Algorithmic Regulation and the Rule of Law

Hildebrandt, Mireille

Published in:
Philosophical Transactions of the Royal Society A: Mathematical, Physical & Engineering Sciences

DOI:
10.1098/rsta.2017.0355

Publication date:
2018

Document Version:
Final published version

Link to publication

Citation for published version (APA):
Hildebrandt, M. (2018). Algorithmic Regulation and the Rule of Law. *Philosophical Transactions of the Royal Society A: Mathematical, Physical & Engineering Sciences*, *376*(2128), [20170355]. https://doi.org/10.1098/rsta.2017.0355

PHILOSOPHICAL TRANSACTIONS A

rsta.royalsocietypublishing.org

Research

Cite this article: Hildebrandt M. 2018 Algorithmic regulation and the rule of law. *Phil. Trans. R. Soc. A* **376**: 20170355. http://dx.doi.org/10.1098/rsta.2017.0355

Accepted: 25 May 2018

One contribution of 13 to a discussion meeting issue 'The growing ubiquity of algorithms in society: implications, impacts and innovations'.

Subject Areas:
cybernetics

Keywords:
written law, code-driven regulation, data-driven regulation, machine learning, cybernetics

Author for correspondence:
Mireille Hildebrandt
e-mail: mireille.hildebrandt@vub.be

Algorithmic regulation and the rule of law

Mireille Hildebrandt[†]

Metajuridica, Faculty of Law and Criminology, Vrije Universiteit Brussel, Pleinlaan 2, 1050 Elsene, Brussels, Belgium

MH, 0000-0003-4558-9149

In this brief contribution, I distinguish between code-driven and data-driven regulation as novel instantiations of legal regulation. Before moving deeper into data-driven regulation, I explain the difference between law and regulation, and the relevance of such a difference for the rule of law. I discuss artificial legal intelligence (ALI) as a means to enable quantified legal prediction and argumentation mining which are both based on machine learning. This raises the question of whether the implementation of such technologies should count as law or as regulation, and what this means for their further development. Finally, I propose the concept of 'agonistic machine learning' as a means to bring data-driven regulation under the rule of law. This entails obligating developers, lawyers and those subject to the decisions of ALI to re-introduce adversarial interrogation at the level of its computational architecture.

This article is part of a discussion meeting issue 'The growing ubiquity of algorithms in society: implications, impacts and innovations'.

1. Regulation by algorithm?

Computational systems increasingly 'infuse' governmental legislation, administration and adjudication. For instance, legislation may at some point be written in a way that is conducive to algorithmic application, administration may be automated, notably administrative decisions, and courts may employ artificial legal intelligence (ALI) to support judgment. This brief essay enquires what this means for the rule of law, raising a number of questions, such as: Are we confronting a conflation of legislation and administration, insofar as legislation becomes self-executing? Could legal judgment at some point be conflated with its prediction? Will

[†] Tenured research professor of 'Interfacing Law and Technology', Vrije Universiteit Brussel (VUB), appointed by the VUB Research Council at the research group of Law Science Technology and Society studies (LSTS), Faculty of Law and Criminology. Part-time full professor 'Smart Environments, Data Protection, and the Rule of Law', Radboud Universiteit, Nijmegen, at the institute of Computing and Information Sciences (iCIS), Science Faculty.

THE ROYAL SOCIETY
PUBLISHING

automated systems run on proprietary software or be too complex to explain? If these systems are not testable, can they be contestable? Do algorithmic decision systems require new types of interpretability, on the nexus of machine learning and law? Does this require a new legal hermeneutics?

To gain some insight into the nature of these questions and their answers, I first discuss two types of algorithmic regulation, namely 'code-driven' and 'data-driven'. This will, in turn, provoke a new perspective on current, modern law as fundamentally 'text-driven'. As algorithmic regulation takes over from human regulation, it becomes important to reconsider how the text-driven nature of modern, positive law determines what is called 'the force of law'. Such force, defined in terms of 'legal effect' and performative speech acts, must be distinguished from the behaviouristic underpinnings of the regulatory paradigm that understands the force of law as a matter of influencing behaviour. After distinguishing law from regulation, this essay will trace the transformation of text-driven *law* into code-driven and data-driven *regulation* [1], demonstrating how this may erode the grammar and the alphabet of modern positive law, while simultaneously pulling the carpet from under the rule of law.

In the final section, I argue that lawyers need to get their act together in the face of major investments in ALI and smart regulation. As I explain, this will require a new hermeneutics based on a proper understanding of the vocabulary and the grammar of machine learning (ML) and distributed ledger technologies. As an example, I argue for 'agonistic machine learning', which should bring adversarial contestation into the heart of the design of ALI.

