CONCLUSION

Can a computer devise a theory of everything? This question is debated in 2020 in an article in the *New York Times* that presents the possibility that a machine equipped with the most sophisticated deep-learning technologies, capable of sifting data for patterns and autonomously discovering basic formulas of physics, could connect the findings into a unified theory of the universe.[1] It would seem a surprising outcome of the debate on the "end of theory" started by Chris Anderson in 2008 in *Wired*,[2] but in fact, the premise is the same: "The end might not be in sight for theoretical physics. But it might be in sight for theoretical physicists."[3] The theory would be produced by self-programming machines and would be incomprehensible to humans and independent of our forms of reasoning. To deny this possibility, according to Tegmark, would be to engage in a form of "carbon chauvinism."[4]

But is it really convenient to put it in these terms? What is the point of an incomprehensible theory, and do we need one? Within days of that *New York Times* article, the possibility of using the autonomous work of algorithms to deal with the classic "protein folding problem" was announced: machine learning was able to predict the three-dimensional shape of a protein given the string of amino acids that compose it, in a way completely different from human research.[5] This development could speed the discovery of new drugs and improve the treatment of viruses and diseases—but not because the algorithms had devised a theory, which in any case would be incomprehensible. Instead, they had learned to reliably predict the shape of proteins by analyzing thousands of known cases and their physical shapes. If the result of their prediction is reliable, is the basic issue really "to have machines that can think like a physicist," as Jesse Thaler wishes?[6] Or is it rather to find a way to communicate efficiently with these incomprehensible machines, in order to use their results and control possible errors or undesired effects?

As I have argued in the previous chapters, the latter is the challenge posed by the development of nontransparent algorithms, capable of learning and programming themselves. Communication is a complex and multifaceted process, and observing the working of algorithms in this perspective discloses a multiplicity of fascinating and difficult questions. Some problems dissolve, others take a different form, and still others arise.

For example, the controversial question at the basis of the Turing test *dissolves*: how do we know if our interlocutor is a human being or a machine? Seventy years after the publication of the article in which the test was proposed,[7] and after countless discussions and comments, the answer did not come from some elaborated version of the test, but simply from interaction with algorithms: in most cases the answer does not matter at all. If anything, it is the machines that have to make sure with some version of captcha,[8] that their interlocutors are people. Rather, what matters is that the communication works, that the partner participates in an interesting, informative, reliable, and even entertaining, way. Normally there is not the time and the motivation to question whether it is a machine or a human being.

Other issues, such as the thorny problem of bias, take another form. Bias has become one of the most discussed and difficult issues to deal with in all practical applications of artificial intelligence.[9] Algorithms are biased, as we know very well, and they produce biased results. Facial-recognition systems are more accurate in identifying white faces than those of other people,[10] programs used to predict crime disproportionately target certain ethnic groups and certain neighborhoods,[11] artificial intelligence chatbots on the web tend to post racist and offensive tweets.[12] In many fields, including insurance, advertising, education, and credit scoring, the use of algorithms and big data can lead to decisions that increase inequality and discrimination.[13] Yet do these outcomes depend on the kind of intelligence that is artificially produced in the algorithms?

Crawford, among many others, claims that "like all technologies before it, artificial intelligence will reflect the values of its creators."[14] Is this really the problem, when we are dealing with machines that program themselves? After all, it is not entirely clear that the designers of the algorithms are actually the creators of the procedures that are put in place. It seems to me that the problem in this case is not so much that the working of the machines reflects the biases of their creators, but on the contrary, that it is biased in large part because their workings do *not* reflect their creators' values. Algorithm designers inevitably have their own prejudices, conscious or not, and in the field of AI they are predominantly white and male.[15] It is very likely, then, that the algorithms themselves will be shaped accordingly. The most significant problem, however, is not that algorithms reflect the biases of their creators—who, granted, do tend to be white men. Rather, algorithmic bias is only one component of the problem. Deeper, and more difficult to manage is what is often labeled as "data bias," which does not depend on the values of the programmers.[16] Instead, it depends on the underlying source of the algorithms' efficiency: the access to the big data they find on the web, which frequently builds upon the uncoordinated input of billions of participants, sensors, and other digital sources. Machines participate in a communication that is neither neutral nor egalitarian, and they learn to work correspondingly, in ways that can be biased very differently from the preferences of their designers.[17] In pursuing the goal of algorithmic justice,[18] then, the most difficult problems are communicative, not cognitive.

