XAI.it 2023 - Preface to the Fourth Italian Workshop on Explainable Artificial Intelligence

Cataldo Musto[1], Riccardo Guidotti[2], Anna Monreale[2], Erasmo Purificato[3] and Giovanni Semeraro[1]

[1]*Department of Computer Science, University of Bari Aldo Moro, Bari, Italy*
[2]*Department of Computer Science, University of Pisa, Pisa, Italy*
[3]*Faculty of Computer Science, Otto von Guericke University Magdeburg, Germany*

Abstract

Artificial Intelligence systems are increasingly playing an increasingly important role in our daily lives. As their importance in our everyday lives grows, it is fundamental that the internal mechanisms that guide these algorithms are as clear as possible. It is not by chance that the recent General Data Protection Regulation (GDPR) emphasized the users' right to explanation when people face artificial intelligence-based technologies. Unfortunately, the current research tends to go in the opposite direction, since most of the approaches try to maximize the effectiveness of the models (e.g., recommendation accuracy) at the expense of the explainability and the transparency. The main research questions which arise from this scenario is straightforward: *how can we deal with such a dichotomy between the need for effective adaptive systems and the right to transparency and interpretability?* Several research lines are triggered by this question: building transparent intelligent systems, analyzing the impact of opaque algorithms on final users, studying the role of explanation strategies, investigating how to provide users with more control in the behavior of intelligent systems. XAI.it, the Italian workshop on Explainable AI, tries to address these research lines and aims to provide a forum for the Italian community to discuss problems, challenges and innovative approaches in the various sub-fields of XAI.

1. Background and Motivations

Nowadays we are witnessing a new summer of Artificial Intelligence, since the AI-based algorithms are being adopting in a growing number of contexts and applications domains, ranging from media and entertainment to medical, finance and legal decision-making. While the very first AI systems were easily interpretable, the current trend showed the rise of opaque methodologies such as those based on Deep Neural Networks (DNN), whose (very good) effectiveness is contrasted by the enormous complexity of the models, which is due to the huge number of layers and parameters that characterize these models.

As intelligent systems become more and more widely applied (especially in very "sensitive" domain), it is not possible to adopt opaque or inscrutable black-box models or to ignore the general rationale that guides the algorithms in the task it carries on Moreover, the metrics that are usually adopted to evaluate the effectiveness of the algorithms reward very opaque methodologies that maximize the accuracy of the model at the expense of the transparency and explainability.

This issue is even more felt in the light of the recent experiences, such as the General Data Protection Regulation (GDPR) and DARPA's Explainable AI Project, which further emphasized the need and the right for scrutable and transparent methodologies that can guide the user in a complete comprehension of the information held and managed by AI-based systems.

Accordingly, the main motivation of the workshop is simple and straightforward: *how can we deal with such a dichotomy between the need for effective intelligent systems and the right to transparency and interpretability?*

These questions trigger several lines, that are particularly relevant for the current research in AI. The workshop tries to address these research lines and aims to provide a forum for the Italian community to discuss problems, challenges and innovative approaches in the area.

2. Accepted Papers

We believe that the program provides a good balance between the different topics related to the area of Explainable AI. Moreover, the program will be further enriched through a keynote given by Giovanni Stilo from University of L'Aquila.

The accepted papers range from the definition of new methodologies to explain the behavior of artificial intelligence systems to the development of new applications implementing the principles of Explainable AI. In total, 8 contributions were accepted at XAI.it 2023 (7 of them included in the proceedings):

1. Caterina Borzillo, Alessio Ragno and Roberto Capobianco - *Understanding Deep RL agent decisions: a novel interpretable approach with trainable prototypes*
2. Simona Colucci, Tommaso Di Noia, Francesco M. Donini, Claudio Pomo and Eugenio Di Sciascio - *Irrelevant Explanations: a logical formalization and a case study*
3. Soumick Chatterjee, Arnab Das, Rupali Khatun and Andreas Nürnberger - *Unboxing the black-box of deep learning based reconstruction of undersampled MRIs*
4. Andrea Tocchetti, Jie Yang and Marco Brambilla - *Rationale Trees: Towards a Formalization of Human Knowledge for Explainable Natural Language Processing*
5. Andrea Colombo, Laura Fiorenza and Sofia Mongardi - A *Flexible Metric-Based Approach to Assess Neural Network Interpretability in Image Classification*
6. Andrea Apicella, Salvatore Giugliano, Francesco Isgro and Roberto Prevete - *SHAP-based explanations to improve classification systems*
7. Muhammad Suffian, Ilia Stepin, Jose Maria Alonso-Moral and Alessandro Bogliolo - *Investigating Human-Centered Perspectives in Explainable Artificial Intelligence*

3. Program Committee

As a final remark, the program co-chairs would like to thank all the members of the Program Committee (listed below), as well as the organizers of the AIxIA 2023 Conference[1].

- Davide Bacciu, *Università di Pisa*
- Valerio Basile, *Università di Torino*
- Ludovico Boratto, *Università di Cagliari*
- Roberta Calegari, *Università di Bologna*
- Roberto Capobianco, *Università di Roma La Sapienza*
- Federica Cena, *Università di Trento*
- Roberto Confalonieri, *Libera Università di Bozen-Bolzano*
- Rodolfo Delmonte, *Università Ca' Foscari*
- Alessandro Giuliani, *Università di Cagliari*
- Kyriaki Kalimeri, *ISI Foundation*
- Antonio Lieto, *Università di Torino*
- Francesca Alessandra Lisi, *Università di Bari*
- Mirko Marras, *Università di Cagliari*
- Ruggero Pensa, *Università di Torino*
- Claudio Pomo, *Politecnico di Bari*
- Roberto Prevete, *Università di Naples Federico II*
- Antonio Rago, *Imperial College London*
- Amon Rapp, *Università di Torino*
- Salvatore Ruggieri, *Università di Pisa*
- Giuseppe Sansonetti, *Roma Tre University*
- Mattia Setzu, *Università di Pisa*
- Fabrizio Silvestri, *Università di Roma La Sapienza*
- Gabriele Tolomei, *Università di Roma La Sapienza*

---

[1]http://www.aixia2023.cnr.it/