3

Code-driven Law: Freezing the Future and Scaling the Past

MIREILLE HILDEBRANDT*

I. Introduction

In this chapter I refer to code-driven law to address legal norms or policies that have been articulated in computer code, either by a contracting party, law enforcement authorities, public administration or by a legislator. Such code can be self-executing or not, and it can be informed by machine learning systems or not. When it concerns codification of contract terms this is often called a 'smart contract', and when it concerns legislation or policies it is called 'smart regulation', especially where the code self-executes when triggered.[1]

My concern in this paper is not with data-driven law, such as prediction of legal judgments or argumentation mining, about which I have written elsewhere.[2] However, because code-driven law may integrate output of data-driven applications, these may nevertheless be relevant. For instance, a smart contract may trigger an increase in the premium of my car insurance after my car has detected a certain threshold of fatigue or risky driving. What interests me here is the fact that in code-driven law the threshold is determined in advance, in the computer code that 'drives' the smart contract. Another example would be a social security fraud detection system that halts benefit payments whenever someone is flagged by the system as probably committing fraud.[3] Again, my

* Mireille Hildebrandt is Research Professor at Vrije Universiteit Brussel on 'Interfacing Law and Technology', and holds the part-time Chair on 'Smart Environments, Data Protection and the Rule of Law' at Radboud University Nijmegen. She was awarded an ERC Advanced Grant in 2018 for 'Counting as a Human Being in the Era of Computational Law' (COHUBICOL), under the HORIZON2020 Excellence of Science program ERC-2017-ADG No 788734, which funded the research for this chapter. See www.cohubicol.com.

[1] P De Filippi and A Wright, *Blockchain and the Law: The Rule of Code* (Harvard University Press, 2018); P Hacker and others (eds), *Regulating Blockchain: Techno-Social and Legal Challenges* (Oxford University Press, 2019); M Finck, *Blockchain Regulation and Governance in Europe* (Cambridge University Press, 2019).

[2] M Hildebrandt, 'Law as Computation in the Era of Artificial Legal Intelligence: Speaking Law to the Power of Statistics' (2018) 68(1) *University of Toronto Law Journal* 12–35.

[3] S Marsh, 'One in Three Councils Using Algorithms to Make Welfare Decisions' *The Guardian* (15 October 2019) www.theguardian.com/society/2019/oct/15/councils-using-algorithms-make-welfare-decisions-benefits accessed 19 January 2020; 'Sweden: Rogue Algorithm Stops Welfare Payments for up to 70,000 Unemployed' (*AlgorithmWatch*, 2019) algorithmwatch.org/en/rogue-algorithm-in-sweden-stops-

interest is in the fact that the threshold for such probability is determined in advance, in the computer code that 'drives' this type of smart regulation. We can actually foresee that many machine learning applications will be used in such a way. First, an algorithm is trained on a – hopefully relevant – dataset, to learn which factors (eg, fatigue, spending) seem to have a strong correlation with a certain behavioural output (risky driving or social security fraud). Then, an algorithm is created that can be applied to individual persons to assess the probability that they will display this behaviour. The infamous COMPAS software operates in that way.

A private company, Northpoint, trained an algorithm on a dataset containing data on recidivism, based on 137 potentially relevant features (factors). The software detected seven of those features as strongly correlated with recidivism, including their different weights (not every feature correlates equally strong). Courts and public prosecutors have invested in the software (which is proprietary), and use it to automatically infer a risk score based on a small set of data and an interview with the person concerned. The output is then used to decide on parole or detention.[4] The software does not self-execute, and judges or prosecutors are fully responsible for the decision. However, the aura of objectivity that is often attributed to computing systems may have a strong influence on the human decision-makers.

In section II I will discuss what code-driven law does, by tracing the kind of questions it raises in different domains of law and by connecting its operations with relevant principles of private, public, constitutional and criminal law. Section III dives deeper into the nature of code-driven normativity.

II. What Code-driven Law Does

What interests me here is what code-driven law 'does' compared to text-driven law.[5] To investigate this, we will investigate different types of code-driven law and inquire how they relate to relevant principles of private, public, constitutional and criminal law.

In the case of a contracting party, code-driven law will probably refer to a smart contract, which is not only articulated in computer code but also self-executing. We could ask about the legal status of smart contracts, that is we can raise the question of whether the code counts as a legal agreement or is merely an expression of what has been agreed upon in speech or in writing. Is such an expression, just like a written agreement, merely evidence of an underlying agreement, or rather, just like with a legal deed, constitutive for the agreement itself?[6] What happens if parties disagree about

welfare-payments/ accessed 19 January 2020; Zack Newmark, 'Cabinet Member Resigns over False Accusations of Fraud by the Tax Authority' (*NL Times*, 2019) nltimes.nl/2019/12/18/cabinet-member-resigns-false-accusations-fraud-tax-authority accessed 19 January 2020.

[4] M Hildebrandt, *Law for Computer Scientists and Other Folk* (Oxford University Press, 2020) chs 10 and 11.

[5] On modern positive law as an affordance of a text-driven information and communication infrastructure (ICI), see ch 1 in M Hildebrandt, *Law for Computer Scientists and Other Folk*, and chs 7 and 8 in M Hildebrandt, *Smart Technologies and the End(s) of Law. Novel Entanglements of Law and Technology* (Edward Elgar, 2015).

[6] JG Allen, 'Wrapped and Stacked' (2018) 14(4) *European Review of Contract Law* 307–43.

the precise meaning of the underlying agreement in view of the operations of the self-executing script?[7] Have they given up their right to go to court about an execution that deviates from what they legitimately expected, considering the circumstances? Should their accession to the contract imply a waiver of any right to claim that the code got it wrong, compared to what they thought they agreed upon?