2. Two types of algorithmic regulation

Algorithmic regulation refers to standard-setting, monitoring and behaviour modification by means of computational algorithms. Such algorithms may be self-executing, meaning that standard-setting integrates with behaviour modification. I call this code-driven regulation. Alternatively, algorithmic regulation may provide decisional support or advice, based on predictive algorithms that basically infer standards to better monitor, predict and influence behaviour. As these inferences are based on data analysis (by means of machine learning) I call this data-driven regulation. Even in the case of decision-*support* instead of decision-*making*, human intervention becomes somewhat illusionary, because those who decide often do not understand the 'reasons' for the proposed decision. This induces compliance with the algorithms, as they are often presented as 'outperforming' human expertise.

Code-driven regulation depends on IFTTT. IFTTT stands for 'if this then that', providing the fundamental logic for all algorithmic decision systems. This type of decisional logic is deterministic, entirely predictable and basically consists of simple or complex decision trees. Whoever determines 'this' as a condition of the 'that' decides the output of the system, which has no discretion whatsoever. Note that we are no longer in the realm of old-school legal artificial intelligence (AI), legal knowledge management or attempts to simulate legal judgment, but dealing with systems that are capable of actually taking decisions that affect legal subjects [2]. This type of regulation usually goes under the heading of 'smart regulation' of 'cryptographic law' [3], as it is currently associated with blockchain applications. The point here is that we are dealing with an entirely deterministic system that is self-executing. Though it may seem that the overdetermination and the lack of discretion imply complete transparency and the absence of interpretability issues, this is not at all the case. Such issues are, instead, hidden in the formalization that precedes the operations of the system [4,5]. An IFTTT decision system in the realm of law enables code-driven regulation, for instance, where the taxation office delegates specific types of decisions to a system capable of checking a series of conditions through feeds from specified databases (e.g. income, bank accounts, real estate and other assets), applying the relevant legal norms (e.g. income tax law, corporate tax), and issuing a decision on the amount of tax to be paid. Such decisions can be explained by referring to the decision trees that have been implemented, but whether this explanation also *justifies* the decision depends on how the legal norms have been translated into computer code. Legal norms are expressed in human language, which is notably ambiguous, and any particular interpretation is, therefore—in principle—contestable. This is not merely cumbersome compared to code-driven regulation, but—on the contrary—also protects us from over-inclusive as well as under-inclusive legal norms.

IFTTT thus enables administrative decision-making, which falls within the scope of the rule of law. Decisions based on code-driven regulation must be comprehensible for the entity that has the competence to take the decision, as such entity is accountable for the legality of the decision. Under the rule of law, such decisions must also be comprehensible for those *subject to the decision*, and they must be justifiable in the sense of satisfying the legal norms that allow for the decision. Administrative decisions taken by code-driven regulation must thus always be contestable on the double basis of: 'the decision is based on legal conditions that *do not apply* because the system got the facts wrong', and 'the decision is based on a *wrong interpretation* of the relevant legal norms'. The latter means that, even if the IFTTT standard applies, one can always appeal that this code-based standard is an incorrect application of the relevant legal norm in the case at hand.

Data-driven regulation is informed by ALI as it integrates data-driven legal tech with code-driven decision systems; data-driven regulation means that the code is informed by the data on which it has been trained instead of being informed by legal experts that have translated their insights into code [6–8]. ALI is mainly based on 'natural language processing', which entails that algorithms are 'trained on' legal text, using techniques like reinforcement learning and backpropagation to mine relevant argumentations, to detect relevant preceding case law, statutes, treaties and doctrinal treatises or to predict the outcome of future case law. ALI is entirely dependent on the choice of training data (legal text), its labelling and curation (marking what is deemed relevant for the output, i.e. a decision about whether or not a legal norm was violated), the hypothesis space that was developed (a set of mathematical functions that aims to connect input data with output data), and the performance metric chosen (percentage of output that is correct according to legal experts) [9]. ALI is not deterministic in the sense that IFTTT is. In point of fact, it contains a new type of discretion, situated in the design choices made when training the algorithms. These choices have many implications as to the reliability, the comprehensibility and the contestability of the output, but, as they are usually made by technology developers and policy makers, they are largely invisible to those who may be subject to decisions based on ALI. Each and every design decision has trade-offs, for instance between speed and accuracy, between generalizability and correctness or between accuracy and interpretability. There is an unwarranted aura of objectivity around the output of ML, notably concerning complex ML systems such as deep learning artificial neural networks that supposedly 'outperform' human experts, though whether this is actually the case depends on the solidity and contestability of the research design. Note that to test such 'outperformance' the expert opinion that is supposedly outperformed is used to decide about such 'outperformance', which raises a number of objections against accepting such judgements without further enquiry [10,11]. Alas, some authors contribute to magical expectations of ML (based on a flawed understanding of the underlying statistics), instead of confronting the bias that is inherent in any learning experience—whether human or machine [12].[1] Next to the problems raised by a new type of discretion that is inherent in the design of ALI, these systems also generate problems for their interpretability. It is important to note that the output of an ALI system can be explained at different levels: (1) at the level of the research design, explaining the choices made and the trade-offs they generate in terms of reliability, generalizability and explainability; (2) at the level of generalized output, consisting of inferred rules that supposedly determine argumentation lines or court decisions; and (3) at the level of individual predictions, targeting a specific case based on algorithms trained on preceding case law. Though explainability at all levels is important, we need to keep in mind that, in the end, an explanation is not the same as a justification; knowing how the algorithm came to its conclusion does not imply that the conclusion is 'in accordance with the law'. ALI is built on simulation, it tries to come as close as possible to what experts would have concluded, had they

