Contents

Mission Statement
The *Erasmus Law Review* seeks to foster independent critical scholarship as relevant to the discipline of law.
The Board of Editors encourages the submission of legally relevant manuscripts by legal scholars and practitioners as well as those versed in other disciplines relevant to law, such as criminology, sociology, political science and economics.
This Open Acces journal usually commissions articles around specific themes, although 'calls for papers' on specific topics might be issued occasionally and will be published on the Review's website. All prospective articles are submitted to double-blind peer review (two reviews per article), and final publication is dependent on the outcome of these reviews.

ISSN: 2210-2671

Nederlands uitgeversverbond
Groep uitgevers voor vak en wetenschap

Computational Methods for Legal Analysis

The Way Forward?

Elena Kantorowicz-Reznichenko*

Abstract

Computational analysis can be seen as the most recent innovation in the field of Empirical Legal Studies (ELS). It concerns the use of computer science and big data tools to collect, analyse and understand the large and unstructured data, such as for instance (legal) text. Given that the text is now the object of analysis, but the methods are (largely) quantitative, it lies in the intersection between doctrinal analysis and ELS. It brings with it not only a great potential to scale up research and answer old research questions, but also to reveal uncovered patterns and address new questions. Despite a slowly growing number of legal scholars who are already applying such methods, it is underutilised in the field of law. Furthermore, given that this method comes from social and computer sciences, many legal scholars are not even aware of its existence and potential. Therefore, the purpose of this special issue is not only to introduce these methods to lawyers and discuss possibilities of their application, but also to pay special attention to the challenges, with a specific emphasis on the ethical issues arising from using 'big data' and the challenge of building capacity to use such methods in law schools. This editorial briefly explains some of the methods which belong to the new movement of Computational Legal Analysis and provides examples of their application. It then introduces those articles included in this special issue. Finally, it provides a personal note on the way forward for lawyers within the movement of Computational Legal Analysis

Keywords: computational legal analysis, empirical legal studies, natural language processing, machine learning

1 Introduction

Traditional doctrinal analysis is the backbone of legal scholars. It entails reading legal documents such as court decisions and legislations, and then offering descriptive, critical, normative or predictive claims about different legal rules or decisions.[1] Empirical Legal Studies (ELS), and especially quantitative ELS, use statistical tools to analyse legal phenomena/issues. This analysis does not need to be performed on legal text. For example, legal questions can be answered by conducting experiments in the lab, online or in the field where a legal situation is simulated, and the actions or choices of actors are recorded,[2] or it can involve vignette studies of legal practitioners, for instance, examining whether cognitive biases affect judges' decisions.[3] ELS also involve the analysis of observational data, using different natural experiment research designs to be able to infer causality, for example, measuring whether an increase in the numbers of police officers enhances deterrence of crimes.[4]

In recent decades, unprecedented technological advances in artificial intelligence (AI) tools and an ever-increasing digitalisation of legal documents can be witnessed. This combination has led to the emergence of a new research stream in law – *Computational Legal Analysis (CLA)*. This type of research lies at the intersection of doctrinal analysis and quantitative ELS. The focus of CLA is the legal text, the same as traditional doctrinal analysis, but it uses computer science and statistical tools to collect, analyse and understand the texts, which are now treated as empirical data.[5] Some of these methods allow the scaling up of the research that has been conducted until now by lawyers using, for example, manual coding of texts. Other methods, as discussed in the next section, make it possible to answer new questions and uncover patterns in legal documents or links between them, which are not easily detectable through hand-coded analysis.

* Elena Kantorowicz-Reznichenko is Professor of Quantitative Empirical Legal Studies at the Rotterdam Institute of Law and Economics, Erasmus School of Law, Erasmus University, Rotterdam. I would like to thank Nina Holvast, Jaroslaw Kantorowicz and Pim Jansen for their useful comments, and Vera Brijer for her editorial assistance.

1. T. Hutchinson, 'The Doctrinal Method: Incorporating Interdisciplinary Methods in Reforming the Law', 8 *Erasmus Law Review* 130 (2015).
2. See for example, C. Engel and M. Kurschilgen, 'Fairness Ex Ante and Ex Post: Experimentally Testing Ex Post Judicial Intervention into Blockbuster Deals', 8 *Journal of Empirical Legal Studies* 682 (2011); E. Kantorowicz-Reznichenko, J. Kantorowicz & K. Weinshall, 'Can We Overcome Ideological Biases in Constitutional Judgments? An Experimental Analysis', *DIIS Working Paper*, https://www.researchgate.net/publication/353333715_Can_We_Overcome_Ideological_Biases_in_Constitutional_Judgments_An_Experimental_Analysis.
3. C. Guthrie, J.J. Rachlinski & A.J. Wistrich, 'Inside the Judicial Mind', 86 *Cornell Law Review* 777-830, 788-791 (2000); C. Guthrie, J.J. Rachlinski & A.J. Wistrich, 'The 'Hidden Judiciary': An Empirical Examination of Executive Branch Justice', 58 *Duke Law Journal*1477, at 1502-4 (2009).
4. R. Di Tella and E. Schargrodsky, 'Do Police Reduce Crime? Estimates Using the Allocation of Police Forces after a Terrorist Attack', 94 *American Economic Review* 115-33 (2004); J. Klick and A. Tabarrok, 'Using Terror Alert Levels to Estimate the Effect of Police on Crime', 48 *Journal of Law & Economics* 267-79 (2005). For ELS in Europe and also more information on qualitative methods used in ELS, see the special issue by P. Mascini and W. van Rossum, 'Empirical Legal Research: Fad, Feud or Fellowship?', *Erasmus Law Review* 2 (2018).
5. M.A. Livermore and D.N. Rockmore, 'Law as Data: Computation, Text, & the Future of Legal Analysis', *Santa Fe Institute Press* xvii (2009).

The purpose of this special issue is to raise the awareness of lawyers and legal scholars to the existence of these new methods. It is meant to present the promises of CLA and its potential uses, its challenges, including ethical ones, and some thoughts on the training of law students and legal scholars in these methods. The starting point is that ELS in general is an important field, and my arguments to support this point can be found in my previous work.[6] The goal of this special issue is to focus on CLA, which I consider as an important innovation and addition to ELS. In this editorial, I first briefly explain which type of research and practical application is possible using computational methods. Next, I introduce the articles in this special issue and explain how they connect to the story of CLA. I end with a personal note on the future of CLA and legal education.

2 What Are CLA Methods, and Which Type of Research Can They Support?

CLA concerns the application of computing capabilities to law as a subject matter.[7] It includes different techniques such as network analysis, machine learning and natural language processing (NLP).[8] Utilising computational power allows researchers to investigate simpler things such as the features of the legal text (for example, length of judicial opinions, linguistic sophistication of judicial opinions) or the prominence of certain decisions by creating a citation network and exploring which cases are the most cited. However, more complicated research questions can also be answered using CLA methods.[9] Computational methods can also be used to simply investigate the evolution of a research field itself[10] or to synthesise fields of research.[11] Here I mention several studies as examples, without entering into the details, to briefly provide a sense of the potential of these methods. A more thorough review of some of these methods – in the context of quantitative text analysis – can then be found in the first article in this special issue by Arthur Dyevre.