In the case of public administration, code-driven law can refer to either a decision-support or a decision-making system that is articulated in computer code, enabling swift execution (based on input from a citizen or a civil servant). In the case of self-execution (decision-making), we could ask whether such decisions have the force of law if taken under the responsibility of a competent government body, that is, we can ask under what conditions a fully automated decision (taken by a software program) even counts as a valid decision of a competent body. We can also ask under what conditions a decision taken by a human person based on a decision-support system nevertheless counts as an automated decision, for instance, because the human person does not really understand the decision and/or lacks the power to deviate from the output of the software.[8]

In the case of the legislature, code-driven law may refer to legislation that is articulated in writing but in a way that anticipates its translation into computer code, or it could relate to legislation written in computer code, which can either self-execute or require human intervention. Could legislation, enacted by a democratic legislature, count as such if it were written in computer code? Or would this depend on whether the legislature and/or its constituency are sufficiently fluent in code? Would code-driven law also refer to legislation, policies and decisions of public administration and judgments of courts that have been made machine-readable, in the sense of being structured with the help of metadata that allow software programs to categorise and frame such legal text, and to apply various types of data analytics such as argumentation mining, prediction of judgments, and search for applicable law?

These are all very interesting and highly relevant questions, relating to core principles of private, administrative and constitutional law. To the extent that policing and sentencing become contingent upon decisions made by software programs that determine the risk that a person has committed or may or even will commit a criminal offence, core principles of the criminal law are at stake.

Private law principles, such as the freedom to contract and the freedom to dispose of one's property raise questions around the constitution of a contract: what information should have been provided by the offering party, what investigations should have been undertaken by the accepting party? When does a lack of information result in the contract being void? Or, has the will of a party been corrupted by duress, fraud or deception, making the contract voidable? How does the law on unfair contract terms apply if it turns out that a party should have known that the terms of service implied their agreement to waive the right to appeal? Does the freedom to conduct a business incorporate the freedom to offer a service on condition that a smart contract is accessed? Does it make a difference whether this concerns a pair of shoes, a car or a health insurance?

[7] M Raskin, 'The Law and Legality of Smart Contracts' (2017) 1(2) *Georgetown Law and Technology Review* 304–41.

[8] M Finck, 'Smart Contracts as Automated Decision Making Under Article 22 GDPR' (2019) 9 *International Data Privacy Law* 1–17.

Public law principles, such as the legality and the fair play principles, aim to ensure that whenever government agencies exercise legal powers they do so in a way that stays within the bounds of the purpose for which they were attributed, while also remaining within the bounds of legitimate expectations that have been raised. To what extent will code-driven law result in what Diver calls computational legalism,[9] confusing rule-fetishism with acting under the rule of law? What happens if citizens are forced to articulate their applications for tax returns, health care, education or social welfare benefits in terms they do not recognise as properly describing their situation? What if their competence to appeal against automated decisions is restricted to what code-driven decision-systems can digest?

Criminal law principles, such as the presumption of innocence, equality of arms, immediacy with regard to the contestation of evidence, and the legality of criminal investigation that requires probable cause, proportionality and a range of other safeguards whenever fundamental rights are infringed in the course of a criminal investigation may be violated in case of, for example, 'smart policing' or 'smart sentencing'.[10] What if one is not made aware of the fact that code-driven systems have raised a flag resulting in invasive monitoring? What if such monitoring is skewed towards black people, or towards those with a criminal record, or towards people with a particular political opinion, taking note that this need not depend on direct discrimination, as it could be the result of flagging based on data that serves as a proxy for this type of bias? What if increased attention to specific groups of people results in them being charged more often, in a way that is disproportional in relation to their actual involvement in criminal offences?

Constitutional principles, such as legality, accountability, transparency and other expressions of the checks and balances of the rule of law are core to constitutional democracies.[11] The rule of law implies that neither the legislature nor public administration get the last word on the meaning (the interpretation and application) of the law. Judgment is reserved for the courts.[12] What if legislation is translated into computer code, that is, disambiguated, and what if at that very moment both its interpretation and application are de facto decided? What should courts decide if a legislature enacts law in the form of code? To what extent is the meaning of the law contestable in a court of law if the law has been disambiguated and caught in unbending rules that only allow for explicitly formulated (and formalised) exceptions? What if courts use the same software as the public prosecutor, or depend on the same legal technologies as Big Law?

What code-driven law does is to fold enactment, interpretation and application into one stroke, collapsing the distance between legislator, executive and court. It has to foresee all potential scenarios and develop sub-rules that hopefully cover all future interactions – it must be highly dynamic and adaptive to address and confront what

[9] LE Diver, 'Digisprudence: The Affordance of Legitimacy in Code-as-Law' (*era*, 2019), era.ed.ac.uk/handle/1842/36567 accessed 19 January 2020.

[10] M Oswald, 'Algorithm-Assisted Decision-Making in the Public Sector: Framing the Issues Using Administrative Law Rules Governing Discretionary Power', Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 376: 2017035 (2018).

[11] M Hildebrandt, 'Algorithmic Regulation and the Rule of Law', Philosophical Transactions of the Royal Society A, 376: 20170355 (2018).

[12] DK Citron, 'Technological Due Process' (Washington University Law Review, 2008) 1249–1313.

cannot easily be foreseen by way of unambiguous rules. If it fails to do so, code-driven law must be subjected to appeal and contestation, based on, for instance, core legal concepts such as 'unreasonableness', 'unacceptable consequences', 'good faith', or 'the circumstances and the context of the case at hand'. This would imply reintegrating ambiguity, vagueness and multi-interpretability into the heart of the law.

III. The Nature of Code-driven Normativity

A. Language, Speech and Text-driven Normativity

In this paper, code-driven normativity refers to behavioural patterns generated by computer code that aims to influence human behaviour. Normativity – as intended here – is close to habits, which are neither moral nor merely regular. Such normativity is in line with how Wittgenstein understood rule following.[13] A prime example is the normative character of language usage (grammar, vocabulary, idiom),[14] which is neither merely regular nor a matter of morality. Following the rules of a particular language or idiom is constitutive of the meaning of the text, and conveying such meaning depends on speaking in a way that others recognise as meaning what is intended. However, meaning depends on connecting the intra-linguistic meaning of a word (its relationship with other words, its connotation) with its extra-linguistic meaning (its reference, or denotation). This reference may concern a tree or a mountain, a table or a car, but also institutions such as a marriage, a university or a legal service. These different types of references demonstrate that extra-linguistic meaning is co-constituted by the intra-linguistic framings that in turn depend on the success they offer language users in navigating their physical as well as institutional world. This is not just a matter of words, but also of grammar (conjugations, the use of pronouns, future and past tense).

