



BOOK REVIEWS

We the Robots? Regulating Artificial Intelligence and the Limits of the Law BY SIMON CHESTERMAN [Cambridge: Cambridge University Press, 2021. xx + 289 pp. Hardcover: USD39.99]

Debates about the law, regulation, and governance of robots, artificial intelligence (“AI”), and machine learning have been dominated by voices from North America, Europe, and Australia. While Singapore has hosted conferences at which these matters have been discussed, and while it is clearly an interested party in relation to these debates, a distinctively Singaporean voice has yet to be heard—or, at any rate, it had not been heard prior to the publication of Simon Chesterman’s stimulating book.

Broadly speaking, when lawyers debate the regulation and governance of AI, we find three particular kinds of conversation. The first conversation (“Law 1.0”, as I term it in Roger Brownsword, *Law 3.0: Rules, Regulation and Technology* (2020)), which is typically led by private lawyers, focuses on the application of traditional legal principles to emerging technologies. Famously, Warren and Brandeis celebrated “the beautiful capacity for growth which characterizes the common law” and which enabled them to articulate a right to privacy that would respond to the unacceptable use of the latest photographic technology of their time (Samuel D Warren and Louis D Brandeis, “The Right to Privacy” (1890)); but, to what extent can we rely on the principles of the common law in the face of today’s technologies? The second conversation (“Law 2.0”) asks whether current regulatory environments are fit for purpose relative to these technologies. Topics in this conversation include whether we are making the right regulatory interventions at the right time, whether legislation is properly connected to the technology and is sustainable, whether regulation is “acceptable” (covering both prudential judgments and questions of legitimacy), whether regulation has found the sweet spot between over-regulation and under-regulation, and whether the law is “effective” (whether laws and regulatory interventions “work”). In both these conversations, technology is, so to speak, “out there”, to be engaged by the law and to be subjected to regulatory control. By contrast, the focus of the third conversation (“Law 3.0”) is on employing technologies, particularly digital technologies (such as LawTech and RegTech), for the better performance of legal and regulatory functions, for channelling conduct, monitoring and securing compliance, managing conflicts and resolving disputes. Chesterman’s book, with a focus on the public regulation of artificial intelligence, does not so much introduce a wholly new conversation as engage robustly and insightfully in all three conversations.



Like many books, the story in Chesterman's book is told in three parts: Challenges; Tools; and Possibilities. However, this is not a book that explicitly employs any of the familiar templates or scaffolds for organizing a conversation about regulating technologies in general or AI in particular. Rather, the framing in this book focuses on three particular characteristics of AI (its speed, autonomy, and opacity)—so, to this extent, it is technology-specific; and, at the same time, in a less technology-specific way, it offers “three discrete ways to view [regulatory dilemmas, namely] through the lenses of practicality, morality, and legitimacy” (at p 244)—these enabling us to focus on “the practical management of risk associated with new technologies, the morality of certain functions being undertaken by machines at all, and the legitimacy gap when public authorities delegate their powers to algorithms” (at p xv).

In his book, Chesterman tells many stories, some more familiar and to be expected than others—including well-known stories of AI and flash crashes, tragic collisions, and bots, but also surprising stories such as that of the pig of Falaise. These are stories that are well told and that are well worth telling. However, what is Chesterman's own story, the story that he is presenting in the book? In his final summarizing chapter, Chesterman explains that in the first part of the book (Challenges), his intention was to “set the stage” (at p 244) by introducing the framing concepts that I have already mentioned. In the second part (Tools), the idea was to “[bring] out the available props in the form of responsibility, personality, and transparency” (at p 244). Then, in the final part (Possibilities), three particular “plot developments” were outlined, these concerning “rules, institutions, and the role that AI might play in regulating itself” (at p 244). Nevertheless, readers might be puzzled as to quite what the storyline is.

The story certainly starts with the three challenges presented by AI, namely: the *speed* of “processing power and connectivity rather than innovation” (at p 29); the emergent *autonomy* of AI enabled machines; and, processual *opacity* (in the sense of “the quality of being difficult to understand or explain” (at p 65)). The critical point, however, is that these incipient characteristics of AI will become even more pronounced unless some controlling or corrective measures are taken. In other words, AI will get even quicker, it will move towards ever greater autonomy, and its operations will become even more complex and opaque. The question at the end of Part One is how regulators should respond to a technology with these particular characteristics and these tendencies.

