

1. Record Nr.	UNINA9911047673103321
Autore	Calvaresi Davide
Titolo	Explainable, Trustworthy, and Responsible AI and Multi-Agent Systems : 7th International Workshop, EXTRAAMAS 2025, Detroit, MI, USA, May 19–20, 2025, Revised Selected Papers // edited by Davide Calvaresi, Amro Najjar, Andrea Omicini, Reyhan Aydogan, Rachele Carli, Giovanni Ciatto, Simona Tiribelli, Kary Främling
Pubbl/distr/stampa	Cham : , : Springer Nature Switzerland : , : Imprint : Springer, , 2026
ISBN	3-032-01399-2
Edizione	[1st ed. 2026.]
Descrizione fisica	1 online resource (352 pages)
Collana	Lecture Notes in Artificial Intelligence, , 2945-9141 ; ; 15936
Altri autori (Persone)	NajjarAmro OmiciniAndrea AydoganReyhan CarliRachele CiattoGiovanni TiribelliSimona ämblingKary
Disciplina	006.31
Soggetti	Machine learning Compilers (Computer programs) Natural language processing (Computer science) Computer science Computers, Special purpose Multiagent systems Machine Learning Compilers and Interpreters Natural Language Processing (NLP) Theory of Computation Special Purpose and Application-Based Systems Multiagent Systems
Lingua di pubblicazione	Inglese
Formato	Materiale a stampa
Livello bibliografico	Monografia
Nota di contenuto	AI Interpretation & Social Dimensions -- Enhancing Surrogate Decision

Trees for Reinforcement Learning with Feature Importance Matrices -- Exploiting Constraint Reasoning to Build Graphical Explanations for Mixed-Integer Linear Programming -- Explaining Autonomous Vehicles with Intention-aware Policy Graphs -- Social Explainable AI: What Is It and How to Make It Happen with CIU?- LLMs & Generative AI in XAI -- Evaluating Prompt Engineering Techniques for Accuracy and Confidence Elicitation in Medical LLMs -- LLM-based Evaluation Methodology of Explanation Strategies -- LLM-based Evaluation Methodology of Explanation Strategies -- Agent Architectures & Inter-Agent Explainability -- Ladder of Intentions: Unifying Agent Architectures for Explainability and Transferability -- Engineering Inter-Agent Explainability in BDI Agents -- Human-XAI Interaction, Quality & Ethics -- Beyond Satisfaction: From Placebic to Actionable Explanations For Enhanced Understandability -- Building a Decolonized AI Ethics and Governance Framework -- Balancing Ethical Persuasion and Explanation Transparency in XAI -- Designing for Effective Human-XAI Inter-action.

Sommario/riassunto

This book constitutes the refereed post-conference proceedings of the 7th International Workshop on Explainable and Transparent AI and Multi-Agent Systems, EXTRAAMAS 2025, held in Detroit, MI, USA, during May 19-20, 2025. The 12 full papers and 1 short paper included in this book were carefully reviewed and selected from 23 submissions. They were organized in topical sections as follows: AI interpretation & social dimensions; LLMs & generative AI in XAI; agent architectures & inter-agent explainability; and human-XAI interaction, quality & ethics.