The way language constitutes a world of institutions, roles and actions determines to a large extent what is possible, feasible or precluded, by shaping what is thinkable and affording what is thought. Interestingly, this determination is never final or complete – precisely because the same text can be interpreted in different ways, meaning that interpretations themselves can always be contested. This implies that the shared world that is constituted by our usage of language is contingent upon our ongoing support. In point of fact our shared, institutional world is performed by way of speech acts that do what they say: 'I declare you husband and wife' does not describe the marriage but performs, concludes, 'makes' it. The same goes for qualifying the complex network of behavioural patterns around higher education as a university. The performative nature of spoken and written speech (use of language) in turn drives a specific normativity, that is typical for the way human beings interact with their shared institutional world and with each other.

[13] L Wittgenstein and others, *Philosophical Investigations* (Wiley Blackwell, 2009); C Taylor, 'To Follow a Rule', in C Taylor, *Philosophical Arguments* (Harvard University Press, 1995) 165–81.

[14] R Bartsch, 'The Concepts "Rule" and "Norm" in Linguistics' (1982) 58(1) *Lingua* 51–81.

The text-driven normativity that concerns us here is based on the attribution of legal effect: if certain legal conditions apply, the law attributes a specific legal effect. For instance, in the case of a contract of sale, if a stipulated consideration is performed, a stipulated price must be paid. The legal effect is neither the consideration nor the payment. The legal effect concerns the fact that two legal obligations come into existence: to perform what is required by the contract. As we have seen above, a whole range of legal norms apply to the interpretation of the terms of the contract, taking into account the concrete circumstances of the case. These legal norms are often framed in terms of essentially contested concepts,[15] which have an open texture,[16] such as reasonableness, equity (in common law jurisdictions), force majeur, foreseeability etc. The multi-interpretability of these concepts generates a normativity of contestability, due to the fact that the potential of contestation is inherent in the nature of text. This is how text-driven normativity affords the core tenet of the rule of law: the contestability of the interpretation given by a party or a public authority. This is also how text-driven normativity affords the other core tenet of the rule of law: the need to perform closure. Once it is clear that such closure is not given with the text, because text does not speak for itself, it becomes clear that interpretation is construction rather than description and thus requires giving reasons for interpreting a legal norm in one way rather than another.[17]

Obviously, the rule of law adds closure by an independent court, instead of closure by a magistrate who is part of public administration. The latter would be rule by law. As Montesquieu said *iudex est lex loqui* (the court speaks the law), thereby countering the absolutist maxim of *rex est lex loqui* (the king speaks the law). As Schoenfeld has argued on historical grounds, the usual understanding of Montesquieu's *bouche de la loi* (the court as mouth of the law) is mistaken.[18] Montesquieu emphasised the court's loyalty to the law rather than to the king. When the court speaks, they do not follow the arbitrary interpretation of the ruler (rule by law by man) but are bound by the law, over which they have the last word (rule of law). We will return to this when discussing the difference between legalism and legality in relation to legal certainty.

B. Computer Architecture, Design and Code-driven Normativity

Clearly, the kind of rule following that is generated by text differs from rule following generated by computer code. What matters here is a set of constraints that are inherent in computer code that do not constrain natural language and text.[19]

The first is the need to formalise whatever requirements are translated into code. Formalisation enables the logical operation of deduction, in the sense of 'if this then that' (IFTTT). Such operations are crucial for automation, which is the core of computing systems. To the extent that formalisation is not possible or questionable, code-driven

[15] WB Gallie, 'Essentially Contested Concepts' (1956) 56 *Proceedings of the Aristotelian Society* 167–98.

[16] HLA Hart, *The Concept of Law* (Oxford University Press, 1994).

[17] FCW de Graaf, 'Dworkin's Constructive Interpretation as a Method of Legal Research' (2015) 12 *Law and Method*.

[18] KM Schoenfeld, 'Rex, Lex et Judex: Montesquieu and La Bouche de La Loi Revisted' (2008) 4 *European Constitutional Law Review* 274–301.

[19] See eg en.wikibooks.org/wiki/Logic_for_Computer_Scientists/Introduction, or a more in-depth discussion by A Arana in her Review of BJ Copeland, O Shagrir and CJ Posy (eds), *Computability: Turing, Gödel,*

architectures cannot be developed or may be unreliable. The second constraint is the need to disambiguate the terms used when formulating the requirements. This constraint is in turn inherent in formalisation, because deduction is not possible if it remains unclear what the precise scope is of the requirements. Disambiguation implies an act of interpretation that should result in a clear demarcation of the consequences of applying the relevant terms. The third constraint is that completeness and consistency cannot be assumed, meaning that the mathematical underpinnings of code-driven systems limit the extent to which claims about the correctness of computer code can be verified.[20] In the context of data-driven models, based on machine learning, a further set of constraints comes to the surface, relating to the limitations inherent in the design of a feature space, the hypothesis space, the articulation of the machine-readable task and the definition of the performance metrics.[21]

To some extent all these constraints are related to the uncertainty that inheres in the future. As to code-driven applications, we must face the limits of our ability to sufficiently foresee how changing circumstances will impact the execution of the code. In data-driven applications this can be summarised in the observation that one cannot train an algorithm on future data. Machine learning has to assume that the distribution of the data on which a learning algorithm has been trained is equivalent with or is a close approximation of the distribution of future data. This assumption, however, is not correct. On the contrary, it is the distribution of future data that machine learning hopes to predict but does not know. Integrating the output of machine learning systems therefore increases the risk that for example self-executing code 'gets things wrong' in the real world it aims to regulate. This is related to the radical uncertainty that defines the future, not in the sense of the future being entirely random or arbitrary but in the sense of its being underdetermined, notably when we concern ourselves with the consequences of human interaction. The radicality must be situated in the fact that this underdetermination cannot be resolved because it defines the human condition.[22]

