

1. Record Nr.	UNINA9910888599203321
Autore	Calvaresi Davide
Titolo	Explainable and Transparent AI and Multi-Agent Systems : 6th International Workshop, EXTRAAMAS 2024, Auckland, New Zealand, May 6–10, 2024, Revised Selected Papers // edited by Davide Calvaresi, Amro Najjar, Andrea Omicini, Reyhan Aydogan, Rachele Carli, Giovanni Ciatto, Joris Hulstijn, Kary Främling
Pubbl/distr/stampa	Cham : , : Springer Nature Switzerland : , : Imprint : Springer, , 2024
ISBN	9783031700743 3031700740
Edizione	[1st ed. 2024.]
Descrizione fisica	1 online resource (247 pages)
Collana	Lecture Notes in Artificial Intelligence, , 2945-9141 ; ; 14847
Altri autori (Persone)	NajjarAmro OmiciniAndrea AydoganReyhan CarliRachele CiattoGiovanni HulstijnJ (Joris) ämlingKary
Disciplina	006.30285436
Soggetti	Multiagent systems 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
Lingua di pubblicazione	Inglese
Formato	Materiale a stampa
Livello bibliografico	Monografia

Nota di contenuto

-- User-centric XAI. -- Effect of Agent Explanations Using Warm and Cold Language on User Adoption of Recommendations for Bandit Problem. -- Evaluation of the User-centric Explanation Strategies for Interactive Recommenders. -- Can Interpretability Layouts Influence Human Perception of Offensive Sentences?. -- A Framework for Explainable Multi-purpose Virtual Assistants: A Nutrition-Focused Case Study. -- XAI and Reinforcement Learning. -- Learning Temporal Task Specifications From Demonstrations. -- Temporal Explanations for Deep Reinforcement Learning Agents. -- An Adaptive Interpretable Safe-RL Approach for Addressing Smart Grid Supply-side Uncertainties. -- Model-Agnostic Policy Explanations: Biased Sampling for Surrogate Models. -- Neuro-symbolic AI and Explainable Machine Learning. -- Explanation of Deep Learning Models via Logic Rules Enhanced by Embeddings Analysis, and Probabilistic Models. -- py ciu image: a Python library for Explaining Image Classification with Contextual Importance and Utility. -- Towards interactive and social explainable artificial intelligence for digital history. -- XAI & Ethics. -- Explainability and Transparency in Practice: A Comparison Between Corporate and National AI Ethics Guidelines in Germany and China. -- The Wildcard XAI: from a Necessity, to a Resource, to a Dangerous Decoy.

Sommario/riassunto

This volume constitutes the papers of several workshops which were held in conjunction with the 6th International Workshop on Explainable and Transparent AI and Multi-Agent Systems, EXTRAAMAS 2024, in Auckland, New Zealand, during May 6–10, 2024. The 13 full papers presented in this book were carefully reviewed and selected from 25 submissions. The papers are organized in the following topical sections: User-centric XAI; XAI and Reinforcement Learning; Neuro-symbolic AI and Explainable Machine Learning; and XAI & Ethics.