One interesting stream of research utilises plagiarism-detection software, or related techniques, which checks for unusual similarities between texts, to investigate the influence of other (legal) texts or people on judicial decisions. For example, different studies use such techniques to identify whether judges write their decisions themselves or whether, more often, they are assisted by clerks;[12] which countries have a stronger influence on WTO decisions;[13] whether texts from international treaties are being copy-pasted into new international agreements;[14] or to what extent Supreme Court judges are influenced by and use the arguments of lower courts.[15] These techniques are useful for lawyers and legal scholars interested in the impact that different actors have on judicial and legislative decisions. A sensational example of the relevance of such an analysis to law can be found in the contested arbitration award imposed on the Russian Federation in the case *Yukos* v *Russia*.[16] Appealing the award in Dutch courts, Russia bought an expert report in which syntactic analysis was used to compare the award decision against the previous writings of the arbitrators on the one hand, and the Tribunal's assistant on the other hand. The linguistic analysis demonstrated that substantial parts of the Tribunal's decision were written by the assistant rather than the arbitrators.[17] Such methods can also be used to examine which texts and groups influenced legislation, for example, whether and how reports of lobby groups are incorporated into the draft legislation.[18]

Another interesting application of computational methods in law is the citation network analysis. It assists lawyers, among others, in identifying landmark cases, detecting trends of precedents' importance over time and exploring potentially overlooked precedents.[19] For example, one study was seeking to identify the most

6. E. Kantorowicz-Reznichenko, 'Lawyer 2.0! Some Thoughts on the Future of Empirical Legal Studies in Europe' in R. van den Bergh, M. Faure, W. Schreuders & L. Visscher (eds.), *Don't Take it Serious: Essays in Law and Economics in Honour of Intersentia* (2018).
7. See, for example, the definition of Computational Legal Studies in R. Whalen, 'The Emergence of Computational Legal Studies: An Introduction', in R. Whalen (ed.) *Computational Legal Studies: The Promise and Challenge of Data-Driven Research*, Edward Elgar Publishing Limited (2020) 1-8 at 2.
8. J. Frankenreiter and M.A. Livermore, 'Computational Methods in Legal Analysis', 16 *Annual Review of Law and Social Science* 39, at 21.2 (2020).
9. *Ibid*., at 21.5
10. E. Kantorowicz-Reznichenko and J. Kantorowicz, 'Law & Economics at Sixty – Mapping the Field with Bibliometric and Machine Learning Tools', *DIIS Working Paper* 2020.
11. S. Kuipers, J. Kantorowicz & J. Mostert, 'Manual or Machine? A Review of the Crisis and Disaster Literature',10 *Risk, Hazards & Crisis in Public Policy* 4, 388-402 (2019) doi: https://doi.org/10.1002/rhc3.12181.
12. S.J. Choi and G.M. Gulati, 'Which Judges Write Their Opinions (and Should We Care)', 32 *Florida State University Law Review* 1077 (2004).
13. M. Daku and K.J. Pelc, 'Who Holds Influence Over WTO Jurisprudence?', 20 *Journal of International Economic Law* 233-55 (2017).
14. T. Allee and M. Elsig, 'Are the Contents of International Treaties Copied and Pasted? Evidence from Preferential Trade Agreements', 63 *International Studies Quarterly* 603-13 (2019).
15. P.C. Corley, P.M. Collins Jr & B. Calvin, 'Lower Court Influence on US Supreme Court Opinion Content', 73 *The Journal of Politics* 31-44 (2011).
16. PCA Case No. AA227, *Yukos Universal Limited (Isle of Man)* v. *Russia*.
17. J. Hepbur, 'Battling $50 Billion Yukos Awards On Two Fronts, Russia Focuses On Claimants' Alleged Fraud And Linguistic Analysis Of Tribunal Assistant's Alleged Role In Drafting Awards, Investment Arbitration Reporter', (2015), www.iareporter.com/articles/battling-50-billion-yukos-awards-on-two-fronts-russia-focuses-on-claimants-alleged-fraud-and-linguistic-analysis-of-tribunal-assistants-alleged-role-in-drafting-awards/ (last visited 7 July 2021).
18. M. Burgess, E. Giraudy, J. Katz-Samuels, J. Walsh, D. Willis, L. Haynes & R. Ghani, 'The Legislative Influence Detector: Finding Text Reuse in State Legislation', *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* 57-66 (2016).
19. D. van Kuppevelt, G. van Dijck & M. Schaper, 'Purposes and Challenges of Legal Citation Network Analysis on Case Law', in R. Whalen (ed.) *Computational Legal Studies: The Promise and Challenge of Data-Driven Research*, Edward Elgar Publishing Limited (2020) 265.

important cases of the Court of Justice of the European Union (CJEU). By 'importance', the authors meant those cases that are the most cited and applied to resolve disputes at hand. To achieve this, the authors used a 'network analysis', where they included all the cases of the CJEU since the beginning of its functioning. The value of this analysis was that they have shown that the cases that are normally assumed to be the most important jurisprudence of the CJEU are not actually the most important in terms of constituting precedents which are applied frequently in later cases.[20] A similar exercise was conducted in respect of other courts. For example, another study investigated the impact of different cases of the European Court of Human Rights (ECtHR) in terms of frequency of citations. The underlying assumption was that these are the cases that can be considered as precedents for general principles of law.[21]

The final application of computational methods for lawyers which I would like to present here are predictive models. It is now possible, and this is increasingly being done, to train a machine learning model, an algorithm, to predict the outcome of specific cases. To explain the method in a very simple way, a dataset, which, for example, consists of court decisions, is randomly divided into a training set and a testing set. The first subset of the data is used to train the algorithm to recognise certain patterns. Once the model is built, it is tested on the other subset of the data to see whether it can predict the outcome in those documents. When the model is sufficiently accurate, it can be used to predict the outcomes of cases that are yet to be resolved (out-of-sample data set).[22] For example, in one study, political scientists who used a statistical model were compared to 83 legal experts in their ability to predict the outcome of upcoming American Supreme Court cases accurately. The outcome was that the statistical model had an accuracy rate of 75% of the cases whereas the legal experts were right 59% of the time.[23] Attempts to predict decisions of courts were also made in Europe, for example, the decisions of the ECtHR.[24]

From the description above, it is clear that predictive models are not only useful for researchers, but also, and maybe even more so, for legal practitioners. Some examples are discussed in detail by Simon Vydra and co-authors in the second article in this issue. Here I would like to briefly present one example. The criminal justice system often involves predictions. For example, the police need to predict in which areas to concentrate their efforts, judges need to predict risk levels of offenders when deciding on the type of arrest (home or jail) and parole decisions heavily depend on the expected level of risk of the convicted offender. Therefore, machine learning–assisted prediction has been found useful in this field.[25] For example, a machine learning model was used in bail decisions in the belief that it can reduce the rate of imprisonment without increasing the risk of crime.[26]