The radical uncertainty of the future is exacerbated by the fact that predictions impact the behaviour they supposedly predict. In economics this is known as the Goodhart effect,[23] the Campbell effect[24] or the Lucas critique[25] and has nicely been summed up by Strathern[26] as: 'When a measure becomes a target, it ceases to be a good

Church, and Beyond (20 March 2015), *Notre Dame Philosophical Reviews*, ndpr.nd.edu/news/computability-turing-gdel-church-and-beyond accessed 30 November 2019.

[20] This connects with Gödel's theorem and the Church-Turing theses, see previous note and eg BJ Copeland, 'The Church-Turing Thesis' in EN Zalta (ed), *The Stanford Encyclopedia of Philosophy* (Spring 2019) plato.stanford.edu/archives/spr2019/entries/church-turing/ accessed 30 November 2019.

[21] TM Mitchell, 'Key Ideas in Machine Learning' in *Machine Learning*, draft for the 2nd edn (2017) 1–11; T Mitchell, *Machine Learning* (McGraw Hill, 1997).

[22] H Arendt, *The Human Condition* (University of Chicago Press, 1958); H Plessner and JM Bernstein, *Levels of Organic Life and the Human: An Introduction to Philosophical Anthropology*, trans. by M Hyatt, (Fordham University Press, 2019).

[23] D Manheim and S Garrabrant, 'Categorizing Variants of Goodhart's Law' (*arXiv*, 2019), http://arxiv.org/abs/1803.04585 accessed 19 January 2020.

[24] DT Campbell, 'Assessing the Impact of Planned Social Change' (1979) 2(1) *Evaluation and Program Planning* 67–90.

[25] RE Lucas, 'Econometric Policy Evaluation: A Critique' (1976) 1 *Carnegie-Rochester Conference Series on Public Policy* 19–46.

[26] M Strathern, '"Improving Ratings": Audit in the British University System' (1997) 5(3) *European Review* 305–21.

measure'. Once a description (measurement) of a certain state of affairs is understood as a prediction it may start functioning as a way to coordinate the behaviour of those whose behaviour is described (measured); if such predictions are then used to influence people they may no longer apply because people change their behaviour in function of the predictions (which they may for instance resist or execute, in divergence of how they would have behaved had such predictions not been employed).

Esposito has framed this effect even more pointedly, where she concludes (in my words) that our present futures change the future present.[27] What she is recounting here is that predictions (our present futures) influence the anticipation of interactions, resulting in an adjustment of actions, thus instantiating a different future present (compared to the one that might have become true if no predictions were employed). Her work also reminds us that whereas we can develop many present futures (predictions, imaginations, anticipations), we have only one future present. Considering the impact of predictions, we may want to exercise prudence when predicting.

Yet another way to state this is that 'the best way to predict the future is to create it'. This adage has been attributed to (amongst others) Gabor, one of the founding fathers of cybernetics, who elaborated:

> We are still the masters of our fate. Rational thinking, even assisted by any conceivable electronic computers, cannot predict the future. All it can do is to map out the probability space as it appears at the present and which will be different tomorrow when one of the infinity of possible states will have materialized. Technological and social inventions are broadening this probability space all the time; it is now incomparably larger than it was before the industrial revolution – for good or for evil.

In other work I have elaborated on this crucial insight, which asserts the counter-intuitive finding that predictions do not reduce uncertainty but rather extend it.[28]

C. Double Contingency and the Radical Uncertainty of the Future

The radical uncertainty that defines the human condition is best explained in terms of what Parssons and Luhmann have coined the 'double contingency' that is inherent in human interaction.[29] This refers to the fact that due to the nature of natural language we are always in the process of anticipating how others anticipate us. To be able to act meaningfully we need to anticipate how others will 'read' our actions, which links the interpretation of text to that of human action.[30] This explains the Goodhart, Campbell and Lucas effects.

[27] E Esposito, *The Future of Futures* (Edward Elgar, 2011).

[28] M Hildebrandt, 'New Animism in Policing: Re-Animating the Rule of Law?' in B Bradford and others (eds), *The SAGE Handbook of Global Policing* (Sage Publishing, 2016) 406–28.

[29] R Vanderstraeten, 'Parsons, Luhmann and the Theorem of Double Contingency' (2007) 2(1) *Journal of Classical Sociology* 77–92; M Hildebrandt, 'Profile Transparency by Design: Re-Enabling Double Contingency' in M Hildebrandt and E. De Vries (eds), *Privacy, Due Process and the Computational Turn: The Philosophy of Law Meets the Philosophy of Technology* (Routledge, 2013) 221–46.

[30] P Ricoeur, 'The Model of the Text: Meaningful Action Considered as a Text' (1973) 5(1) *New Literary History* 91–117.

Parssons and Luhmann both emphasised the radical uncertainty that is generated by what they called the 'double contingency' of human interaction, where I try to anticipate how you will read my actions while you try to anticipate how I will read your actions. They emphasise that this demands consolidation and stabilisation, by way of the institutionalisation of specific patterns of behaviour that form the background against which interactions are tested as meaning one thing or another. Luhmann thus explains the existence of social systems as a means to achieve *Kontingenzbewaltigung*, reducing complexity and uncertainty to a level that is productive instead of merely confusing. Without endorsing Luhmann's depiction of social systems as autopoietic I think we can take from his work the crucial insight that the mode of existence of human interaction is anticipatory, forever reaching out into a future world of human interaction that we cannot control. Legal certainty plays a pivotal role in stabilising and consolidating the mutual double anticipation that defines human interaction within a specific jurisdiction, without freezing the future based on a scaling of the past.