[1] See section 2.7.3 'The Futility of Bias-Free Learning' on pp. 42ff. in [9].

been asked—it does not understand the decisions it makes or prepares; in a sense, it is parasitic on the human expertise it tries to approximate.

3. Law and regulation

To make sense of algorithmic regulation we must first come to terms with the relationship between law and regulation. The term 'regulation' has two different meanings that should be distinguished, though they are not mutually exclusive. They concern a different level of analysis. On the one hand, regulation is used to describe any attempt to influence the behaviour of a population, whether by means of law, by force, by nudging or by means of surreptitious manipulation. This way of using the term entails an external perspective on law. It is external because it 'thinks' in terms of behaviour instead of action and in terms of information instead of meaning. On the other hand, regulation may refer to the policy rules enacted by 'regulators', in the sense of administrative bodies endowed with regulatory or decisional competences by a legislator. Under this definition, a regulator is not a legislator, but an administrative body.[2] In the context of data protection, the US Federal Trade Commission is usually referred to as the regulator. This US terminology has contaminated EU discourse, where data protection authorities and the European data protection supervisor are now often qualified as regulators. The distinction between legislating (the prerogative of Parliament or Congress) and regulating (based on the attribution of decisional or regulatory competence by the legislator to an administrative body) is grounded in an internal perspective of law, assuming familiarity with concepts such as legality, which requires that government action is always based in law. Note that this type of regulation binds the regulator (the government body that acts based on competences attributed by the legislator) because those affected by its actions should be able to count on the regulator abiding by its own policy rules. The binding force of such policy rules can be indirect, based on the principles of trust and legal certainty, or direct, if the legislator has delegated regulatory competence to a governmental agency.

Though it is critical for lawyers to understand external perspectives on law, it is also critical to keep check of the internal perspective that is linked with constitutional protection and the checks and balances of the rule of law. Perhaps even more important, lawyers should keep in mind that the grammar of law is based on the concept of legal effect (internal perspective), which must be distinguished from mere enforcement or social consensus (external perspective).

I will call the external perspective on regulation 'cybernetic regulation' and the internal perspective 'legal regulation'. In Anglo-American discourse, the law is often seen as a subset of cybernetic regulation, alongside, for instance, market forces, social norms and techno-regulation or architecture (as Lawrence Lessig famously wrote) [14]. Indeed, the term 'regulation' stems from cybernetic and systems theory. In her seminal article on regulation, Julia Black defines regulation as [15]:

> [R]egulation is the sustained and focused attempt to alter the behaviour of others according to defined standards or purposes with the intention of producing a broadly identified outcome or outcomes, which may involve mechanisms of standard-setting, information-gathering and behaviour modification.

The advantage of this definition is that it allows her to render visible that law is 'just' one way of regulating society, while many other ways of regulating human behaviour may be highly effective even if they are not conducive to the checks and balances of the rule of law (for instance, no requirement of general application, publicity, contestability, etc.). Black's definition is based on the cybernetic theory that defines regulation as a combination of standard setting,

[2] From an external perspective, legislation is a specific type of cybernetic regulation. Some authors call both legislation and the rules enacted by other governmental bodies 'regulation'. For my argument this makes no difference, but as it increases confusion between cybernetic regulation and law I will avoid such terminology here. In my dictionary law is not regulation, because it does not treat people as pawns whose behaviour must be modified but as subjects capable of giving reasons for their actions. See also [13].

monitoring and behaviour modification [16]. Cybernetics shares its history with AI, having been 'invented' by one of the founding fathers of AI, Norbert Wiener, who was interested in a system's remote control over the behaviour of another system. He noted that this depends on a clear goal (standard setting), information about whether the goal is achieved (monitoring) and successful attempts to influence the behaviour of the other system (behaviour modification). His crucial contribution to AI, and core to cybernetics, is the emphasis on feedback loops, which imply the gaining of information about how the other system responds to one's attempts to influence it, followed by adapting one's own behaviour to be more successful [17]. The feedback loop enables adaptive behaviour, depending on the acuity of the observing agent that is trying to achieve its goal.

