

1. Record Nr.	UNISA996472069203316
Autore	Holzinger Andreas
Titolo	XxAI - Beyond Explainable AI [[electronic resource]] : International Workshop, Held in Conjunction with ICML 2020, July 18, 2020, Vienna, Austria, Revised and Extended Papers
Pubbl/distr/stampa	Cham, : Springer International Publishing AG, 2022
ISBN	3-031-04083-X
Descrizione fisica	1 online resource (397 p.)
Collana	Lecture Notes in Computer Science ; ; v.13200
Altri autori (Persone)	GoebelRandy FongRuth MoonTaesup MüllerKlaus-Robert SamekWojciech
Soggetti	Artificial intelligence Machine learning
Lingua di pubblicazione	Inglese
Formato	Materiale a stampa
Livello bibliografico	Monografia
Note generali	Description based upon print version of record.
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.