IV. Legal Certainty and the Nature of Code

A. Legalism and Legality; Consistency and Integrity

Legal certainty can be understood in two ways. Some authors equate it with consistency, which assumes that legal systems are coherent and complete. This usually goes with a legalistic understanding of the rule of law, where 'rules are rules' and 'facts are facts'. The discussion above should clarify that this is an untenable position that ignores the role played by natural language and the open texture of legal norms. Text-driven normativity simply does not afford the logical and deductive coherence such legalism assumes. In his doctorate thesis Diver has built on this pivotal insight by qualifying code-driven law as a form of computational legalism.[31] This seems a salient qualification, notably insofar as such 'law' claims perfect execution (where the enactment of the law includes both its interpretation and its implementation).

A less naïve understanding of legal certainty instead emphasises the integrity of the law, which is both more and less than consistency. Many of the misunderstandings around Dworkin's *Empire of Law*, where he explains the concept of the integrity of law, stem from conflating his 'integrity' with logical constituency (which would turn law into a closed system and the judge into a master of logical inference). The integrity of law could be understood as referring both to the coherence of the legal system (the intra-systematic meaning of legal norms), and to the moral implications of their legal effect (their extra-systematic meaning, which is performative as it reshapes the shared institutional world). The moral implications, however, do not depend on the 'subjective' opinion of the deciding judge but on the 'implied philosophy' that is given with law's complex interaction between intra- and extra-systematic meaning. It is crucial to understand the fundamental uncertainty that sustains the dynamic between internal coherence and the performative nature of attributing legal effect. Integrity is therefore

[31] Diver, 'Digisprudence'.

more than consistency, where it needs to achieve closure under uncertainty, and it is less than consistency, where it relies on an implied philosophy that must take into account both the justice and the instrumentality of the law (next to legal certainty).[32] This connects with Dworkin's constructive interpretation,[33] which emphasises that the right interpretation is not given but must be constructed as part of the refined but robust fabric of legal meaning production.

In light of the above, legal certainty faces two challenges. First, it needs to sustain sufficient consistency to enable those subject to law to foresee the consequences of their actions. This is not obvious, due to the impact of changing circumstances that may destabilise common sense interpretations of legal norms. The terms of a contract may seem clear and distinct, but in the case of unexpected events a reasonable interpretation may unsettle mutual expectations and require their reconfiguration. The fact that written law affords such reconfiguration is not a bug but a feature of text-driven normativity, because it enables to calibrate and consolidate such mutual expectations in a way that is in line with past and future decisions – thus also weaving a fabric of legitimate mutual expectations that holds in the course of time. Text-driven law is adaptive in a way that would be difficult to achieve in code-driven law (which relies on a kind of completeness that is neither attainable nor desirable).

The second challenge, which is deeply connected with the first, concerns the fact that legal certainty is not the sole constitutive aim of the law. If it were, perhaps computational legalism would work. Totalitarian and populist ideologists may vouch for this, hoping to construct the ideal legal system that enforces by default with no recourse to independent courts. Law's empire, however, is also built on two other constitutive aims: those of justice and instrumentality. Even though these aims may be incompatible in practice, the law should align them to the extent possible. The mere fact that legal certainty, justice and instrumentality are what Radbruch coined as 'antinomian',[34] and require decisions whenever they cannot be aligned, does not imply that when one goal overrules another in a particular case the others are disqualified as constitutive goals. In other words, any practice or theory that systematically resolves the tension between these three goals reduces the rule of law to either legal certainty (legalism), justice (natural law) or instrumentality (politics). As discussed above, legalism does not actually provide for certainty, as it builds on the mistaken assumption that future events will not impact the interpretation of a legal norm. Similar things can be said about justice and instrumentality. If any of them is taken to systematically overrule the others, they lose their fitting.

Justice refers to equality, both in the sense of proportionality (punishment should, eg, be attributed in proportion to the severity of the crime, compensation paid in proportion to, eg, the damage suffered) and in the sense of distribution (treating equal

[32] M Hildebrandt, 'Radbruch's Rechtsstaat and Schmitt's Legal Order: Legalism, Legality, and the Institution of Law' (2015) 2(1) *Critical Analysis of Law*, cal.library.utoronto.ca/index.php/cal/article/view/22514 accessed 24 March 2015.

[33] Graaf, 'Dworkin's Constructive Interpretation as a Method of Legal Research'.

[34] G Radbruch, 'Five Minutes of Legal Philosophy (1945)' (2006) 26(1) *Oxford Journal of Legal Studies* 13–15; G Radbruch, 'Legal Philosophy', in K Wilk (ed), *The Legal Philosophies of Lask, Radbruch, and Dabin* (Harvard University Press, 1950) 44–224; M Hildebrandt, 'Radbruch's Rechtsstaat and Schmitt's Legal Order'; M Hildebrandt, 'The Artificial Intelligence of European Union Law' (2020) 21(1) *German Law Journal* 74–79.

cases equally and unequal cases unequally to the extent of their inequality). These two types of justice, which inform legally relevant justice, have been coined corrective and distributive justice by Aristoteles and it should be clear that they interact;[35] to decide on distributive justice one needs a decision on corrective justice and vice versa. The equality that defines justice has a direct link with legal certainty, since it enables to foresee how one's case would be treated and thus helps to foresee the consequences of one's actions. On top of this it is crucial to remember that in law what matters in decisions that define what counts as either equal or unequal cases, will always be how this affects individuals.[36] Law requires governments to treat each and every person under their rule with equal respect and concern.[37] This grounds both the rule of law (individual human rights) and democracy (one person one vote) and their interaction (majority rule cannot overrule individual rights).

