Data-driven prediction of judgment. Law's new mode of existence?
Mireille Hildebrand[*]

Abstract

This chapter enquires into the upcoming domain of data-driven 'law', that is, into the use of big data analytics and predictive technologies as a means to *inform* the law. I will argue that this may transform the 'mode of existence' of law, due to the novel 'affordances' of data-driven systems. In the first part I will investigate the promises of legal decision-making based on the mathematical assumptions of machine learning, opening the black box of algorithmic 'insights' at the level of the underlying research design. In the second part I will examine the nature of modern positive law as text-driven law, by highlighting the performative nature of legal effect and how this relates to the force of law. Finally, I will identify some of the challenges presented by data-driven 'law' in terms of legal protection. This will result in a proposal to integrate 'by design' approaches into law and the rule of law, clarifying *how* and *why* 'legal protection by design' is not equivalent with 'legal by design' or 'techno-regulation'.

Keywords: data-driven law, prediction of judgment, machine learning, natural language processing, text-driven law, rule of law, legal protection, speech act theory

1.   Introduction

Though some may still be dreaming of law *as* literature or of law *as* logic,[1] the legal services market is currently swamped by the advertorial lure of big data analytics, promising cheap and easy solutions to many of the inefficiencies and the ineffectiveness of professional legal advice and decision-making. Big law firms are investing in the development, purchase and employment of data-driven legal technologies,[2] facing the challenges of big tech companies

---

[*] Mireille Hildebrandt is a tenured research professor of 'Interfacing law and technology' at Vrije Universiteit Brussel (VUB), appointed by the VUB Research Council, co-director of the Research Group Law, Science, Technology and Society studies (LSTS). She is also part-time full professor of 'Smart environments, data protection and the rule of law' at the Science Faculty of Radboud University, in the Institute of Computing and Information Sciences (iCIS). This paper has been inspired by research done in the context of the ERC Advanced Grant research project Counting as a Human Being in the Era of Computational Law (COHUBICOL), which is funded from 2019-2024 by the European Research Council (ERC) under the HORIZON2020 Excellence of Science program ERC-2017-ADG No 788734, see www.cohubicol.com.

[1] Though scepticism against law as logic has been around for some time, e.g. Duncan Kennedy, 'The Disenchantment of Logically Formal Legal Rationality, or Max Weber's Sociology in the Genealogy of the Contemporary Mode of Western Legal Thought', *Hastings Law Journal* 555 (2004): 1031–76. A more interesting perspective may have been that of James Boyd White, *Justice as Translation: An Essay in Cultural and Legal Criticism* (Chicago: University of Chicago Press, 1990). This is not to claim that logic plays no role in legal reasoning, on the contrary, I would say it constrains the decision space by requiring a justification based on valid legal norms. For such justification, however, logic is just one condition.

[2] William D. Henderson, 'From Big Law to Lean Law', *International Review of Law and Economics*, Unlocking the Law: Building on the Work of Professor Larry E. Ribstein, 38 (1 June 2014): 5–16, https://doi.org/10.1016/j.irle.2013.06.001. Richard Susskind, *The End of Lawyers?: Rethinking the Nature of Legal Services*, Revised edition (Oxford: Oxford University Press, 2010). Note that many legal scholars who advocate legal tech have a background in 'old school' law and economics. For new inroads see *Critical Analysis of Law (CAL)* 5 (1) 2018, the special issue on 'new economic analysis of law' https://cal.library.utoronto.ca/index.php/cal/issue/view/1977.

as new competitors regarding the control over legal knowledge.[3] The old(er) paradigm of 'legal knowledge management' is traded for a belief in opaque systems whose advocates pledge new types of transparency and unprecedented speed when searching, analysing, summarising and building upon big data in law.[4] Such data consists of documentation in the realm of evidence (e-disclosure), case law and legislation to detect patterns in legal reasoning (argumentation mining), or case law and regulation to predict the outcome of future cases (prediction of judgement).[5] Based on the idea that machine-readable access to legally relevant data will bring unprecedented foresight to both lawyers and their clients, some authors even herald 'data-driven law' as the new *path of the law*.[6] In other words, predictive legal technologies may transform the way that law-as-we-know-it exists, disrupting the *mode of existence* of modern positive law. Such disruption is a concern because it may transform or diminish both the legal protection and the instrumentality offered by law and the rule of law.

In the second part of this chapter we will enquire into the lure of data-driven law, starting with a discussion of predictive legal technologies (based on machine learning) and their mathematical assumptions, zooming in on a specific type (called natural language processing), followed by a more in-depth discussion of predictive software applied to the judgements of the European Court of Human Rights. The second part should thus provide us with a first taste of what these technologies afford as 'legal' technologies.

The third part is dedicated to an in-depth enquiry into the *affordances* of text-driven law, to better understand what enabled modern law's mode of existence and what we stand to lose if law comes to depend on data-driven technologies. This will involve an analysis of 'natural language and speech', highlighting the implications of the shift from oral to written speech acts, while paying keen attention to the notion of a 'speech act' as key to a proper understanding of the force of law. This feeds into a discussion of 'structure and action' to mark the multi-interpretability of human action, the contestability this implies and how this relates to the concepts of legal certainty and the rule of law. The third part, finally, clarifies how the text-driven nature of modern law enabled some of the core tenets of the rule of law, notably its positivity (the notion of legal effect) and contestability (based on the multi-interpretability of text and human action). I will argue that both are dependent on the performativity of legal decisions and address the difference between the *performativity* of legal norms and the *performance* of predictive legal technologies.

The fourth part will draw some conclusions and advocate the notion of *legal protection by design* as a means to ensure that legal protection is preserved or reinvented in the era of computational 'law'.

---

[3] Though this is not necessarily a matter of big data analytics, see e.g. https://www.ibm.com/blogs/internet-of-things/iot-spotlight-on-lawtech/.

[4] An in-depth survey of previous and current legal analytics: Kevin D. Ashley, *Artificial Intelligence and Legal Analytics: New Tools for Law Practice in the Digital Age* (Cambridge University Press, 2017).

[5] Michael A. Livermore and Daniel N. Rockmore, eds., *Law as Data: Computation, Text, & the Future of Legal Analysis*, Seminar, book 3 (Santa Fe: SFI Press, 2019).

[6] Benjamin Alarie, 'The Path of the Law: Towards Legal Singularity', *University of Toronto Law Journal* 66, no. 4 (11 November 2016): 443–55.

2. The Lure of Data-Driven 'Law'

A. *A tasting of predictive legal technologies*

i. *Anticipation as the heart of the law*

In this chapter the focus will be on technologies used to predict the outcome of a court case, or prediction of judgement (PoJ). This may remind us of Oliver Wendell Holmes seminal work on *The Path of the Law*, who wrote that:[7] 'The prophecies of what the courts will do in fact, and nothing more pretentious, are what I mean by the law'. Though I often hear people quoting Holmes as saying that 'law is what the courts say', he instead situated the law in the *anticipation* of court decisions, tasking lawyers with the role of providing those subject to the law with a reasonably predictable path of the law. This anticipation may assuage but cannot resolve the fundamental uncertainty about 'what the courts will do in fact'. Instead, it prepares the road for 'legal certainty' by helping those subject to law to estimate the legal consequences of their actions, such as a legal obligation to pay compensation, to perform a contractual obligation, a duty to refrain from interference with another's property, or punishability due to violation of the criminal law. Interestingly, Holmes speaks of what the courts do rather than what they say; he seems on par with Austin's *How to do things with words*.[8]

Legal scholars trained in the assumptions of 'law and economics' often claim that the success of predictive technologies asserts Holmes' prophetic foresight about the direction of the law,[9] as he also claimed that:[10] '[f]or the rational study of the law the blackletter man may be the man of the present, but the man of the future is the man of statistics and the master of economics'. One could surmise that the combination of machine learning (a modulation of statistics) and nudge theory (a modulation of behavioural economics) may one day qualify as the new hermeneutics of legal professionals.[11] Instead of engaging in the close reading of legal texts to argue their case, lawyers may soon be asked to outsource their 'reading' to sophisticated software systems. The claim is that these systems are capable of anticipating what a court would decide, based on the processing of incredible amounts of legally relevant data. If it is true that law must be situated in such anticipation, their predictions could be framed as the new way to establish positive law. Though I believe that this is not so, it is crucial to gain a proper understanding of the difference that makes a difference between human and machine anticipation.

In this part (2) of the chapter we will investigate software capable of what has been coined as 'distant reading' by way of natural language processing (NLP),[12] which is a specific type of machine learning (ML). This will be done under B (on the mathematics of ML), under C (on the nature of meaning in NLP) and under D (taking an example of using NLP to predict

---

[7] Oliver Wendell Holmes, 'The Path of the Law', *Harvard Law Review* 110 (1997): 991–1009.

[8] J.L. Austin, *How to Do Things with Words*, 2nd ed. (Boston: Harvard University Press, 1975).

[9] Ian Ayres, *Super Crunchers:Why Thinking-by-Numbers Is the New Way to Be Smart* (New York: Bantam Books, 2007).

[10] Holmes, 'The Path of the Law'.

[11] Karen Yeung, '"Hypernudge": Big Data as a Mode of Regulation by Design', *Information, Communication & Society* 20, no. 1 (2 January 2017): 118–36, https://doi.org/10.1080/1369118X.2016.1186713.

[12] Franco Moretti, *Graphs, Maps, Trees. Abstract Models for a Literary History* (London: Verso, 2005).

judgements of the European Court of Human Rights). First (under A), however, we will look into some predictive technologies that are not even concerned with content, while nevertheless claiming to reliably predict the outcome of court cases.

ii.   Using machine learning technologies to predict legal judgments

To highlight that a prediction of judgment (PoJ) is not necessarily concerned with substance (content),[13] we will briefly discuss PoJ based on sensor and behavioural data. This should warn us against the trap of thinking that data-driven systems perceive, reason and cognize as we do, just because they manage to arrive at similar conclusions.