After briefly presenting CLA's potential, I would like to point out one important limitation.[27] Especially with predictive models, one could be tempted to interpret prediction results as offering causal links. With traditional empirical methods, we are usually interested in investigating how A causes B. For example, we seek to understand how the new directive on digital copyright will affect creators' incentives when contracting their copyrights, or how changing employees' protection rules will affect the flexibility of the labour market and the behaviour of employers. Computational methods, on the other hand, are very effective in providing prediction of outcomes. As I have shown in the previous section, by utilising large data on past behaviour we can predict future behaviour. However, this does not enable us to immediately refer to causal links between different factors. In other words, the fact that, for example, some combination of words or other factors predicts a certain outcome does not mean that these factors *cause* this outcome.[28] One illustrative example can be found in the above-mentioned study on ECtHR decisions. Among the words/combinations of words that predicted a decision by the court that a violation took place was the date October 2007.[29] Clearly, there is no causal link between the date and the question of whether there was violation or not. Because machine learning techniques are not normally suitable for inference questions, to fully utilise their potential, researchers should identify those questions that can be meaningfully answered by prediction and classification.[30] Having said that, one should note

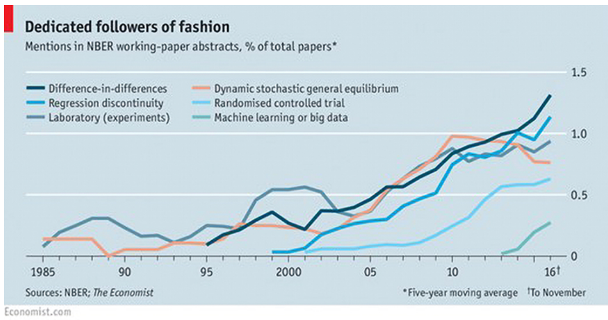
20. M. Derlén and J. Lindholm, 'Goodbye van Gend en Loos, Hello Bosman? Using Network Analysis to Measure the Importance of Individual CJEU Judgments', 20 *European Law Journal* 667-87 (2014).
21. H.P. Olsen and M. Esmark, 'Needles in a Haystack: Using Network Analysis to Identify Cases that are Cited for General Principles of Law by the European Court of Human Rights', in R. Whalen (ed.) *Computational Legal Studies: The Promise and Challenge of Data-Driven Research*, Edward Elgar Publishing Limited (2020) 293-311.
22. R. Copus, R. Hübert & H. Laqueur, 'Big Data, Machine Learning, and the Credibility Revolution in Empirical Legal Studies', in Michael A. Livermore and Daniel N. Rockmore (eds.) *Law as Data: Computation, Text & the Future of Legal Analysis*, Santa Fe: Santa Fe Institute of Science (2019) 21-37, at 31-3.
23. A.D. Martin, K.M. Quinn, T.W. Ruger & P.T. Kim, 'Competing Approaches to Predicting Supreme Court Decision Making', 2 *Perspectives on Politics* 4, 761-767 (2004).
24. M. Medvedeva, M. Vols & M. Wieling, 'Using Machine Learning to Predict Decisions of the European Court of Human Rights', 28 *Artificial Intelligence and Law* 237-66 (2020). (The prediction is within their data set, and not necessarily already advanced enough to forecast a future decision, even though that is the eventual goal with such methods.)
25. Copus et al., above n. 22, at 31-3.
26. J. Kleinberg, H. Lakkaraju, J. Leskovec, J. Ludwig & S. Mullainathan, 'Human Decisions and Machine Predictions', 133 *The Quarterly Journal of Economics* 237-93 (2018).
27. This is not to say there are no other limitations, and some will be covered in the second article in this special issue, which discusses the ethical side of using computational methods.
28. M. Dumas and J. Frankenreiter, 'Text as Observational Data', Michael A. Livermore and Daniel N. Rockmore (eds.) *Law as Data: Computation, Text & the Future of Legal Analysis*, Santa Fe: Santa Fe Institute of Science (2019) 59-70, at 59-65.
29. Presentation of the paper M. Vols, 'Using machine learning to predict decisions of the European Court of Human Rights', ATLAS AGORA Summer School, Erasmus School of Law, Erasmus University Rotterdam (25 June 2021).
30. Dumas and Frankenreiter, above n. 28, at 63.

that the stream of causal machine learning is gaining more and more traction in academic literature.[31]
Predictions are easier to do than causal inference. And sometimes we would like to predict something rather than insisting on understanding the causes (for instance, which type of offenders are expected to reoffend, without identifying the exact causes, allows us to focus resources on such offenders).[32] Furthermore, given the complexity of causal inference, it might be suggested to first use a prediction model in order to establish that a certain variable can predict changes in the other, and only then turn one's attention into the work of causal inference.[33]

3 What Is This Special Issue About?

Computational analysis as such is not new. It has already been implemented in other fields like social sciences and digital humanities. See, for example, Figure 1 for the increase in the use of machine learning in economics research.

Figure 1 The Use of Different Quantitative Methods by Economists



Source: Economists Are Prone to Fads, And the Latest Is Machine Learning: Big Data Have Led to the Latest Craze in Economic Research, *The Economist*, 24 November 2016.

However, it is a relatively new movement for the legal world and seems for now to be overlooked by legal scholars, especially in Europe. Despite a slowly growing group of legal computational scholars, the number of legal scholars who implement those methods in their research, or are even aware of their existence, is still limited. Therefore, the main purpose of this special issue is to simply open up the discussion, and to raise awareness of CLA among lawyers in Europe and to invoke their curiosity. Most lawyers and legal scholars are not skilled in computational methods. However, they are the experts in law. They are in the best position to understand the legal texts and to identify the interesting questions that can be answered using those texts. Therefore, CLA can benefit greatly from turning lawyers into users of computational methods. How to put it into practice is discussed in the next section.
The first article in this special issue – *Text-mining for Lawyers: How Machine Learning Techniques Can Advance our Understanding of Legal Discourse* by Arthur Dyevre – reviews in a non-technical manner some of the available computational methods (with the emphasis on the text-as-data approaches) and how they can be used in legal research. Even though this editorial also briefly reviewed some methods, this article enters into more detail about how those methods are applied and provides examples of the conclusions that can be derived from such research.
Computational methods require access to large amounts of data and the algorithms used to reach different conclusions often lack explainability and transparency (i.e., how the algorithm reached its outcome). This is of particular concern when computational methods are used not only in research but also in practice, for example, by decision makers in the criminal justice system and by financial institutions. Therefore, the second article in this special issue – *Big Data Ethics: A Life Cycle Perspective* by Simon Vydra, Andrei Poama, Sarah Giest, Alex Ingrams and Bram Klievink focuses on the ethical issues in using this data. In particular, it looks at the cycle of using big data, and which particular concerns are raised at each stage. For example, one of the common concerns with using predictive models to assess offenders' risk of recidivism is the inherent bias in past data (e.g., if one of the factors used by the algorithm for prediction of an increased risk is the ethnicity of the offender). This article is an important aid for remaining mindful of the concerns CLA brings with it, and for seeking to address those concerns.
The final article – *Teaching Technology to (Future) Lawyers* by Mikołaj Barczentewicz addresses the 'elephant in the room': how are lawyers and legal scholars supposed to be able to apply these methods? The traditional law school curriculum, especially in Europe, does not provide for training in statistical or computational methods. This in principle limits the ability of legal scholars as well as practitioners to use these methods themselves. This article therefore discusses different models of training that can be introduced in order to enable such usage, at least to some extent.
These articles together attempt to present a more complete picture of CLA. Besides stressing the promises of these methods for legal research and practice, this special issue does not shy away from the challenges. However, stressing these challenges by no means suggests that promoting CLA among lawyers is a lost cause. It simply sheds light on the aspects that need to be addressed in order to better achieve such a goal. In the next section, I provide some thoughts on how, in my opinion, some of those challenges should be addressed and how CLA can be promoted among lawyers.