Instrumentality refers to how law serves policy goals determined by the legislature. The latter not only defines the legality principle that requires a legal competence for public administration to act lawfully, but also enables to serve a range of policy goals (full employment, sustainable environments, competitive markets, healthcare, social welfare, crime reduction, education, etc). The point here is that under the rule of law the legal norms that configure the space of lawful action are instrumental in a way that also safeguards the other goals of the law. In that sense legal norms are always both constitutive and limitative of lawful interaction. They allow or enable certain actions but also limit them (eg, by requiring that contracts are performed, but are not valid when they serve an illegitimate goal).

As Waldron has argued,[38] legal certainty is not only important because it contributes to foreseeability and trust, but also because it builds on the contestability of legal norms. Precisely because their interpretation (including their validity in light of other legal norms) can be contested, their force is more robust than the force of mechanical application or brute enforcement could ever be. This relates to the primacy of procedure in the substance of the rule of law, which like any text cannot speak for itself. Without litigation, due process (US) and fair trial (Europe), and an independent judiciary, legal certainty and the rule of law lose their meaning.

B. The Nature of Code-driven Law: Inefficiencies and Ineffectiveness

Legal interpretation is constructive and has performative effect. Legal interpretation, based on text-driven normativity, institutes the adaptive nature of text-driven law. The force of law is based on a complex interplay between the demands of legal certainty,

[35] J Waldron, 'Does Law Promise Justice?' (2001) 17(3) *Georgia State University Law Review*.

[36] Waldron, 'Does Law Promise Justice?'.

[37] See references to Dworkin and a discussion of their impact in Stefan Gosepath, 'Equality' in EN Zalta (ed), *The Stanford Encyclopedia of Philosophy* (Spring 2011), plato.stanford.edu/archives/spr2011/entries/equality/ accessed 1 December 2019.

[38] J Waldron, 'The Rule of Law and the Importance of Procedure' (2011) 50 *Nomos* 3–31, 19. See also M Hildebrandt, 'Law as Computation in the Era of Artificial Legal Intelligence' 21–22.

justice and instrumentality. It is more than mechanical application of disambiguated rules and more than brute enforcement based on the monopoly of violence that grounds the rule of law in most constitutional democracies. The force of law is robust due to procedures in front of independent courts that engage with contestation while providing closure.

The force of code differs from the force of law. The act of translation that is required to transform text-driven legal norms into computer code differs from the constructive interpretation typically required to 'mine' legal effect from text-driven legal norms in the light of the reality they aim to reconfigure. The temporal aspect is different, because code-driven normativity scales the past; it is based on insights from past decisions and cannot reach beyond them. The temporality also differs because code-driven normativity freezes the future; it cannot adapt to unforeseen circumstances due to the disambiguation that is inherent in code. Instead it can accommodate a range of additional rules that apply under alternative conditions, implying complex decision trees that hope to map future occurrences. This mapping is by definition underdetermined, not because of a lack of knowledge but due to the radical uncertainty that is the future (see above Sections IIIB and IIIC).

If machine learning is involved as input into the decision tree, some may argue that this affords adaptiveness, for instance by triggering new interpretations based on learning algorithms trained, validated and tested on, for example, streaming data. This, however, still requires specifying the behavioural response of the code-driven system, for instance by way of specified input thresholds. In many ways this will make the legal system more complicated and cumbersome, as it requires endlessly complex decision trees based on the identification of relevant future circumstances.

In the end, such coding efforts will forever lag behind the myriad relevant future circumstances that can be captured by legal concepts endowed with an open texture, as these are flexible and adaptive 'by nature', while nevertheless constrained due to the institutional settings of an independent judiciary that has the last word on their interpretation. On top of this particular inefficiency – and concomitant ineffectiveness – code-driven law has other inefficiencies that may involve massive externalisation of costs. If smart regulation is based on an inventory of legal norms that has been mined from statutory and case law, a small army of relatively cheap legal experts (students? paralegals?) is required to label relevant factors (features) in the body of relevant text. A redistribution of labour will be enacted: those who define the feature space and those who sit down to qualify text elements in terms of these features. The power of interpretation will reside with those who design the feature space, though in the end those who actually label the data may unintentionally disrupt such framing based on their own (mis)understandings. These problems may be solved – guess what – by even more automation, hoping to write code that automates the identification of relevant features as well as the labelling process. Obviously, this will push decisions on interpretation even deeper into the design of the code.

As long as such systems are 'under the rule of law', their output remains contestable in a court of law. Considering the drawbacks of the disambiguation that is at the heart of code-driven normativity, this could lead to a surge in litigation, with litigants claiming that their rights have been violated because the system mis-qualifies their actions when freezing the interpretation of the norm, or contending that it fails to take into account higher law, such as international human rights or constitutional law. A surge in

litigation would make the employment of these systems less efficient (also considering the investment they require and the huge costs of maintenance, considering both security threats and other bugs). We should not be surprised when the legislature decides to restrict contestation, based on legal assumptions (in public administrative law) and waivers (in private law). This would create a decisional space outside the rule of law, enabling consistent application of arbitrary norms (as these norms cannot be tested against the architecture of legal norms they are a part of).

V. 'Legal by Design' and 'Legal Protection by Design'

A. Legal by Design

In their article on 'Legal By Design: A New Paradigm for Handling Complexity in Banking Regulation and Elsewhere in Law',[39] Lippe, Katz and Jackson observe that (at 836):

> in many instances, the growth of legal complexity appears to be outpacing the scalability of an approach that relies exclusively or in substantial part on human experts and the ability of the client to absorb and act on the advice given.

This argument has been heard before, for instance in the realm of healthcare (we need remote healthcare and robot companions to address the rising costs of care), suggesting that code-driven architectures will be more efficient and effective in solving relevant problems than the employment of human beings.[40] As argued above, I believe this is an untenable position.

In this subsection I will trace the meaning of 'legal by design' (LbD) and confront it with a concept I coined a long time ago, namely 'legal protection by design' (LPbD),[41] which should not be confused with LbD. The distinction should not only clarify that we need LPbD rather than LbD, but also allow us to inquire to what extent 'compliance by design',[42] 'enforcement by design',[43] 'technological management',[44] 'technoregulation,[45] or 'computational law' could support LPbD,[46] enhancing rather than diminishing human agency, and challenging rather than scaling the past, even though such techno-legal 'solutions' may result in freezing dedicated parts of our shared future.

