BOOK REVIEW

Algorithmic Regulation edited by Karen Yeung and Martin Lodge (2019) Oxford University Press, Oxford, hardback £74, 294pp. ISBN: 9780198838494

Is regulation an end in itself? Is it (or should it be) a normative concern? Does regulation encompass law, overlap it or is it distinct phenomenon from law? And what vision of law and governance should it represent? These questions have swirled in this domain for decades and are contested both inside and outside the field. But with the rise of the 'algorithm', an often-misunderstood and much-abused term, we are seeing – as Lodge and Menneckin point out in chapter 8 – a flattening of the discussion. In this flattening process, many of the political and conceptual nuances of the field are collapsed under the weight of a cybernetic, neoliberal model of governance that views citizens as 'end-users', 'customers' and 'rule-takers', rather than full participants in a societal structure that seeks, even if it doesn't always or even often succeed, to empower them as the authors of their own inter-connected lives.

The European Commission recently published its draft proposal on a new artificial intelligence (AI) regulation, an instrument that will undoubtedly impact many of the themes explored in this volume, even if it is likely that the enacted version of the proposal will differ significantly from the current text (see European Commission, 2021**).** In that respect, some of what is discussed here has been overtaken by events, but that is perhaps one of the values of the book: it foreshadows many of these issues, and whatever happens with the development of the law, it stands as an excellent snapshot of contemporary thinking on algorithmic regulation at this moment in time (or 2019, at any rate). Scholars of the future will look back at this tumultuous period with fascination, trying to understand the reasons why the laws we made were the laws we made. With its wide spectrum of both themes and worldviews, *Algorithmic Regulation* will provide them with many threads on which to pull.

The remainder of this review briefly summarizes the main contributions of the book's chapters, which are grouped into three parts: normative concerns, public sector applications and governing algorithmic systems. I conclude with some thoughts on the overarching contribution of the volume and on the desirability of algorithmic regulation more generally.

In the introductory chapter, editors Yeung and Lodge set the context for the rest of the volume. Identifying the notion of algorithmic regulation (hereafter AR), that they adopt from Yeung's own work, namely:

> decision-making systems that regulate a domain of activity in order to manage risk or alter behaviour through continual computational generation of knowledge from data emitted and directly collected (in real time on a continuous basis) from numerous dynamic components pertaining to the regulated environment in order to identify and, if necessary, automatically refine (or prompt refinement of) the system's operations to attain a pre-specified goal. (Yeung, 2018, p.5)

Yeung and Lodge frame the debate around several primary concerns: a lack of discussion of algorithms in public law as compared with other fields, such as data protection; the need for caution against both naïve techno-solutionism and a techno-determinism which prioritizes unfettered innovation and/or assumes that regulatory responses will invariably be slow or manipulated for the benefit of Big Tech (p.7); the tendency to view innovation as an unqualified good (although some later chapters seem to adopt this perspective); and recognition of the insights that science and technology studies (STS) can provide on the social embeddedness of technology, AR being simultaneously both technological and social (p.8) (alas, this particular strand of analysis is not picked up in any detail later in the book).

DOI:10.13169/prometheus.37.4.0387

As the authors identify, the deeper question is one of legitimacy, not of the effectiveness (however that might be measured, and according to whose perspective) of individual approaches or applications.[1] Much AR is built around a particularly capitalist rationality that is assumed to be a given. It is not, and the question is what mechanisms of counter-balance allow for: are they part of the same logic or do they reflect the need to protect something more fundamental, a kind of meta-logic of social ordering? Yeung and Lodge's concerns about the threats AR poses to the fundaments of liberal constitutional order, including fundamental rights, democracy and individual dignity and freedom, are crucial. Tackling these issues requires cross-disciplinary expertise from law, politics, applied philosophy, computer science, organizational sociology and public administration.

In the second chapter, Yeung asks a fundamental question: what is it that we are concerned with when we worry about decision by machine? (p.21). She characterizes concerns about AR along three dimensions: (i) the decision-making process, (ii) the outcome or outputs of that process and (iii) the predictive personalization of services, each of which has implications at the individual, group and societal level. This provides a useful matrix for structuring analyses of AR, and for identifying areas that have received comparatively little attention so far.