31. See for example, H. Farbmacher, M. Huber, L. Lafférs, H. Langen & M. Spindler, 'Causal Mediation Analysis with Double Machine Learning', *arXiv preprint arXiv:2002.12710*(2020). https://arxiv.org/abs/2002.12710.
32. Copus et al., above n. 22, at 50.
33. Dumas and Frankenreiter, above n. 28, at 63.

4 A Personal Note on the Way Forward

Here I would like to focus on the particular question of capacity: can lawyers fully utilise these tools given their limited expertise with such methods? Such discussion should differentiate between the current legal scholars, who have not been trained in these methods, and the future generations of lawyers and legal scholars.

Given lawyers' specialised knowledge of law and legal institutions, they can greatly benefit from, and also contribute to, the development of the field of CLA. They are in the best position to identify the relevant questions, provide the initial annotations and coding for training the models later, etc. However, most of the current generation of legal scholars, and especially the more senior ones, are not well equipped not only to use computational methods themselves, but also to understand them in a way that will allow them to come up with good research designs. Despite the small but growing group of legal scholars who are being trained in CLA, it cannot be expected that all legal scholars will become experts in computational methods, just as it was not reasonable to assume with the traditional empirical methods. Not only are the initial investment costs very high, but computational methods are also rapidly evolving and require constant expansion of expertise. It seems more reasonable to follow the logic of comparative advantages and solve the capacity 'problem' through collaboration rather than trying to capture everything. This new development (CLA) should be viewed as a great opportunity for legal scholars and methodology experts (who can come from different fields such as social sciences, digital humanities and computer sciences) to join forces. Therefore, the way forward for the current group of legal scholars is to facilitate such collaborative projects.

How can this be done? Even though one's own expertise in computational methods is not required for legal scholars in such collaborative projects, a basic understanding of the tools is necessary. As has been discussed in this editorial piece, there is a plethora of ideas that can be explored using these tools. However, in order to identify the research questions that are suitable for these techniques, the lawyers need to understand the possibilities and the limitations of the techniques. Law faculties could offer training programmes to lawyers to just understand the logic and the intuition behind each of the available methods. The participants in such programmes will not be required to develop any programming skills themselves or understand statistical models. The focus will be on a non-technical training, in which different examples can be used to demonstrate the nature of these methods. Once legal scholars understand what these methods are about, they can come up with their ideas derived from their deep understanding of the substantive legal fields. At this point, legal scholars can start collaborating with the methodology experts. The training programmes will enable legal scholars to communicate with, for example, computer scientists using the proper jargon, thus allowing an easy and well-informed discussion. The methodology experts can then help by developing the necessary models and tools to execute the research. They can also comment on the potential limitations and add to the design of the research.

The expansion of the domain experts (legal scholars and lawyers) involved in this context can advance not only the research itself, but also the development of tools especially adjusted to the legal field. For example, methodology experts and engineers can develop NLP tools adjusted to law in different languages (given that national systems around the world have their unique legal languages). Furthermore, pulling together the routine efforts of lawyers while utilising computation power can save tremendous time and overlapping efforts. For example, annotating and coding case law in the course of legal analysis is a routine labour-intensive task made by many law students, lawyers and legal scholars. If such efforts are pooled together, machine learning models can be trained to make such annotations at scale. This will avoid duplication of effort; it will allow building a golden standard of annotation, which will serve not only research but also education. Better annotation software can then be developed, allowing for further annotation of similar legal documents. Finally, such an exercise can also benefit scholars from other fields who treat legal text as an additional object of research but lack the domain expertise to complete all the work themselves.

The proposal to develop such a training programme does not suggest that there will not be existing legal scholars who will choose to obtain that expertise themselves even in their advanced stage of legal career. Such legal scholars already exist, and this is a most welcome practice since they can enjoy both worlds. But if the entire field is to be built only on this small group, many opportunities for further developments will be missed. Therefore, I would suggest that investment should also be made to promote the understanding of the methods among the larger group of legal scholars who might resist (for obvious reasons) full scale (re)training.

The second group on which I would like to comment is the future generations of legal scholars/lawyers, who enter the law schools. An interesting discussion of, and suggestions for, the different opportunities and pitfalls of combining legal and technological education are put forward by Mikołaj Barczentewicz in the last article of this special issue. Therefore, here I only offer some general thoughts.

The increasing importance of technological literacy in general, and the future promise of CLA in particular, should render reforms in the law schools' curricula as an important and necessary step. However, the change needs to be made in such a way that complements the traditional methods of analysis rather than trying to replace them. Such an approach will take into account not only the interests of the students (it is doubtful that all incoming legal students will be interested in also acquiring quantitative skills), but also of the labour market. It is not a new insight that law schools' curricula are heavily driven by market demands.[34] Even though there are already advances in legal services that utilise big data and information technologies, currently the legal labour market needs people who are trained in analysing and applying legal rules and cases. These skills are adequately provided by doctrinal education in law schools.[35] However, given the capacity of computational methods to complement legal analysis and make legal as well as scholarly work more efficient, soon enough acquiring such skills will provide an advantage to the graduating students. Already nowadays, top law firms employ lawyers with specific technology expertise. Moreover, in the future, technological literacy might even become indispensable for the legal field, but this is yet to be seen. An increased usage of legal text as data might also promote more extensive digitisation of legal documents by the respective authorities.[36]

In order not to lag behind, law schools can introduce parallel tracks (as some law schools already do, see Barczentewicz on this issue).[37] Alongside the standard law programme, an honours programme can be offered. In the latter, the students will receive in addition to the standard training in law, training in empirical methods (e.g., statistics, econometrics) as well as in computational methods (e.g., programming). Students can then choose for themselves whether to follow the technical training, thus, assuring self-selection of motivated and capable students. In the next step, a research master's in law with a focus on empirical legal studies in general and CLA in particular can be offered. Such a programme will build the human capital necessary to further develop the field. Furthermore, it will build in-house capacity in law schools, which will enable development and promotion of educational programmes without using external methodology experts. In the future, ELS in general and CLA in particular will be able to take a more prominent place in legal education if the legal labour market (practice and academia) can utilise the new incoming human capital.