[39] P Lippe, DM Katz and D Jackson, 'Legal by Design: A New Paradigm for Handling Complexity in Banking Regulation and Elsewhere in Law' (2015) 93(4) *Oregon Law Review* 833–52.

[40] R Susskind and D Susskind, *The Future of the Professions: How Technology Will Transform the Work of Human Experts* (Oxford University Press, 2015).

[41] M Hildebrandt, 'Legal Protection by Design: Objections and Refutations' (2011) 5(2) *Legisprudence* 223–48. See also M Hildebrandt and L Tielemans, 'Data Protection by Design and Technology Neutral Law' (2013) *Computer Law & Security Review* 509–21.

[42] N Lohmann, 'Compliance by Design for Artifact-Centric Business Processes' (2013) 38(4) *Information Systems, Special section on BPM 2011 conference* 606–18.

[43] N García and AP Kimberly, 'Enforcement by Design: The Legalization of Labor Rights Mechanisms in US Trade Policy', CIDE Working Paper Series, División de Estudios Internacionales (2010).

[44] R Brownsword, 'Technological Management and the Rule of Law' (2016) 8(1) *Law, Innovation and Technology* 100–40.

[45] R Leenes, 'Framing Techno-Regulation: An Exploration of State and Non-State Regulation by Technology' (2011) 5(2) *Legisprudence* 143–69.

[46] M Genesereth, 'Computational Law. The Cop in the Backseat' (2015,) logic.stanford.edu/complaw/complaw.html accessed 9 October 2016.

Let's first note that Lippe, Katz and Jackson are referring to a very specific subsection of what is best called 'legal services', rather than 'law'. Their article concerns regulation of banking and the claim or assumption of the authors is that such regulation has become so complex that those addressed are at a loss as to compliance. This is a bit funny, of course, considering the choices that have been made by the financial sector in advancing their own interests.[47] The article seems entirely focused on business to business (B2B) relationships, treating them as if they only concern the businesses involved, though we all know decisions made around financial markets affect many individuals whose lives may be disrupted due to decisions by those who couldn't care less (about them). This is the first caveat; a healthy network of financial markets is not merely the private interest of financial institutions. The second caveat is that the proposed LbD suggests a smooth path towards compliance with legal norms deemed overly complicated, whereas the article abstracts from the underlying goals. These goals involve the public interest in a way that may not align with the interests of those running the financial sector. We can be naïve about this, or turn a blind eye, but this will not do when investigating the role of code-driven law in the light of LbD solutionism.

The authors refer to the Massive Online Legal Analysis (MOLA) and recount (at 847):

> The MOLA process is conceptually similar to processes that have been used for almost two decades to address and solve extremely large, complex mathematical and scientific problems. IBM developed one of the best organized efforts – the World Community Grid – to conduct massive and complex research in a variety of areas, including cancer research, clean air studies, AIDS investigations, and other health-related projects. As described on its website, the 'World Community Grid brings together people from across the globe to benefit humanity by creating the world's largest non-profit computing grid … by pooling surplus processing power from volunteers' devices'.

There is no evidence that MOLA's solutions 'work', no reference to serious, independent verification of the findings (which findings?), let alone any attempt at falsification.[48] One is tempted to quote Cameron:[49]

> It would be nice if all of the data which sociologists require could be enumerated because then we could run them through IBM machines and draw charts as the economists do. However, not everything that can be counted counts, and not everything that counts can be counted.

[47] The financial crisis of 2008 demonstrated that compliance is perhaps not enough. To prevent global catastrophes will require the right kind of rules, instigating the right type of incentives, coupled with transnational enforcement. On the interest of the financial sector in achieving compliance 'It Knows Their Methods. Watson and Financial Regulation' (*The Economist*, 22 October 2016), www.economist.com/news/finance-and-economics/21709040-new-banking-rules-baffle-humans-can-machines-do-better-it-knows-their-methods accessed 6 June 2017.

[48] The World Community Grid enables 'anyone with a computer, smartphone or tablet to donate their unused computing power to advance cutting-edge scientific research on topics related to health, poverty and sustainability'. It 'brings together volunteers and researchers at the intersection of computational chemistry, open science and citizen science – three trends that are transforming the way scientific research is conducted'. The Grid provides computational power, it does not impose any specific methodology. See www.worldcommunitygrid.org/about_us/viewAboutUs.do.

[49] W Cameron, Informal Sociology, a Casual Introduction to Sociological Thinking (Random House, 1963) at 13.

Lippe, Katz and Jackson continue (at 847):

> As with the World Community Grid, MOLA breaks a large, data rich, and complex legal project into small pieces that can be assigned to individual attorneys for completion. Those small, individual solutions, when combined with thousands of other individual solutions, result in a cost-effective solution to the overarching larger project.
>
> More concretely, the solution is specified as an approach meant to convert a contract into a pointable data object where the contract memorializes the set of rights and obligations that are attendant to that agreement. To attain this goal, [financial, mh] institutions will need to:
>
> 1. Collect the set of all agreements held by a bank.
> 2. Identify each counterparty from those agreements (and third party where available).
> 3. Develop a model of counterparty risk which would include both an individual and systematic (ecosystem) component.
> 4. Determine the nature of resource (financial) flows attendant to each counterparty.
> 5. Convert each contract into a pointable data object, which allows its contents to be immediately memorialized in a balance sheet or other relevant IT system.
> 6. Offer the ability for key decision makers to query a system and run various scenarios in which some sort of aggregate or systematic risk could be the output.