The cybernetic understanding of regulation makes a lot of sense, but as observed above it takes an external perspective on law, missing out on the domain specificity of law and the force of law. The fact that lawyers may take such an external perspective does not turn it into an internal perspective; it rather means that such lawyers fail to distinguish between a sociological or cybernetic account of law and the domain-specific operations of legal normativity which is closely aligned with the concept and the rule of law. Or, it means that lawyers are capable of switching between an internal and an external perspective—which I believe to be a very good exercise. To better understand both law and the rule of law, we need a mitigated internal perspective that clarifies where and how the operations of law differ from the operations of cybernetic regulation.

From the perspective of the rule of law 'legal regulation' is a subset of law and as such it is constrained by the legality principle, meaning that such regulation is *both more and less* than the cybernetic understanding of regulation purports. Law, including 'legal regulation', is focused not merely on the modification of behaviour, but on coordinating, prohibiting and enabling action *in a way that addresses individuals that are subject to law as capable of giving reasons for their actions and in a way that respects their autonomy*. This implies that *not anything goes* in terms of behaviour modification, as those whose behaviour is to be modified must be *treated as the authors of their own actions.* Scholars such as Paul Ricoeur and Neil MacCormick understand the law in terms of speech act theory [18–21], meaning that the force of law depends on the attribution of legal effect whenever specified legal conditions apply. Such legal effect cannot be understood in terms of 'influencing'; it is not a matter of causality. Instead, the legal effect creates 'institutional facts', such as ownership, contract and human rights. Legal effect means that certain acts 'count as' or 'qualify as', for example, murder or personal data processing. Institutional legal facts are based on the performative nature of speech acts and notably of legal text. In that sense modern, positive law is based on text—it is a *text-driven law*.

Modern law is the set of rules and principles that determine positive law; they establish what 'counts as' or 'qualifies as' a violation of a legal norm or a legal right. The rules and principles that constitute modern positive law are generated by the binding sources of the law: legislation, case law and treaties, in combination with doctrine, fundamental principles and customary law. They are enacted by legislators and courts (that produce the binding sources of the law) and applied and thus interpreted by government authorities. In a constitutional democracy, that interpretation can be challenged and the final word on how the law must be interpreted is with independent, impartial courts. This entails that government must be based on specified attributions of legal competence, whereas legal subjects are free to act unless constrained by law. Government bodies are thus always constrained by the legality principle; they may only act if competent and only in accordance with the rules and principles that govern discretion under the rule of law. Since legal certainty requires that citizens can foresee how discretion will be used, there is a space for 'legal regulation' which basically clarifies the policies developed by the regulator on how it will use its competences. As Dworkin argued:[3]

> [D]iscretion is not the absence of principles or rules; rather it is the space between them.

[3] [22], referring to [23].

As I explained in other work,[4] without rules or standards the concept of discretion makes no sense; the mere fact of being bound by a particular authority creates the need to judge whether a norm applies, what decision it calls for and how it should be performed. The rule-bound nature of discretion makes possible a discussion about the interpretation and employment of discretionary competences; it allows a learning curve by requiring those who intervene to give reasons for their actions if called to account. Those reasons are—in part—the norms that regulate their behaviour as public officials, but in the end, those reasons also include the situated interpretation of those norms. In that sense, discretion is not close to but the opposite of arbitrary rule.

From the perspective of the rule of law, regulation can be understood in the context of the difference between law and administration. Administration concerns decisions and actions of governmental bodies; administration is not law but it does fall under the rule of law (legality principle). As administration implies more or less discretion, it requires 'legal regulation', which can be seen as a subset of law that enhances legal certainty with regard to administrative decision-making.

4. Algorithmic regulation under the rule of law: from agnostic to agonistic machine learning

Currently, under the reign of modern positive law, we have text-driven law to determine what counts as lawful or unlawful. Core to text-driven law is its performative nature: *it does what it says*, by attributing legal effect when certain legal conditions are met. In this way, text-driven law makes our shared world predictable, as it requires that government agencies provide reasons based on the applicable legal norms, thus making government actions and decisions explainable and contestable. Once code-driven and data-driven regulation enter the playing field, we need to consider whether they are 'merely' a form of cybernetic regulation that requires adherence to the legality principle if employed by regulatory bodies, or whether at some point we could speak of code-driven and data-driven *law*. For now, that would seem far-fetched [18,19]. In the last part of this contribution, I will briefly present one way of bringing *data-driven regulation* under the rule of law, by notably reinstating the adversarial core of the rule of law for data-driven decision-making.

