BOOK REVIEW

The Doctor and the Algorithm. Promise, Peril, and the Future of Health AI S. Scott Graham (2022) 272pp., £33 hardback, Oxford University Press, New York, ISBN 978-0-19-764446-1

Building on the promise of the new smart technologies being developed worldwide, and their interoperability with each other, we can say that we live in what Mireille Hildebrandt (2016, p.77) calls the 'onlife world'. Within this new environment, we are going to find a myriad of interactions that present us with lots of opportunities – but also risks. The posterchild that depicts these opportunities and threats is artificial intelligence (AI) and machine learning (ML) in particular. In the context of the new spring of AI, we can find works on this technology that, on many occasions, try to offer us multidisciplinary and complete views of Calabresi's Cathedral.[1] Their outputs vie to be the next candidate for the new single and unified theory of AI. These works are fascinating but, in our very dynamic world, certain industries need more refined approaches. One of these industries is the healthcare industry.

For centuries, the medical profession has been the primary source of analogies and inspiration for inventors and academics who have tried to explain the role of naturoids in our societies.[2] For instance, Aristotle described the organic functioning of living beings, including humans, building on proto-mechanical automata theories, as we can verify in his *Politics* and *De Generatione Animalium*. More recently, John von Neumann (1966) has developed his *Theory of Self-Reproducing Automata,* trying to find patterns that could allow us to develop and/or identify artificial life. It should not be surprising, then, that the healthcare industry is playing an active role in the development and deployment of AI, as the reader can verify in Scott Graham's book.

This work is most timely. Using the same dynamic systems theory employed to develop AI solutions in the Big Data era, it is possible to argue that the original Hippocratic oath, which stated that prognostics should be the first art of medicine, was developed following the same model of interaction between inputs and outputs. Of course, this is a highly complex art, highlighting the elements that configure the need for both natural and artificial intelligences in conditions of imperfect information and scarce resources. In the field analysed by the author, this means that we cannot talk about the development of perfect and neutral rules, given that we have to consider that every patient is different and that these very differences will foster the development of various biases within these smart architectures. For instance, Scott Graham presents the case of physicians who could take as granted certain false beliefs about biological differences among patients with a range of backgrounds. Furthermore, we could argue that these biases can be found in the processes of development and selection of the technologies that are being used to address the needs of institutions, medical professionals and patients. This selection of technology has to be addressed and analysed further, building on the recommendations and conclusions offered by the author.

These biases are to be found throughout the value chains that support both the hard and the soft elements of these systems. In stark contrast to most enthusiasts of the technologies covered by the book, the author makes clear that he is 'not that optimistic about the promise of health AI'.

---

[1]Calabresi (1972, p.1090) argues that to describe reality, one has to present and analyse different arguments, just as one would do to understand Monet's Cathedral. If you want to understand it, you have to see all Monet's paintings of the cathedral at Rouen.

[2]This term is employed to describe 'any real artifact coming from the attempt to reproduce natural instances' (Negrotti, 2012).

DOI:10.13169/prometheus.39.1.0051

I agree with this position, but I would have enjoyed a deeper analysis of the value chains behind these applications, given the economic and political implications that the author presents in the subsequent chapters. Yes, we can agree, in the context of the covid-19 pandemic, that we witnessed, and even experienced, the benefits of deploying different AI systems to forecast the spread of the disease, develop new drugs to tackle it and identify signals of health anomalies across different databases. However, 'algorithms are opinions embedded in code' (O'Neil, 2017, p.21) that, by themselves, cannot avoid the emergence of market imperfections. One might even argue that they foster the emergence of these imperfections in information asymmetries.

