

EXTRAAMAS 2026 Program

Room: **APHRODITE B** |

[Web Link](#)

Welcome and Opening Remarks

09:00 – 09:05

SESSION 1 Foundations & Limits of Explainability in Agent(s/ic) Systems

09:05 – 10:30

Chair: Prof. Davide Calvaresi

09:05–09:35

KEYNOTE Re-evaluating Responsible AI in the AI-First Age

Rachele Carli

09:35–09:50

FULL PAPER Making AI Memory Transparent: Explainability for Persistent State in Agentic Systems

Matthew Stewart

09:50–10:05

FULL PAPER Predicting the Cost of Explanations in Optimisation Problems

Roger Lera-Leri et al.

10:05–10:20

FULL PAPER The Illusion of Explainability with LLMs and LLM-Agents

Sara Montese, Sergio Alvarez-Napagao, Victor Gimenez-Abalos, Ulises Cortés

10:20–10:30

DEMO Explainable Agentic Oversight: A Multi-Agent System for Transparent Groundwater Regulatory Decision Support

Ayyub Alzahem et al.

Coffee Break 10:30 – 11:00

SESSION 2 Trust, Ethics, and Human-Centred Explainable AI

11:00 – 12:30

Chair: Dr. Rachele Carli & Dr. Amro Najjar

11:00–11:15

SHORT PAPER Social Dynamics of Trust Formation in XAI Systems: A Multi-Agent Approach

Narendar Kumar et al.

11:15–11:30

FULL PAPER Perceived Trustworthiness of Large Language Models and Attitudes Toward Them

Muaadh Noman et al.

11:30–11:45

FULL PAPER Auditing LLM Multi-Agent Systems via Protocolized Interaction Traces

Azzedine Idir Ait Said et al.

11:45–12:00

FULL PAPER Bias Mitigation in Quantum Machine Learning

Alessandro Marchetti et al.

12:00–12:15

FULL PAPER Designing a Trustworthy Intelligent Agent for Legal Reasoning

Alessia Donata Camarda et al.

12:15–12:30

FULL PAPER Emergence of Secondary Norm from Distributed Interaction of Autonomous Agents

Edoardo Zappia et al.

Lunch Break 12:30 – 14:00

SESSION 3 Methods for Explainability: Argumentation, Learning, and Hybrid AI

14:00 – 15:30

Chair: Prof. Kary Främling

14:05–14:35

KEYNOTE Human–AI Collaboration: From Support to Dependency and What Design Can Do

Prof. Dr. Raian Ali

14:35–14:50

FULL PAPER ArgVLLMs: Structuring Multimodal Reasoning

Tianyi Ma, Francesca Toni, Soteris Demetriou

14:50–15:05

FULL PAPER Towards Trustworthy Temporal Event Reasoning

Sukriti Bhattacharya et al.

15:05–15:20

FULL PAPER M2–PALE: Explaining Multi–Agent MCTS–Minimax Hybrids

Yiyu Qian et al.

15:20–15:30

FULL PAPER Critic–Driven Voronoi–Quantization for Distilling Deep RL Policies

Senne Deproost et al.

Coffee Break 15:30 – 16:00

SESSION 4 Systems and Applications of Explainable Multi–Agent Systems

16:00 – 17:45

Chair: Dr. Amro Najjar

16:00–16:15

FULL PAPER Evolutionary Rule Extraction from Corporate Default Prediction Models

Desiré Fabbretti et al.

16:15–16:30

FULL PAPER User–Adapted and Interactive Explainability with Knowledge Graphs

Kary Främling

16:30–16:45

FULL PAPER Translation as a Decision Space: A Multi–Agent Perspective

Hasan Alkhder et al.

16:45–17:00

FULL PAPER Rashomon Memory

Albert Sadowski, Jaroslaw A. Chudziak

17:00–17:15

FULL PAPER Talk2Prolog: from Controlled Natural Language to Prolog

Laura Fuciarelli et al.

17:15–17:30

FULL PAPER Toward Explanatory Equilibrium: Verifiable Reasoning as a Coordination Mechanism under Asymmetric Information

Feliks Bařka and Jaroslaw A. Chudziak

17:30–17:45

FULL PAPER Exploring Vision Features with Sparse Autoencoders

Marlin Lee et al.

Closing Remarks and Thanks

17:45 – 17:50