The question of process legitimacy is an important one, with a rich background in public law literature. The concern is not just that the outcome is in some sense good, but that it is arrived at in a way that both respects the governing ideals of society, and also reflexively sustains systems and institutions of governance so that citizens can have faith in them. Process-based concerns identified by Yeung include the lack of an identifiable human responsible for the decision; a lack of participation in the algorithmic decision-making process or the ability to contest its outcomes; the use of unlawful or discriminatory variables in AR systems; the lack of reason-giving (as opposed to merely explanation); and, perhaps most viscerally, the dehumanization of those affected by the decisions being made.[2]

Taken together, these process concerns demonstrate a normative concept of regulation that is closer to legality and the rule of law, with the commitments to equal respect and checks and balances that these represent; indeed, the chapter later highlights the threats posed to these by prediction and personalization. Yeung highlights, for example, the question of dignity and the importance of protecting the individual's ability to make moral choices (pp.30–1), which she in turn connects to the important theme of communication as a constitutive element of social relations. Without recognition of situated experience, which includes ensuring that those affected feel listened to and have meaning attributed to their experiences, the notion of regulation quickly devolves into an instrumental mechanism of control, the veil of technocratic objectivity obscuring a very particular (and ideological) view of how society ought to be structured and governed.

The analysis does not end with the process, however; results are important too. Here Yeung considers erroneous and inaccurate decisions, biased and discriminatory decisions that are unjust and the ethical concerns that surround the use of algorithmic systems that mimic humans unbeknownst to those who are interacting with them.

The third pillar of Yeung's framework is prediction and personalization. What does it mean for an AR system to optimize its performance against some goal, chosen according to the (commercial) logic of its designer, when the effects of that optimization have a direct impact on the action possibilities of the citizen? This is not simply nudged according to the ostensibly benign creed of libertarian paternalism: it is hyper-nudged into conforming with an expectation that might bear little resemblance to any notion of self-actualization, moral contemplation or practice of social solidarity (pp.34–7). The results of recent Faustian bargains struck between political parties and large social

---

[1]This is the focus of my own work on 'digisprudence', in which I argue that architectural design must be made tractable to the law if it is to be deemed acceptable in a democracy, and that processes of technology production can be mobilized in service of this goal (see Diver, 2022).

[2]Consider, e.g., the Post Office Horizon scandal, a 20-year miscarriage of justice caused by the erroneous outputs of computational systems (see BBC News, 2021).

media platforms would seem to confirm the role of AR in various high-profile political maladies that have struck over the past few years (see Geoghegan, 2020).

Where to go from here? Yeung suggests framing ethical concerns in the legal language of justice – rights, wrongs and harms – the law being, of course, an enforceable mechanism that might just have the teeth to address the 'chronic asymmetry of power between those who design, own, and implement these algorithmic decision-making systems . . . and the individuals whose lives they affect' (p.38). Yeung discusses the deficiencies of relying on the contested notion of risk, rather than concretely specified rights and wrongs – of and to individuals, groups and society as a whole – that can in turn assist in the legal attribution of responsibility and all the reflexive consequences this entails. Yeung's analysis does a real service to the discourse on regulation by retaining focus on the harms that are deleterious to the kinds of society we want to build and maintain, in contrast to analyses more commonly built around detached, technocratic notions of risk that fail to recognize the need to identify and sustain a normative conception of what governance for the social good should look like.