In their article 'Emotional Arousal Predicts Voting on the U.S. Supreme Court',[14] Dietrich, Enos and Sen discussed their use of 'vocal pitch data' (sensor data) of Justices during oral arguments in the US Supreme Court (SC) to predict what they term the 'voting behaviour' of the Justices. They summarise their findings in stating that the 'results show that the higher emotional arousal or excitement directed at an attorney compared to his or her opponent, the less likely that attorney is to win the Justice's vote (p < 0,001)'. Interestingly, the predictive value of their experiment was 1,79% higher that the famous algorithm developed by Katz, Bommarito, and Blackman in 2014,[15] suggesting that emotional inferences based on vocal pitch were more predictive than the 95 variables used by Katz, Bommarito II, and Blackman.

Three years later Katz, Bommarito II, and Blackman published their latest findings in a seminal article,[16] claiming their algorithm provides for predictions that generalise beyond the data on which it was trained, is consistent across time and across different Justices, and notably also generalises to future cases (this is called 'out of sample' testing in ML-speak). The so-called accuracy of their predictions is said to be 70,2 % for the outcome of the case and 71,9 % for the voting behaviour of individual Justices.[17] This links back to the fact that – in line with the limits of predictive technologies – they had to seek out questions that allow for discrete answers:[18]

---

[13] In ML terms this is now called 'content'. ML thinks in terms of content and meta-data, as we (lawyers) think in terms of substance and procedure (which is not to say that meta-data is equivalent with procedure). In relation to PoJ we have content data (text of judgements), sensor data (e.g. voice pitch of judges), behavioural data (e.g. voting behaviour of judges), and meta-data (e.g. date, case-number, court, judges, outcome). Note that the categories are debatable and may overlap.

[14] Bryce J. Dietrich, Ryan D. Enos, and Maya Sen, 'Emotional Arousal Predicts Voting on the U.S. Supreme Court', *Political Analysis* 27, no. 2 (April 2019): 237-243.

[15] Daniel Martin Katz, Michael J. Bommarito II, and Josh Blackman, 'Predicting the Behavior of the Supreme Court of the United States: A General Approach', *ArXiv:1407.6333 [Physics]*, 23 July 2014, http://arxiv.org/abs/1407.6333.

[16] Daniel Martin Katz, Michael J. Bommarito II, and Josh Blackman, 'A General Approach for Predicting the Behavior of the Supreme Court of the United States', *PLOS ONE* 12, no. 4 (12 April 2017): e0174698, https://doi.org/10.1371/journal.pone.0174698.

[17] Accuracy is a technical term in ML, it refers to the ratio between the number of correct predictions and the total number of predictions. Other metrics are precision, which refers to the ratio between true positives and the sum of true and false positives, and recall, which refers to the ratio between true positives and the sum of true positives and false negatives. Especially where the real-world implications of a positive or negative prediction matter, it becomes very important to know precision and recall. To seriously assess the reliability of a predictive system all three metrics must be provided.

[18] Katz, Bommarito II, and Blackman, 'A General Approach for Predicting the Behavior of the Supreme Court of the United States'.

> While many questions could be evaluated, the Court's decisions offer at least two discrete prediction questions: 1) will the Court as a whole affirm or reverse the status quo judgment and 2) will each individual Justice vote to affirm or reverse the status quo judgment?

Though a lawyer may find these the least interesting and the least informative questions, for a machine that is at its best when crunching numbers this is a wonderful chance to show how it can outperform human experts, distracted as they may be by the substance (content) of the case law.

Katz, Bommarito II, and Blackman trained their algorithm on the Supreme Court Database (SCDB), containing the machine-readable text of all court cases decided by the SC, where two hundred and forty metadata were added (as relevant variables), such as:[19]

> chronological variables, case background variables, justice-specific variables, and outcome variables. Many of these variables are categorical, taking on hundreds of possible values; for example, the ISSUE variable can take 384 distinct values. These SCDB variables form the basis for both our features and outcome variables.

To train an algorithm, a 'machine readable task' must be developed. This is done in terms of a so-called 'target variable', in this case a variable that refers to the outcome of the case (reversal or affirmation of the verdict that is appealed). The *outcome variable* must be correlated with relevant *input variables*. All this goes back to the need to develop a type of questions that can be formalised to enable number crunching, instead of questions that require acuity in terms of ambiguous concepts with an open texture. Next, a subset of the available variables is selected as potentially relevant *features*, enriched with additional features that were 'engineered' by the authors, e.g. those that relate to the Circuit courts whose decisions are appealed, those related to whether or not oral arguments were heard, and those specifying the historical reversal rate (per Justice).

After thus curating and enriching the *training data*, a model is constructed to detect potentially relevant correlations between the *feature set* and the *target variable*. The model used here is a so-called 'random forest classifier', i.e. a set of *statistical learners* (trees) that correlate features with the target variable and then take the average over the set (the forest).[20] Please note that the term 'statistical learners' refers to the use of mathematical functions that best 'capture' the correlation. In ML this set is usually called the *hypothesis space*. Whereas lawyers may think hypotheses are statements such as 'if intent is not proven conviction is not likely', in data-driven legal technologies a hypothesis is a *mathematical function* that depicts a very precise mathematical relationship between a specific combination of variable inputs (with different weights) and a specified target variable.

As indicated above, the results of thus training the algorithm are presented in terms of the accuracy of the model: in 70,2 % the model correctly predicted the outcome of the case, which depends on the voting of all the Justices, and in 71,9 % the model correctly predicted

---

[19] Katz, Bommarito II, and Blackman.

[20] Taking the average over the set reduces potential *overfitting*, which refers to developing a model that is highly accurate on the training data but not on new data (meaning it does not generalise).

the votes of the individual Justices. The model does not *explain* why Justices voted in one way or another (a correlation is not a cause), nor does it *justify* the decision of the SC (it does not engage in legal reasoning). The only thing it does is predict binary decisions made by the SC and by individual Justices.

To argue why this is interesting at all, Katz, Bommarito and Blackman compare the accuracy of their predictions with the predictive force of other models, which is called comparing against a *baseline* or *null model*.[21] One such model would be to flip a coin, though that does not sound very impressive as it refers to a random distribution of outcomes. According to the authors, the legal community agrees that the best bet is to predict reversal, because – statistically – this is what the SC more often decided during the previous 35 terms: 57% of Justices voted reverse, and 63% of SC decisions were reverse. Interestingly, this baseline does not hold if one goes further back in time, whereas their own model does hold sway. After adding some constraints, they come up with three baseline models[22] and conclude that their predictions 'outperform' the most relevant one by 5%.

Though this may sound like a minor advantage, they point out that: 'Indeed, with respect to markets, given judicial decisions can impact publicly traded companies … even modest gains in prediction can produce significant financial rewards'. This raises the question of the aim of this type of legal technology. Is the lure of their accuracy connected with the intent to use the measure they provide as a target (for financial gain)? In that case we may be in for some very unwelcome dynamics, taking into account the so-called Goodhart effect that 'predicts' that:[23] '[w]hen a measure becomes a target, it ceases to be a good measure'. This can be due to attempts to game the measures, but just as well because human beings are always in the process of anticipating the consequences of their actions. Both those using the measure as a target (an insurance company anticipating whether they will win a case) and those being targeted (judges with access to prediction of their voting behaviour) will probably change their actions in view of the predictions. Even if those changes are small, they may have far reaching consequences.[24]

B.   *The mathematical assumptions of machine learning*

iii.   *Learning as a machine*

In his textbook Tom Mitchell's describes ML as drawing on:[25] 'ideas from a diverse set of disciplines, including artificial intelligence, probability and statistics, computational complexity, information theory, psychology and neurobiology, control theory, and philosophy.' For a machine to learn, a well-defined learning problem must be articulated,

---

[21] A baseline is any model that one hopes to outperform, e.g. the most frequent output variable, a random outcome, the median of the training set, or simply another ML model you want to outperform. A null model is based on the assumption that the features one wishes to test do not influence the target variable.

[22] Paul Cilliers, *Complexity and Postmodernism: Understanding Complex Systems*, 1 edition (London; New York: Routledge, 1998).

[23] Marilyn Strathern, '"Improving Ratings": Audit in the British University System', *European Review* 5, no. 3 (July 1997): 305–21, https://doi.org/10.1002/(SICI)1234-981X(199707)5:3<305::AID-EURO184>3.0.CO;2-4.

[24] As human society is highly complex, small changes have big consequences, see

[25] Thomas Mitchell, *Machine Learning*, 1 edition (New York: McGraw-Hill Education, 1997).

because computing systems can only operate when given formalised *machine-readable tasks*:[26]

> A well-defined learning problem requires a well-specified task, performance metric, and source of training experience.
>
> Designing a machine learning approach involves a number of design choices, including choosing the type of training experience, the target function to be learned, a representation for this target function.

This implies a number of decisions, including the making of a number of assumptions that may or may not be valid. Many of these decisions relate to issues on the nexus of logic and statistics, where grotesque mistakes can be made.[27] In ML these mistakes are difficult to assess because they are hidden in the design process. The most blatant and hazardous, though also inevitable and reasonable assumption concerns *the distribution of training data and future data*:[28]

> We shall see that most current theory of machine learning rests on the crucial assumption that the distribution of training examples is identical to the distribution of test examples. Despite our need to make this assumption in order to obtain theoretical results, it is important to keep in mind that this assumption must often be violated in practice.

iv.   *No machine learning without mathematical assumptions*

This links up with another – even more fundamental – assumption of ML, which is that of an underlying mathematical reality that 'maps' human intercourse. Without that assumption ML makes no sense. It is the ground on which all ML research design stands, and therewith the point of departure for all ML applications.[29] Though one can be agnostic about whether such mapping implies either causality between the mathematics and the real world, or neo-Platonic realism, it remains an assumption that requires critical attention. Legal technologies that employ ML techniques, or have been built using ML techniques, depend on this assumption and require keen attention to its implications. For instance, the *limits* that are inherent in mathematical and computational inquiry should be foregrounded when trusting predictive software, taking into account that such limits may impact the *reliability* of these techniques, especially if there are no institutional checks and balances to detect and counter unsubstantiated claims regarding the functionality and validity of their outputs.

