1.	Record Nr.	UNINA9910678248603321
	Autore	Sreevallabh Chivukula Aneesh
	Titolo	Adversarial Machine Learning : Attack Surfaces, Defence Mechanisms, Learning Theories in Artificial Intelligence / / by Aneesh Sreevallabh Chivukula, Xinghao Yang, Bo Liu, Wei Liu, Wanlei Zhou
	Pubbl/distr/stampa	Cham : , : Springer International Publishing : , : Imprint : Springer, , 2023
	ISBN	9783030997724 3030997723
	Edizione	[1st ed. 2023.]
	Descrizione fisica	1 online resource (314 pages)
	Disciplina	005.8
	Soggetti	Artificial intelligence Data protection Artificial Intelligence Data and Information Security
	Lingua di pubblicazione	Inglese
	Formato	Materiale a stampa
	Livello bibliografico	Monografia
	Nota di bibliografia	Includes bibliographical references.
	Nota di contenuto	Adversarial Machine Learning -- Adversarial Deep Learning -- Security and Privacy in Adversarial Learning -- Game-Theoretical Attacks with Adversarial Deep Learning Models -- Physical Attacks in the Real World -- Adversarial Defense Mechanisms -- Adversarial Learning for Privacy Preservation.
	Sommario/riassunto	A critical challenge in deep learning is the vulnerability of deep learning networks to security attacks from intelligent cyber adversaries. Even innocuous perturbations to the training data can be used to manipulate the behaviour of deep networks in unintended ways. In this book, we review the latest developments in adversarial attack technologies in computer vision; natural language processing; and cybersecurity with regard to multidimensional, textual and image data, sequence data, and temporal data. In turn, we assess the robustness properties of deep learning networks to produce a taxonomy of adversarial examples that characterises the security of learning systems using game theoretical adversarial deep learning algorithms. The state-of-the-art in adversarial perturbation-based privacy protection mechanisms is

also reviewed. We propose new adversary types for game theoretical objectives in non-stationary computational learning environments. Proper quantificationof the hypothesis set in the decision problems of our research leads to various functional problems, oracular problems, sampling tasks, and optimization problems. We also address the defence mechanisms currently available for deep learning models deployed in real-world environments. The learning theories used in these defence mechanisms concern data representations, feature manipulations, misclassifications costs, sensitivity landscapes, distributional robustness, and complexity classes of the adversarial deep learning algorithms and their applications. In closing, we propose future research directions in adversarial deep learning applications for resilient learning system design and review formalized learning assumptions concerning the attack surfaces and robustness characteristics of artificial intelligence applications so as to deconstruct the contemporary adversarial deep learning designs. Given its scope, the book will be of interest to Adversarial Machine Learning practitioners and Adversarial Artificial Intelligence researchers whose work involves the design and application of Adversarial Deep Learning.