Chapter 3 takes a welcome excursion into the research designs of AR systems. Scantamburlo *et al*. argue that decision-makers must internalize and apply normative principles of justice, lawfulness and the protection of rights. They are somewhat equivocal, however, in their suggestion that 'there is thus no guarantee that intelligent algorithms will necessarily internalize accurately, or apply effectively, the relevant normative principles' (p.51). Even to couch this as an aim is to invite the possibility that they ever could internalize and apply anything normative, which I think is a dangerous path to follow. Nevertheless, the argument that legitimate AR requires critical assessment of the 'interplay of key technical and normative concepts' is an important one, highlighting the need to develop an internal perspective on the technology that considers its production and not simply its effects once it is operating in the world. Following a useful summary of machine learning technology, and an eye-opening analysis of the research design of HART (the infamous tool used by Durham police to make custody decisions), the authors conclude by suggesting four normative benchmarks for automated decision-making (ADM): (i) prediction accuracy, (ii) fairness and equality, (iii) transparency and accountability and (iv) informational privacy and freedom of expression. The claim made that high accuracy alone is normatively insufficient is welcome, particularly given the tendency of AI marketers to tout their systems' accuracy as the only relevant measure of value. Is it possible, however, to design in the authors' other benchmarks, without 'computationalising', and thus changing, their fundamental character? On the one hand, the authors suggest that the increased use of ADM is inevitable (p.75), but simultaneously they argue that it is essential that such systems 'are made remain (*sic*) capable of incorporating and acting upon appropriate normative principles and obligations' (p.76). If the latter turns out not to be possible, what should our response be? Eventually we will have to grapple with the notion that we should simply outlaw systems that fail to meet such benchmarks. Indeed, this may be on the horizon: see the prohibited and 'high risk' classes of AI application in the Commission's proposed AI regulation (European Commission, 2021, articles 5 and 6).

Criado and Such's chapter considers the challenges inherent in forcing real-world square pegs into algorithmic round holes. First, they set out forms of discrimination recognized in law (direct discrimination, including explicit/implicit and intentional/unintentional distinctions, and indirect discrimination, or disparate impact), before considering how these manifest in the ontological characteristics of AR systems. Like other software, models seek unambiguous specifications of real-world phenomena, but are incapable of representing socially constructed features that are not susceptible to such reduction, such as what makes a job candidate good or bad. As the authors note, proxies and data biases can be reflexively magnified by reinforcement learning, while there are monetary incentives to crowbar expensive machine learning (ML) models into areas that were not considered in their research design and training. Indeed, to the old adage 'all models are wrong, but some are useful' might be added 'and some might even be harmful'. In that vein, the authors highlight the importance of importing legal definitions of harmful discrimination into the world of AR. But definitions alone are not enough for legality, and the chapter's argument in favour of developing

automated methods risks collapsing the processes of law into merely a question of requirements specification and box-ticking. This concern aside, the chapter provides a valuable review of the literature on algorithmic discrimination, and highlights the importance of engagement between law and ML research design.

Danaher's chapter reflects his familiar analytic style in considering the numerous ways AR systems inveigle their way into the everyday fabric of our existence. His appreciation of the profoundly normative role of socio-technical artefacts as more than simply tools is valuable in this area, *pace* the Silicon Valley 'tech bros' whose dissembling use of that word obscures how their commercial interests are deeply implicated in the design of the products they make 'freely' available. In that vein, the central question for the chapter is the effect of AR on autonomy. He adopts Raz's definition of the latter, which requires that individuals have (i) the rationality and ability needed to plan actions in service of achieving specified goals, (ii) an adequate range of choices and (iii) freedom from coercion and manipulation in making these choices. Danaher considers AR's impact to lie more on the second and third conditions through the limitation of our 'choice architecture' in ways that reflect commercial expediency rather than our preferred self-concept (p.106). Interestingly, he adopts the lens of neo-republication theory and the idea of non-domination as a requirement of autonomy, noting the ability of algorithms to micro-dominate our lives (pp.108–9). But is this more a question of degree than of kind? Danaher's position is somewhat ambivalent; he notes that while the 'scope, scape, and speed' of AR are novel, it is not possible yet to tell what the global effects will be (p.112). So, too, are his prescriptions ambivalent, on the one hand suggesting that individuals have the 'ultimate power to accept or reject the influence of these algorithmic tools' (p.113) (this may be true of the author, but is it true of everyone else?) while on the other acknowledging that individuals may need help to resist AR built by companies 'deeply committed to practices that are, shall we say, not always respectful of autonomy' (p.113). Closing appeals to Frischmann and Selinger's 'right to be off'' (Frischmann and Selinger, 2018) and to the notion of a legal right to attention protection are welcome. I'm not sure, however, that Danaher's centring of autonomy, or at least a notion of it that relies so heavily on rationality, is ultimately an equitable basis for protecting individuals against algorithmic overreach.