34. A. Dyevre, 'Fixing Europe's Law Schools', 25(1) *European Review of Private Law* 151-168 (2017); D. Hazel, M. Partington & S. Wheeler, *Law in the Real World: Improving Our Understanding of How Law Really Works, Final Report and Recommendations*(2006), at 29; R. Cooter, 'Maturing into Normal Science: The Effect of Empirical Legal Studies on Law and Economics', 2011(5) *University of Illinois Law Review* 1475-84.
35. Hazel et al., above n. 34, at 29.
36. For instance, currently, Dutch courts publish only a small portion of their decisions. An increased number of people applying CLA might create a demand-driven supply for legal texts. This can increase transparency and allow for important research to be conducted. Of course, a risk exists that precisely this will lead to the opposite reaction. In France, for example, the Government banned the publication of statistical information about judges' decisions. See www.artificiallawyer.com/2019/06/04/france-bans-judge-analytics-5-years-in-prison-for-rule-breakers/ (last visited 8 August 2021). This decision followed the discontent of judges from NLP and machine learning companies who used public data to analyse patterns of specific judges' decisions.
37. For additional reading on different ideas how quantitative methodology can be introduced in legal educations, see also D. M. Katz, 'The MIT School of Law? A Perspective on Legal Education in the 21st Century', 2014(5) *University of Illinois Law Review* 101-42; A. Dyevre, *The Future of Legal Theory and The Law School of the Future*. Antwerpen: Intersentia (2015).

Text-mining for Lawyers: How Machine Learning Techniques Can Advance our Understanding of Legal Discourse

Arthur Dyevre*

Abstract

Many questions facing legal scholars and practitioners can be answered only by analysing and interrogating large collections of legal documents: statutes, treaties, judicial decisions and law review articles. I survey a range of novel techniques in machine learning and natural language processing – including topic modelling, word embeddings and transfer learning – that can be applied to the large-scale investigation of legal texts

Keywords: text mining, machine learning, law, natural language processing

1 Introduction

Much of the information of interest to lawyers and legal scholars comes in the form of texts. Whether they are briefs, contracts, court rulings, law review articles, legislative acts, treaties, newspapers or blog posts, all are either legal documents themselves or documents about the law. Retrieving, analysing, commenting, relating and expounding these documents has been the bread and butter of legal practice and legal scholarship alike for centuries.

Lawyers deal in words, and the law can be viewed as a vast and complex network of interrelated texts, as illustrated in Figure 1. The function of this discourse is not only to announce legal rules and how they apply to a particular set of facts but also to explain or summarise them in more succinct or more accessible language – which is understood to be one of the core functions of traditional, doctrinal scholarship.

While the study of legal texts is at least as old as academic legal scholarship, what is new is that a whole range of text mining techniques have emerged to assist the legal community in navigating and analysing the ever-expanding sea of legal and law-related documents. These techniques rely on recent advances in machine learning and natural language processing.

The media hype about artificial intelligence (AI) occasionally leads to exaggerated claims about the capabilities of these techniques. Except for the simplest legal tasks, robot lawyers are not yet around the corner. Nor are fully automated robot judges (provided that robot judges are even desirable, which is, at least, questionable). However, even if the media hype (sometimes amplified by legal scholars) paints a misleading picture of what AI can achieve, it would be at least equally wrong to dismiss these techniques as irrelevant to legal practice or legal scholarship. This is true even for those who see themselves as hardcore black-letter law scholars. The now famous Gartner Hype Cycles tell us that perceptions of AI advances oscillate between peaks of inflated expectations and troughs of disillusionment before reaching a plateau of productivity.[1]

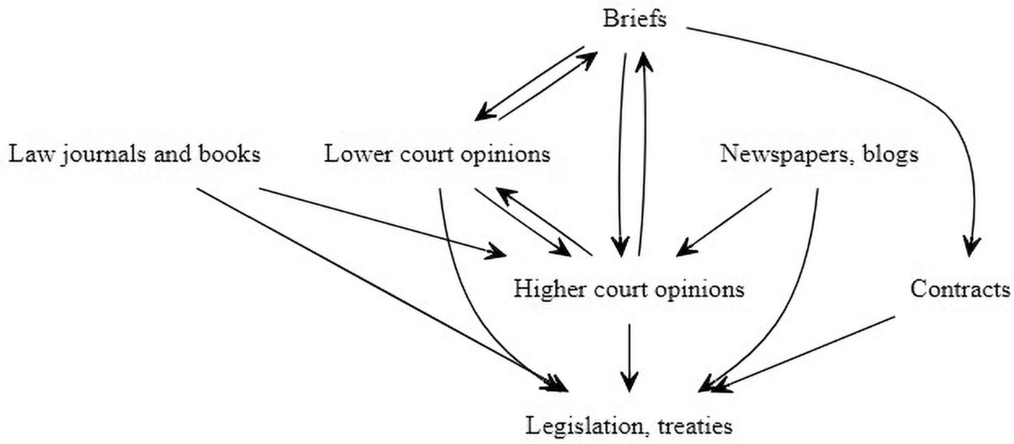
Researchers with experience in text-mining applications in the legal domain recognise that text-mining techniques cannot (yet) fully replace careful human reading. Yet these technologies are already sufficiently mature and progressing at a breakneck pace to deliver substantial advances. While increasingly popular in the interdisciplinary fields of law and economics, empirical legal studies and law and politics,[2] text-mining methods are also directly relevant to the work of doctrinal legal scholars. Indeed, one way to view them is as augmented doctrinal reality.

The present contribution aims to introduce these techniques to jurists who are unfamiliar with machine learning and natural language processing or who may only have a faint notion of the use that these tools can be put to. To this end, I shall first describe how data-harvesting methods can be deployed to gather large collections of legal documents. I will then proceed to explain how text is transformed into input data for text-mining tasks. Next, I will offer an overview of the text-mining techniques themselves, distinguishing supervised and unsupervised methods and walking the reader through a

* Arthur Dyevre is Professor at the KU Leuven Centre for Empirical Jurisprudence, Leuven, Belgium. arthur.dyevre@kuleuven.be. I am grateful to Dr. Nicolas Lampach, Dr. Timothy Yu-Cheung Yeung, Monika Glavina, Kyra Wigard and Nusret Ipek for their invaluable research assistance. I acknowledge financial support from European Research Council Horizon 2020 Starting Grant #638154 (EUTHORITY).

1. *See*www.gartner.com/smarterwithgartner/5-trends-drive-the-gartner-hype-cycle-for-emerging-technologies-2020/ (last visited 2 March 2021).
2. For a review *see* J. Frankenreiter and M.A. Livermore, 'Computational Methods in Legal Analysis', 16 *Annual Review of Law and Social Science* 39-57 (2020); for reflections and illustrations of the use of machine learning and natural language processing methods in empirical legal studies *see* M.A. Livermore and D.N. Rockmore, *Law as Data: Computation, Text, & the Future of Legal Analysis* (2019).