I believe that reducing the problem of ML application to their black box nature or to the potential bias they may create or reinforce distracts attention from the range of trade-offs that are inherent in any ML *research design*. Whereas lawyers and computer scientists are

---

[26] Mitchell.

[27] Gerd Gigerenzer, 'Statistical Rituals: The Replication Delusion and How We Got There', *Advances in Methods and Practices in Psychological Science* 1, no. 2 (1 June 2018): 198–218, https://doi.org/10.1177/2515245918771329; Tal Yarkoni, 'The Generalizability Crisis', preprint (PsyArXiv, 22 November 2019), https://doi.org/10.31234/osf.io/jqw35.

[28] Mitchell, *Machine Learning*.

[29] Dan McQuillan, 'Data Science as Machinic Neoplatonism', *Philosophy & Technology* 31, no. 2 (1 June 2018): 253–72, https://doi.org/10.1007/s13347-017-0273-3.

exploring explainable AI on the one hand,[30] and seeking ways to ensure fair computing on the other,[31] the more fundamental question about what kind of questions these systems can actually address is hardly touched upon.

C. *Natural language processing as 'distant reading'*

v. *Distant reading of legal text*

Having discussed several examples of PoJ technologies based on voice pitch, prior voting behaviours, and other metadata, we will now focus on the more obvious legal ML technology, namely 'natural language processing' (NLP). Though NLP is again about detecting mathematical patterns in data, the relevant data here is legal text; the mathematics concerns the content, not the metadata.[32]

In 2000, literary historian Franco Moretti explained how computational techniques can be used map and mine an immensely large corpus of texts for potentially relevant patterns. His book was called 'distant reading',[33] and has captured the imagination of those versed in what has been called the 'digital humanities', by uncovering heretofore unknown patterns within and without the literary canon. To give just one example we can quote Clement on her findings after 'distant reading' the infamously arduous novel *The Makings of Americans* by Gertrude Stein:[34]

> The particular reading difficulties engendered by the complicated patterns of repetition in The Making of Americans by Gertrude Stein make it almost impossible to read this text in a traditional, linear manner. However, by visualizing certain patterns and looking at the text 'from a distance' through textual analytics and visualizations, we are enabled to make readings that were formerly inhibited. Initial analysis on Making within the MONK (metadata offer new knowledge) project (http://www.monkproject.org/) has yielded evidence which suggests that the text is intricately and purposefully structured. Using text mining to retrieve repetitive patterns and treating each as a single object makes it possible to visualize and compare the three dimensions upon which these repetitions co-occur—by length, frequency, and location—in a single view. Certainly, reading The Making of Americans in a traditional way appears to have yielded limited material for

---

[30] Margot E. Kaminski, 'The Right to Explanation, Explained', SSRN Scholarly Paper (Rochester, NY: Social Science Research Network, 15 June 2018), https://papers.ssrn.com/abstract=3196985; Alexandra Chouldechova and Aaron Roth, 'The Frontiers of Fairness in Machine Learning', *ArXiv:1810.08810 [Cs, Stat]*, 20 October 2018, http://arxiv.org/abs/1810.08810; Julia Powles, 'The Seductive Diversion of "Solving" Bias in Artificial Intelligence', Medium, 7 December 2018, https://medium.com/s/story/the-seductive-diversion-of-solving-bias-in-artificial-intelligence-890df5e5ef53.

[31] Alexandra Chouldechova and Aaron Roth, 'The Frontiers of Fairness in Machine Learning', *ArXiv:1810.08810 [Cs, Stat]*, 20 October 2018, http://arxiv.org/abs/1810.08810; Julia Powles, 'The Seductive Diversion of "Solving" Bias in Artificial Intelligence', Medium, 7 December 2018, https://medium.com/s/story/the-seductive-diversion-of-solving-bias-in-artificial-intelligence-890df5e5ef53.

[32] Note that the distinction between content and metadata is itself a new way of understanding legal text.

[33] Franco Moretti, *Distant Reading*, 1 edition (London; New York: Verso, 2013); Kathryn Schulz, 'What Is Distant Reading?', *The New York Times*, 24 June 2011, https://www.nytimes.com/2011/06/26/books/review/the-mechanic-muse-what-is-distant-reading.html. Stefan Jänicke et al., 'On Close and Distant Reading in Digital Humanities: A Survey and Future Challenges', 2015, https://doi.org/10.2312/eurovisstar.20151113.

[34] Tanya E. Clement, '"A Thing Not Beginning and Not Ending": Using Digital Tools to Distant-Read Gertrude Stein's The Making of Americans', *Literary and Linguistic Computing* 23, no. 3 (1 September 2008): 361–81, https://doi.org/10.1093/llc/fqn020. 361 (the quotation is the full abstract).

> scholarly work, but reading the text differently, as an object of pairings or as parts of combinations, ultimately works in contrast to the supposition that the text is only meaningful to the extent that it defeats making meaning. A distant view of the text's structure allows us to read the text as an object that becomes, as it continues to turn in on itself with a centrifugal force, a whole history without beginning or ending.

The humanities have been known for their methodological rigour in 'close reading' and philological inquiry, supposedly thus uncovering the meaning of literary and other texts. Computational 'distant reading' confronts us with an entirely different perspective on our ability to map and navigate complex textual sources. This has heralded a new era of purportedly enhanced comprehension of textual corpora, raising a number of interesting questions about the extent to which these computational X-rays are capable of grasping a deeper, novel, more relevant or merely different meaning of text. Especially where the sheer volume precludes close reading by any individual human scholar. When investigating 'the meaning and mining of legal text' in a chapter of a Berry's *Understanding humanities*, I referred to the relevance of previous work by Moretti and Clement:[35] 'Rather than painstakingly reading one's way into a mass of individual cases to see how codes and statutes are interpreted, helped by legal treatises on the relevant legal doctrine, the developers of the inference machines discussed above hope that they will at some point provide a more immediate oversight regarding whatever may be the factual and legal issues at stake.'

vi.   Caveats when engaging with distant reading of legal text

Eight years later, Livermore and Rockmore, in their *Law as Data*, again refer to Moretti, suggesting that 'the potential utility of distant reading for many different types of legal scholarship is great',[36] quoting a number of studies that employ a variety of NLP methodologies to actually mine legal text. Their overview, however, does not address the new kind of questions that are raised by NLP and does not provide the caveats lawyers should take into account when outsourcing the 'reading' of legal text. Clearly, the output of the computational methodologies presents us with a new kind of 'text': trees, graphs and scattergrams, or, output based on the latest approaches in NLP, such as the fantastic bi-directional encoder representations from transformers (BERT),[37] which has revolutionised both search and machine translation. This will require what I have called a 'new hermeneutics', capable of sustaining legal certainty in the face of oftentimes inscrutable claims about the reliability and relevance of NLP's output for the law. Based on the work of

---

[35] M Hildebrandt, 'The Meaning and Mining of Legal Texts', in *Understanding Digital Humanities: The Computational Turn and New Technology*, ed. D.M. Berry (London: Palgrave Macmillan, 2011), 153. Referring to Moretti, *Graphs, Maps, Trees. Abstract Models for a Literary History*; Tanya Clement et al., 'How Not to Read a Million Books', Harvard University, Cambridge, MA, 2008, http://www3.isrl.illinois.edu/~unsworth/hownot2read.html.

[36] Michael A. Livermore et al., 'Distant Reading the Law', in *Law as Data: Computation, Text, and the Future of Legal Analysis* (Santa Fe: SFI Press, 2019), 19.

[37] Jacob Devlin et al., 'BERT: Pre-Training of Deep Bidirectional Transformers for Language Understanding', *ArXiv:1810.04805 [Cs]*, 24 May 2019, http://arxiv.org/abs/1810.04805.

Sculley and Pasanek,[38] I articulated five caveats for legal scholars that still seem highly relevant:[39]

> First, the collaborative effort of computer engineers and lawyers who sit down together to mine legal data bases should focus on making their assumptions about the scope, function and meaning of the relevant legal texts explicit.
>
> Second, they should use multiple representations and methodologies, thus providing for a plurality of mining strategies that will most probably result in destabilising any monopoly on the interpretation of the legal domain.
>
> Third, all trials should be reported instead of 'cherry-picking' those results that confirm the experimenters' bias. As they suggest, at some point failed experiments can reveal more than supposedly successful ones. This is particularly important in a legal setting, because the legal framework of constitutional democracy should prevent individual outliers and minority positions from being overruled by dominant frames of interpretation.
>
> Fourth, in a legal setting the public interest requires transparency about the data and the methods used to make the data-mining operations verifiable by other joint ventures of lawyers and software engineers.
>
> This connects to their fifth recommendation, regarding the peer review of the methodologies used. The historical artifact of a legal system that is relatively autonomous in regard to both morality and politics, and safeguarded by an independent judiciary has been nourished by an active class of legal scholars, and practitioners willing to test and contest the claimed truth of mainstream legal doctrine. Only a similarly detailed and agonistic scrutiny of the results of data-mining operations can sustain the fragile negotiations of the rule of law.

Since writing this, I have developed the notion of agonistic machine learning,[40] to refer to ways and means to ensure both the *testing* and the *contesting* of truth claims made 'on behalf of' machine learning applications. More recently I added a plea to introduce a pertinent requirement of preregistration of ML research design by developers and producers of ML systems, in combination with strict liability for damage caused by the employment of such systems, where such liability should target those who put these systems on the market (as they stand to profit).[41] Though it may not be obvious whether the use of NLP as a legal technology has caused damage, it seems crucial that technologies capable of influencing or even determining legal advice and legal decision-making must obey the highest standards in terms of testability and contestability to thus ensure their reliability.

---

[38] D Sculley and Bradley M Pasanek, 'Meaning and Mining: The Impact of Implicit Assumptions in Data Mining for the Humanities', *Literary and Linguistic Computing* 23, no. 4 (2008): 409–24.

