refers to the ability to provide understandable and interpretable explanations of how AI systems make decisions. The combination of observability and explainability allows users and stakeholders to understand how input data influences the automated decisions. This transparency is essential for building trust in AI systems, especially in critical applications, such as healthcare, finance, cybersecurity, legal system and government services. Developing both capabilities is beyond the scope of foundation of AI models that further involves system design for services and cloud computing, software engineering and data management. The topics covered in this workshop are supplementary to CASCON's main program. In particular, this workshop offers opportunities for researchers, practitioners, and academics to share insights, discoveries, and best practices that connect observability and explainability of AI systems.

II. Workshop Topics

We invite workshop talks on a focused topics related to observability and explainability in AI systems, including but not limited to:

- **Observability in AI Systems**: Techniques for monitoring and logging AI behaviors, tools for real-time analysis of AI systems, metrics for assessing system performance, and frameworks for AI system observability.
- **Explainability in AI**: Methods for generating interpretable models, techniques for explaining AI decisions, human-centered approaches to AI explainability, and evaluation metrics for explainability.
- **Integrating Observability and Explainability**: Approaches to combining observability and explainability to enhance AI system transparency, methods for aligning explainability with observability metrics, and frameworks for AI system management.
- **Ethical and Social Implications**: The role of observability and explainability in ensuring ethical AI, regulatory requirements, and impacts of opaque AI systems.
- **Case Studies and Applications**: Real-world examples of observability and explainability in AI systems, challenges encountered, and lessons learned in deploying observable and explainable AI.

III. Workshop Audience

The target audience for this workshop comprises a group of academic scholars and industry professionals in the AI field. This includes:

- **Data Scientists**: Individuals who specialize in analyzing and interpreting complex data to help organizations make informed decisions. They typically have strong backgrounds in statistics, programming, and machine learning.
- **Machine Learning Engineers**: Professionals who design, build, and deploy machine learning models. They possess deep knowledge of algorithms, data structures, and software engineering principles.
- **System Developer**: Professionals who design, build and deploy AI-as-a-Service. They have technical background

in delivering scalable service oriented and cloud based systems.
- **AI Ethics Researchers**: Scholars and practitioners focused on the ethical implications of AI technologies. They often have interdisciplinary expertise, such as philosophy, law, and social sciences.

The participants are expected to have varying levels of familiarity with observability and explainability in AI:

- **Beginners**: Some attendees may be new to the concepts of observability and explainability, seeking a knowledge base and practical guidelines.
- **Intermediate**: The major portion of the audience will have a working experience of these concepts and are looking to deepen their knowledge and apply best practices in their work.
- **Advanced**: Experienced professionals and researchers who are already implementing observability and/or explainability techniques to a certain extend. They are interested in exploring advanced methodologies.

IV. Workshop Organization

This workshop is a full-day workshop with invited talks from academic researchers, domain experts and industry practitioners in the field of healthcare, finance, telecommunication, and digital twins through the collaboration network of the organizers. We will also invite PhD and Postdoc researchers to join a session for graduate research. We will organize a panel session to discuss in-depth the challenges, and directions and collaboration opportunities in multidisciplinary research. Example relevant questions are as follows:

- How can we design AI systems that provide near real-time observability into their decision-making processes?
- What are the best practices for explaining complex AI models to non-expert users?
- How can observability and explainability be balanced with performance and efficiency in AI systems?
- What role do observability and explainability play in the ethical deployment of AI in sensitive domains such as healthcare and finance?

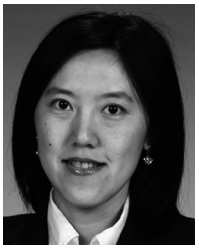
We plan to have five to six scholar and industry talks and another three to five students' research presentation.

V. Workshop Organizers' Background

The organizers have established track record of organizing events. Dr. Hamou-Lhadj was general chair of The System Analysis and Modeling Conference (SAM'21), co-located with MODELS, co-chair of ICSE'19 Posters Track, co-chair of ICPC'19 Negative Results Track, co-chair of SANER'15 Industry Track, co-chair of MODELS'15 Workshop Track, founder and co-chair of the International Workshop on Software Faults (IWSF), co-located with ISSRE for seven years, founder and co-chair of the InternationalWorkshop on Program Comprehension for Dynamic Analysis (PCODA), co-located with WCRE (now SANER). PCODA run for seven years (2005-2010, 2014). Hamou-Lhadj served on many program committees of major conferences including SANER, ICSME, MODELS, ICPC, WCRE among others.

Dr. Liu was the Track Chair of early career researchers Forum IEEE International Conference on Software Architecture (ICSA), 2022. Chair of Canada Consortium for Software Engineering Research (CSER), Fall, 2018. Workshop Chair of International Conference of Software Architecture (ICSA), 2018. Tools Demo Chair of European Conference on Software Architecture (ECSA) 2016. Finance Chair of Working IEEE / IFIP Conference on Software (WICSA) 2015 & ACM SIGSOFT Symposium on Component-Based Software Engineering (CBSE) 2015 & ACM SIGSOFT Conference on the Quality of Software Architectures (QoSA) 2015. Yan has regularly served on program committees of conferences and workshops. She is now the associate editor of IEEE Transactions on Service Computing.

Both organizers have been highly active researchers in the fields of AI system observability and explainability. The organizers edited a special issue on the related topics to IEEE Software published in January 2024. Other publication venues include ICSE, IEEE Software, IEEE Transactions of Cloud Computing, IEEE Transactions on Software Engineering, Journal of Software and Systems, Journal of Expert Systems with Applications, IEEE Big Data, IEEE Service Engineering to name a few.



Dr. Yan Liu is a Full Professor and Gina Cody Research and Innovation Fellow at Concordia University (2023-2024). Before the faculty position, Yan worked as a Senior Research Scientist at the National ICT Australia (NICTA) laboratory and US Department of Energy Pacific Northwest National Laboratory with ten years of experience with large-scale software systems. As a tenured faculty, Yan's research is generously funded by NSERC Discovery Grants, Quebec FRQNT New Research Award, and MITACS and industry collaborators in the domains of telecommunication, health care, senor networks, NLP for public services, cloud game servers, and digitization of building architecture design. Yan has two US patents granted. Her recent research is on the foundation of explainable AI and service computing for XAI-as-a-Service. Contact her at yan.liu@concordia.ca

Dr. Abdelwahab Hamou-Lhadj is a Professor and a Gina Cody Research and Innovation Fellow at Concordia University, Montreal, Canada. He is also an Affiliate Researcher at NASA JPL, Caltech. His research interests are in AIOps, Software Observability and Monitoring, Software Engineering, and Model-driven Engineering. He has been the principal investigator for several projects with various companies. Many tools that were developed in his lab were successfully transferred to the industry, and are currently used by thousands of developers. Dr. Hamou-Lhadj published several papers and gave numerous talks on the topics of observability of software systems and AIOps. He served on the organization and program committees of major software engineering conferences such as ICSE, MODELS, SANER, ICPC, ICSME, and ICEIS. He is currently an Associate Editor of IEEE Transactions on Reliability. Dr. Hamou-Lhadj is a Senior Member of IEEE, and a long-lasting member of ACM.