Lecture Notes in Artificial Intelligence 12688

Subseries of Lecture Notes in Computer Science

More information about this subseries at http://www.springer.com/series/1244

Davide Calvaresi · Amro Najjar ·
Michael Winikoff · Kary Främling (Eds.)

Explainable and Transparent AI and Multi-Agent Systems

Third International Workshop, EXTRAAMAS 2021
Virtual Event, May 3–7, 2021
Revised Selected Papers

Springer

Editors
Davide Calvaresi (ID)
University of Applied Sciences and Arts
Western Switzerland
Sierre, Switzerland

Amro Najjar (ID)
University of Luxembourg
Esch-sur-Alzette, Luxembourg

Michael Winikoff (ID)
Victoria University of Wellington
Wellington, New Zealand

Kary Främling (ID)
Umeå University
Umeå, Sweden

Preface

Recent advancements in eXplainable Artificial Intelligence (XAI) are generating new understanding and opportunities. The increasingly complex intelligent agents/robots rely on often opaque machine learning-based algorithms. Explaining such mechanisms is a chief priority to enhance their acceptability, avoid failures, foster trust, and comply with relevant (inter)national regulations.

The 2021 edition of the EXplainable and TRAnsparent AI and Multi-Agent Systems (EXTRAAMAS) workshop continues the successful track initiated in 2019 at Montreal and followed by the 2020 edition in New Zealand (which was virtual due to the COVID-19 pandemic circumstances). In particular, EXTRAAMAS 2021 set the following aims: (i) to strengthen the common ground for the study and development of explainable and understandable autonomous agents, robots, and Multi-Agent Systems (MAS), (ii) to investigate the potential of agent-based systems in the development of personalized user-aware explainable AI, (iii) to assess the impact of transparent and explained solutions on the user/agent behaviors, (iv) to discuss motivating examples and concrete applications in which the lack of explainability leads to problems, which would be resolved by explainability, and (v) to assess and discuss the first demonstrators and proof of concepts paving the way for the next generation of systems. EXTRAAMAS 2021 received 32 submissions. Each submission underwent a rigorous single-blind peer review process (three to five reviews per paper). Eventually, 20 papers were accepted (19 full papers and 1 short paper), which are contained in this volume. Due to COVID-19 restrictions, the workshop and AAMAS (the hosting conference) were held online rather than in London, UK. For each paper, the authors performed live video presentations that, with their consent, are available on the EXTRAAMAS website[1]. Moreover, EXTRAAMAS 2021 included two keynotes: "Social and Ethical Responsibilities of Computing and the Role of Explainability and Transparency" given by Prof. Julie Shah and "Explainable Reasoning in the Face of Contradictions: From Humans to Machines" given by Prof. Emeritus Dov Gabbay, and two panels: the industrial panel "Explainable Agents - Escape from the Ivory Tower" and the panel dedicated to the European Project EXPECTATION which considers XAI in distributed and heterogeneous agent-based recommender systems.

We would like to thank the industrial chair, publicity chairs, and Program Committee for their valuable work. We also thank the authors, presenters, and participants. Particular emphasis goes to Julie Shah and Dov Gabbay for their fantastic keynotes, and to Johanna Björklund, Tathagata Chakraborti, Kristijonas Cyras, Elizabeth Sklar, Andrea Omicini,

---

[1] https://extraamas.ehealth.hevs.ch/archive.html.

Reyhan Aydogan, and Leon Van der Torre for their participation in very enthusiastic discussion panels.

Organization

General Chairs

Davide Calvaresi	University of Applied Sciences Western Switzerland, Switzerland
Amro Najjar	University of Luxembourg, Luxembourg
Michael Winikoff	Victoria University of Wellington, New Zealand
Kary Främling	Umeå University, Sweden

Industrial Chair

Timotheus Kampik	Umeå University, Sweden

Publicity Chairs

Sviatlana Höhn	University of Luxembourg, Luxembourg
Giovanni Ciatto	University of Bologna, Italy

Advisory Board

Tim Miller	University of Melbourne, Australia
Michael Schumacher	University of Applied Sciences Western Switzerland, Switzerland
Virginia Dignum	Umeå University, Sweden
Leon van der Torre	University of Luxembourg, Luxembourg

Program Committee

Natasha Alechina	Utrecht University, The Netherlands
Amr Alyafi	Grenoble Electrical Engineering, France
Cleber Jorge Amaral	Federal University of Santa Catarina, Brazil
Kim Baraka	Carnegie Mellon University, USA
Suna Bensch	Umeå University, Sweden
Grégory Bonnet	University of Caen, France
Joost Broekens	Leiden University, The Netherlands
Jean-Paul Calbimonte	HES-SO Valais-Wallis, Switzerland
Tathagata Chakraborti	IBM Research AI, USA

Nicolas Cointe	Delft University, The Netherlands
Kristijonas Cyras	Ericsson Research, Sweden
Jérémie Dauphin	University of Luxembourg, Luxembourg
Alaa Daoud	École Nationale Supérieure des Mines de Saint-Étienne, France
Dustin Dannenhauer	Parallax Advanced Research, USA
Lora Fanda	University of Applied Sciences Western Switzerland, Switzerland
Michael W. Floyd	Knexus Research, USA
Stephane Galland	UTBM, France
Önder Gürcan	CEA, France
Maike Harbers	Rotterdam University of Applied Sciences, The Netherlands
Brent Harrison	Georgia Institute of Technology, USA
Salima Hassas	Claude Bernard Lyon 1 University, France
Helen Hastie	Heriot-Watt University, UK
Thomas Hellström	Umeå University, Sweden
Koen Hindriks	Virje Universiteit Amsterdam, The Netherlands
Sviatlana Höhn	University of Luxembourg, Luxembourg
Isaac Lage	Harvard University, USA
Beishui Liao	Zhejiang University, China
Tomer Libal	University of Luxembourg, Luxembourg
Brian Lim	National University of Singapore, Singapore
Avleen Malhi	Thapar Institute of Engineering & Technology, India
Niccolò Maltoni	University of Bologna, Italy
Réka Markovich	University of Luxembourg, Luxembourg
Viviana Mascardi	University of Genova, Italy
Laëtitia Matignon	Claude Bernard Lyon 1 University, France
Yazan Mualla	UTBM, France
Juan Carlos Nieves Sanchez	Umea University, Sweden
Sana Nouzri	University of Luxembourg, Luxembourg
Andrea Omicini	University of Bologna, Italy
Marrina Paolanti	Polytecnic University of Marche, France
Gauthier Picard	École Nationale Supérieure des Mines de Saint-Étienne, France
Patrick Reignier	LIG, France
Francisco Javier Rodríguez Lera	University of León, Spain
Stefan Sarkadi	King's College London, UK
Giovanni Sartor	Università di Bologna and European University Institute, Italy

Sarath Sreedharan	Arizona State University, USA
Cesar A. Tacla	Federal Technological University of Paraná, Brazil
Silvia Tulli	Instituto Superior Tecnico and INESC-ID, Portugal
Rob Wortham	University of Bath, UK
Deshraj Yadav	Georgia Institute of Technology, USA
Jessie Yang	University of Michigan, USA

Contents

XAI Applications

XAI Logic and Argumentation

Decentralized and Heterogeneous XAI