1.	Record Nr.	UNISA996490355103316
	Titolo	Explainable and transparent AI and multi-agent systems : 4th international workshop, EXTRAAMAS 2022, virtual event, May 9-10, 2022, revised selected papers / / edited by Davide Calvaresi [and three others]
	Pubbl/distr/stampa	Cham, Switzerland : , : Springer, , [2022] ©2022
	ISBN	3-031-15565-3
	Descrizione fisica	1 online resource (242 pages)
	Collana	Lecture Notes in Computer Science ; ; v.13283
	Disciplina	006.3
	Soggetti	Intelligent agents (Computer software)
	Lingua di pubblicazione	Inglese
	Formato	Materiale a stampa
	Livello bibliografico	Monografia
	Nota di bibliografia	Includes bibliographical references and index.
	Nota di contenuto	Intro -- Preface -- Organization -- Contents -- Explainable Machine Learning -- Evaluation of Importance Estimators in Deep Learning Classifiers for Computed Tomography -- 1 Introduction -- 2 Importance Estimators -- 3 Evaluation Methods -- 3.1 Model Accuracy per Input Feature Perturbation -- 3.2 Concordance Between Importance Scores and Segmentation -- 3.3 XRAI-Based Region-Wise Overlap Comparison -- 4 Results -- 4.1 Model Accuracy per Input Feature Perturbation -- 4.2 Concordance Between Importance Scores and Segmentation -- 4.3 XRAI-Based Region-Wise Overlap Comparison -- 5 Discussion -- References -- Integration of Local and Global Features Explanation with Global Rules Extraction and Generation Tools -- 1 Introduction -- 2 State of the Art -- 3 Methodology -- 4 Results and Analysis -- 4.1 S1 - ECLAIRE -- 4.2 S2 - ExpL and ECLAIRE -- 4.3 S3 - CIU and ECLAIRE -- 4.4 S4 - ExpL CIU and ECLAIRE -- 5 Discussion -- 6 Conclusions -- A Appendix Feature Description -- References -- ReCCoVER: Detecting Causal Confusion for Explainable Reinforcement Learning -- 1 Introduction -- 2 Related Work -- 2.1 Structural Causal Models (SCM) -- 2.2 Explainable Reinforcement Learning (XRL) -- 3 ReCCoVER -- 3.1 Extracting Critical States -- 3.2 Training Feature-Parametrized Policy -- 3.3 Generating Alternative Environments -- 3.4 Detecting Causal Confusion -- 4 Evaluation Scenarios and Settings -- 5

Evaluating ReCCoVER -- 6 Results -- 6.1 Taxi Environment -- 6.2 Minigrid Traffic Environment -- 7 Discussion and Future Work -- References -- Smartphone Based Grape Leaf Disease Diagnosis and Remedial System Assisted with Explanations -- 1 Introduction -- 1.1 Motivation -- 2 Related Work -- 3 Proposed Methodology -- 4 Implementation -- 4.1 Dataset -- 4.2 Image Augmentation -- 4.3 Model Training -- 4.4 Contextual Importance and Utility- Image -- 4.5 Mobile App.

5 Experimental Evaluation -- 5.1 Experimental Setup -- 5.2 Results and Analyses -- 6 Discussion -- 7 Conclusions and Future Work -- References -- Explainable Neuro-Symbolic AI -- Recent Neural-Symbolic Approaches to ILP Based on Templates -- 1 Introduction -- 2 Background -- 2.1 First-Order Logic -- 2.2 ILP -- 3 Neural-Symbolic ILP Methods -- 3.1 ILP -- 3.2 NTPs -- 3.3 ILPCamp -- 3.4 MetaAbd -- 4 A Comparison Based on Four Characteristics -- 4.1 Language -- 4.2 Search Method -- 4.3 Recursion -- 4.4 Predicate Invention -- 5 Open Problems and Challenges -- 6 Conclusion -- References -- On the Design of PSyKI: A Platform for Symbolic Knowledge Injection into Sub-symbolic Predictors -- 1 Introduction -- 2 Knowledge Injection Background -- 2.1 Constraining Neural Networks -- 2.2 Structuring Neural Networks -- 2.3 Workflow -- 3 A Platform for Symbolic Knowledge Injection -- 3.1 Overall Architecture -- 3.2 Technology -- 4 Case Study -- 4.1 Code Example -- 4.2 Results -- 5 Conclusion -- References -- Explainable Agents -- The Mirror Agent Model: A Bayesian Architecture for Interpretable Agent Behavior -- 1 Introduction -- 2 Background -- 2.1 The Mirror Agent Model -- 2.2 Previous Work -- 2.3 Informative Intention Verbalization -- 2.4 Legible Behavior -- 3 Generating Explanations -- 3.1 Background -- 3.2 Explanation Model -- 3.3 Preliminary Results -- 4 Conclusions -- References -- Semantic Web-Based Interoperability for Intelligent Agents with PSyKE -- 1 Introduction -- 2 State of the Art -- 2.1 Symbolic Knowledge Extraction -- 2.2 PSyKE -- 2.3 Semantic Web -- 2.4 Owlready -- 3 Interoperability via PSyKE -- 3.1 Output Rules in SWRL Format -- 3.2 Propositionalisation -- 3.3 Relationalisation -- 3.4 Semantic Web for SKE: Pros and Cons -- 4 An Example: The Iris Data Set -- 5 Open Issues -- 6 Conclusions -- References.

