Teresa Numerico, *Big data e algoritmi. Prospettive critiche*, Carocci, Rome, 2021.

The expression "big data" means everything and nothing. Literally, it conveys the idea of a large quantity, but it tells us little about quality. Among jurists and software engineers, the definition is not univocal, even though there is the agreement that "big data" refers more to the processing than to data type. In fact, a large amount of information alone is insufficient to warrant talk about big data, as technological tools are needed to manage and analyse it.

Major tech companies such as Amazon, Apple, Facebook, Google, and Microsoft created big data when they first had to cope with the explosion of information capable of revolutionizing the entire industry. Names, phone numbers, addresses, biometric data, pictures, geolocations, payments, credit card numbers, searches, comments, and much else are recorded and stored on servers in real time through smartphones and personal computers, providing significant statistics not only upon aggregate analysis but also when examined individually.

Teresa Numerico's book focuses on big data from a critical, philosophical perspective and helps us understand the problems and opportunities it creates.

Big data and algorithms build correlations, regularities, and quantifications to propose interpretations of social phenomena based on mathematical automatisms. However, as Numerico explains, it is an illusion to think that the automated understanding of habits and events can be objective and neutral: "Artificial intelligence technologies aspire to define what has been and to anticipate the future, but they have been invented and developed by human beings and therefore preserve their genius, instability, prejudices, often even arrogance."

Numerico, an associate professor of logic and philosophy of science at Roma Tre University, clarifies that relying on algorithms to make decisions in uncertain contexts, such as those found in real life, allows those who define the criteria to do so in the shadow of the opacity of the technical device, even at the risk of unfair outcomes.

The book alerts readers to the urgent problems of the current data revolution: "Those who are the object of automatic decision-making processes must obtain explicitly shared explanations for their choices, whether public or private. Algorithmic evaluations cannot be the alibis for a new *latinorum* of the international techno-aristocracy."

Regarding the structure of the book, chapter one introduces the genealogy of artificial intelligence and its correlation with human memory, through the history of philosophy, up to cybernetics and the human-machine relationship. Chapter two addresses the issue of big data between technique and politics, analysing the definition of big data and its ambivalence by bringing ontological, epistemological, and critical considerations to bear on the matter. The author poses the question: "Who owns, controls, exploits personal data?", reflecting on privacy, data protection, and "surveillance capitalism," as theorized by Shoshana Zuboff in The Age of Surveillance Capitalism: *The Fight for a Human Future at the New Frontier of Power* (Profile Books Ltd, 2018).

Chapter three focuses on data correlations and unintended consequences, treating some fascinating case studies: Facebook experiments in data science, the failure of Google Flu Trends, Cambridge Analytica, and predictions of behaviour based on social data.

This case-studies-analysis takes us to the issue at the heart of Numerico's research: *the illusory neutrality of the algorithm*. With chapter four the analysis turns to philosophy and ethical issues related to the use of technology; the author asks who is responsible for algorithmic decisions and explains when we can and when we cannot speak of algorithmic neutrality. Chapter six gives further definitions of algorithms, the concept of fairness, and the obstacles on the road to true algorithmic fairness. It underlines the key role of big data in medicine and its great mission for the future: naturalizing inequalities through a new digital welfare.

Chapter seven discusses habit, addiction, homophily, and segregation, all issues that prompt a study of "artificial intelligence employees." Face recognition and discrimination in discussions on social networks belong to this section.

According to the author, it is not correct to speak of "risks" related to artificial intelligence. The most successful A.I. techniques are those based on machine learning and deep learning, whose operation depends

on the availability of large amounts of data, both structured and unstructured. The obstacles to be overcome include the volume of data required for the processing mechanisms to function and the preservation of competitive advantage.

Data become inaccessible because no one is cognitively able to control what is inside the databases used to train the machine learning systems, employed afterward for A.I. Consequently, there is an asymmetrical relationship between those who hold the data (large internet companies or third-party actors) and those to whom the data belong and who provide them in exchange for the use of internet services. The persons whose data are processed have no control over the interpretations devised by data holders, which have no legitimacy and are subject to error, arbitrariness, misunderstanding, and exploitation. The algorithm, then, is not neutral.

Numerico offers a detailed discussion of predictive mechanisms. The first is the principle of induction, which holds that if a situation occurred in the past, it will tend to recur. This approach does not go unchallenged, for it is a conservative representation of society. From this principle and from the data on women's employment, which is lower than men's, one would infer that men are better than women at finding jobs and more efficient workers. Put simply, the combination of the discriminatory structure of society and the principle of induction will tend to maintain, indeed amplify, the status quo. One might speak of erroneous predictions, but this is not entirely true. The fact is that these predictions tend to be normative and verify the reality they are "suggesting" with their recommendations.

Concerning the data in machine learning, the principle of similarity presides over interpretative inferences of pattern recognition: objects of analysis that are similar in some characteristics will tend to behave similarly in specific pre-determined contexts.

And yet, even in this case, discriminatory logic is found lurking. If I say that people belonging to a particular ethnic group are all slackers, you jump up and call me a racist. But if the algorithm says that people who live in this neighbourhood have a high probability of not repaying their debts, no one balks at this. However, things are different when we consider that some people who cannot afford to pay higher rents consequently live in certain areas. The residents of a given neighbourhood are only similar in that they lack purchasing power, but they are discriminated against and cannot buy goods on credit because some of their neighbours have not repaid their debts.

In short, how we group people based on similarities and differences is a consequence of judgments and interpretations. That such interpretations may be given in algorithmic and quantified forms does not exempt

them from the arbitrary exercise of subjective evaluation mechanisms, which must always be very clear when attributing to algorithms the ability to make decisions that affect people's lives in important ways, such as access to healthcare or welfare, jobs, or ascription of the likelihood of criminal recidivism.

The goal, then, is to achieve true algorithmic fairness. How? The principle driving the algorithmic mechanism must be clear; the method for creating the software must be accessible; the database used to train the algorithm and the criteria used to classify and label the data must be available.

To ensure algorithmic fairness, the evaluation process must be controllable, and it must be possible to establish data justice. The individual's right to an explanation for the decision must also be protected, and this right must be clearly set out, just as we expect judges or public officials to issue readable and understandable judgments for which they are fully responsible. It is hard to see why evaluation performed by a machine should take place in a way that reduces the guarantees for those concerned.

Critical studies on algorithms are the only way to understand the subjective and interpretative nature of every decision. This does not mean abolishing algorithmic tools, which would be anachronistic, but understanding that a judgment is not purely the result of a choice of representation and interpretation of the phenomenon to be evaluated. Teresa Numerico's book enlightens us about the issues at the core of the debate on A.I. technologies and our future.

Alessandro Albanese Ginammi
Università per Stranieri di Perugia, Italy