Figure 1 *Law as text*



bunch of examples from the EUTHORITY Project (www.euthority.eu). Finally, because I expect that many readers might be interested in learning some of the reviewed techniques, I will say a few words about practical software implementation.

The present article addresses mainly continental European legal scholars. Because of this deliberate focus, the discussion deliberately excludes tasks and questions – such as contract review or document assembly – that are important in legal practice,[3] but of lesser relevance to academic legal research, as traditionally understood in continental Europe. Nor do I engage matters such as causal inference that are central to the integration of text-mining and machine learning approaches in empirical legal studies and law and economics.[4] Furthermore, my aim is to introduce text-mining methods in terms that my target audience (hopefully) will find understandable. For this reason I eschew mathematical notation and technical jargon to focus on the underlying conceptual intuitions with the help of concrete illustrations. Obviously, this comes at the cost of precision. But I hope that this sacrifice earns the benefit of lowering the barrier to access. It is also worth mentioning at the outset that the scope of the present review is, by its very nature, limited. Text-mining and natural language processing have become vast fields, currently progressing at a breakneck pace possibly unmatched in any other field of scientific inquiry. So to pretend that this survey is, in any sense, comprehensive would be silly.

The present contribution assumes that legal scholars, with or without prior training in statistics or empirical methods, can become not just intelligent consumers but also active users of this panoply of powerful techniques. Readers interested in applying computational textual methods will find some pointers in the section on 'Learning Text-Mining Methods'.

2 Harvesting Legal Texts

Computerised text-mining methods require that texts be in digital form. Luckily, millions of legal documents are now available at a few clicks in electronic repositories and legal databases. The degree of exhaustiveness of these repositories varies widely from jurisdiction to jurisdiction. At best, judicial databases offer access to all published decisions. Often, it will only be to a subset of these decisions, with older rulings typically less likely to make the cut. Because the universe of documents is somewhat smaller, legislative databases usually fare better, although, here too, there are jurisdictional and cross-national disparities.[5] As official gazettes are increasingly published digitally, they potentially represent a treasure trove of legal data.

When documents are not available in digital format, it is still possible to convert them to this format using scanning combined with Optical Character Recognition (OCR). OCR works better with more recent, undamaged texts than with old dusty casebooks or well-worn legal treatises. However, the technology has made huge strides, thanks mainly to machine learning (which helps guess semi-erased words or phrases). It is now even possible to digitalise handwritten documents,[6] opening up new possibilities for legal historians to scour old manuscripts.

When done manually, assembling a large collection of legal documents for a text-mining project can be excruciatingly time-consuming (try to download all European Court of Justice decisions since 1954). However, data-harvesting techniques can make this step considerably easier. Using libraries designed for this purpose in popular programming languages like R and Python, it is possible to download the entire content of EUR-Lex (the EU law database) with less than five lines of code.

3. Efforts to automate these tasks have been an important focus of the emerging Legal Tech scene, *see* R. Dale, 'Law and Word Order: NLP in Legal Tech', 25 *Natural Language Engineering* 211-17 (2019).
4. *See* Livermore and Rockmore, above n. 3.
5. At the European Union level, EUR-Lex is fairly comprehensive with regard to both legislative acts and case law. EU law – EUR-Lex, https://eur-lex.europa.eu/homepage.html (last visited 9 November 2020). National databases are typically less complete.
6. Digitize Your Notes With Microsoft Computer Vision API | Nordic APIs |, Nordic APIs (2017), https://nordicapis.com/digitize-your-notes-with-microsoft-vision-api/ (last visited 9 November 2020).

Figure 2 Converting text into document-term matrix

Text 1

'The law shall apply to agreements between non-resident firms.'

Text 2

'Firms are required to abide by the terms of the present law.'

	the	firms	shall	are	apply	...
Text 1	1	1	1	0	1	...
Text 2	2	1	0	1	0	...

Document-term matrix

Web scraping, as the method is commonly referred to, is now the main go-to technique for collecting data in social scientific disciplines.[7] As scientists in various fields, including physics and medicine, have turned to text-mining methods to summarise vast collections of peer-reviewed articles, publishers (notably Oxford University Press and Elsevier) have made their journal collections available. Note, though, that the terms and conditions of commercial and non-commercial databases may sometimes explicitly prohibit web scraping, while there remain some uncertainties about when web scraping may be prohibited even for non-profit, purely academic research purposes.

3 From Text to Data

When we read a text our brain parses it applying our knowledge of semantics, syntax and context. In any language, the stock of words is finite, but syntactic rules allow the construction of infinitely many sentences from this finite vocabulary. Moreover, humans are able to communicate more than they say or write by taking the context into account. This is why we ascribe different meanings to the sentence 'I would like a table' when uttered in a restaurant and when uttered in a furniture shop.[8] While the language of legal documents – including contracts, statutes and judicial opinions – can diverge, sometimes significantly ('Any proviso to the contrary notwithstanding'), from everyday language, these basic principles of linguistic cognition and interpersonal communication are equally valid in the legal domain as in other areas of human activity.

Text-mining methods do not parse texts quite the same way the human brain does. Instead, these methods typically involve a good deal of complexity reduction. This may seem surprising to those less well-versed in machine learning. But even the most advanced natural language processing algorithms are still based on statistical principles. Texts are represented as numbers, in which the algorithms look for patterns. The ability to detect patterns depends on the amount of textual data and the sophistication of the algorithm, but the basic principle remains the same, including for the most cutting-edge techniques. In that sense, it is not entirely wrong to say that machine learning algorithms are still quite dumb. Yet their power stems from their ability to leverage the brute force of computing to arrive at useful (and sometimes surprisingly good) approximations.

Until recently, most text-mining methods relied on what is known as the bag-of-words (BOW) approach. To see what this amounts to, let us assume that we have a corpus with two texts, Text 1 and Text 2, as in Figure 2. The BOW approach involves converting texts into sequences of word counts and corpora to document-term matrices. The sequence of word counts representing a text is called a 'vector'. This vector contains counts of all the words occurring in that text and zeros for the words occurring in the other texts but not in that particular text. For Text 1 the zeros will represent all the words that appear in Text 2 but not in Text 1 and vice-versa. In a large corpus spanning a vocabulary of millions of words, the vector of word counts representing a text will contain mostly zeros – accounting for all the words that occur in other texts but not in the one under consideration.

To keep some phrases such as 'European Union' or 'Court of Justice' together instead of treating their component words as distinct lexemes, it is possible to throw some bigrams or trigrams into the document-term matrix. Think of an n-gram as a contiguous sequence of words. A bigram is a sequence of two words; a trigram a sequence of three words, and so on. Turned into a bigram 'European Union', for example, becomes 'European_Union', whereas 'Court of Justice' becomes the trigram 'Court_of_Justice'. These n-grams can then be processed just as individual words (unigrams).