Case-Based Reasoning via Comparing the Strength Order of Features -- 1 Introduction -- 2 Strength of Atomic Features -- 3 Discussion and Future Work -- 4 Conclusion -- References -- XAI Measures and Metrics -- Explainability Metrics and Properties for Counterfactual Explanation Methods -- 1 Introduction -- 2 Related Work -- 3 Methodological Framework -- 3.1 Counterfactual Methods -- 3.2 Deriving XAI Metrics for Counterfactual Explanation Methods -- 4 System Design and Implementation -- 4.1 Pipeline Structure and Workflow -- 4.2 Dataset -- 4.3 Experiments -- 5 Results and Analysis -- 5.1 Results -- 6 Conclusion and Future Work -- References -- The Use of Partial Order Relations and Measure Theory in Developing Objective Measures of Explainability -- 1 Introduction -- 1.1 Background on Explainability -- 1.2 Purpose of the Paper -- 2 Related Work -- 2.1 The Need for an Objective Measure of Explainability -- 2.2 Explanation Selection -- 3 The Use of Partial Order Relations in the Domain of Explainability -- 3.1 Some Background on Partial Order Relations -- 3.2 Explainability as a Partial Order Relation -- 4 The Use of Measure Theory in the Domain of Explainability -- 4.1 Some Background on Measure Theory -- 4.2 Assigning a Measure to an Explanation -- 5 Unifying Partial Order Relations and Measure Theory Through the Law of Diminishing Returns -- 5.1 Background on the Law of Diminishing Returns -- 5.2 Application of the Law of Diminishing

Returns in the Context of Explainability -- 6 Conclusion -- References -- AI and Law -- An Evaluation of Methodologies for Legal Formalization -- 1 Introduction and Motivation -- 2 Literature Overview of Recent Legal Formalization Approaches -- 3 Overview of Evaluation Methods of Legal Formalization -- 4 What Constitutes a ``Good'' Formalization? -- 4.1 Correctness -- 4.2 Transparency -- 4.3 Comprehensibility.

4.4 Multiple Interpretations Support -- 5 The Legal Experts Evaluation Methodology Proposal -- 6 Conclusions -- References -- Risk and Exposure of XAI in Persuasion and Argumentation: The case of Manipulation -- 1 Introduction -- 2 Background and State of the Art -- 2.1 Explainable AI -- 2.2 Explainable AI Through the Lens of Legal Reasoning -- 2.3 Persuasion and Manipulation: The Impact on Self-authorship -- 3 Legal Entanglements Beyond XAI -- 3.1 Can Data-Driven AI Systems be Actually Transparent for Non-expert Users? -- 3.2 Can the Mere Fact of Giving an Explanation Make the System Safer? -- 3.3 Can the Explanation Make the User ``Really'' Aware of the Dynamic of the Interaction? -- 3.4 Desiderata for a Not Manipulative XAI -- 4 Conclusions and Future Works -- References -- Requirements for Tax XAI Under Constitutional Principles and Human Rights -- 1 Introduction -- 2 Requirements for Tax XAI under Constitutional Principles -- 3 Requirements for Tax XAI under the European Convention on Human Rights (ECHR) -- 3.1 Right to a Fair Trial (Article 6 ECHR) -- 3.2 Right to Respect for Private and Family Life (Article 8 ECHR) -- 3.3 The Prohibition of Discrimination (Article 14) in Conjunction with Other Provisions of the ECHR and its Protocols -- 4 Preliminary Proposals to Meet the Explanation Requirements under the Constitutional Principles and the ECHR -- 5 Concluding Remarks -- References -- Author Index.