Proceedings of the 55th Hawaii International Conference on System Sciences | 2022

Introduction to the Minitrack on Explainable Artificial Intelligence (XAI)

Christian Meske
Ruhr-Universität Bochum
christian.meske@rub.de

Babak Abedin
Macquarie University
babak.abedin@mq.edu.au

Mathias Klier
University of Ulm
mathias.klier@uni-ulm.de

Fethi Rabhi
UNSW Australia
f.rabhi@unsw.edu.au

Abstract

Artificial Intelligence (AI) already reached or even surpassed the task performance of humans in various domains. However, most AI systems are still "black boxes" that are difficult to comprehend – not only for developers, but also for users and decision-makers. Explainable Artificial Intelligence (XAI) aims at intelligent systems that are highly performant and enable their users to understand, appropriately trust, and scrutinize them.

1. Introduction

Today, Artificial Intelligence (AI) already reached or even surpassed the task performance of humans in various domains. However, most AI systems are still "black boxes" that are difficult to comprehend – not only for developers, but also for users and decision-makers [1]. In addition, the development and use of AI is associated with many risks and pitfalls like biases in data or predictions based on spurious correlations ("Clever Hans" phenomena), which eventually may lead to malfunctioning or biased AI and hence technologically driven discrimination.

This is where research on Explainable AI (XAI) comes in. XAI is aimed at AI systems that are highly performant and enable their users to understand, appropriately trust, and scrutinize them [2]. Hence, XAI refers to "the movement, initiatives, and efforts made in response to AI transparency and trust concerns, more than to a formal technical concept" [3, p. 52140]. Often, XAI is designed in a user-centric manner so that users are empowered to scrutinize AI [4]. In this spirit, XAI is used to evaluate, to improve, to learn from, and to justify AI, in order to eventually be able to manage AI [1, 5].

With a focus on decision support, this minitrack builds on the last year's minitrack [6] and explores and extends research on how to establish explainability of intelligent black box systems – machine learning-based or not. We received many high-quality submissions, of which we were able to accept five.

2. Papers

In the first paper, "An Interpretable Deep Learning Approach to Understand Health Misinformation Transmission on YouTube", the authors propose a novel interpretable deep learning, Generative Adversarial Network based Piecewise Wide and Attention Deep Learning (GAN-PiWAD) to predict health misinformation transmission in social media. The artifact captures the interactions among multi-modal data, offers unbiased estimation of the total effect of each feature, and models the dynamic total effect of each feature. Empirical evaluations indicate that GAN-PiWAD outperforms strong baseline models.

The second paper, "Visual Interpretability of Image-based Real Estate Appraisal", examines visual interpretability methods in real estate appraisal. It applies a two-stage modeling approach combining Regression Activation Maps (RAM) for a Convolutional Neural Network (CNN) and a linear regression for the overall prediction. Results of experiments based on 62.000 family homes in Philadelphia suggest that the CNN learns aspects related to vegetation and quality aspects of the house from exterior images, thereby improving the predictive accuracy of real estate appraisal.

The third paper, "Intelligent Decision Assistance Versus Automated Decision-Making: Enhancing Knowledge Work Through Explainable Artificial Intelligence", is based on a literature review of the two research streams Decision Support Systems (DSS) and automation. The authors conceptualize a new class of DSS, namely Intelligent Decision Assistance (IDA), to support knowledge workers without influencing them through automated decision-making. A first validation on the impacts of IDA is provided based on empirical studies in the literature.

URI: https://hdl.handle.net/10125/79514
978-0-9981331-5-7

In the fourth paper, "Validation of AI-based Information Systems for Sensitive Use Cases: Using an XAI Approach in Pharmaceutical Engineering", the authors develop an XAI based validation approach for AI in sensitive use cases that facilitates the understanding of the system's behavior. Interviews in a case study in pharmaceutical manufacturing reveal that the approach is suitable to collect the required evidence for a software validation while requiring additional efforts compared to a traditional software validation.

In the fifth and last paper, "Detecting and Understanding Textual Deep Fakes in Online Reviews", the authors propose a novel artifact capable of detecting textual deep fakes to then explain their peculiarities by using XAI. They conduct an evaluation for the case of an online review data set and find that XAI enables further investigation to develop a better understanding of the algorithm's decisions which turns out to be particularly useful for unclear predictions.

3. References

[1] C. Meske, E. Bunde, J. Schneider, and M. Gersch (2021): Explainable Artificial Intelligence: Objectives, Stakeholders and Future Research Opportunities. Information Systems Management, pp. 1-11.

[2] A. Abdul, J. Vermeulen, D. Wang, B.Y. Lim, and M. Kankanhalli (2018): Trends and Trajectories for Explainable, Accountable and Intelligible Systems: An HCI Research Agenda. Proceedings of the CHI Conference on Human Factors in Computing Systems 2018.

[3] A. Adadi and M. Berrada (2018): Peeking Inside the Black-Box: A Survey on Explainable Artificial Intelligence (XAI). IEEE Access, 6, pp. 52138-52160.

[4] M. Förster, M. Klier, K. Kluge, and I. Sigler (2020): Fostering Human Agency: A Process for the Design of User-Centric XAI Systems. Proceedings of the International Conference on Information Systems 2020.

[5] B. Abedin (2021): Managing the tension between opposing effects of explainability of artificial intelligence: a contingency theory perspective. Internet Research. https://doi.org/10.1108/INTR-05-2020-0300

[6] C. Meske, B. Abedin, I. Junglas and F. Rabhi (2021). Introduction to the Minitrack on Explainable Artificial Intelligence (XAI). Proceedings of the 54th Hawaii International Conference on System Sciences (p. 1262).