As discussed above, ALI simulates and predicts legal decisions, and if combined with code-driven regulation it can actually *make* decisions based on such simulation. As indicated, ALI hinges on a series of design choices that determine its validity, its reliability and also its contestability. The choice of a performance metric, which will determine its predictive accuracy, the choice of the training data and the development of the hypotheses space, all define the boundaries of the mathematical optimization to be achieved. As such, ALI generates a double interpretability problem. First, the legal text must be selected and translated into machine-readable data, by means of labelling and curation (removing irrelevant text). This is an act of translation and interpretation. Obviously, this can be done in different ways, depending on a whole number of contingencies, for instance, based on which text is publicly available. Second, each design decision entails a particular way of framing the problem that must be solved, for instance, the problem of predicting case law. Such framing implies an act of interpretation and generates a machinic interpretation. To grasp the assumptions and implications of the machinic interpretation, which will inform subsequent predictions or decisions of the ALI, we will need to develop a new hermeneutics, based on a new vocabulary.

If we take the article by Aletras *et al.* [25] on predicting judicial decisions of the ECHR as an example [25], we can highlight the difference made by the relevant design choices and demonstrate how they inform machinic interpretation. Their ML system was trained to 'predict' case law of the European Court of Human Rights (ECHR), based on a selection of relevant judgments. The researchers restricted the training data to published judgments, the assumption being 'that text extracted from published judgments bears a sufficient number of similarities with, and can, therefore, stand as a (crude) proxy for, applications lodged with the Court, as

[4] This paragraph is taken from [24], p. 417.

well as for briefs submitted by parties in pending cases' ([25]: 2/19). This is an example of opting for 'low hanging fruit' (easy to obtain training data), which raises some issues, as it implies that the system draws its conclusions based on the Court's articulation of the facts of the case. As the authors note, the Court probably formulates the facts in a way that is conducive to fit the conclusion ([25]: 5/19). This entails that no conclusions can be drawn about correlations between the facts of the case and the outcome, and although the authors seem to be aware of this, they nevertheless do conclude that the facts of the case have high predictive value. If the cases that were deemed inadmissible or struck out beforehand had been taken into account, this may have made a difference to the assessment. This potential difference now remains invisible. The conclusion that the facts of the case, rather than the applicable law, determine the outcome is also problematic because whenever the Court itself decides that a case is inadmissible, the judgment has no section on law, which means that the results are skewed.[5] The system was restricted to violations of articles 3, 6 and 8 of the ECHR, as this provided more balanced data (equal amount of violation/non-violation cases). This restriction, however, makes invisible how a case is framed, as many cases are framed as violations of more than one human right (for instance, combining articles 3, 6 and 8 with articles 5, 7, 9, 10 and 14 of the ECHR). Furthermore, the algorithm was trained only on existing case law, using 90% for training and 10% for validation—no out-of-sample testing was reported. This means that the system actually did not engage in *prediction* but remained in the realm of *describing* a historical dataset.[6]

It should be clear that data-driven legal tech is *not agnostic* in the sense of being unbiased, objective and neutral in its prediction of case law. To predict the outcome of a case, a system must be designed based on a series of design decisions that are not obvious and these decisions each have various types of trade-offs: speed may result in lower accuracy; high accuracy may correlate with less interpretability; huge datasets may result in a manifold of spurious correlations; whereas performance matrices may misrepresent how the system does in terms of the real world [27]. When employing data-driven legal tech we must, therefore, ensure that these design choices are the result of *agonistic debates* between data scientists, expert lawyers and the lay people who are subject to such decisions based on this kind of legal tech. It is only when such agonism nourishes the research design of machine learning that we can hope to build and employ systems that are reliable and contestable in real-world applications [28,29].

The notion of agonism derives from Chantal Mouffe's democratic theory [30,31], which complements and enhances both representative and deliberative versions of the democratic theory [32]. The idea is that robust, sustainable decision-making requires agonistic, adversarial debates between experts, policy makers and those who suffer the direct and indirect effects of decisions [32]. Ignoring dissent results in weak and unreliable decision-making; inviting dissent will broaden the hypotheses space in the metaphorical sense, and will better ground what machine learning experts call 'the ground truth'.[7] Inviting dissent will also demonstrate 'equal concern and respect' [23], which I dare say is the core of both democracy and the rule of law. Agonism should not be confused, however, with antagonism (saying 'b' whenever the other says 'a'); agonism means to turn enemies into adversaries, vouching for a decision-making process that takes into account the concerns of those who will be affected. The same idea has emerged in constructive technology assessment [34], which aims to involve 'users' in the early design stages of technological innovation. Constructive technology assessment builds on the fact that resilient and sustainable systems, that hold up in real-world environments, require constructive resistance and contestation throughout their design.

[5] As the authors admit ([25]: 11/19). See also [26].

[6] This is a crucial point, which marks the difference between the original, historical dataset on which an algorithm is trained (training data plus validation data) and the out-of-sample data used to test the algorithm after it has been trained. Though some may argue that validation and test data are one and the same thing, this is incorrect: validation data are historical data, whereas test data are future data from the perspective of the original dataset.

[7] 'Ground truth' refers to the output data with which the input data must be correlated, see [33].