In many applications, it is also common to remove so-called 'stopwords' – articles and prepositions like 'the', 'but', 'to', etc. – and to convert all words to lower case.[9] The resulting document-term matrix is the basic input

7. N.J. DeVito, G.C. Richards and P. Inglesby, 'How We Learnt to Stop Worrying and Love Web Scraping', 585 *Nature* 621-2 (2020).
8. D. Sperber and D. Wilson, *Relevance: Communication and Cognition* (1996).
9. Some text-mining tasks such as authorship identification require a distinct approach to pre-processing. Indeed, because pronouns and prepositions are markers of personal style, it is common to restrict the document-matrix to this class of words and to exclude nouns, verbs and adjectives.

 doi: 10.5553/ELR.000191 - ELR 2021 | No. 1

of many popular text-mining methods, such as latent semantic analysis (LSA) or topic modelling.

This modus operandi may strike many as a crude simplification. Yet, crude as it may be, this simplification can nonetheless produce useful results, as we shall see.

It is easy to see, however, that progress in modelling language and improvements in the performance of downstream applications – in law just as in other fields – ultimately entailed bringing the field beyond the BOW paradigm to develop richer representations of vocabularies while capturing more of the context and rules of syntax.

As we will see, static word embedding models such as Word2Vec have taken a significant step in that direction by representing words by their co-occurrence associations. These methods reflect the emergence of new paradigm building on notions from distributional linguistics, notably the intuition that a word is defined by the company it keeps.

Cutting-edge methods like transformers have taken the field several steps further into this new paradigm. Pre-trained on giant corpora, transformer models like Google's BERT (Bi-directional Encoder Representation Transformer) rely on a contextualised representation of word usage, enabling them to handle polysemy and to parse the reference of pronouns – a remarkable achievement that constitutes a major milestone in the development of AI language models.

Note, however, that while these novel techniques do not require converting raw texts to a document-term matrix, they still require texts to be in digitalised, machine-readable format.

4 Unsupervised Techniques

Computer scientists and machine learning scholars typically speak in terms of tasks –information retrieval, clustering, summarising, forecasting, etc. – or in terms of whether the method or algorithm operates with human-labelled documents or not – supervised versus unsupervised.

Translated into more familiar language, information retrieval is what jurists do when they search a document collection for a specific set of documents: e.g. entering a list of keywords into a database search engine to retrieve all judicial rulings addressing a particular issue. Similarly, clustering is what lawyers do when they try to sort out documents into categories: e.g. the themes to which law review articles relate or the topics coming up in judicial rulings. Turning long documents into more easily digestible summaries is also something that lawyers do on a routine basis. Prediction is something that one may not intuitively associate with texts. Yet words, too, whether from legal briefs or other textual inputs, can also serve to predict events or behaviours.

Techniques referred to as 'supervised' are those that necessitate human-labelled documents. They operate by seeking patterns correlated with human annotations, and their ability to predict how humans would annotate unlabelled documents is the measure of their performance. 'Unsupervised' techniques, on the other hand, do not require manually labelled textual input. However, the output they generate requires human interpretation or validation.

Some techniques and machine learning algorithms have been specifically designed for particular tasks. Yet several methods, some supervised, others unsupervised, may sometimes come into consideration for the same task, in which case the optimal choice should ultimately depend on the specific research question of interest to the legal analyst.

4.1 Word Cloud

Word cloud plots are arguably one of the most familiar and simplest text-mining methods. A word cloud simply plots words according to their aggregate frequency in the document-term matrix. Illustrated in Figure 3 is a slightly more sophisticated word cloud, known as a 'comparison cloud'.[10] It is based on a corpus compiling all European Court of Justice rulings up to 2015 (over 12,000 documents). Plotted are not the most frequent words in the overall corpus but the words that are most distinctive of the three main procedures: annulments (Art. 263 Treaty on the Functioning of the European Union (TFEU)); infringements (Art. 258 TFEU) and preliminary rulings (Art. 267 TFEU).

Our comparison cloud suggests that 'undertakings' is more distinctive of annulment proceedings (maybe because European Commission competition decisions reach the Court via this procedural channel), whereas 'agreement' and 'sugar' are more characteristic of, respectively, infringement and preliminary rulings.

Word clouds are popular and easy to interpret but are rather crude tools when it comes to detecting more granular patterns. In some applications pre-processing steps, such as restricting the document-term matrix to certain parts of speech (e.g. nouns or adjectives) may help make them more informative. But limitations remain.

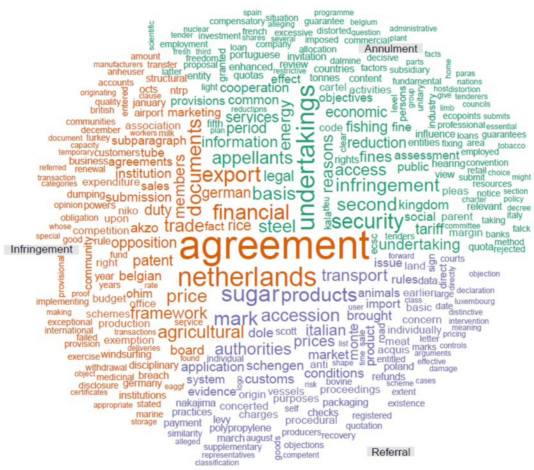
4.2 Latent Semantic Analysis and Principal Component Analysis

A notch more advanced are principal component analysis (PCA) and latent semantic analysis (LSA). Both are closely related and relatively old statistical techniques to arrange large arrays of data into more interpretable patterns. In the field of text mining, they fundamentally serve as unsupervised clustering methods to explore how texts and their words relate to each other.

PCA and LSA both work by seeking to represent the high-dimensional variations in word usage – a corpus and the documents it comprises vary in as many ways as the number of words in its vocabulary – into something

10. The size of a word reflects its deviation from their average across documents. Suppose $p_{i,j}$ is the rate at which word *i* occurs in document *j* and p_I its average rate across documents ($\frac{\Sigma_i p_{i,j}}{number\ of\ documents}$). Word size is determined by the maximum deviation ($max(p_{i,j} - p_I)$).

Figure 3 Comparison word cloud of infringement, annulment and preliminary rulings



more easily interpretable (and cognitively manageable) for the human brain. The output of both statistical procedures are a smaller number of dimensions on which words and documents are arrayed to facilitate the identification of meaningful patterns of relatedness.

The patterns of interest and the words expressing them depend on the specific task. PCA, for example, has been used to identify the authorship of *The Federalist Papers*.[11] But both methods can also be used to cluster legal documents around themes if one of the generated dimensions allows such an interpretation.

Figures 4 and 5 illustrate the use of LSA to explore oscillations in the position of the German Federal Constitutional Court over European from the 1960s and up to 2020. The corpus comprises 26 rulings, whose lengths vary from a little more than 1,000 to more than 20,000 token words.[12]

Here the basic interpretive assumption with which the output of the algorithm was approached is that variations in jurisprudential stance should be reflected in the use of words related to statehood and the internal market, with greater divergence in vocabulary manifesting greater jurisprudential divergence.

