Develop Explainable AI Services on Cloud Computing and Open Source Models

Zerui Wang
Department of Electrical and Computer Engineering
Concordia University
Montréal, Québec, Canada
zerui.wang@concordia.ca

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

***Abstract*—The objective of this tutorial is to introduce systematic methods with supplementary tools and computing models for developing explainable AI. As AI increasingly penetrates various application domains, practitioners need more comprehensive approaches for model quality assessments and explainability. This tutorial bridges the gap between theoretical XAI techniques and their practical implementation in cloud services and open-source models. We explore a range of XAI methodologies, from model-agnostic to model-specific approaches, examining their applicability across diverse AI models and deployment scenarios. The tutorial introduces design principles for integrating XAI into AI services and facilitating early adoption in MLOps lifecycles. Through experimental evaluations and case studies on both cloud AI services and open-source models, we demonstrate the effectiveness of these approaches in real-world scenarios. We discuss strategies for pipeline automation, provenance data management, and optimizing the operational overhead of XAI integration. Additionally, we explore the robustness of XAI methods under adversarial conditions. This comprehensive exploration equips participants with insights into XAI's role in AI systems, practical implementation skills, and critical perspectives on future research directions, ultimately benefiting the development of explainable and robust AI systems.**

***Index Terms*—explainable AI, cloud computing, open source model, software quality, robustness**

I. Tutorial Goals and Relevance

As AI increasingly influences software systems, the demand for explainability has become essential, particularly in sensitive domains [1]. We address this challenge by providing an automatic solution for implementing and evaluating XAI methods across diverse AI models and services [2]–[4].

The goals of this tutorial are:

1) To provide an understanding of state-of-the-art XAI methods and their theoretical foundations.
2) To demonstrate practical implementation of XAI techniques for both cloud services and open-source models.
3) To explore strategies for integrating XAI into the MLOps lifecycle and assess its impact on model development.
4) To analyze the computational costs and performance implications of XAI in production environments.
5) To investigate the robustness of XAI methods under adversarial conditions.

The tutorial covers a range of XAI techniques, including model-agnostic and model-specific explanation methods [5]. We review the mathematical foundations of these methods, discussing their strengths, limitations, and appropriate use cases. Particular attention is given to the challenges of applying these techniques to black-box cloud AI services versus transparent, open-source models.

The tutorial is dedicated to the practical aspects of XAI implementation. Participants gain hands-on experience in applying XAI methods to real-world AI systems. The tutorial also includes the evaluation of XAI methods. We introduce and discuss various quality attributes for assessing the model with explanations. An important aspect of the quality attributes is the examination of model robustness and XAI resilience under adversarial conditions. We provide insights into developing more resilient XAI.

The relevance of this tutorial to the CASCON community is multifaceted. For researchers, it offers insights into the recent XAI techniques and their theoretical underpinnings, potentially inspiring new research directions in AI transparency. Practitioners gain knowledge for implementing XAI in real-world systems, enhancing their ability to develop more trustworthy AI solutions. The tutorial's focus on both cloud services and open-source models makes it particularly valuable for attendees working across different AI systems.

In conclusion, this tutorial aims to bridge the gap between theoretical XAI research and their practical implementation, addressing the XAI need in the AI community.

II. Tutorial Structure

Here, we introduce the segments of the half-day tutorial.

A. Segment 1: Theoretical Foundations (45 minutes)

1) Introduction to XAI: XAI methods response to the increasing complexity and opacity of AI systems [1]. As detailed in study [2], the AI service, particularly through cloud platforms, has intensified the need for explainable. XAI aims to make AI decisions by providing explanations for feature contributions.

2) AI Quality Requirements and the Role of XAI: Our study has uncovered inconsistencies and risks in cloud AI services. XAI addresses quality concerns in cloud AI services by providing explanations and evaluations.

3) Challenges of XAI: There is operational complexity in adapting XAI techniques to diverse model architectures. Implementing XAI faces several significant challenges.

4) XAI Tools and Frameworks: Our research provides a comprehensive comparison of existing XAI frameworks including IBM's Explainability 360 [6], Microsoft's InterpretML [7], Vertex XAI [8], and OmniXAI [9]. We analyze their strengths, limitations, and suitability for different platforms. We offer integration with both cloud AI services and open-source models, support parallel XAI pipelines, and incorporate other XAI frameworks as well as evaluation metrics for assessing explanation.

B. Segment 2: XAI Service Operation (45 minutes)

1) Essential Components: We introduces a service-oriented approach to XAI. The core components, as outlined in our methodology section, include:

- Data Processing Microservice: Ensures correct data formatting and applies adversarial attack conditions.
- AI Model Microservice: Encapsulates and deploys pre-trained AI models.
- XAI Method Microservice: Offers various XAI algorithms for generating explanations.
- Evaluation Microservice: Aggregates results and evaluates defined quality attributes.

