1. Record Nr. UNINA9910561298803321

Autore	Holzinger Andreas
Titolo	xxAI - Beyond Explainable AI : International Workshop, Held in Conjunction with ICML 2020, July 18, 2020, Vienna, Austria, Revised and Extended Papers / / edited by Andreas Holzinger, Randy Goebel, Ruth Fong, Taesup Moon, Klaus-Robert Müller, Wojciech Samek
Pubbl/distr/stampa	Cham : , : Springer International Publishing : , : Imprint : Springer, , 2022
ISBN	3-031-04083-X
Edizione	[1st ed. 2022.]
Descrizione fisica	1 online resource (397 p.)
Collana	Lecture Notes in Artificial Intelligence, , 2945-9141 ; ; 13200
Altri autori (Persone)	GoebelRandy FongRuth MoonTaesup MullerKlaus-Robert SamekWojciech
Disciplina	006.31
Soggetti	Artificial intelligence Machine learning Artificial Intelligence Machine Learning
Lingua di pubblicazione	Inglese
Formato	Materiale a stampa
Livello bibliografico	Monografia
Note generali	Description based upon print version of record.
Nota di contenuto	Editorial -- xxAI - Beyond explainable Artificial Intelligence -- Current Methods and Challenges -- Explainable AI Methods - A Brief Overview -- Challenges in Deploying Explainable Machine Learning -- Methods for Machine Learning Models -- CLEVR-X: A Visual Reasoning Dataset for Natural Language Explanations -- New Developments in Explainable AI -- A Rate-Distortion Framework for Explaining Black-box Model Decisions -- Explaining the Predictions of Unsupervised Learning Models -- Towards Causal Algorithmic Recourse -- Interpreting Generative Adversarial Networks for Interactive Image Generation -- XAI and Strategy Extraction via Reward Redistribution -- Interpretable, Verifiable, and Robust Reinforcement Learning via Program Synthesis -- Interpreting and improving deep-learning models with reality checks -- Beyond the Visual Analysis of Deep Model Saliency -- ECQ^2:

Quantization for Low-Bit and Sparse DNNs -- A whale’s tail - Finding the right whale in an uncertain world -- Explainable Artificial Intelligence in Meteorology and Climate Science: Model fine-tuning, calibrating trust and learning new science -- An Interdisciplinary Approach to Explainable AI.-Varieties of AI Explanations under the Law - From the GDPR to the AIA, and beyond -- Towards Explainability for AI Fairness -- Logic and Pragmatics in AI Explanation.

Sommario/riassunto

This is an open access book. Statistical machine learning (ML) has triggered a renaissance of artificial intelligence (AI). While the most successful ML models, including Deep Neural Networks (DNN), have developed better predictivity, they have become increasingly complex, at the expense of human interpretability (correlation vs. causality). The field of explainable AI (xAI) has emerged with the goal of creating tools and models that are both predictive and interpretable and understandable for humans. Explainable AI is receiving huge interest in the machine learning and AI research communities, across academia, industry, and government, and there is now an excellent opportunity to push towards successful explainable AI applications. This volume will help the research community to accelerate this process, to promote a more systematic use of explainable AI to improve models in diverse applications, and ultimately to better understand how current explainable AI methods need to be improved and what kind of theory of explainable AI is needed. After overviews of current methods and challenges, the editors include chapters that describe new developments in explainable AI. The contributions are from leading researchers in the field, drawn from both academia and industry, and many of the chapters take a clear interdisciplinary approach to problem-solving. The concepts discussed include explainability, causability, and AI interfaces with humans, and the applications include image processing, natural language, law, fairness, and climate science.