[39] Hildebrandt, 'The Meaning and Mining of Legal Texts': 155.

[40] Mireille Hildebrandt, 'Privacy As Protection of the Incomputable Self: Agonistic Machine Learning', *Theoretical Inquiries in Law*, Special Issue on The Problem of Theorizing Privacy, 20, no. 1 (forthcoming 2019), https://papers.ssrn.com/abstract=3081776.

[41] Mireille Hildebrandt, 'Preregisration of Machine Learning Research Design. Against P-Hacking', in *BEING PROFILED:COGITAS ERGO SUM* (Amsterdam Univesrity Press, 2018).

D.   The case of predicting the European Court of Human Rights

vii.   Using supervised NLP for PoJ

Having explained the affordances of legal technologies that aim to predict the outcome of court cases, highlighting the need to verify and falsify the claims made regarding the behaviour of these technologies, I will now discuss a European example of PoJ. Other than the previous examples, under A, the predictions are based on NLP instead of sensor data (such as vocal pitch) or behavioural metadata (such as voting behaviour of individual judges). This should give us a clear example of 'distant reading' of legal text.

In 2016 Aletras et al. published the findings of their application of NLP on judgements of the ECtHR.[42] After training an algorithm on the published texts of judgements relating to a limited set of articles of the ECHR, they concluded that their algorithm managed to correctly predict a binary outcome (violation or no violation) with 79 % accuracy. Based on their findings, they also suggest that their algorithm confirmed that it is the facts of the case rather than issues of law that decide the outcome, thus supposedly validating core tenets of legal realism. In other work I have explained that the conclusions are brittle, based on a number of assumptions, most of which don't fly.[43] The most problematic assumptions were that (1) the text of published judgements is proper proxy for the underlying briefs and evidence, and that (2) the rendering of the facts as related in the Court's judgement is an independent variable. The first assumption is problematic because many relevant data are missing (e.g. sections on law in cases deemed inadmissible by the Court). One could say that mistaking the published judgement for the relevant data demonstrates the 'survival bias' or the 'low hanging fruit bias'.[44] The second assumption is clearly incorrect as any lawyer knows that the Court will formulate the facts such that they are well attuned to the verdict (as the authors actually acknowledged). Though in machine learning it is necessary to make a number of assumptions to enable the computational research design, it is not a good idea to draw real life conclusions based on the outcome.

viii.   Using unsupervised NLP for PoJ

In 2019, Chalkidis, Androutsopoulos, and Aletras published new research into the use of NLP on legal text, this time based on the employment of so-called neural nets. They speak of 'legal judgement prediction', which they define as 'the task of automatically predicting the outcome of a court case, given a text describing the case's facts'.[45] The paper 'sells' the idea of modelling POJ by claiming that '[s]uch models may assist legal practitioners and citizens, while reducing legal costs and improving access to justice'. They add that '[l]awyers and judges can use them to estimate the likelihood of winning a case and come to more consistent and informed judgments, respectively'. This again implies that the research aims for people

---

[42] Nikolaos Aletras et al., 'Predicting Judicial Decisions of the European Court of Human Rights: A Natural Language Processing Perspective', *PeerJ Computer Science* 2 (24 October 2016): e93, https://doi.org/10.7717/peerj-cs.93.
[43] M Hildebrandt, 'Algorithmic Regulation and the Rule of Law.', *Philosophical Transactions of the Royal Society A* 376 (2018), http://dx.doi.org/10.1098/rsta.2017.0355.
[44] On the 'cognitive' biases of ML systems see https://www.geckoboard.com/best-practice/statistical-fallacies/.
[45] Ilias Chalkidis, Ion Androutsopoulos, and Nikolaos Aletras, 'Neural Legal Judgment Prediction in English', *ArXiv:1906.02059 [Cs]*, 5 June 2019, http://arxiv.org/abs/1906.02059. 1.

to use the *measure* that is the outcome of the research as a *target* for cost reduction, 'better' judgements and better access to justice. This type of claims is questionable, more fit for consultants than scientists, and highly problematic in view of the Goodhart effect referred to above. Obviously, the ability to predict the outcome of cases, when based on statistical correlations, will not necessarily result in better judgements. The authors seem to be aware of this, in their conclusions, to which I will return after recounting their research.

The research presented in 2019 differs from that in 2016. First, the dataset is extended from 600 to 11.500 cases. Second, the binary prediction (violation, no violation) is no longer limited to a restricted set of articles but concerns violations of any article (including those in the Protocols). Third, the prediction not only concerns whether or not the Court decided a violation, but also whether the case is important on a scale of 1 – 4 (the scores were provided by the Court). Fourth, the methodology differs substantially from that used in 2016, as the authors tested four types of 'neural models', including a modulation of the famous BERT model that involves highly complex and layered processing of word sequences, taking into account the way potentially relevant wordings are embedded in sentences and textural corpora. Using the modulated BERT model, the accuracy for the binary violation task is now 82 %, more than 10% better than that of the previous research.

ix. The accuracy – reliability trade-off

As with many neural networks, high accuracy combines with a lack of interpretability.[46] Whereas the previous research allowed one to detect which features correlated with the target variable, the neural nets operate as a blackbox and hide potentially relevant features, their interrelations and their weights. As the authors assert in the conclusions, the model does not provide an explanation for its predictions.[47] Instead of speaking of an explanation, however, they refer to a justification, pointing out that judges will need a 'justification' for the model's predictions. An explanation of how the system reached its conclusion, however, does not provide a justification in the legal sense, which would indeed be of interest for the law. Chalkidis, Androutsopoulos, and Aletras plan 'to expand the scope of this study by exploring the automated analysis of additional resources (e.g., relevant case law, dockets, prior judgments) that could be then utilized in a multi-input fashion to further improve performance and justify system decisions'.[48] Once again, what they call 'a justification' of system decisions is not a justification in the legal sense, but the verification of a causal influence of specified input features on the system's output variable (violation or no violation). The law is, however, not about the causality between specific arguments and their

---

[46] The accuracy-interpretability trade-off suggests that though one cannot explain the outcome, in the case of neural networks it is nevertheless highly accurate. This is, however, not necessarily the case, and the lack of interpretability makes it hard to check, see notably Rich Caruana et al., 'Intelligible Models for HealthCare: Predicting Pneumonia Risk and Hospital 30-Day Readmission', in *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '15 (New York, NY, USA: ACM, 2015), 1721–1730, https://doi.org/10.1145/2783258.2788613.

[47] Chalkidis, Androutsopoulos, and Aletras, 'Neural Legal Judgment Prediction in English'. 5, also referring to Sarthak Jain and Byron C. Wallace, 'Attention Is Not Explanation', in *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)* (NAACL-HLT 2019, Minneapolis, Minnesota: Association for Computational Linguistics, 2019), 3543–3556, https://doi.org/10.18653/v1/N19-1357.

[48] Chalkidis, Androutsopoulos, and Aletras, 'Neural Legal Judgment Prediction in English', 5.

conclusion, which would entail a category mistake.[49] The law requires that a legal decision (the attribution of legal effect, such as 'these facts qualify as a violation') is based on the fulfilment of specified legal conditions, which in turn requires a reflective equilibrium between qualifying the facts in terms of the relevant legal norm and the relevant legal norm in terms of the facts.[50] Other than the predictive NLP systems assume, the facts are not given. As the French say: '*les faits sont faits*' (the facts are made).[51] Not in the sense of 'made up', but in the sense that what counts as the criminal offence of murder depends on the double focus of looking at the facts from the perspective of the criminal law and looking at the criminal law from the perspective of the facts. There is a feedback loop between how we read the facts and how we read the law, which marks the difference between *logical consistency* and what Dworkin called the *integrity of the law*.[52] Consistency would freeze both the law and our understanding of the facts, whereas integrity leaves room for the multi-interpretability of both, while still requiring a decision – grounded in the shared experience of being human in a shared jurisdiction. This 'shared world' dimension is connected with empathy and common sense, tacit knowledge and the need to anticipate the institutional world we navigate.[53] This dimension, however, is not shared by the NLP systems that cannot but restrict themselves to text *as data*. In short, these systems have a very poor grasp of the world that grounds the ECtHR, which is also the world the Court must decide upon in its verdicts.

3.   Text-driven law: what we stand to lose

A.   *Language and speech*

x.   *The nature of text-driven law*

In light of the statistical underpinnings of machine learning, as discussed in part 2, let us return to Holmes prophecy. A prophecy implies uncertainty. It may refer to an educated guess, to trained intuition, while a machine learning expert may associate a prophecy with feedforward, backpropagation, or reinforcement mechanisms. A guess can be explained by way of the underlying psychology, or by way of the sociological patterns that prevail in a particular society. It can also be explained with the help of model interpretability, if the prophecy was provided by an NLP prediction algorithm. In the context of the law, however, the validation of the prophecy must be developed in the space of interpretation and argumentation; it is not a matter of an explanation (in the causal or motivational sense) but of a justification. One can understand law and the rule of law as an institution that restricts the decision space of judges: whatever their inclination, the law constrains the kind of reasons

---

[49] Gilbert Ryle, *The Concept of Mind* (New York: Barnes & Noble, 1949).

[50] John Rawls, *A Theory of Justice*, vol. Original (Cambridge, Mass.: Belknap Press, 2005).

[51] Barbara J. Shapiro, *A Culture of Fact: England, 1550–1720*, 1 edition (Ithaca, NY: Cornell University Press, 2003).

[52] Ronald Dworkin, *Law's Empire* (Glasgow: Fontana, 1991); Francisca Christina Wilhelmina de Graaf, 'Dworkin's Constructive Interpretation as a Method of Legal Research.', *Law and Method*, no. 12 (2015), https://doi.org/10.5553/REM/.000012.

