



2024

Prof. Davide Calvaresi, Dr. Amro Najjar, Prof. Andrea Omicini, Prof. Kary Framling

Auckland, 6.05.2024

[9:05 – 10:00] – Morning Opening Keynote Title: Enhancing Outcomes with User–Centered Explainable AI: Insights from two case studies Speaker(s): Dr. Dena Al–Thani
[10:00 – 11:45] – Session I: Usec–Centered XAI Session Chair: Dr. Amro Najjar [10:00 – 10:15] – Title: People Attribute Purpose to Autonomous Vehicles When Explaining Their Behavior Authors: Balint Gyevnar, Stephanie Droop and Tadeq Quillien [10:15 – 10:30] – Title: Evaluation of the User–centric Explanation Strategies for Interactive Recommenders Authors: Berk Buzcu, Emre Kuru, Davide Calvaresi and Reyhan Aydogan
[10:30 – 11:00] – COFFE BREAK
[11:00 – 11:15] – Title: Can Interpretability Layouts Influence Human Perception of Offensive Sentences? Authors: Thiago Freitas dos Santos, Nardine Osman e Marco Schorlemmer [11:15 – 11:30] – Title: Effect of Agent Explanations Using Warm and Cold Language on User Adoption of Recommendations for Bandit Problem Authors: Selim Karaoglu, Marina Katoh, Titash Majumdar, Ethan Beaird, Feyza Hafizoglu and Sandip Sen [11:30 – 11:45] – Title: A Framework for Explainable Multi–purpose Virtual Assistants: A Nutrition–Focused Case Study Authors: Berk Buzcu, Yvan Pannatier, Reyhan Aydogan, Michael Ignaz Schumacher, Jean–Paul Calbimonte and Davide Calvaresi
[11:45 – 12:25] – Session II: XAI and Ethics Session Chair: Dr. Reyhan Aydogan [11:45 – 12:05] – Title: The Wildecard XAI: from a Necessity, to a Resource, to a Dangerous Decoy Authors: Rachele Carli and Davide Calvaresi [12:05 – 12:25] – Title: Explainability and Transparency in Practice: A Comparison Between Corporate and National AI Ethics Guidelines in Germany and China Authors: Timo Speith and Jing Xu
[12:30 – 14:00] – LUNCH BREAK

[14:00 – 14:55] – Afternoon Opening Keynote Title: Toward Human–Centered Explainable AI Speaker(s): Prof. Brian Lim
[15:00 – 17:40] – Session III: XAI and Reinforcement Learning Session Chair: Prof. Kary Framling [15:00 – 15:20] – Title: Learning Temporal Task Specifications From Demonstrations Authors: Mattijs Baert, Sam Leroux and Pieter Simoens [15:20 – 15:40] – Title: Temporal Explanations for Deep Reinforcement Learning Agents Authors: Mark Towers, Yali Du, Christopher Freeman and Timothy Norman [15:40 – 16:00] – Title: An Adaptive Interpretable Safe–RL Approach for Addressing Smart Grid Supply–side Uncertainties Authors: Sumanta Dey, Praveen Verma, Pallab Dasgupta and Soumyajit Dey [16:00 – 16:20] – Title: Explainable Multi–Objective Reinforcement Learning via Bi–Objective Policy Clustering Authors: Zuzanna Osika, Jazmin Zatarain Salazar, Frans A. Oliehoek and Pradeep K. Murukannaiah [16:20 – 16:40] – Title: Invited paper Authors: Vaibhav Unhelkar
[16:40 – 17:00] – COFFE BREAK
[17:00 – 17:15] – Title: Model–Agnostic Policy Explanations: Biased Sampling for Surrogate Models Authors: Bryan Lavender and Sandip Sen [17:15 – 18:00] – Session IV: Neuro–symbolic AI and Explainable Machine Learning Session Chair: Prof. Davide Calvaresi [17:15 – 17:30] – Title: Explanation of Deep Learning Models via Logic Rules Enhanced by Embeddings Analysis, and Probabilistic Models Authors: Victor Hugo Contreras Ordoñez, Michael Schumacher, and Davide Calvaresi [17:30 – 17:45] – Title: Towards interactive and social explainable artificial intelligence for digital history Authors: Richard Albrecht, Amro Najjar, Igor Tchappi and Joris Hulstijn [17:45 – 18:00] – Title: py ciu image: a Python library for Explaining Image Classification with Contextual Importance and Utility Authors: Kary Främling, Ioan–Vlad Apopei, Gustav Grund Pihlgren and Avleen Malhi [18:00 – 18:05] – Concluding Remarks Speaker(s): Prof. Davide Calvaresi, Dr. Amro Najjar, Prof. Kary Framling