With the chapter from Veale and Brass, the volume shifts to consider the role of AR in managing and delivering public services. They posit three levels of analysis and evaluation – macro, meso, and street level (p.122) – and contrast automation of decision-making with its augmentation, noting the difficulties that arise from assuming a task to be rote, with its resulting formalization in an AR design inevitably crystallizing one of many possible interpretations of how it should be performed. Connecting back to normative themes discussed earlier in the volume, the authors note the necessity of judgement in provision of public services that seek to be distributed equitably and effectively (p.124), which in turn requires not simply explanation of a decision, but its justification according to some overarching philosophy of provision (p.131). At the macro governmental level, coordination is needed across bodies involved in public sector AR to provide codes of best practice, stronger rights (such as those in the General Data Protection Regulation, GDPR) that can provide enforceable legal mechanisms for protection and redress, and public research centres that can ensure governmental technology strategy is properly informed and self-directing rather than led by the whim of Big Tech. The meso level targets the interplay between policy and implementation, where a balance must be struck between ministerial and civil service pronouncements of policy and their translation into the design of an AR system. Methods for achieving this include version control, testing and internal and external peer review (p.134), although the authors make the important point that these must not substitute a technical standard of effectiveness for a normative one: 'even when a model seems to "work", it may not "work" for everyone' (p.134). On the micro (street) level, tensions arise between ADM and the decisions of frontline bureaucrats at point of delivery. Ultimately, the questions raised by these various dimensions highlight the need for cross-disciplinary research across public administration, computer science, sociology, law and anthropology.

Griffiths's analysis is built around case studies of two failed public sector AR approaches: the prediction of National Health Service (NHS) inspection outcomes by the Care Quality Commission in England and Wales, and the prediction of the outcomes of reviews of higher education institutions by the Quality Assurance Agency. Griffiths's exposition of the controversies around these two systems, including details of their design, is an excellent warning to those who believe complex social institutions can be modelled using contemporary machine learning approaches. This is something he implicitly acknowledges when he suggests that 'judgements of care quality are subjective, intangible, and difficult to capture by indicators alone' (p.162). One might assume this should disqualify many AR applications from the outset. Griffiths seeks nevertheless to identify several normative requirements that might legitimate the use of AR. As articulated, these seem to me to hole AR below the waterline, at least in the contexts identified in the chapter: Griffiths argues that data must be 'timely, robust, granular in terms of the unit being assessed (e.g., a hospital ward or department rather than a collection of hospitals)' (p.171), but what are the assumptions inherent in identifying the 'correct' level of granularity? Furthermore, data must be simultaneously 'not too specific (e.g., the budget shortfall or surplus of a hospital, rather than spending on a specific type of catheter)' and 'of sufficient volume so that an effective model can be developed based on the underlying patterns in the data set' (p.171). How can we balance the inductive and contingent nature of ML's *ex post* pattern recognition with *ex ante* judgements of granularity and specificity that necessarily consider context at the point of assessment? The latter form of judgement is surely central to the exercise of effective public administration. It seems that, in the end, the sheer number of hurdles that Griffiths suggests an AR system would need to clear for its design to be acceptable points clearly toward the elephant in the AR room, namely the conclusion that such systems should not be employed for such purposes at all lest their precariously assembled designs produce much more uncertainty than they take away.

Lodge and Mennicken's chapter demonstrates a sensitivity to this more fundamental question. Their analysis identifies how the push for AR has flattened an otherwise multi-dimensional and contested discourse on the nature and purposes of regulation. The authors describe several problematic and inter-related concerns: first, discussions about AR tend to obscure different views about the nature and purpose of regulation (indeed, this can be detected throughout this volume); second, AR transforms existing practices in ways that are not immediately apparent; third, AR creates new, additional administrative problems that are distinctive to the new paradigm; and lastly, AR highlights the fundamental question, who will regulate the regulators? (p.180). The presumption of a shared worldview obscures important questions of political priorities – what they are, who holds them and why – which, in turn, leads to the presumption that control by an objective, cybernetic AR is self-evidently worth pursuing (pp.181–5).