Clearly, contestability is at the heart of the rule of law, notably its procedural core [35]. There is more to law than predictability (legal certainty) and expediency (instrumentality) [36,37]; law embodies a set of values that circle around 'equal respect and concern' for each individual person, a maxim that also grounds democracy (one person, one vote) and prevents democracy from slipping into tyranny of the majority. The centrality of the adversarial procedure is meant to contribute to better decision-making (both in administration and in court), and this is precisely why the integration of ALI requires a new hermeneutics and a new adversarial design process.

In a short but seminal article, Hofman *et al.* [38] distinguish between exploratory and confirmatory machine learning. They define the first as a playing field where:

> researchers are free to study different tasks, fit multiple models, try various exclusion rules and test on multiple performance metrics. When reporting their findings, however, they should transparently declare their full sequence of design choices to avoid creating a false impression of having confirmed a hypothesis rather than simply having generated one. Relatedly, they should report performance in terms of multiple metrics to avoid creating a false appearance of accuracy.

One could say that exploratory research is open-ended, provides room to experiment, try out, play around, mainly generating a set of promising hypotheses that in turn may also reconfigure other dimensions of the research design: rearticulating the tasks, trying out different datasets, alternative ways of curating them, etc. The point Hofman *et al*. make is that those who engage in exploratory research should put their cards on the table by clearly reporting on the experimental character of their work, not pretending confirmation where exploration is at stake. They follow up with a description of confirmatory machine learning [38]:

> To qualify research as confirmatory, however, researchers should be required to preregister their research designs, including data preprocessing choices, model specifications, valuation metrics and out-of-sample predictions, in a public forum such as the Open Science Framework (https://osf.io). Although strict adherence to these guidelines may not always be possible, following them would dramatically improve the reliability and robustness of results, as well as facilitating comparisons across studies.

In many instances this will require research into the extent to which correlations indicate causality [38], though in ALI things are not that simple, as correlations in this field are tied up with meaning attribution, more precisely with argumentation that relates to the implied philosophy [39] in the relevant legal framework and other types of reasoning that depend on human experience out in the real world (not its simulation within the contours of a dataset, however 'big'). If we take up the recommendations of Hofman *et al*., we have the beginnings of an agonistic research design. Note that their concern is not so much the implications of bad research designs for those subject to decision systems based on bad design, but a loyalty to the methodological integrity of machine learning. Such methodological integrity (scientifically robust machine learning) supports the kind of contestation that is also core to the rule of law, even if they are not equivalent. Transparent exploratory and properly tested confirmatory machine learning can be a means to provide feedback to lawyers, clients, prosecutors, courts.[8] In previous work I made six recommendations for the use of ML in law, five of which were based on the work of Sculley & Pasanek [41] on the use of ML in literary theory [42] (p. 157):

> First, assumptions about the scope, function and meaning of the relevant legal texts should be made explicit. Second, a multiplicity of representations and methodologies must be used to prevent the appearance of objectivity that is so often attached to computing. Third, when developing a legal knowledge system all trials should be reported, whether they confirm or

[8] Cf. [40], where a similar position is taken with regard to the use of ML for the grading of student papers.

> refute a designer's original expectations. Fourth, the public interest requires transparency about the data and the methods used. Fifth, a sustained dialogue about different mining methods should enrich doctrinal debates. I would add a sixth recommendation that requires legal theory and legal and political philosophy to engage with the implications of the not-reading of legal texts. Instead of deferring to data scientists or resisting any kind of automation, there is an urgent need for research into the epistemological, practical and political implications of different types of mining methods.

This would entail, for example, developing a sensitivity analysis, modulating facts, legal precepts, claims; it should be a playground for well-designed experimentation, for developing new insights, argumentation patterns, for testing alternative approaches, for detecting missing information (facts, legal arguments), helping to improve the outcome of cases. If developed in this way ALI can improve the acuity of human judgement, instead of merely replacing it. And *if* there are good arguments to replace human judgement with ALI, it should not be confused with law. Delegating governmental decision-making to data-driven legal tech is a form of regulation in the legal sense of that term. In other words, it cannot be more than a kind of administration—the difference with the law is crucial, critical and pertinent.

5. Concluding remarks

Algorithmic regulation as a concept is deeply indebted to cybernetic or regulatory theory. It refers to the idea of controlling a population by means of feedback mechanisms, based on the threefold requirement of standard-setting, monitoring and behaviour modification. As such it is grounded in a behaviourist perspective on human intercourse and displays an external perspective on human action. In this brief essay, I distinguish between code-driven algorithmic regulation, based on self-executing code that is necessarily deterministic, and data-driven algorithmic regulation, based on machine learning and statistical inferences that may be unpredictable. This raises the question of whether law is also a matter of regulation in the cybernetic sense, and, if so, whether algorithmic regulation could replace or support legal regulation. Though we can analyse law through the lens of regulatory theory, this assumes an external perspective on law, which cannot grasp the meaning of the law as based on a very specific understanding of human action and interaction. From the internal perspective on law, legal regulation is not merely a matter of influencing or controlling the behaviour of a population. It concerns, on the contrary, a perspective that sees law as addressing individuals as being the author of their actions, meaning that they are capable of giving reasons for their actions. This aligns with an understanding of legal effect in terms of speech act theory rather than cybernetic theory, acknowledging the performative nature of legal attributions.