These components work in concert to provide a comprehensive XAI solution adaptable to various AI deployment scenarios.

2) Parallel XAI Pipelines: Our framework supports parallel XAI pipelines to address XAI's computational challenges. This allows for efficient processing of multiple XAI tasks simultaneously, significantly reducing overall computation time for complex AI systems.

3) Deployment and SDK: We offers a streamlined deployment process and a comprehensive SDK. The SDK provides a high-level interface for configuring and executing XAI tasks, making it accessible to both researchers and practitioners. Our deployment strategy leverages containerization for easy scaling and cross-platform compatibility.

C. Segment 3: XAI Application Scenarios and Future Directions (45 minutes)

1) XAI Application Scenarios: This segment explores practical applications of XAI across various domains.

a) Cloud AI Service Evaluation and Discoveries: This work integrates XAI with Azure AI services [10]. Our findings demonstrate that XAI techniques can uncover inconsistencies in model predictions across different cloud platforms.

b) Adversary Perturbation for Open Model AI Services: We conduct comprehensive experiments apply various adversarial perturbations [11] at different severity levels. We observe that certain models, such as Vision Transformers [12], exhibited higher robustness.

2) Future Directions - XAI 2.0: We propose a roadmap for XAI's future, aligned with XAI 2.0 principles: Extending XAI to new AI types, including cloud services and complex models like Vision Transformers. Introducing metrics for explanation consistency and stability across models and datasets. Enabling flexible adjustment of models and XAI methods in microservices. This simplified version reduces the word count while maintaining the key points of the future directions for XAI.

REFERENCES

[1] D. Gunning and D. Aha, "Darpa's explainable artificial intelligence (xai) program," *AI magazine*, vol. 40, no. 2, pp. 44–58, 2019.

[2] Z. Wang, Y. Liu, and J. Huang, "An open api architecture to discover the trustworthy explanation of cloud ai services," *IEEE Transactions on Cloud Computing*, 2024.

[3] Z. Wang, Y. Liu, A. Arumugam Thiruselvi, and A. Hamou-Lhadj, "Xaiport: A service framework for the early adoption of xai in ai model development," in *Proceedings of the 2024 ACM/IEEE 44th International Conference on Software Engineering: New Ideas and Emerging Results*, 2024, pp. 67–71.

[4] Z. Wang and Y. Liu, "Cloud-based xai services for assessing open repository models under adversarial attacks," *IEEE SSE24*, 2024.

[5] J. Huang, Z. Wang, D. Li, and Y. Liu, "The analysis and development of an xai process on feature contribution explanation," in *2022 IEEE International Conference on Big Data (Big Data)*, 2022, pp. 5039–5048.

[6] V. Arya, R. K. Bellamy, P.-Y. Chen, A. Dhurandhar, M. Hind, S. C. Hoffman, S. Houde, Q. V. Liao, R. Luss, A. Mojsilović *et al.*, "One explanation does not fit all: A toolkit and taxonomy of ai explainability techniques," *arXiv preprint arXiv:1909.03012*, 2019.

[7] Microsoft. (2023) Interpret ml models using azure machine learning. Accessed: October 3, 2023. [Online]. Available: https://learn.microsoft.com/en-us/azure/machine-learning/how-to-machine-learning-interpretability?view=azureml-api-2

[8] Google. (2023) Google cloud platform. Accessed: 2023. [Online]. Available: https://cloud.google.com

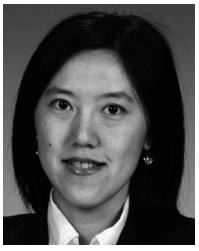
[9] W. Yang, H. Le, S. Savarese, and S. C. Hoi, "Omnixai: A library for explainable ai," *arXiv preprint arXiv:2206.01612*, 2022.

[10] Microsoft. (2023) Azure cloud services. Accessed: 2023. [Online]. Available: https://azure.microsoft.com

[11] D. Hendrycks and T. Dietterich, "Benchmarking neural network robustness to common corruptions and perturbations," *arXiv preprint arXiv:1903.12261*, 2019.

[12] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly *et al.*, "An image is worth 16x16 words: Transformers for image recognition at scale," *arXiv preprint arXiv:2010.11929*, 2020.

Zerui Wang is a Ph.D. candidate in the Department of Electrical and Computer Engineering at Concordia University. His research interests include Explainable AI (XAI), cloud AI services, and applied AI. Zerui's work has been published in IEEE Transactions on Cloud Computing and international conferences such as ICSE. He focuses on enhancing explainable AI in cloud computing environments. Contact him at zerui.wang@concordia.ca.



Dr. Yan Liu is a Full Professor and Gina Cody Research and Innovation Fellow at Concordia University. 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 work is defining an evaluation framework for explanation consistency. Contact her at yan.liu@concordia.ca