

# **Lecture Notes in Artificial Intelligence**

**12688**

Subseries of Lecture Notes in Computer Science

## Series Editors

Randy Goebel

*University of Alberta, Edmonton, Canada*

Yuzuru Tanaka

*Hokkaido University, Sapporo, Japan*

Wolfgang Wahlster

*DFKI and Saarland University, Saarbrücken, Germany*

## Founding Editor

Jörg Siekmann

*DFKI and Saarland University, Saarbrücken, Germany*


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

### *Editors*

Davide Calvaresi   
University of Applied Sciences and Arts  
Western Switzerland  
Sierre, Switzerland

Michael Winikoff   
Victoria University of Wellington  
Wellington, New Zealand

Amro Najjar   
University of Luxembourg  
Esch-sur-Alzette, Luxembourg

Kary Främling   
Umeå University  
Umeå, Sweden

ISSN 0302-9743                      ISSN 1611-3349 (electronic)  
Lecture Notes in Artificial Intelligence  
ISBN 978-3-030-82016-9              ISBN 978-3-030-82017-6 (eBook)  
<https://doi.org/10.1007/978-3-030-82017-6>

LNCS Sublibrary: SL7 – Artificial Intelligence

© Springer Nature Switzerland AG 2021

This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

The publisher, the authors and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, expressed or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

This Springer imprint is published by the registered company Springer Nature Switzerland AG  
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

# 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<sup>1</sup>. 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,

---

<sup>1</sup> <https://extraamas.ehealth.hevs.ch/archive.html>.

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

May 2021

Davide Calvaresi  
Amro Najjar  
Michael Winikoff  
Kary Främling

# 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

Cesar A. Tacla

Silvia Tulli

Rob Wortham

Deshraj Yadav

Jessie Yang

Arizona State University, USA

Federal Technological University of  
Paraná, Brazil

Instituto Superior Tecnico and INESC-ID,  
Portugal

University of Bath, UK

Georgia Institute of Technology, USA

University of Michigan, USA

# Contents

## XAI and Machine Learning

To Pay or Not to Pay Attention: Classifying and Interpreting Visual Selective Attention Frequency Features .....	3
<i>Lora Fanda, Yashin Dicente Cid, Pawel J. Matusz, and Davide Calvaresi</i>	
GridEx: An Algorithm for Knowledge Extraction from Black-Box Regressors .....	18
<i>Federico Sabbatini, Giovanni Ciatto, and Andrea Omicini</i>	
Comparison of Contextual Importance and Utility with LIME and Shapley Values .....	39
<i>Kary Främling, Marcus Westberg, Martin Jullum, Manik Madhikermi, and Avleen Malhi</i>	
ciu.image: An R Package for Explaining Image Classification with Contextual Importance and Utility .....	55
<i>Kary Främling, Samanta Knapič, and Avleen Malhi</i>	
Shallow2Deep: Restraining Neural Networks Opacity Through Neural Architecture Search .....	63
<i>Andrea Agiollo, Giovanni Ciatto, and Andrea Omicini</i>	
Visual Explanations for DNNs with Contextual Importance .....	83
<i>Sule Anjomshoae, Lili Jiang, and Kary Främling</i>	
Towards Explainable Recommendations of Resource Allocation Mechanisms in On-Demand Transport Fleets .....	97
<i>Alaa Daoud, Hiba Alqasir, Yazan Mualla, Amro Najjar, Gauthier Picard, and Flavien Balbo</i>	

## XAI Vision, Understanding, Deployment and Evaluation

A Two-Dimensional Explanation Framework to Classify AI as Incomprehensible, Interpretable, or Understandable .....	119
<i>Ruben S. Verhagen, Mark A. Neerincx, and Myrthe L. Tielman</i>	
Towards Explainable Visionary Agents: License to Dare and Imagine .....	139
<i>Giovanni Ciatto, Amro Najjar, Jean-Paul Calbimonte, and Davide Calvaresi</i>	

Towards an XAI-Assisted Third-Party Evaluation of AI Systems: Illustration on Decision Trees .....	158
<i>Yongxin Zhou, Matthieu Boussard, and Agnes Delaborde</i>	
What Does It Cost to Deploy an XAI System: A Case Study in Legacy Systems .....	173
<i>Sviatlana Höhn and Niko Faradouris</i>	
<b>XAI Applications</b>	
Explainable AI (XAI) Models Applied to the Multi-agent Environment of Financial Markets .....	189
<i>Jean Jacques Ohana, Steve Ohana, Eric Benhamou, David Saltiel, and Beatrice Guez</i>	
Toward XAI & Human Synergies to Explain the History of Art: The Smart Photobooth Project .....	208
<i>Egberdien van der Peijl, Amro Najjar, Yazan Mualla, Thiago Jorge Bourscheid, Yolanda Spinola-Elias, Daniel Karpati, and Sana Nouzri</i>	
Assessing Explainability in Reinforcement Learning .....	223
<i>Amber E. Zelvelder, Marcus Westberg, and Kary Främling</i>	
<b>XAI Logic and Argumentation</b>	
Schedule Explainer: An Argumentation-Supported Tool for Interactive Explanations in Makespan Scheduling .....	243
<i>Kristijonas Čyras, Myles Lee, and Dimitrios Letsios</i>	
Towards Explainable Practical Agency: A Logical Perspective .....	260
<i>Nourhan Ehab and Haythem O. Ismail</i>	
Explainable Reasoning in Face of Contradictions: From Humans to Machines .....	280
<i>Timotheus Kampik and Dov Gabbay</i>	
Towards Transparent Legal Formalization .....	296
<i>Tomer Libal and Tereza Novotná</i>	
Applying Abstract Argumentation to Normal-Form Games .....	314
<i>You Cheng, Beishui Liao, and Jieting Luo</i>	

## Decentralized and Heterogeneous XAI

EXPECTATION: Personalized Explainable Artificial Intelligence for Decentralized Agents with Heterogeneous Knowledge .....	331
<i>Davide Calvaresi, Giovanni Ciatto, Amro Najjar, Reyhan Aydoğar, Leon Van der Torre, Andrea Omicini, and Michael Schumacher</i>	
<b>Author Index</b> .....	345