6th International Workshop on EXplainable and TRAnsparent AI and Multi-Agent Systems (EXTRAAMAS)
in conjunction with AAMAS 2024, Auckland, New Zealand, 6-9 May 2024

Running since 2019, EXTRAAMAS is a well-established workshop and forum on EXplainable and TRAnsparent AI and Multi-Agent Systems. It aims to discuss and disseminate research on explainable artificial intelligence, with a particular focus on intra/inter-agent explainability and cross-disciplinary perspectives. In its 6th edition, EXTRAAMAS identifies four particular focus topics with the ultimate goal of strengthening cutting-edge foundational and applied research. This, of course, comes in addition to the workshop's main theme, focusing, as usual, on XAI fundamentals. The four tracks for this year are:

- **Track 1: XAI in symbolic and subsymbolic AI:** the “AI dichotomy” separating symbolic AKA classical AI from connectionism AI has persisted for more than seven decades. Nevertheless, the advent of explainable AI has accelerated and intensified the efforts to bridge this gap since providing faithful explanations of black-box machine learning techniques would necessarily mean combining symbolic and subsymbolic AI. This track aims to discuss the recent works on this hot topic of AI.
    - **Track chair: Giovanni Ciatto, University of Bologna, Italy.**
- **Track 2: XAI in negotiation and conflict resolution:** Conflict resolution (e.g., agent-based negotiation, voting, argumentation, etc.) has been a prosperous domain within the MAS community since its foundation. However, as agents and the problems they are tackling become more complex, incorporating explainability becomes vital to assess the usefulness of the supposedly conflict-free solution. This is the main topic of this track, with a special focus on MAS negotiation and explainability.
    - T**rack Chair: Reyhan Aydoğan: Ozyegin University, Turkey**
- **Track 3: Prompts, Interactive Explainability and Dialogues:** Appropriate everyday explanations about automated decision-making are context-dependent and interactive. An explanation must fill a 'gap' in the apparent knowledge of the user in a specific context. However, dynamic user modelling is hard. Explanatory dialogue allows designers to try out partial explanations and fine-tune or adjust the explanations based on feedback. This potential for dynamic adjustment can only be redeemed if the system has appropriate interactive capabilities, such as context modelling, user modelling, initiative handling, topic management and grounding. The rapid evolution of LLM and Chatbots has sparked a debate on how to make good use of the interactive capabilities of these new models for explainable AI. The use of LLM also has risks, especially concerning reliability. This triggers relevant methodological questions. How to ensure LLM use reliable data for answering? How to evaluate research based on black-box models? What are good techniques for prompt engineering? In this research track, we welcome new ideas as well as established research outcomes, on the wider topic of Interactive or Social Explainable AI.
    - **Track chair: Joris Hulstijn, University of Luxembourg**
- **Track 4: XAI in Law and Ethics:** complying with regulation (e.g. GDPR) is among the main objectives for XAI. The right to explanation is key to ensuring transparency of ever more complex AI systems dealing with a multitude of sensitive AI applications. This track discusses works related to explainability in AI ethics, machine ethics, and AI and law.
    - **Track chair: Rachele Cari, University of Bologna, Italy**

This year, EXTRAAMAS will feature a keynote delivered by **Brian Lim** (title TBD)

All accepted papers are eligible for publication in the Springer Lecture Notes of Artificial Intelligence conference proceedings (after revisions have been applied).

Important Dates	**Submission link**
Paper submission: 01/03/2024 Notification of acceptance: 25/03/2025 Early registration deadline: April 5, 2024. Workshop: 06-07/05/2023 Camera-ready (Springer post-proceedings): 10/06/2023	https://easychair.org/conferences/?conf=extraamas2024

EXTRAAMAS Tracks

Track1: XAI in symbolic and subsymbolic AI	**Track3: Prompts, Interactive Explainability and Dialogue**
• XAI for Machine learning • Explainable neural networks • Symbolic knowledge injection or extraction • Neuro-symbolic computation • Computational logic for XAI • Multi-agent architectures for XAI • Surrogate models for sub-symbolic predictors • Explainable planning (XAIP) • XAI evaluation	• Interactive capabilities for XAI • Arguments for persuasive explanations • Context modelling • User modelling • Initiative handling • Topic modelling • Grounding and acknowledgement • Prompt engineering • Research methodology for LLM applications • Responsible LLM applications
Track2: XAI in negotiation and conflict resolution	**Track4: (X)AI in Law and Ethics**
• Explainable conflict resolution techniques/frameworks • Explainable negotiation protocols and strategies • Explainable recommendation systems • Trustworthy voting mechanisms • Argumentation for explaining the process itself • Argumentation for explaining and supporting the potential outcomes • Explainable user/agent profiling (e.g., learning user's preferences or strategies) • User studies and assessment of the aforementioned approaches • Applications (virtual coaches, robots, IoT)	• XAI in AI & Law • Fair AI • XAI & Machine Ethics • Bias reduction • Deception and XAI • Persuasive technologies and XAI • Nudging and XAI • Legal issues of XAI • Liability and XAI • XAI, Transparency, and the Law • Enforceability and XAI • Culture-aware systems and XAI

Workshop Chairs

Dr. Davide Calvaresi, HES-SO, Switzerland research areas: *Real-Time Multi-Agent Systems, Explainable AI, BCT, eHealth,* mail: davide.calvaresi@hevs.ch, web page, Google.scholar	Dr. Amro Najjar, University of Luxembourg, Luxembourg research areas: Multi-Agent Systems, Explainable AI, AI mail: amro.najjar@uni.lu, Google Scholar
Prof. Kary Främling, Umeå & Aalto University Sweden/Finland, research areas: *Explainable AI, Artificial Intelligence, Machine Learning, IoT* mail: Kary.Framling@cs.umu.se, web page, Google Scholar	Prof. Andrea Omicini research areas: *Artificial Intelligence, Multi-agent Systems, Soft. Engineering* mail: andrea.omicini@unibo.it, web page, Google Scholar

Track Chairs	**Advisory Board**
Dr. Giovanni Ciatto, University of Bologna, Italy – giovanni.ciatto@unibo.it Prof. Rehyan Aydogan, Ozyegin University, Turkey – reyhan.aydogan@ozyegin.edu.tr Rachele Carli, University of Bologna – rachele.carli2@unibo.it Joris HULSTIJN: University of Luxembourg – joris.hulstijn@uni.lu	Dr. Tim Miller, University of Melbourne Prof. Leon van der Torre, UNILU Prof. Virginia Dignum, Umea University Prof. Michael Ignaz Schumacher

Primary Contacts Davide Calvaresi - davide.calvaresi@hevs.ch and Amro Najjar - amro.najjar@list.lu