Following an excursion into the epistemology of algorithms, including data quality, bias and the vexed question of what indicators actually denote success, the authors turn to the important issue of meta-regulation. They identify four primary means of achieving this (pp.195–6): first, centralized oversight can develop standards which might include algorithm auditing and internal testing; second, the defining of procedures for dealing with bias through de-discrimination, coupled with explanations that can facilitate external scrutiny; third, the prescription of maximum transparency on operations of algorithmic regulation to facilitate scrutiny and indeed competition; and fourthly, reliance on disciplinary standards, where designers of algorithms are subject to fiduciary duties akin to those in such professions as law and accountancy.

In the end, the fundamental question is one of legitimacy, and the compatibility of AR with the constitutional order. Can a balance be struck between the need for contestability that this entails, and the notional benefits of AR? A potential Catch-22 arises: contestability requires transparency, but transparency might allow the system to be gamed (p.197). As the authors note, AR in general lacks 'enhanced ethics, sustained oversight, and appropriate legal frameworks to establish procedural standards and prescribe degrees of scrutiny' (p.197). It might be that ultimately machine learning algorithms are not the best, a good or even valid medium through which to achieve the aims of public administration in a constitutional democracy.

Andrews's chapter opens the final part of the book, 'Governing algorithmic systems', by identifying two primary concerns: first, how does the use of AR by government and enterprise undermine existing laws; and second, to what extent is there a risk of human intelligence being undermined, or controlled, by intelligent machines? As Andrews rightly asserts, the question of whether and how democratic governments can exercise power is central in these debates (p.204), particularly when contrasted with the size and power of Big Tech firms, some of which have themselves been compared with sovereign states in recent years. Andrews argues that governance readiness requires states to have discursive power – the ability to frame the debate, and to influence both cognitive and normative views – but that this has been ceded in recent years to Big Tech (p.206). Governments have, however, a unique ability to legitimize their positions by convening relevant parties in the development of policy (experts, think tanks, professional bodies). They must use this capacity to redress the informational imbalance with technology firms which have direct access to whatever is state-of-the-art (p.209). Deliberative and representative democracy requires nothing less.

The penultimate chapter, by Lohr *et al.,* takes a somewhat instrumental view of law, one that clings to the utilitarian and market-driven assumptions of much AR and of law and economics discourse (and so highlights the range of perspectives reflected, both explicitly and implicitly, in *Algorithmic Regulation*). The authors propose a four-layer approach to conceptualizing AR: first, dealing with the risks and benefits of AR using existing legal mechanisms, and especially contract; second, using corporate governance to control AI systems; third, developing sector-specific laws to regulate risks in particular sectors, such as banking and autonomous vehicles; and fourth, grappling with 'more diffuse horizontal harms to society' by means of 'a light-touch regulatory framework' (pp.224–5). Diffuse harms are those that 'sector-specific regulators would not be in a position to detect and regulate' (p.225). This is a troubling position. One might think that harms of this kind – undetectable by those with narrow domain expertise, but nevertheless significant enough to affect society more broadly – might be appropriate targets for the opposite approach, one based around precaution and greater regulation, at least initially while the nature and reflexive effects of the harm are being identified and made sense of. Later in the chapter, the authors consider the role that contracts can play in moderating AI risk, discussing the intellectual property value of a trained model, attribution of liability where AR outputs cause harm, auditing the provenance of a model in vertical supply chains, as well as using contractual provisions to ensure compliance with regulatory instruments, such as the GDPR (pp.234–9). The authors also highlight the requirements implied by corporate boards' fiduciary duties of care and diligence, and the bearing these might have on any decision to employ AR (p.240).