In Part Two, Chesterman's recourse to the “available props” answers this question somewhat obliquely. One thought is that regulators might try to arrest these incipient developments or even turn them back. Regulators might try to slow down AI, or restrict its autonomy, or make its operations more explainable. We can treat the chapter on transparency as an exercise of this kind: if regulators were to require more transparency, then that should alleviate concerns about opacity. However, it is not so simple. Regulators would need to ask what exactly it is that is to be made transparent, when, and to whom. Not only might transparency prove to be counter-productive if it enables regulatees to game the system, but it will not suffice to prevent “inferior, impermissible, or illegitimate decisions” (at p 169). Another thought is that, if regulators are unable to arrest the incipient development of AI, then they need to have a fallback position and particularly so as AI becomes



progressively more autonomous (and comparable to autonomous humans). At that point, the live question will be whether we should treat AI as the bearer of legal duties or the holder of legal rights—and, at that point, we have an invitation for the discussions of responsibility and personality that take up the other two chapters in this part of the book.

In Part Three, the first line of plot development, concerns rules. Here, we get down to serious regulatory business with some important reflections on why, when and how to regulate, but all this in the shadow of the thought that it is not a lack of rules that is the problem. Rather the problem is engaging with “distributed systems operating with humans in, out of, and over the ‘loop’ across jurisdictional boundaries” (at p 194). The problem, in other words, recurs to a certain interpretation of “speed” (*qua* distributed technologies and cross-jurisdictional features) and the obvious question raised by AI autonomy. Quite rightly, Chesterman anticipates that, if we are to rise to this challenge, it will need both “new ways of thinking about regulation and, in particular, new institutions to implement it” (at p 194). Cue, the second line of plot development: new institutions. Here, Chesterman is largely content to leave basic safety and risk management to industry standard-setting; but, drawing inspiration from the International Atomic Energy Agency he makes a significant (and, I suspect, against the grain) proposal for an International AI Agency that would set global red lines in relation to AI; and he also proposes that an AI ombudsperson, or the like, should *inter alia* plug whatever regulatory gaps are found. The third line of plot development joins the Law 3.0 conversation about the various ways in which technologies such as AI might assist or take over legal and regulatory functions. Like many others, Chesterman is sceptical about how far we can go with the idea of robot judges or converting legal texts into machine readable and actionable formats, but his particular interest is in relying on a supervisory AI to monitor the operations of AI. In the same way that we talk about “privacy by design” (or blockchain and cryptography for opacity) might we design out harmful AI applications? While there is some mileage in this thought, the more that the technology takes over the regulatory work, the greater the risk that humans are not in control and that, neither in the foreground nor in the background is there a human who can clearly be held to account for harms caused by the AI. Our reservation, as Chesterman rightly discerns, is that law is an essentially human enterprise; it is far from perfect; but, even if governance by AI might be more efficient and effective, it just is not human.

To return to my earlier remarks, readers will find that all three law and technology conversations figure in the book. While Chesterman joins the Law 3.0 conversation in the closing chapter of the book, in general, the discussions belong to Law 1.0 and Law 2.0 conversations.

In relation to the Law 1.0 conversation, Chesterman’s thinking is in tune with that of Warren and Brandeis. Thus, he emphasises that “the underlying [legal] principles are sound” (at p 86); and, a persistent and pervasive theme is that there are fewer legal gaps than we tend to assume (see *eg*, the discussion of tort law in managing the risk of physical injury presented by autonomous vehicles and allocating compensatory responsibility; the remarks at p 120 in the chapter on personality; and at pp 217-220). As he repeats in his concluding remarks, “[f]or the most part, existing laws and institutions can deal with these challenges” (at p 244).



That said, a great deal of the book is concerned with the Law 2.0 question of how we might get the regulatory environment right. On the one hand, Chesterman thinks that we should not exaggerate the problem: the regulatory environment is in pretty good shape, or it can be got into shape without this involving rocket science. On the other hand, we need to be prepared for cases where we have “undesirable harms [caused by AI] that fall outside existing regimes of public control” (at p 86). In the middle part of the book the implication (on my reading of those chapters) is that the regulatory choice is between controlling AI or adjusting to living with it; and, in the third part of the book, it seems that new approaches and new institutions will be required.

The story is a continuing one but does it give reason to think that we will flourish in our AI-enabled worlds? That is hard to say. We might feel reassured by Chesterman’s view that the regulatory environment is to a considerable extent in good shape or that it is easily put in shape. That said, we might wonder why so much work has to be put into making this point. Might it be that, too often, our scattered intelligence about the governance of emerging technologies is so poor that we do not appreciate that we actually do know how to regulate a particular technology or particular technological characteristics? On the other hand, where the regulatory environment is not in good shape, we might worry that we appreciate all too well that we do not know how to regulate the technology (such as technologies that are distributed and cross-border) or we might suspect that there are cases where we do not know whether or what regulatory steps we should be taking (and AI that becomes progressively more autonomous might well be in this category).

In sum, there is a great deal to like about this book. It is good to welcome a Singaporean voice (and a new angle of approach) to the debates about the regulation of AI; and, it is terrific to see Chesterman leading the way in proposing new global regulatory institutions. Let’s hope that the story does end well and that we can save humanity from itself.

ROGER BROWNSWORD

Professor of Law

King’s College London and Bournemouth University