[53] Mireille Hildebrandt, 'The Artificial Intelligence of European Union Law', *German Law Journal* 21, no. 1 (January 2020): 74–79, https://doi.org/10.1017/glj.2019.99.

they can provide and therewith the set of decisions they can take. That is where the naïve interpretation of legal realism fails; it is not 'the facts' as given parameters that decide the case, because courts are not free to decide in whatever way suits them. Even the fabrication of the facts is bound by the law of evidence. Their construction must be reliable, and their truth claims must verifiable as well as falsifiable, though the extent to which this is the case will depend on the legal domain.[54]

In this part we address the question of whether the anticipation of legal decision-making by the courts that is core to positive law, implies that prediction of judgements by ML applications should be qualified as law.[55] If they were to qualify as law, we would have to attribute legal effect to the output of prediction machines. If not, they might instead count as public administration[56] or technological management,[57] and in that case, they must be brought under the rule of law - implementing checks and balances such as the right level of transparency, to achieve their contestability. This makes the question of whether or not PoJ qualifies as law of prime importance. To approach this question we will investigate the nature of text-driven law, its affordances and how they align with the rule of law as a system of checks and balances that counters the law of the jungle.[58] To come to terms with the shift from current law to data-driven 'law' we need to re-assess the nature of modern positive law *as text-driven law*, embodied in a specific technology, i.e. the printing press. Precisely because the textual nature of contemporary law seems obvious, we may overlook the critical affordances of text-driven infrastructures, taking for granted what may be on the verge of a major transformation. To investigate the assumptions implied by law's dependence on text and their implications, I will employ the work of Paul Ricoeur,[59] a continental (French) philosopher in the traditions of phenomenology and hermeneutics, who integrated relevant parts of analytical philosophy in his work, more notably speech act theory. His writings on interpretation align with historical and anthropological work on the relationship between speech, writing and printing press on the one hand, and the size, structure and organisation of

---

[54] In private law the autonomy of the parties may reduce the role of the court in establishing the facts of the case, to the extent that the parties agree. In criminal law the court may has a duty to establish the material truth, even if it differs from what prosecutor and defence agree upon (especially in civil law jurisdictions).

[55] See also Kerr's argument that Holmes focused on prediction to enable protection of those subject to law rather than enabling protection of either the state or other big players against individuals. Ian Kerr, 'Chapter 4: Prediction, Preemption, Presumption: The Path of Law After the Computational Turn', in M. Hildebrandt and B. Van den Berg (eds.) *Privacy, Due Process and the Computational Turn: The Philosophy of Law Meets the Philosophy of Technology* (Abingdon, Oxon, [England] ; New York: Routledge, Taylor & Francis, 2013), 91–120.

[56] Hildebrandt, 'Algorithmic Regulation and the Rule of Law.'

[57] Roger Brownsword, 'Technological Management and the Rule of Law', *Law, Innovation and Technology* 8, no. 1 (2 January 2016): 100–140, https://doi.org/10.1080/17579961.2016.1161891.

[58] Chapter 8 in Mireille Hildebrandt, *Smart Technologies and the End(s) of Law. Novel Entanglements of Law and Technology* (Cheltenham: Edward Elgar, 2015); Mireille Hildebrandt, 'Law As an Affordance: The Devil Is in the Vanishing Point(s)', *Critical Analysis of Law* 4, no. 1 (3 April 2017), http://cal.library.utoronto.ca/index.php/cal/article/download/28154.

[59] Paul Ricoeur, 'The Model of the Text: Meaningful Action Considered as a Text', *New Literary History* 5, no. 1 (1973): 91–117. Austin, *How to Do Things with Words*; John R. Searle, *Speech Acts: An Essay in the Philosophy of Language*, 34th. print (Cambridge: Cambridge University Press, 2011); Neil MacCormick and Ota Weinberger, *An Institutional Theory of Law: New Approaches to Legal Positivism* (Dordrecht; Boston; Hingham, MA, U.S.A.: D. Reidel, 1986).

society on the other.[60] Don Ihde, the founding father of philosophy of technology in the empirical tradition, has worked extensively with Ricoeur's work, demonstrating its relevance for understanding the constitutive nature of information and communication infrastructures (ICIs).[61] After investigating 'natural language and speech' in this section (A), to highlight the nature of oral and written speech acts, we will discuss 'structure and action' to mark the nature of meaningful action (under B), followed by conclusions about the nature of text-driven law and the rule of law (C).

xi.   Language system and language usage

In line with linguistic theory, more precisely semiology, Ricoeur distinguishes between *langue* (language system or code) and *parole* (speech, discourse, or language usage):[62]

> [Speech] is always realized temporally and in a present, whereas the language system is virtual and outside of time. Whereas language lacks a subject (in the sense that the question 'Who is speaking?' does not apply at its level), [speech] refers to its speaker by means of a complex set of indicators such as the personal pronouns. We will say that the 'instance of speech' is self-referential.

Though language is virtual in the sense of providing myriad possibilities to express oneself without being itself an expression,[63] I would not qualify it as outside of time. Language is both the *condition of possibility for* and *the result of* speech, meaning it is necessarily in constant flux, helping us to navigate a changing environment. What Ricoeur rightly points out is that language is not something we invent as individuals, instead, we are 'thrown into' a language that enables us to speak while constraining what we can say, and how. In that sense language is 'given' and outside of time. With 'outside of time' Ricoeur refers to the intra-linguistic meaning of a term that is restricted to the other terms it refers to. If we ask for the meaning of a term, the answer will be given by the use of other terms, thus creating a web of meaning. However, to navigate the real world, we need to simultaneously grasp the extra-linguistic meaning of a term. That is, its reference to our material and institutional environment. Speech, therefore, is an event that brings together the intra- and extra-linguistic meaning, thus producing a 'world', which depends on an interpretive community to make sense.

If we apply Ricoeur's distinction between language and speech to the law, we can affirm that a legal decision is always realized temporally and in a present, whereas the legal system is virtual and to some extent given. It is 'outside of time' in the sense that its intra-systematic

---

[60] Walter Ong, *Orality and Literacy: The Technologizing of the Word* (London/New York: Methuen, 1982); Jack Goody, *The Logic of Writing and the Organization of Society* (Cambridge [Cambridgeshire]; New York: Cambridge University Press, 1986); Elisabeth Eisenstein, *The Printing Revolution in Early Modern Europe* (Cambridge New York: Cambridge University Press, 2005).

[61] Don Ihde, *Technology and the Lifeworld : From Garden to Earth*, The Indiana Series in the Philosophy of Technology (Bloomington: Indiana University Press, 1990); Hildebrandt, *Smart Technologies and the End(s) of Law. Novel Entanglements of Law and Technology*.

[62] Ricoeur, 'The Model of the Text: Meaningful Action Considered as a Text'. The translation I quote uses 'discourse' for *parole*; I prefer to use 'speech'.

[63] Virtuality and actuality are used here in a sense similar to that introduced by Deleuze and Guatari, as elaborated by Pierre Lévy, *Becoming Virtual. Reality in the Digital Age* (New York and London: Plenum Trade, 1998).

coherence remains within the bounds of the legal system; legal concepts are explained by referring to other legal concepts. The legal system lacks a subject (in the sense that the question 'Who is speaking?' does not apply at its level), whereas a legal decision refers to its speaker by means of a complex set of indicators such as the personal pronouns. We will say that the 'instance of legal decisions' is self-referential – where the speaker may be a legislature, a court, public administration, a corporation, a natural person or any other legal subject. A legal decision refers to an extra-systematic reality that is configured by the intra-systematic web of references (terms, rules and principles) thus deciding both the meaning of the law and its application.

Ricoeur finds that:

> Whereas language is only the condition for communication, for which it provides the codes, it is in [speech] that all messages are exchanged. In this sense, [speech] alone has not only a world, but an *other* - another person, an interlocutor to whom it is addressed.

We can once more apply this to the law, by confirming that a legal decision is always about something. It refers to a world which it claims to order and decide. And, whereas the legal system is only the condition for legal effect, for which it provides the codes, it is in legal decision-making that legal effect is attributed. In this sense, a legal decision creates not only a world and has an author, but also addresses an *other* - another person, an interlocutor. The legislature addresses those under its jurisdiction; courts address the parties, the lower court, the public prosecutor and the defendant; a party to a contract address the other party; the owner who transfers property addresses all others that should refrain from interference with that property. The problem with, for instance, NLP or distant reading of legal text, is that computing systems only have access to the intra-systematic web of legal references.

xii.   From speaking to writing

Speech (language usage) is a matter of speaking or writing. Ricoeur developed some incisive insights into the consequences of moving from speaking to writing, by distinguishing three types of distantiation: (1) a distantiation of meaning, because the author cannot control the meaning of their inscription, (2) a distantiation from the ostensive reference, because the reader may no longer share the same space and time as the author, moving from a shared situation to a shared world; and (3) a distantiation from the interlocutor, meaning that the author addresses a potentially unlimited audience. The triple distantiation implies that the meaning of a text cannot be defined by the intention of the author. This nicely ties in with the difference between the German *Verstehen*, in the sense of understanding a *speaker's intent*, and the German *Auslegung*, in the sense of understanding the *meaning of the text*. An oral speech act enables those addressed to check directly with the author, whether they got them right; it also enables the author to exercise some control over their audience, telling them when they misunderstand. In written speech the author may not be present or even dead, and the lapse in time or the distance in space may require that the reader makes up their own mind about the meaning of the text. Ricoeur explains that the shift from spoken to written speech therefore necessitates the emancipation of the text from the tutelage of the author, instantiating the *autonomous meaning of the text* – based on a reflective equilibrium between

the author's intent, the reader's response and the way the text has been embedded in the contemporary context of the reader. This accords with the idea of an 'autonomous law' that is not under the control of the legislature, as its meaning is ultimately decided by the court, taking into account that its interpretation must make sense to those under its jurisdiction, while also respecting the *telos* of the law in the wider context of the legal system. As Dworkin argued,[64] this is directly connected with the difference between rules and principles and the nature of discretion as bound by the *implied philosophy of the law*. The rise of an autonomous law and the *rule of law* as opposed to the rule by law by men can thus be traced to the affordances of the printing press that reinforced and extended the triple distantiation described above, which basically forced the rise of an autonomous, positive law that can no longer be controlled by the arbitrary will of an undivided sovereign.