Graham opens chapter 1 ('How to make an AI') by stating that when we think about AI, we tend to visualize it as intellectual maps and instructions; however, in opposition to ideas that tend to show and highlight a path, the AI world could be labelled a 'black box' society in which many individuals and companies that ordinary users cannot identify will play a role (Pasquale, 2015). This is one of the most important points developed in the work. AI systems do not exist in isolation. They are the output of very long and complex value chains while, at the same time, acting as the input for more complex agents and systems. So, if an AI deployed in the healthcare industry hurts a patient, we cannot conclude immediately that the doctor is liable for this unfortunate outcome. We have to develop more complex analyses and include principles to find an answer within these algorithmic black boxes to see who is responsible for this damage.

In chapter 1, the author narrows the scope of his argument to the area known as deep medicine, which relies on developing a specific technique: machine learning. Most works on markets in disciplines that go beyond informatics tend to present very complex explanations how these technologies work, increasing the black box effect. However, in this book, the reader will find simple definitions and descriptions that will aid understanding of machine learning and its uses within the medical field. For this purpose, Scott Graham uses the case of cancer diagnostics, which is not restricted simply to either cancer or no cancer, but has to ask which type of cancer and develop the four main activities that are behind the operation of machine learning: 1) data curation, 2) feature engineering, 3) model training and 4) benchmarking.

Building on these four activities, the author opens chapter 2 ('Digital oracles') by arguing that, following the spirit of the Hippocratic oath, these systems could act as digital oracles that could, in turn, complement another set of four activities that have defined Western medicine: 1) diagnosis, 2) prognosis, 3) treatment and 4) prevention. The difference that these oracles will introduce to our medicine is that they can rely on Big Data and its three Vs (velocity, volume and variety) to present individualized outputs that can address the needs of individual patients almost immediately. Furthermore, taking advantage of their value chains and the intellectual property rights that support them, we can talk about permanent monitoring of the patient through the introduction of the systems under analysis in an internet of things (IoT) environment in which different smart systems and providers can address different elements of the patient's environment. In so doing, potential unhealthy behaviours can be identified and access to a constant stream of high-quality data can be obtained.

There are challenges described in many works on the role of the General Data Protection Regulation (GDPR) in the development of AI that are related to the development of trustworthy AI beyond the traditional expert programs built around decision trees. This has been the main concern for regulators, as one can verify from efforts of the US Food & Drug Administration (2019) which highlight the differences between locked algorithms and learning and adaptative algorithms. The author agrees that the algorithms that lie behind the deep medicine model introduce some elements that were neither considered nor designed for these adaptative technologies, and that could have catastrophic consequences 'especially in life-critical applications'. On this point, machine learning systems have been criticized for their lack of transparency; however, it could be that a category error is at fault (Pasquale, 2015, p.16). By overestimating the capabilities of AI and claiming that a system's actions cannot be understood by hospitals, doctors and patients following a simple formula, we ascribe values to mechanical automata and not to the humans who designed, built and

applied them (Kroll, 2018, p.2). Furthermore, this inscrutability is not a result of technical complexity, but of power dynamics in the choice of how to use tools whose inner workings are not fully understood. This fact shows how information asymmetries can be obstacles to the diffusion of knowledge within the discipline and, of course, in medical practice.

With these asymmetries in mind, Scott Graham introduces chapter 3 ('How to make it as an AI') with a very interesting story about the work of James Lind's *Treatise on Scurvy*, famous for its length and its complexity, features that reduced its impact on publication. In the context of AI, this is very important given that the number of companies and developers that can put in place a system that can create a vaccine to face covid-19 or label a cancer is significantly reduced. This means that only a handful of natural and legal persons can set the rules for the development of these systems and, consequently, exercise their market power on medical centres and patients. In answer to this problem, the author refers to the concept of 'open science' which comprises open access, open data, open source and open reproducible research. Through open science, it is possible to distribute the know-how behind these intelligent systems to improve the quality of the solutions offered, and thus to address the real needs of patients.