To the extent that algorithmic regulation becomes part of legislative, judicial or other practices of law, we need to make sure that it is not merely compatible with the rule of law, but actually integrates its core principles. In the final part of this essay, I argue for an agonistic approach to ALI, that weaves the adversarial core of the rule of law into the research design of machine learning. Instead of taking the output of ALI for granted, we—the lawyers—and us—the people—must learn *to speak law to statistics*. Instead of being agnostic about the reliability of ALI, we must develop agonistic, adversarial practices when building ALI systems. From the perspective of computer science, this is not a hindrance but a prerequisite for a robust research design. As one of the founding fathers of the science of networks—Duncan Watts—explains, machine learning requires a flexible exploratory research design that fosters playing around with datasets, curation, modelling and testing, before moving on to confirmatory learning that requires rigorous testing in ways that are geared to falsification rather than verification. Proper testing implies built-in contestation—in science as in law [43–46].

Data accessibility. This article has no additional data.

Competing interests. I declare I have no competing interests.

Funding. I received no funding for this study.

References

1. Brownsword R. 2016 Technological management and the rule of law. *Law Innov. Technol.* **8**, 100–140. (doi:10.1080/17579961.2016.1161891)
2. Szabo N. 1997 Formalizing and securing relationships on public networks. *First Monday* **2** (9). (doi:10.5210/fm.v2i9.548)
3. Raskin. 2017 The law and legality of smart contracts. *Georgetown Law Technol. Rev.* **1**, 304–341. See https://ssrn.com/abstract=2959166.
4. Surden H. 2017 *Values embedded in legal artificial intelligence*. University of Colorado Law Legal Studies Research Paper No. 17-17. See https://ssrn.com/abstract=2932333.
5. Wright A, De Filippi P. 2015 *Decentralized blockchain technology and the rise of Lex Cryptographica*. See https://ssrn.com/abstract=2580664.
6. Surden H. 2014 Machine learning and law. *Washington Law Rev.* **89**, 87–115. See https://ssrn.com/abstract=2417415.
7. Hildebrandt M. 2017 Law as computation in the era of artificial legal intelligence. Speaking law to the power of statistics. *Univ. Toronto Law J.* **68**(suppl. 1), 12–35. (doi:10.3138/utlj.2017-0044)
8. Katz DM. 2013 Quantitative legal prediction—or—How I learned to stop worrying and start preparing for the data-driven future of the legal services industry. *Emory Law J.* **62**, 909–966. See https://ssrn.com/abstract=2187752.
9. Mitchell T. 1997 *Machine learning*, 1st edn. New York, NY: McGraw-Hill Education.
10. Cabitza F. 2016 The unintended consequences of chasing electric zebras. In *IEEE SMC Interdisciplinary Workshop on the Human Use of Machine Learning, Venice, Italy, 16 December*.
11. Cabitza F. 2017 Breeding electric zebras in the fields of medicine. See https://arxiv.org/abs/1701.04077.
12. Hildebrandt M. 2017 Learning as a machine. Crossovers between humans and machines. *J. Learn. Anal.* **4**, 6–23. (doi:10.18608/jla.2017.41.3)
13. Gutwirth S, De Hert P, De Sutter L. 2008 The trouble with technology regulation from a legal perspective. Why Lessig's 'optimal mix' will not work. In *Regulating technologies* (eds R Brownsword, K Yeung), pp. 193–218. Oxford, UK: Hart.
14. Lessig L. 2006 *Code*, version 2.0. New York, NY: Basic Books.
15. Black J. 2002 Critical reflections on regulation. *Aust. J. Legal Philos.* **27**, 1–35. See http://www.austlii.edu.au/au/journals/AUJlLegPhil/2002/1.pdf.
16. Wiener N. 1988 *The human use of human beings: cybernetics and society*. Cambridge, MA: Da Capo Press.
17. Hayles NK. 1999 *How we became posthuman. Virtual bodies in cybernetics, literature, and informatics*. Chicago, IL: University of Chicago Press.
18. Austin JL. 1975 *How to do things with words*, 2nd edn. Boston, MA: Harvard University Press.
19. Searle JR. 1969 *Speech acts, an essay in the philosophy of language*. Cambridge, UK: Cambridge University Press.
20. MacCormick N, Weinberger O. 1986 *An institutional theory of law: new approaches to legal positivism*. Dordrecht, The Netherlands: D. Reidel.
21. Ricoeur P. 2003 *The just*. Chicago, IL: University of Chicago Press.
22. Evans T, Harris J. 2004 Street-level bureaucracy, social work and the (exaggerated) death of discretion. *Br. J. Soc. Work* **34**, 871–895. (doi:10.1093/bjsw/bch106)
23. Dworkin R. 1978 *Taking rights seriously*, 5th edn. Cambridge, MA: Harvard University Press.
24. Hildebrandt M. 2016 New animism in policing: re-animating the rule of law? In *The SAGE handbook of global policing* (eds B Bradford, B Jauregui, I Loader, J Steinberg), pp. 406–428. Beverley Hills, CA: SAGE.
25. Aletras N, Tsarapatsanis D, Preoţiuc-Pietro D, Lampos V. 2016 Predicting judicial decisions of the European Court of Human Rights: a natural language processing perspective. *PeerJ Comput. Sci.* **2**, e93. (doi:10.7717/peerj-cs.93)
26. Pasquale F, Cashwell G. 2018 Prediction, persuasion, and the jurisprudence of behaviourism. *Univ. Toronto Law J.* **68**(suppl. 1), 63–81. (doi:10.3138/utlj.2017-0056)
27. Caruana R, Lou Y, Gehrke J, Koch P, Sturm M, Elhadad N. 2015 Intelligible models for healthcare: predicting pneumonia risk and hospital 30-day readmission. In *Proc. 21st ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining, Sydney, Australia, 10–13 August*, pp. 1721–1730. New York, NY: ACM. (doi:10.1145/2783258.2788613)