xiii.   Speech act theory and institutional facts

Having thus clarified the difference between positive law and legal decision making, we can now explore their relationship by moving into speech act theory. Invoking Austin,[65] Ricoeur distinguishes between three different types of speech, highlighting the fact that speaking is acting: (1) *locutionary acts* that equate with propositional acts, for instance describing that you 'are getting married'; (2) *illocutionary acts* or *performative speech acts*, for instance declaring you 'husband and wife', meaning that from that moment onwards you will be qualified as 'husband and wife'; and (3) *perlocutionary acts* that are meant to exert influence, for instance urging you 'to get married'. Clearly, legal decisions have illocutionary force, because they generate 'legal effect'. They are performative speech acts (as in, 'I declare you "husband and wife"'), *legal decisions do what they say*. The importance of this simple statement can hardly be overstated. Legal decision-making does not *cause* legal effect, it *constitutes* such effect. Mistaking the illocutionary act for a perlocutionary act means mistaking the *instrumentality* of law (its ability to constitute institutional facts) for a utilitarian or behaviourist *instrumentalism* (the ability to nudge or force people into certain behaviours). *Institutional facts* are facts that result from a performative speech act and they are often opposed to brute facts.[66] A stone or a pregnancy may be considered as 'brute facts', whereas a marriage, a university, money or the rule of law are 'institutional facts' – they depend on a 'language game' that qualifies specific 'things' as such,[67] while this qualification or 'counting as' defines the institution.

The performative effect of oral or written speech acts cannot be achieved by way of computer code, even when such code is said to 'outperform' human experts. The code has different affordances, and 'doing things with words' is not one of them. This – obviously – does not mean that computer code is not capable of forcing or luring people into specific behaviours. On the contrary, many automated decision systems are very good in doing just that. They

---

[64] Dworkin, *Law's Empire*.

[65] Austin, *How to Do Things with Words*.

[66] John Searle, *The Construction of Social Reality* (New York: The Free Press, 1995). Even more interesting, building on e.g. Wittgenstein, without however employing the terminology of speech act theory Peter Winch, *The Idea of a Social Science* (London and Henley: Routledge & Kegan Paul, 1958).

[67] The concept of language games was instituted by Ludwig Wittgenstein and G. E. M. Anscombe, *Philosophical Investigations : The German Text, with a Revised English Translation*, vol. 3rd (Malden, MA,: Blackwell Pub., 2003).

remain, however, in the realm of perlocutionary acts,[68] which means that the effect they have is causal rather than constitutive, they generate a rule of machines instead of a rule by law. If data-driven regulation were to become law, we would return to the rule by machines by men, even though we should not overestimate the control that those in charge of the machines have over their own invention.

B. Structure and Action

xiv. From text to action

Ricoeur then moves into the issue of meaningful action, as a sequel to his discussion of the meaning of text. As discussed above, the meaning of text derives from the interplay between a given language system (structure) and the speech it enables (action). This means that text (as a type of speech) is understood as an action, whose meaning depends both on the language system (the structure) and on how it is used (the action). Simultaneously, the distantiation between author, text and reader requires an iterant act of interpretation on the side of the readers, who are, however, constrained by both the language system and the way others use it. In other words, though the meaning of a specific speech act requires interpretation, there are constraints that consolidate and stabilise meaning, to allow people to actually understand each other.

When speaking of action, Ricoeur notes that for the same reason, the meaning of an action requires what he calls fixation. It must be stabilised, to be recognisable as a specific type of action. For instance, the gesture of 'thumbs up' has meaning, but this meaning is not *caused* by the physical behaviour of keeping your thumb up. Just like speech acts, actions such as 'thumbs up' require a certain 'autonomisation of action', such that similar behaviour invites similar interpretation, recognition as being of a certain type, with a specific meaning that qualifies the gesture as an encouragement or approval. For *behaviour* to count as a certain type of *action* it has to fit with sedimented mutual expectations, the behaviour must be 'legible' as such an action. The implication is that human action is 'an open work',[69] its meaning is underdetermined, and just like with the interpretation of words, human action is multi-interpretable. This also means that the author of the action (the actor) cannot define the meaning of their action all by themselves. If a person puts up their middle finger, they cannot claim that it was their intention to give a 'thumbs up' and therefor others are wrong to interpret the gesture as a sign of contempt or aggression. The objective meaning of an action is something other than the subjective intention of the author, and it may be construed in various ways, depending on the habits of the interpretive community they navigate. The problems of the right understanding cannot, therefore, be solved by a simple return to the alleged intention of the actor, it must be situated in the context, the culture or the institutional environment of those affected by the act. Problems arise when the context, culture and institutional environment of those affected are not shared with the actor, who could perhaps not foresee the consequences of their behaviour.

---

[68] Note that the difference, as made by Austin in his Austin, *How to Do Things with Words*. is notoriously confusing. I have taken the liberty of developing my own interpretation, which also builds on the work of Judith Butles, as e.g. in her Judith Butler, *Giving an Account of Oneself* (New York: Fordham University Press, 2005).
[69] Ricoeur, 'The Model of the Text: Meaningful Action Considered as a Text': 103.

xv.   Anticipation and validation

This generates a dialectic of anticipation and validation, since it is always possible to relate the same gesture in different ways to, for instance, different types of gestures that play out as cornerstones of a specific context. For instance, putting up one's little finger may either be interpreted as a reference to the thumbs up, or to the middle finger. Actions must be understood based on the 'plurivocity' that is inherent in their context, culture or institutional environment, which is not the same as saying that they are 'polysemous'. They may indeed have different meanings (just like words, such as 'pupil'), but what Ricoeur is after here is the fact that the meaning of an action is interwoven with the context in which was taken, and that shifts in meaning within other parts of that context may result in a different meaning for the action itself. Ricoeur suggest that:[70]

> To show that an interpretation is more probable in the light of what is known is something other than showing that a conclusion is true. In this sense, validation is not verification. Validation is an argumentative discipline comparable to the juridical procedures of legal interpretation. It is a logic of uncertainty and of qualitative probability.

This is in turn connected with the fact that though there is often more than one way to give meaning to an action, this does not imply that all possible interpretations are equally valid. Acknowledging that author's intent does not suffice does not imply that it does not matter or that anything goes. Presenting or claiming that a certain action must be read in one way rather than another, however, opens the floor for contestation; the interpretation that is put forward will have to be defended and argued if an attempt is made to defeat it. Ricoeur reiterates his reference to the law:[71]

> Only in the tribunal is there a moment when the procedures of appeal are exhausted. But it is so only because the decision of the judge is implemented by the force of public power. Neither in literary criticism nor in the social sciences is there such a last word. Or, if there is any, we call that violence.

The force of public power is part of what lawyers call the 'positivity' of the law.[72] It is part of the monopoly of violence that makes positive law possible, including the protection it offers against the arbitrary powers of a sovereign. Despite this aspect of violence that is inherent in legal decision-making, legal effect and the force of law ultimately depend on *the performative effect of meaningful action*, which should not be confused with either physical causality or subjective motivation.

This, in turn, has many implications for the operations of data-driven 'law'. Distant reading offers a mathematical (probabilistic) compression of legal text, but the patterns do not represent causality. They map e.g. conceptual overlap and the way speech acts are coordinated, but these mathematical patterns do not engage in speech acts themselves. They

---

[70] Ricoeur, 107.
[71] Ricoeur, 110.
[72] Gustav Radbruch, 'Legal Philosphy', in *The Legal Philosophies of Lask, Radbruch and Dabin. Translated by Kurt Wilk ... Introduction by Edwin W. Patterson*, Twentieth Century Legal Philosophy Series 4 (Boston & London: Harvard University Press, 2014), 44–224.

are about something (law), but they are not that something (law).[73] The anticipation that situates law entails what Ricoeur calls a 'qualitative probability',[74] not the quantitative probability of number crunching machines.

C.   Text-driven Law
xvi.   The nature of legal effect
Modern positive law is a system or interrelated legal norms. Such norms basically attribute a specified legal consequence if specified legal conditions apply. This legal consequence is not 'caused' by the fulfilment of the conditions but is constituted by the performative effect of their 'being the case'.[75]

The normativity of positive law depends on: (1) the performative nature of written and unwritten law,[76] (2) the mutuality between primary and secondary rules,[77] (3) the state's monopoly of violence,[78] which in turn depends (4) on international law that institutes internal and external sovereignty, which depend on each other.[79] This implies that positive law is not merely a matter of brute force (monopoly of violence), and/or mechanical application, because brute force, here, depends on the performativity of international law, while mechanical application hinges on the performative nature of the written and unwritten speech acts that 'make' the law.[80]

The fact that 'mechanical application' hinges on the performative nature of (un)written speech acts can be explained by the insights gained from the previous sections. If objective meaning is something other than the subjective intentions of either the author or the reader, it may be construed in various ways. This generates a dialectic of anticipation and validation, as it is always possible to relate the same content in different ways to other content within the same context.[81] Such – fundamentally text-driven – *multi-interpretability* is what 'makes' modern positive law. It implies a type of contestability that requires argumentation, as neither logic nor brute force would do. Logic can be used to test the soundness of argumentation but cannot decide which of several sound arguments must be chosen; brute force can enforce but not legitimate. The argumentative nature of current law is an affordance of its textual

---

[73] Mireille Hildebrandt, 'Law as Information in the Era of Data-Driven Agency', *The Modern Law Review* 79, no. 1 (1 January 2016): 1–30, https://doi.org/10.1111/1468-2230.12165.

[74] Ricoeur, 107 (see above, note 70). 'The Model of the Text: Meaningful Action Considered as a Text'.David Aldous, 'Probability, Uncertainty and Unpredictability', *Probability and the Real World* (blog), n.d., https://www.stat.berkeley.edu/~aldous/Real-World/phil_uncertainty.html.