The chapter concludes with some further comments on the limits of omnibus AI regulation – 'a general AI regulator would be ill-equipped to conduct risk analyses for the different kinds of AI applications that could potentially cause harm' (p.243) – and the issue of diffuse AI harms. For the latter, the authors suggest that key performance indicators for measuring societal harms be developed, akin to the approach suggested by the European Commission in relation to online disinformation. Contrary to the authors' preference for sector-specific regimes, it remains to be seen whether it will be possible to detect such diffuse harms without some form of omnibus regulator that has the resources, expertise and high-level cross-domain perspective necessary to analyse all potentially relevant areas of concern. This debate may in the end be moot, given the provisions in the EU's newly minted AI regulation for national supervisory authorities and a European artificial intelligence board (see European Commission, 2021, articles 30 and 56).

In the final chapter of the volume, Bygrave picks up his own 2001 analysis of the 1995 Data Protection Directive's provisions on automated profiling, comparing them with the equivalent regime in the GDPR. While the former made few waves in the courts, the latter is likely to have much greater impact, says Bygrave, not least because of confusion around its drafting.[3] Bygrave applauds the focus in the GDPR on data protection by design, but laments the provisions' lack of clarity for those

[3]Does the term 'including' in article 22 mean involving or as well as? The difference is significant. Similarly, does article 22(1) define a general prohibition, or a specific right that the data subject must invoke?

who would implement them – a problem that has been highlighted in privacy engineering circles for many years now. This is compounded if one draws a distinction between consumer-facing products and services, which might be comparatively simple to design in a compliant fashion, and the internet infrastructure beneath it all, governed as it is by transjurisdictional, technocratic engineering bodies, such as the internet engineering task force, which are subject to few mechanisms of legal, ethical or political accountability (p.257). Bygrave ends with the refreshing observation that data protection is but one dimension of the broader AR research landscape, which requires sensitivity beyond issues of personal data, and indeed beyond data *per se*, to encompass more fundamental normative questions about the nature of democratic constitutional order (p.260).

One sometimes gets the sense that the deployment of algorithmic regulation is seen as non-negotiable, and its deficiencies as temporary obstacles that simply require a few tweaks on the road to full effectiveness. Mixing technical and organizational compromises into an AR system might indeed allow us to achieve something that approximates the institutions and structures they replace. There is a profound risk, however, that the system's apparent effectiveness will be short-lived and skin-deep. These institutions and structures have been developing in theory and in practice over hundreds of years, and while AR might seem superficially better than what went before, its black box might harbour structural shifts whose consequences are not immediately apparent – quite apart from the usual concerns around transparency and explainability. These unforeseen reflexivities potentially undermine both what AR on its own terms claims to achieve, and the very conditions of possibility of democratic society itself. It's not clear that a Rube Goldberg approach to making AR work is more beneficial than the (admittedly messy and imperfect) processes and institutions it seeks to displace, at least not for those unenamoured with the political and economic logic AR often represents. At times, it seems AR is an answer in search of a question, which might not be a problem were the stakes not so high.

Apart from the breadth and high substantive quality of its individual chapters, *Algorithmic Regulation* is a fascinating volume in that it crosses a wide spectrum of positions on these questions, both implicitly and explicitly. This makes it an invaluable marker of the current state of thinking in this fundamental and rapidly developing field, the next chapter of which is likely to be dominated by the European Commission's proposed AI regulation.

References

BBC News (2021) 'Convicted Post Office workers have names cleared', 23 April, available at https://www.bbc.com/news/business-56859357 (accessed April 2021).

Diver, L. (2022) *Digisprudence: Code as Law Rebooted*, Edinburgh University Press, Edinburgh.

European Commission (2021) *Proposal for a Regulation on a European Approach for Artificial Intelligence*, available at https://digital-strategy.ec.europa.eu/en/library/proposal-regulation-european-approach-artificial-intelligence (accessed November 2021).

Frischmann, B. and Selinger, E. (2018) *Re-Engineering Humanity*, Cambridge University Press, Cambridge.

Geoghegan, P. (2020) *Democracy for Sale: Dark Money and Dirty Politics*, Head of Zeus, London.

Yeung, K (2018) 'Algorithmic regulation: a critical interrogation', *Regulation & Governance,* 14, 4, pp.505–23.

Laurence Diver
Faculty of Law and Criminology
Vrije Universiteit Brussel
laurence.diver@vub.be