2023

Dr. Davide Calvaresi, Dr. Amro Najjar, Prof. Andrea Omicini, Prof. Kary Framling
Dr. Reyhan Aydogan, Dr. Giovanni Ciatto, Prof. Yazan Mualla, Rachele Carli

London, 29.05.2023

Session 1 - Explainable Agents: **10:45 - 12:30** CEST

Session Chairs: Dr. Reyhan Aydoğan & Prof. Kary Främling

[10.45 – 11.00] – Ahmad Alelaimat, Aditya Ghose and Hoa Khanh Dam.
Mining and Validating Belief-based Agent Explanations

[11.00 – 11.15] – Michael Winikoff and Galina Sidorenko.
Evaluating a mechanism for explaining BDI agent behaviour

[11.15 – 11.30] – Balint Gyevnar, Cheng Wang, Christopher G. Lucas, Shay B. Cohen and Stefano V. Alb
Causal Social Explanations for Stochastic Sequential Multi-Agent Decision-Making

[11.30 – 11.45] – Giovanni Ciatto, Matteo Magnini, Berk Buzcu, Reyhan Aydogan and Andrea Omicini.
A General-Purpose Protocol for Multi-Agent based Explanations

[11.45 – 12.00] – Joris Hulstijn, Igor Tchappi, Amro Najjar and Reyhan Aydoğan.
Metrics for Evaluating Explainable Recommender Systems

[12.00 – 12.15] – Yifan Xu, Joe Collenette, Louise Dennis and Clare Dixon.
Dialogue Explanations for Rules-based AI Systems

[12.15 – 12.30] – Saaduddin Mahmud, Samer Nashed, Claudia Goldman and Shlomo Zilberstein.
Estimating Causal Responsibility for Explaining Autonomous Behavior

Session 3 - Explainable AI and Law: **16:45 - 17:15** CEST

Session Chairs: Dr. Reyhan Aydoğan & Rachele Carli

[16.45 – 17.00] – Balint Gyevnar and Nick Ferguson.
Aligning Explainable AI and the Law: The European Perspective

[17.00 – 17.15] – Rachele Carli and Davide Calvaresi.
Reinterpreting Vulnerability to Tackle Deception in Principles-Based XAI for Human-Computer Interaction.

Session 2 - Explainable Machine Learning: **14:00 - 16:45** CEST

Session Chairs: Dr. Giovanni Ciatto & Dr. Amro Najjar

[14.00 – 14.15] – Andrea Agiollo, Pradeep Kumar Murukannaiah, Luciano Cavalcante Siebert and Andrea Omicini.
The Quarrel of Local Post-hoc Explainers for Moral Values Classification in Natural Language Processing

[14.15 – 14.30] – Kary Främling.
Counterfactual, Contrastive and Hierarchical Explanations with Contextual Importance and Utility

[14.30 – 14.45] – Federico Sabbatini and Roberta Calegari.
Bottom-Up and Top-Down Workflows for Hypercube- and Clustering-based Knowledge Extractors

[14.45 – 15.00] – Victor Contreras, Andrea Bagante, Niccolò Marini, Michael Schumacher, Vincent Andrearczyk and Davide Calvaresi.
Explanation Generation via Decompositional Rules Extraction for Head and Neck Cancer Classification

[15.00 – 15.15] – Umanta Dey, Sharat Bhat, Pallb Dasgupta and Soumyajit Dey.
Imperative Action Masking for Safe Exploration in Reinforcement Learning

[15.15 – 15.30] – Fumito Uwano and Keiki Takadama.
Reinforcement Learning in Cyclic Environmental Change for Non-Communicative Agents: A Theoretical Approach

[15.30 – 15.45] – Ahmad Alelaimat, Aditya Ghose and Hoa Khanh Dam.
Leveraging Imperfect Explanations for Plan Recognition Problems

[15.45 - 16.15] - Coffee Break

[16.15 – 16.30] – Matija Franklin, David Lagnado, Chulhong Min, Akhil Mathur and Fahim Kawsar.
Using Cognitive Models and Wearables to Diagnose and Predict Dementia Patient Behaviour

[16.30 – 16.45] – Andreas Kontogiannis and George Vouros.
Inherently Interpretable Deep Reinforcement Learning through Online Mimicking