[75] Which is itself a qualification that qualifies as a speech act, *les faits sont faits*.

[76] MacCormick and Weinberger, *An Institutional Theory of Law*.

[77] H.L.A. Hart, 'Law as the Union of Primary and Secondary Rules', in *The Concept of Law*, 2nd ed. (Oxford: Clarendon Press, 1994).

[78] Harold Berman, *Law and Revolution. The Formation of the Western Legal Tradition* (Cambridge Massachusetts and London, England: Harvard University Press, 1983).

[79] Jeremy Waldron, 'THE RULE OF INTERNATIONAL LAW', *Harvard Journal of Law & Public Policy* 30, no. 1 (2006): 15–30.

[80] Bruno Latour, *The Making of Law: An Ethnography of the Conseil d'Etat*, 1 edition (Cambridge, UK ; Malden, MA: Polity, 2009).

[81] This is one of the crucial insights of Wittgenstein, who concludes that at some point a decision must be made, see xxx, Wittgenstein and Anscombe, *Philosophical Investigations: The German Text, with a Revised English Translation*.

embodiment and defines the kind of certainty it offers. As Waldron has argued,[82] legal certainty is not only about providing foresight, trust and stability but simultaneously about the need to argue for whatever legal decision is be taken, acknowledging that both the applicability and the application of the relevant legal norm can always be contested. To make this merry-go-round between contestation and consolidation work in practice, an institutionalised system of checks and balances is required. For instance, by establishing and sustaining independent courts. These courts will have room for discretion though not for arbitrary decision-making – as they need to stay within the decision-space offered by written law, precedent and – paradoxically – also by future case law.

xvii.   *Legal decision-making as performative speech acts*

To demonstrate the importance of thinking in terms of *performatives*—decisions with a performative effect—I will inquire into the Google Spain v Costeja decision of the Court of Justice of the EU (CJEU).[83] One of the questions raised was whether a search engine must be qualified as a controller in the sense of the (then applicable) Data Protection Directive,[84] which defined (just like the current GDPR) a data controller as the entity that determines the purpose and the means of the processing of personal data. The AG recounts the following:

> (84) The internet search engine service provider merely supplying an information location tool does not exercise control over personal data included on third-party web pages. The service provider is not 'aware' of the existence of personal data in any other sense than as a statistical fact web pages are likely to include personal data. In the course of processing of the source web pages for the purposes of crawling, analysing and indexing, personal data does not manifest itself as such in any particular way.

The AG finds that a search engine cannot be qualified as a controller:

> (90) An opposite opinion would entail internet search engines being incompatible with EU law, a conclusion I would find absurd.

This was not at all a controversial position. The idea that a search algorithm is a neutral tool that provides an objective listing of relevant search results, that is not 'controlled' by the search engine provider may seem naïve but has its attractions (similar to those that would qualify the predictions of legal technologies as objective outputs). The hidden nature of the algorithm probably contributes to its technocratic aura, distracting attention from the choices made when designing the search engine and the invisible trade-offs involved (in the case of PoJ algorithms the use of terms such as 'accuracy' invites a similar trust in the neutrality of the software). Though, just checking the paper published by Brin and Page in 1997,[85] would already clarify that even at that time it was apparent that commercialising the service would

---

[82] JEREMY WALDRON, 'THE RULE OF LAW AND THE IMPORTANCE OF PROCEDURE', *Nomos* 50 (2011): 3–31.
[83] CJEU, 13 May 2014, C-131/12 *('the right to be forgotten')*
[84] D 95/46/EC, now replaced by the GDPR Regulation (EU) 2016/679.
[85] Sergey Brin and Lawrence Page, 'The Anatomy of a Large-Scale Hypertextual Web Search Engine', *Computer Networks and ISDN Systems* 30, no. 1–7 (April 1998): 107–17, https://doi.org/10.1016/S0169-7552(98)00110-X. 109 and Appendix A.

'corrupt' its employment. The authors of the PageRank algorithm, who are now running the Google company, observed in 1997, when still at Stanford University:

> At the same time, search engines have migrated from the academic domain to the commercial. Up until now most search engine development has gone on at companies with little publication of technical details. This causes search engine technology to remain largely a black art and to be advertising oriented (see Appendix A in the full version). With Google, we have a strong goal to push more development and understanding into the academic realm.

And, in '8. Annex A, Advertising and Mixed Motives' they explain in salient detail why they:

> (…) expect that advertising funded search engines will be inherently biased towards the advertisers and away from the needs of the consumers.

It would be interesting to see whether developers of legal technologies have similar insights. Let's note the straightforward decision of the Court, in deciding without further ado, diametrically opposed to the Opinion of the AG, that

> (33) It is the search engine operator which determines the purposes and means of that activity and thus of the processing of personal data that it itself carries out within the framework of that activity and which must, consequently, be regarded as the 'controller' in respect of that processing pursuant to Article 2(d).

And, despite the fact that the AG had advised that:

> 104. The rights to rectification, erasure and blocking of data provided in Article 12(b) of the Directive concern data, the processing of which does not comply with the provisions of the Directive, in particular because of the incomplete or inaccurate nature of the data.

The CJEU decided that:

> 92. As regards Article 12(b) of Directive 95/46, the application of which is subject to the condition that the processing of personal data be incompatible with the directive, it should be recalled that, as has been noted in paragraph 72 of the present judgment, such incompatibility may result not only from the fact that such data are inaccurate but, in particular, also from the fact that they are inadequate, irrelevant or excessive in relation to the purposes of the processing, that they are not kept up to date, or that they are kept for longer than is necessary unless they are required to be kept for historical, statistical or scientific purposes.

> 93. It follows from those requirements, laid down in Article 6(1)(c) to (e) of Directive 95/46, that even initially lawful processing of accurate data may, in the course of time, become incompatible with the directive where those data are no longer necessary in the light of the purposes for which they were collected or processed. That is so in particular where they appear to be inadequate, irrelevant or no longer relevant, or excessive in relation to those purposes and in the light of the time that has elapsed.

> 94 Therefore, if it is found, following a request by the data subject pursuant to Article 12(b) of Directive 95/46, that the inclusion in the list of results displayed following a search made on the basis of his name of the links to web pages published lawfully by third parties and containing true information relating to him personally is, at this point in time, incompatible with Article 6(1)(c) to (e) of the directive because that information appears, having regard to all the circumstances of the case, to be inadequate, irrelevant or no longer relevant, or excessive in relation to the purposes of the processing at issue carried out by the operator of the search engine, the information and links concerned in the list of results must be erased.

The judgement constitutes three *performatives* with far reaching consquences: (1) the decision to qualify indexing as 'processing of personal data, (2) the decision to qualify a search engine as a data controller, and (3) the decision that the right to have data erased if inaccurate, inadequate, irrelevant or excessive, and the right to object to the processing in case of the f-ground, mean that the right of the data subject to be dereferenced must be in principle be respected. The court adds that in case of the f-ground such dereferencing would probably be required if: (a) the legitimate interest of the controller is an economic interest, and if (b) the public interest of access to that information does not overrule the rights of the data subject, noting that this may depend on the celebrity of the data subject. On top of that we should take into account all the decisions that the CJEU did not take, whereas the AG advised them. These implied decisions can *a contrario* also be seen as *performatives*, for instance, (1) the fact that the newspaper lawfully published the data and is allowed to continue processing them (indeed must do so) does NOT imply that indexing by the search engine is therefore lawful, and (2) the fact that the search engine is based on automated processing, without any intent to specifically process personal data, does not imply that they are not a controller.

The decisions made by the Court cannot be reduced to mechanical application (the life of the law is not logic), nor can they be conflated with Schmittian decisionism (the decision space is restricted against arbitrary choice of action). They are speech acts that decide the reconfiguration of the backend systems of commercial search engines, in the face of the kind of multi-interpretability that is inherent in text-driven normativity.

xviii.   The performance of PoJ software and its performative effect
What would the Court have decided if supposedly 'intelligent' software had estimated that Google is (or is not) to be considered a controller, with an 'accuracy' that is higher than some supposedly relevant 'baseline'. If doctors may at some point be compelled by their insurance company to motivate digressions from computational support systems, will courts be held to supplementary motivation if they digress from predictions on points of law? Should lawyers 'buy into' the discourse that human lawyers are subjective, biased and limited by their bounded rationality, whereas machine learning systems will soon outwit them due to their objective output (assuming potential bias can be 'corrected' by those versed in the relevant subfield of CS)? Must we accept that 'distant reading' of legal text will provide a grasp over far more relevant legal textual corpora than any human lawyer could ever hope to achieve?

Ricoeur's work can help to change the frame of reference from that of subjectivity (of human lawyers) versus objectivity (of machines) to that of a dynamic interaction between anticipation and validation, which equates to an iteration of anticipation, contestation, argumentation and consolidation that culminates in closure. Data-driven prediction engines can trace, mine and 'read' our anticipatory interactions in the domain of legal decision-making, but their anticipations are of another kind than our own. They are mathematical mappings, not a way to navigate their own institutional environment. They simulate our past behaviours, they scale our past, but not to face their future (let alone ours). They have no future, no past and no present. They have nothing to lose. Whereas some might think that makes 'them' more objective, in point of fact it makes them dangerously unreliable because they have no clue as to what matters.[86]

xix.   What we stand to lose
In this third part I have argued that modern positive law and the rule of law are historical artefacts, contingent upon the technologies of the printing press, sharing the affordances of their text-driven embodiment. Whereas we often hear that law is slow to catch up with new technologies, always 'regulating' after the fact, and thus incapable of dealing with disruptive innovation, this chapter takes another point of departure. After probing the lure of data-driven technologies for the law and the mathematical and statistical assumptions they build on in the second part, we conducted an in-depth assessment of the affordances of text-driven law. This provides us with a better understanding of law's current mode of existence, while resisting an essentialist take on the way law exists. The analysis should generate an acuity as to what we wish to *preserve*,[87] in terms of legal protection, legal certainty and the rule of law, noting that we cannot take for granted that data-driven 'law' will afford similar protection, similar certainty and similar checks and balances.