28. Hildebrandt M. In press. Privacy as protection of the incomputable self: agonistic machine learning. *Theor. Inq. Law* **20** (1).
29. McQuillan D. 2018 People's councils for ethical machine learning. *Social Media Soc.* **4**, 2056305118768303. (doi:10.1177/2056305118768303)
30. Mouffe C. 2000 *The democractic paradox*. London, UK: Verso.
31. Mouffe C. 2000 *Deliberative democracy or agonistic pluralism*. See https://www.ihs.ac.at/publications/pol/pw_72.pdf.
32. Hildebrandt M, Gutwirth S. 2007 (Re)presentation, pTA citizens' juries and the jury trial. *Utrecht Law Rev.* **3**, 24. (doi:10.18352/ulr.35)
33. Nadim T. 2016 Blind regards: troubling data and their sentinels. *Big Data Soc.* **3**, 2053951716666301. (doi:10.1177/2053951716666301)
34. Rip A. 2003 Constructing expertise: in a third wave of science studies? *Soc. Stud. Sci.* **33**, 419–434. (doi:10.1177/0306312703033006)
35. Waldron J. 2010 *The rule of law and the importance of procedure*. NYU School of Law, Public Law Research Paper No. 10-73. See https://ssrn.com/abstract=1688491.
36. Radbruch G. 2014 Legal philosophy. In *The legal philosophies of Lask, Radbruch, and Dabin*, Cambridge, MA: Harvard University Press. (doi:10.4159/harvard.9780674493025)
37. Hildebrandt M. 2015 Radbruch's *Rechtsstaat* and Schmitt's legal order: legalism, legality and the institution of law. *Crit. Anal. Law* **2**, 42–63. See https://cal.library.utoronto.ca/index.php/cal/article/download/22514/18311.
38. Hofman JM, Sharma A, Watts DJ. 2017 Prediction and explanation in social systems. *Science* **355**, 486–488. (doi:10.1126/science.aal3856)
39. Dworkin R. 1991 *Law's empire*. Glasgow, UK: Fontana.
40. Paruchuri V. 2013 On the automated scoring of essays and the lessons learned along the way. *Viks Blog*. See http://www.vikparuchuri.com/blog/on-the-automated-scoring-of-essays/.
41. Sculley D, Pasanek BM. 2008 Meaning and mining: the impact of implicit assumptions in data mining for the humanities. *Lit. Linguist. Comput.* **23**, 409–424. (doi:10.1093/llc/fqn019)
42. Hildebrandt M. 2011 The meaning and mining of legal texts. In *Understanding digital humanities: the computational turn and new technology* (ed DM Berry). London, UK: Palgrave Macmillan.
43. Yeung K. 2017 Algorithmic regulation: a critical interrogation. *Regul. Gov.* (doi:10.1111/rego.12158)
44. Yeung K. 2017 The withering of freedom under law? Blockchain, transactional security and the promise of automated law enforcement. In *A reinvention of ethics in the digital age* (eds PH Otto, E Graf). Berlin, Germany: iRightsmedia.
45. Dewey J. 1927 *The public & its problems*. Chicago, IL: Swallow Press.
46. Pearl J. 2009 *Causality: models, reasoning and inference*, 2nd edn. Cambridge, UK: Cambridge University Press.