One of the most exciting and vital contributions in this book can be found in chapter 4 ('The search for ground truth'). In addition to recounting an exciting story about the Lazarus syndrome, this chapter introduces its readership to the term 'groundtruthing' which can be paired with the analysis of machine learning presented in chapter 1. Accordingly, Graham describes groundtruthing as 'the art for the research practices involved in cultivating a data set against which a prototype of AI will be measured'. This is a significant contribution given that it highlights some practices that could be employed to face the biases referred to above. For instance, to avoid an AI trained to diagnose like one specific doctor, we can involve more than one doctor and researchers in the labelling of data to reach a consensus that will be translated into better decision-making that will address the needs of patients beyond a single socio-cultural niche.

With echoes of science fiction, the author presents chapter 5, which I would have preferred to see as chapter 2. Findings and developments in health AI tend to be exaggerated through the use of specific language and platforms to distribute, and to find, new content on the current and future states of these applications (see, e.g., Bootle, 2019). These exaggerations could have an impact on the way we see the future of our economies and the way we react to the changes fostered by the introduction of AI in industries such as healthcare. An illustration of this can be found in the everyday use of the argument that AI will save money, for instance, through the development of cheaper drugs. To evaluate the extent of hype in deep medicine, Graham introduces a system by which one can identify exaggerations using the groundtruthing techniques developed in chapter 4.

Chapter 6 ('Ethics, justice, and health AI') presents an analysis of the challenges that emerge from the market exercised by the stakeholders behind the development and deployment of machine learning as a complement to the medical profession in the Fourth Industrial Revolution. Among the elements considered is the existence of biases, transparency that is difficult to achieve and control over the data being processed throughout the value chains that support deep medicine. As a transition to Chapter 7 ('Regulating health AI'), this chapter acknowledges that we have to work on the ethical principles that will act as the cornerstones for the development of these technologies and their regulatory frameworks. These, as the reader will infer, are supported by interaction among the market, civil society and the state.

Though the book is not enriched by analysis of such documents as the GDPR and the draft of the European AI Act, it is still a very good introduction to a specific industry that is very close to us. We can understand – as potential and current patients – the technological and social dynamics that lie behind every decision as the output of a human doctor. This book will help the reader to reduce the information asymmetries that result from exaggerations about the virtues of AI and the fears that push us to avoid the best solution available in the market. *The Doctor and the Algorithm* gives us the tools to take part in discussions and even in policy-making related to the new healthcare regulations that will soon be put in place around the world.

References

Bootle, R. (2019) *The AI Economy. Work, Wealth and Welfare in the Robot Age,* Nicholas Brealey, London.

Calabresi, G. (1972) 'Property rules, liability rules, and inalienability: one view of the cathedral', *Harvard Law Review*, 85, pp.1089–1128.

Hildebrandt, M. (2016) *Smart Technologies and the End(s) of Law,* Edward Elgar, Cheltenham.

Kroll, J. (2018) 'The fallacy of inscrutability', *Philosophical Transactions of the Royal Society*, 376, pp.1–14.

Lind, J. (1753) *Treatise on Scurvy*, Edinburgh.

Negrotti, M. (2012) *The Reality of the Artificial. Nature, Technology and Naturoids,* Springer, Heidelberg.

O'Neil, C. (2017) *Weapons of Math Destruction. How Big Data Increases Inequality and Threatens Democracy,* Broadway Books, New York.

Pasquale, F. (2015) *The Black Box Society. The Secret Algorithms that Control Money and Information,* Harvard University Press, Cambridge MA.

US Food & Drug Administration (2019) *Proposed Regulatory Framework for Modifications to Artificial Intelligence/Machine Learning (AI/ML)-Based Software as a Medical Device (SaMD)*, available at https://www.fda.gov/media/122535/download (accessed October 2019).

von Neumann, J. (1966) *Theory of Self-Reproducing Automata,* University of Illinois Press, Urbana, IL.

Israel Cedillo Lazcano
Universidad de las Américas Puebla
israel.cedillo@udlap.mx