Moving from the assumptions of text-driven anticipation to data-driven predictions, we may be embracing (1) the idea that human interaction can be reduced to mathematical relationships, (2) formalisation and disambiguation of both facts and norms, such that they can be made machine-readable, (3) the reconfiguration of legal-decision making in terms of correlations and causalities, mapping vocal pitch, political affiliation, or 'word embeddings' as influencers to the binary outcome of a court case, (4) a distinction between content and meta-data in place of that between substance and procedure, (5) the idea that computational systems are more objective and less biased (provided the training data have been debiased) than human judges, (6) replacement of interpretation, contestation and argumentation with regard to both the facts of a case and the applicability and the application of legal norms, by an algorithm that was trained on data that is considered relevant and sufficiently complete, with some kind of explanation of why the algorithm came to its conclusion, (7) a type of

---

[86] Embodiment matters, as roboticists Pfeifer and Bongard demonstrate in their work on 'understanding by building' artificial intelligence. Full intelligence will depend on an agent being capable of navigating and surviving in their 'real world' environment. Rolf Pfeifer and Josh Bongard, *How the Body Shapes the Way We Think. A New View of Intelligence* (Cambridge, MA - London, England: MIT Press, 2007).

[87] In Maryanne Wolf, *Proust and the Squid: The Story and Science of the Reading Brain* (Icon Books Ltd, 2008). Maryanne Wolf highlights what she believes we should preserve as key affordances of the 'bookish' mind, based on extensive research into the neuroscience of the human brain.

'distant reading' that in point of fact collapses the distantiation between author, text, meaning and reader that was instantiated by the shift from oral to written speech, and thereby we might thus give up (8) the core tenets of the rule of law, notably practical and effective remedies to contest claims of validity regarding both legal norms and legally relevant facts. We may lose the *mode of existence of law* where legal protection is *part of* law's instrumentality rather than an add-on; because this instrumentality is built on legal effect rather than merely brute force, and because legal certainty thrives on the contestability of facts and norms, not on the mechanical application of logical rules. We may even lose the kind of human agency that is constituted by text-driven law, because the radical ambiguity that is inherent in the open texture of legal concepts and legal norms affords a kind of discretion that has to be removed in the process of formalisation that is inherent in data-driven predictions.

The final question is whether this is inevitable. Do I confess to technological determinism in summing up all that we stand to lose when shifting from text- to data-driven law?

4.   A new path for the law: legal protection by design?

Technological determinism is a trap. It reduces our acuity in relation to the affordances of specific types of technology instead of enhancing it.[88] What we should face, however, is the possibility that due to the assumptions and trade-offs inherent in their design, specific data-driven technologies determine our choice architecture in ways that diminish our agency. This is not merely a technological question, but involves serious consideration of the political economy that informs the design, the operations and the uptake of a technology. If the market is organised such that companies compete by way of algorithmic optimisation of advertising revenue in ways that prioritise extreme content and fake news, it may be very difficult to solve that problem at the level of the technology. The relevant technological solutions will easily require various types of algorithmic censure and moderation, which has radical drawbacks for the freedom of information. Instead, the business model that invites such optimisation should be eradicated, noting that a market that invites this kind of behaviour also invites unhealthy concentration of economic power, requiring a much-needed rethinking of competition law.[89] In short, whether a technology overdetermines individual behaviour and societal interaction is an empirical question that must be addressed by doing more rather than less research. Both the idea that technology is an autonomous force that will necessarily diminish our agency and the idea that technology is in the end a force for good, are naïve and lazy shortcuts to unsubstantiated conclusions.

In other work I have advocated the idea of legal protection by design (LPbD) as way of integrating legal norms at the level of fundamental rights law in the research design and the

---

[88] A detailed analysis of the notion of technological determinism in relation to the law can be found in chapter 8 Hildebrandt, *Smart Technologies and the End(s) of Law. Novel Entanglements of Law and Technology*.
[89] See e.g. Lina Khan, 'The Ideological Roots of America's Market Power Problem', *Yale Law Journal* 127 (2018), https://www.yalelawjournal.org/forum/the-ideological-roots-of-americas-market-power-problem; Lina Khan, 'The New Brandeis Movement: America's Antimonopoly Debate', *Journal of European Competition Law & Practice* 9, no. 3 (1 March 2018): 131–32, https://doi.org/10.1093/jeclap/lpy020.Mireille Hildebrandt, 'Primitives of Legal Protection in the Era of Data-Driven Platforms', *Georgetown Law Technology Review* 2, no. 2 (2018): 252–73.

architecture of data-driven systems (and environments). For instance, in the Google Spain v Costeja case the CJEU required that a search engine provider implements a specified legal right into the design of its system. It is not just that the search engine had to delist a particular search result. Rather, the search engine's backend system had to be reconfigured in such a way that requests similar to that of Mr Costeja could be honoured. Art. 7.3 of the GDPR requires that consent for the processing of personal data (for a specific purpose) can be withdrawn as easily as it has been provided. This is not merely a matter of adding a button to a website, but a matter or reengineering the backend system (that is not visible to end-users but crucial for putting an end to the processing). Under art. 25 GDPR such design choices become part of a general legal obligation for data controllers to ensure that rights of data subjects and legal obligations for controllers are by default and by design integrated into systems that process personal data. This is called data protection by design, and could be termed a first instance of legal protection by design. Interestingly, the obligation is articulated as an obligation to 'implement appropriate technical and organisational measures, such as pseudonymisation, which are designed to implement data-protection principles, such as data minimisation, in an effective manner and to integrate the necessary safeguards into the processing in order to meet the requirements of this Regulation and protect the rights of data subject'. The formulation demonstrates that data protection by design is not about 'compliance by design', which is not feasible but also not desirable. It is about a duty for controllers to design their systems in ways that minimise infringements of fundamental rights, while integrating the checks and balances that are inherent in the rule of law. Similarly, LPbD is not about the 'automation of compliance'. It is not about creating an online hyperconnected environment that limits our choice architecture to what is lawful. This would be the inverse of LPbD and closer to 'legal by design' or techno-regulation. It would come close to a technology that actually determines us, not because technology necessarily does so but because a particular technology has been designed to do so.

PoJ could be used in a way that aligns with 'legal by design' strategies. For instance, public administration could reject the use of legal remedies if their PoJ software predicts that the complainant will lose anyway. Courts with a huge backlog might start processing cases via PoJ software and resort to abbreviated procedures to speed up decision making. Insurance companies may use PoJ software to deny claims and force their clients to sign a contract where they waive their right if the software predicts that they will lose their case. Some legal scholars have argued that this would solve many problems and improve legal decision-making.[90] I hope that this chapter has convinced the reader this would require a radically different understanding of what counts as law.

LPbD, however, has relevance for PoJ and other data-driven technologies in the legal domain. First of all, the term 'legal' in LPbD does not refer to 'regulation' in the sense of attempts to influence human behaviour, but to law and the rule of law. More specifically, it

---

[90] Michael Genesereth, 'Computational Law. The Cop in the Backseat', White Paper (CodeX: The Center for Legal Informatics Stanford University, 2015), http://logic.stanford.edu/complaw/complaw.html; Paul Lippe, Daniel Martin Katz, and Dan Jackson, 'Legal by Design: A New Paradigm for Handling Complexity in Banking Regulation and Elsewhere in Law', *Oregon Law Review* 93, no. 4 (2015), http://papers.ssrn.com/abstract=2539315.

refers to safeguarding effective and practical human rights, and instituting checks and balances that provide countervailing powers against big players (including the state). This would mean, first of all, that PoJ software requires not only the highest standards in terms of the underlying research design, but also its testability and contestability, based on a proper understanding of the requirements of confirmatory research design in machine learning (including its preregistration in e.g. the Open Science Foundation).[91] Second, it would limit and restrict the use of PoJ to decision support, and – for instance – require that those making a decision based on the software (1) understand how it came to its conclusion, and (2) has the legal power to make a decision that overrules the prediction,[92] without an additional burden of argumentation. Third, LPbD would require that the curriculum of law schools is extended to integrate proper training in the data fallacies that hamper both the research design and the outcome of PoJ, while also allowing students to play around with the software until they understand the underlying assumptions and their implications for the claims made in terms of e.g. accuracy. Clearly, these types of requirements rule out the use of proprietary software, let alone the use of non-disclosure agreements (NDAs), when contracting with providers of PoJ software. This, in turn, means that lawyers must become aware of the accuracy – reliability trade-off, as discussed above: high accuracy that correlates with diminished or absent interpretability implies problems with reliability. Precisely because we do not know why the system concludes as it does, we cannot be sure its conclusions are valid. That is why we should not buy into the narrative that proprietary software may be more opaque but will nevertheless be more accurate. The latter is an unsubstantiated claim, with high risk consequences.

Lawyers should get their act together before technology developers – whether or not with the best of intentions – take over and transform law's mode of existence. Technology developers, in the meantime, should help lawyers understand the limits as well as the potential of PoJ.

---

[91] Jake M. Hofman, Amit Sharma, and Duncan J. Watts, 'Prediction and Explanation in Social Systems', *Science* 355, no. 6324 (3 February 2017): 486–88, https://doi.org/10.1126/science.aal3856.

[92] Cp. the EDPB (formerly Art. 29 WP) *Guidelines on Automated individual decision-making and Profiling for the purposes of Regulation 2016/679*, WP251, 3 October 2017, at 10 on the question of whether human intervention 'counts' as such: 'The controller cannot avoid the Article 22 provisions by fabricating human involvement. For example, if someone routinely applies automatically generated profiles to individuals without any actual influence on the result, this would still be a decision based solely on automated processing. To qualify as human intervention, the controller must ensure that any oversight of the decision is meaningful, rather than just a token gesture. It should be carried out by someone who has the authority and competence to change the decision. As part of the analysis, they should consider all the available input and output data.'