In what sense could this 'solution' be understood as a LbD solution? This depends on how one understands 'design'. The authors suggest (at 843):

> As such, 'design' describes object creation, manifested by an agent, to accomplish a goal or goals, where the object satisfies a set of requirements, and its creation is subject to certain fixed constraints. Used in this traditional sense, the design 'object' is a physical one, the agent is a human being (the designer), the goal is the purpose of the design exercise (move this large object from here to there), the set of requirements include material specifications (use only found objects), and the constraints are things such as available found materials (stone and wood). Thus, the first rudimentary wheel was not invented, but designed.

Beyond this physical context, legal design based on MOLA supposedly enables a financial institution to reinterpret the fixed constraints imposed by financial regulation:

> Using new technology and alternative approaches to organize legal information can expand the available options well beyond what are initially seen as fixed constraints.

Why and how this would amount to 'legal by design' is not clear to me, but it does sound like a scheme that allows financial institutions to document compliance based on hybrid systems that integrate code- and data-driven search and assessment, assuming for no apparent reason that such an assessment is reliable, or at least more reliable then human auditing. Taking into account the feedback loops discussed in Section IIIC this is not at all obvious, and the question comes to mind how human auditors could check whether or not the system is getting things right. If the 'regulatory environment' is too complex for any human auditor, how could we assume that the exercise of decoding (written law) and recoding (in code-driven output) reliably frames these complexities?

A second attempt to achieve LbD is the use of blockchain-based smart contracts, which add self-execution to a decisional system; above and in other work I have

explained why this cannot be LbD because the code is fixed and will inevitably turn out to be over- and underinclusive when applied.[50]

VI. Legal Protection by Design

LPbD embodies an entirely different approach. It is not focused on achieving compliance or enforcement of whatever legal norms, but targets the articulation of legal protection into the ICT infrastructure of code- and data-driven environments. The point of departure is not the translation of a written legal norm into computer code but an inquiry into the way a data- and code-driven environment affects the substance of fundamental legal principles and rights. Based on this assessment LPbD seeks ways to prevent diminished legal protection by intervening in the design of the relevant computational architecture, where design refers to the joint constructive work of whoever make, build, assembly, and construct such architectures. This may involve engineers, computer scientists, lawyers and other domain experts, as well as those who will suffer the consequences of the decisions mandated to these architectures.

The most important principles that need articulation in the design phase are those that safeguard the checks and balances of the rule of law. This refers for example to legality and fair play for systems employed by public administration, proportionality, accountability, transparency, access to justice, and a series of more specific fundamental rights, such as the presumption of innocence, non-discrimination, privacy, freedom of speech and fair trial. As to code-driven decision systems we need to acknowledge that these principles and rights cannot all be articulated in these systems in a straightforward, scalable manner. They require bespoke architectures, targeting specific contexts, taking into account vulnerable groups or individuals, potential redistribution of risks and benefits, and further consequences for public goods such as safety, trust, trustworthiness, fairness, and expediency – all fine-tuned to the relevant context (of eg, education, policing, healthcare, medical interventions, welfare benefits, employment conditions, and access to all of these).

Nevertheless, we can surmise that at the very least these systems must afford an effective right to appeal against automated decisions, to obtain a meaningful explanation of the logic that informs them and to be given a legal justification for the decision by those who employ the system (note that an explanation of the software is not at all equivalent to a justification of the decision). Of similar importance is that for such rights to be effective, claims that they were violated must be heard by an independent court. These rights have been developed in the context of the General Data Protection Regulation, notably in the prohibition of fully automated decisions that have a significant effect on those whose data are being processed. A major and often cross-disciplinary response has erupted to this prohibition and to a range of concomitant obligations. A new type of cross-disciplinary doctrinal discourse has developed around, for instance, the legal obligation to provide a meaningful explanation on the one hand and 'explainable machine learning' on the other, often co-authored by lawyers, computer scientists

[50] M Hildebrandt, *Law for Computer Scientists and Other Folk* ch 10.2.

and philosophers. In a broader sense, targeting algorithmic decision systems beyond personal data and beyond the jurisdiction of the EU, new subdomains in computer science have emerged, notably XAI (explainable AI) and FAccT (fair, accountable and transparent) computing.

The goal of the collaboration of computer scientists and lawyers in this new strand of computational-legal doctrine is not to develop code-driven compliance, but – on the contrary – to ensure that computer architectures incorporate fundamental safeguards against bias, invasion of privacy, incomprehensible decisions, unreliable assessments, and against an effective denial of access to justice. LPbD must be situated as the primary goal of this new doctrinal development; instead of investing in replacing law with automation, LPbD demands cross-disciplinary investment in keeping the rule of law on track – into the capillaries of code-driven architectures. In relation to code-driven law, this will require a straightforward acknowledgement that such 'law' is not law but public administration or technological management, asserting the need to build LPbD into the technical architectures of code-driven law, and doing the usual – but now cross-disciplinary – doctrinal work on how specified risks must be assessed, mitigated or redistributed.

VII. Finals: the Issue of Countability and Control

As noted above 'not everything that can be counted counts, and not everything that counts can be counted'.[51] I would add that 'not everything that matters can be controlled, and not everything that can be controlled matters'. If we take these admonitions seriously, we may want to hesitate when considering investments in code-driven regulation.

Law is not computable in any final sense, because due to its text-driven multi-interpretability it can be computed in different ways and these different ways will make a difference for those subject to law. In a constitutional democracy such design choices belong to 'the people' and to the courts, not to arbitrary software developers in big tech or big law.

[51] On the need to protect the incomputable self from overdetermination by algorithmic systems: M Hildebrandt, 'Privacy As Protection of the Incomputable Self: Agonistic Machine Learning' (2019) 20(1) *Theoretical Inquiries in Law*, Special Issue on The Problem of Theorizing Privacy 83–121. The solution is not to reject computation but to acknowledge its limitation and its framing problem, see also M Hildebrandt, 'The Issue of Bias. The Framing Powers of Ml' in M Pellilo and T Scantamburio (eds), *Machine Learning and Society: Impact, Trust, Transparency* (MIT Press, 2020), https://papers.ssrn.com/abstract=3497597 accessed 5 January 2020.