Finally, other problems *arise* when the focus shifts to communication. Practical experience with the use of algorithms for specific tasks, now accumulated in many fields, has almost inadvertently led to the emergence of diverse, and extremely complex, issues related to their involvement in communication. In law, for example, "mechanical jurisprudence"[19] is already a reality: computational systems of legal reasoning are capable of exploring legal databases, discovering patterns, identifying the relevant rules, applying them, generating arguments—and even explaining their chain of reasoning to the users.[20] However, the problems that arise and animate the debate do not concern the fact that the author of the reasoning is a machine. As Canale and Tuzet claim, “Jurisdictional motivation does not consist in the psychological account of the process that led to the decision, but in the indication of the legal reasons that justify it.”[21]

It is communication that must work, and it is not simple. How can the fundamental ambiguity of legal argumentation be reproduced in communication with algorithms?[22] To account for the inevitable variety of cases and contexts, legal arguments are “typically vague and ambiguous,”[23] and legal communication must be “susceptible of more than one reasonable interpretation.”[24] The task of lawyers—as Garfinkel claims—is to make ambiguous the interpretations of facts and laws.[25] For algorithms, however, ambiguity is notoriously a challenge. Machines not only struggle with understanding the ambiguity of human communication, they struggle harder to *generate* ambiguous communication—that is, to use in competent ways the ambiguity required by legal arguments.

The debate about the difference between explanation and interpretation in law reflects this difficulty.[26] What machines do to make their decision transparent—as demanded by explainable AI—is to illustrate in detail the procedural steps through which their decisions were produced. This requires a “decision analysis in microscopic refinement,” far beyond what is produced in communication between human beings.[27] For effective communication among humans, it is sufficient to regulate “the presentation, not the production of the decision.”[28] Interpretation can and often must remain vague, because it is “not concerned with how we understand or produce texts, but with how we establish the acceptability of a specific reading thereof.”[29] To explain their decisions, lawyers and judges must provide a convincing account, which does not necessarily imply that they reconstruct all passages of their reasoning—and their recipients interpret them as they choose. When coping with algorithms, instead, interpretation often coincides with explanation—without the required space for vagueness and without using ambiguity. Paradoxically, one could say that the problem of interpretation in legal argumentation—even and precisely when dealing with algorithms whose processes are hidden from human intelligence—is not that the machine does not explain enough, but that it explains too much, and too precisely. The problem is not how the machine works, but how it communicates.

Problems of this kind, which affect all sectors of society in specific forms, cannot be grasped, let alone dealt with, without an adequate theory of communication and a thorough knowledge of different social domains. The analysis of the most pressing problems related to the use of algorithms in our society is not only a technical issue, but first of all a communicative issue—an issue of artificial communication.

This is a section of doi:10.7551/mitpress/14189.001.0001

Artificial Communication

How Algorithms Produce Social Intelligence

By: Elena Esposito

Citation:

Artificial Communication: How Algorithms Produce Social Intelligence
By: Elena Esposito
DOI: 10.7551/mitpress/14189.001.0001
ISBN (electronic): 9780262368865
Publisher: The MIT Press
Published: 2022

The open access edition of this book was made possible by generous funding and support from the MIT Libraries.

The MIT Press

The MIT Press would like to thank the anonymous peer reviewers who provided comments on drafts of this book. The generous work of academic experts is essential for establishing the authority and quality of our publications. We acknowledge with gratitude the contributions of these otherwise uncredited readers.

This book was set in Stone Serif and Avenir by Jen Jackowitz. Printed and bound in the United States of America.

Library of Congress Cataloging-in-Publication Data

Names: Esposito, Elena, author.
Title: Artificial communication : how algorithms produce social intelligence / Elena Esposito.
Description: Cambridge, Massachusetts : The MIT Press, [2022] | Series: Strong ideas series | Includes bibliographical references and index.
Identifiers: LCCN 2021013271 | ISBN 9780262046664 (hardcover)
Subjects: LCSH: Telecommunication—Social aspects. | Artificial intelligence—Social aspects. | Online identities. | Social intelligence.
Classification: LCC HM851 .E765 2022 | DDC 303.48/33—dc23
LC record available at https://lccn.loc.gov/2021013271

10 9 8 7 6 5 4 3 2 1