XAI.it 2020 - Preface to the First Italian Workshop on Explainable Artificial Intelligence

Cataldo Musto[1], Daniele Magazzeni[2]
, Salvatore Ruggieri[3], and Giovanni Semeraro[1]

[1] Università degli Studi di Bari 'Aldo Moro', Italy
{cataldo.musto,giovanni.semeraro}@uniba.it
[2] JP Morgan AI Research, London, United Kingdom
daniele.magazzeni@jpmorgan.com
[3] University of Pisa, Italy
salvatore.ruggieri@unipi.it

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 first 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 Dino Pedreschi from University of Pisa.

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, 14 contributions were accepted at XAI.it 2020:

1. Nazanin Fouladgar, Marjan Alirezaie and Kary Främling - *Decision Explanation: Applying Contextual Importance and Contextual Utility in Affect Detection*
2. Luca Capone and Marta Bertolaso - *A Philosophical Approach for a Human-centered Explainable AI*
3. Matteo Baldoni, Cristina Baroglio, Roberto Micalizio and Stefano Tedeschi - *Is Explanation the Real Key Factor for Innovation?*
4. Federico Maria Cau, Lucio Davide Spano and Nava Tintarev - *Considerations for Applying Logical Reasoning to Explain Neural Network Outputs*
5. Bruno Apolloni and Ernesto Damiani - *Learning simplified functions to understand*
6. Pierangela Bruno, Cinzia Marte and Francesco Calimeri - *Understanding Automatic COVID-19 Classification using Chest X-ray images*

7. Ivan Donadello and Mauro Dragoni - *SeXAI: Introducing Concepts into Black Boxes for Explainable Artificial Intelligence*
8. Luca Marconi, Ricardo Anibal Matamoros Aragon, Italo Zoppis, Sara Manzoni, Giancarlo Mauri and Francesco Epifania - *Approaching Explainable Recommendations for Personalized Social Learning: the current stage in the educational platform "WhoTeach"*
9. Roberta Calegari, Andrea Omicini and Giovanni Sartor - *Argumentation and Logic Programming for Explainable and Ethical AI*
10. Stefania Costantini and Valentina Pitoni - *Towards a Logic of "Inferable" for Self-Aware Transparent Logical Agents*
11. Fabio Massimo Zanzotto, Dario Onorati, Pierfrancesco Tommasino, Andrea Santilli, Leonardo Ranaldi and Francesca Fallucchi - *Pat-in-the-loop: Syntax-based Neural Networks with Activation Visualization and Declarative Control*
12. Francesco Craighero, Alex Graudenzi, Fabrizio Angaroni, Fabio Stella and Marco Antoniotti - *Understanding Deep Learning with Activation Pattern Diagrams*
13. Laura Giordano, Daniele Theseider Dupré and Valentina Gliozzi - *Towards a Conditional interpretation of Self Organizing Maps*
14. Carmelo Ardito, Yashar Deldjoo, Eugenio Di Sciascio and Fatemeh Nazary - *Interacting with Features: Visual Inspection of Black-box Fault Type Classification Systems in Electrical Grids*

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 AI*IA 2020 Conference[4].

– Luca Maria Aiello, *Nokia Bell Labs*
– Davide Bacciu, *Università di Pisa*
– Matteo Baldoni, *Università di Torino*
– Valerio Basile, *Università di Torino*
– Federico Bianchi, *Università Bocconi - Milano*
– Ludovico Boratto, *EURECAT*
– Roberta Calegari, *Università di Bologna*
– Federica Cena, *Università di Torino*
– Cristina Conati, *University of British Columbia*
– Roberto Confalonieri, *Free University of Bozen-Bolzano*
– Stefano Ferilli, *Università di Bari*
– Fabio Gasparetti, *Roma Tre University*
– Alessandro Giuliani, *Università di Cagliari*
– Riccardo Guidotti, *Università di Pisa*
– Andrea Iovine, *Università di Bari*
– Kyriaki Kalimeri, *ISI Foundation*

---

[4] https://aixia2020.di.unito.it/

- Antonio Lieto, *Università di Torino*
- Francesca Lisi, *Università di Bari*
- Alessandro Mazzei, *Università di Torino*
- Anna Monreale, *Università di Pisa*
- Stefania Montani, *Università Piemonte Orientale*
- Andra Omicini, *Università di Bologna*
- Marco Polignano, *Università di Bari*
- Amon Rapp, *Università di Torino*
- Gaetano Rossiello, *IBM Research*
- Giuseppe Sansonetti, *Roma Tre University*
- Lucio Davide Spano, *Università di Cagliari*
- Fabio Stella, *Università Milano-Bicocca*
- Stefano Teso, *Katholieke Universiteit Leuven*
- Fabio Massimo Zanzotto, *Università di Roma Tor Vergata*