Figure 4 shows the extent to which selected clusters of words tend to appear in the same decisions. Unlike in word cloud plots such as the one depicted in Figure 3, the position of words has a precise meaning here. Vertical and horizontal axes denote separate dimensions, while the position of words is itself related to the documents in which they occur. If two documents share many words that are close to each other on a dimension, these documents will also be close to each other on that particular dimension. For example, 'sovereignty of the people' (*Volkssouveränität*), 'constitutional identity' (*Verfassungsidentität*), 'enumerated powers' (*Einzelermächtigung*) and '*ultra vires*' are close to each other on Dimension 1. These words are also more closely associated with the Court's more Eurosceptic judgments, like *Maastricht* and *Lisbon*. 'Duty to refer' (*Vorlagepflicht*), 'direct' (*unmittelbar*), 'effect' (*Wirkung*), 'export' (*Ausfuhr*), 'good' (*Ware*) form another separate cluster on the same dimension on the right-hand side. These words are also more closely associated with integration-friendly rulings, like *Kloppenburg*, *Banana* or *Lütticke*.

If we interpret Dimension 1 as Europhilia, the document positions associated with Dimension 1 can be interpreted as capturing the rulings' expressed position over European integration. Figure 5 depicts document positions on Dimension 1 over time. It shows that the resulting scaling is highly consonant with the conventional doctrinal wisdom. The *ECB Ultra Vires* ruling, in which the German Court declared the Court of Justice decision in *Weiss* ultra vires, clearly scores as the most Eurosceptic ever. Other rulings, such as *Maastricht*, *Lisbon* and *OMT*, which borrow the same state-centric sovereignty rhetoric, are also on the more Eurosceptic side, in keeping with the conventional wisdom.

A recent article compared the performance of eight algorithms, including LSA, in mapping the evolution of the German Court's case on European integration. The positions ascribed to the decisions by the algorithms were evaluated against scholarly accounts and legal expert ratings. A variant of LSA (correspondence analysis) performed best against scholarly accounts in law journals, achieving a 75% pairwise correlation.[13]

4.3 Topic Modelling

A more recent technique specifically designed for clustering and automated classification is topic modelling.[14] Suppose you have a large amount of legal texts and you want to get a sense of the themes and topics they pertain to. Instead of asking you to come up with a list of cate-

11. D.I. Holmes, 'Authorship Attribution', 28 *Computers and the Humanities* 87-106 (1994).
12. For a discussion and assessment of the performance of LSA and other text-mining methods to map jurisprudential change, *see* A. Dyevre, 'The Promise and Pitfall of Automated Text-Scaling Techniques for the Analysis of Jurisprudential Change', *Artificial Intelligence and Law* 1-31 (2020).
13. *Id.*
14. For a non-technical introduction *see* D.M. Blei, 'Probabilistic Topic Models', 55 *Communications of the ACM* 77–84 (2012).

Figure 4 Frames and phraseology of German constitutional rulings on Europe

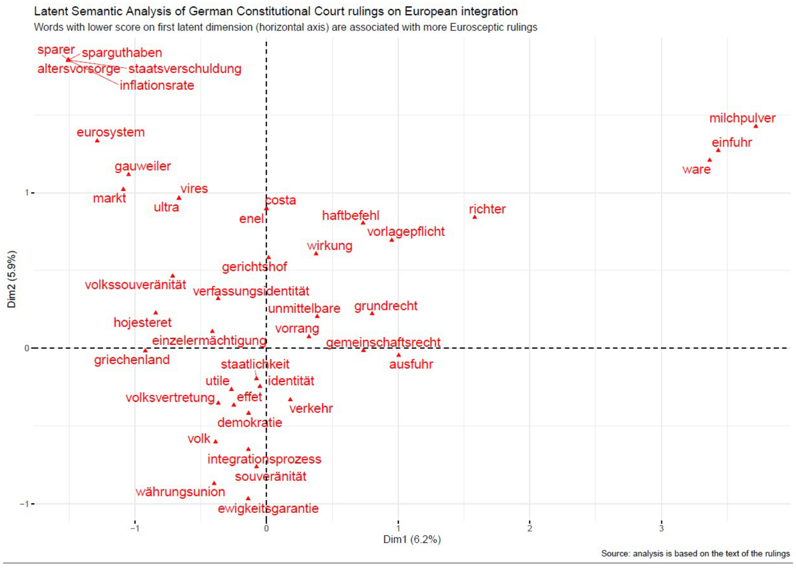
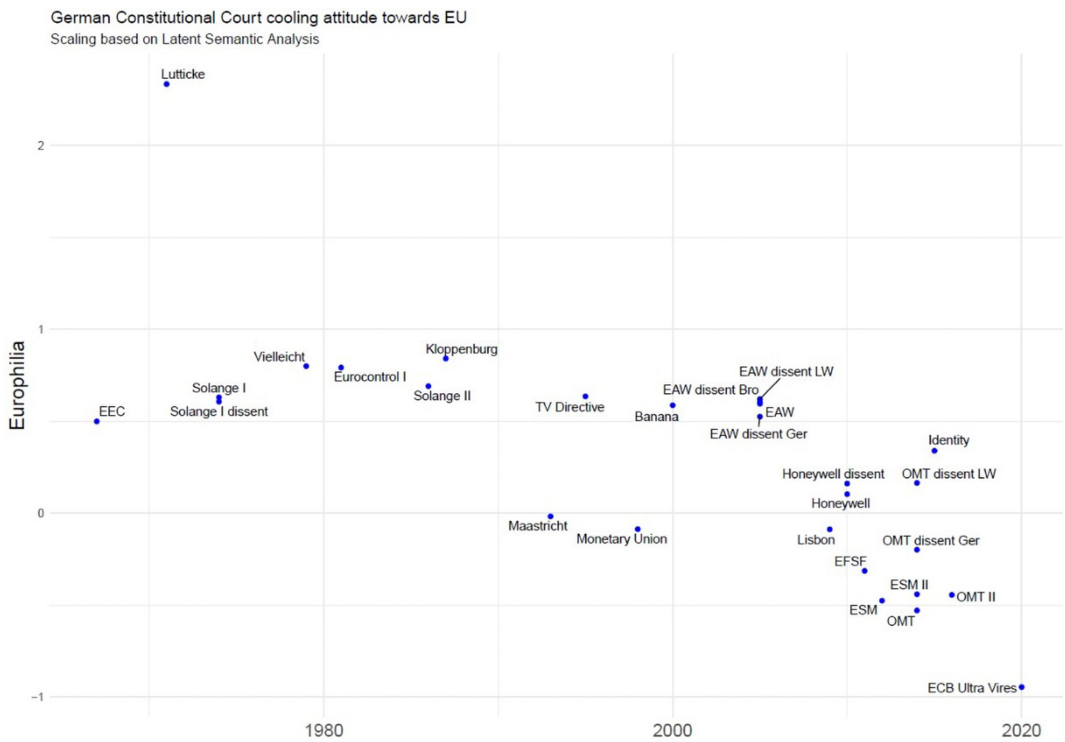


Figure 5 Evolution of the German Constitutional Court's stance on European integration based on Dimension 1 of LSA



gories or a classificatory scheme, topic modelling generates the categories and sorts out the documents accordingly after you have specified how many topics you wanted. At least, this is how the method is supposed to work.

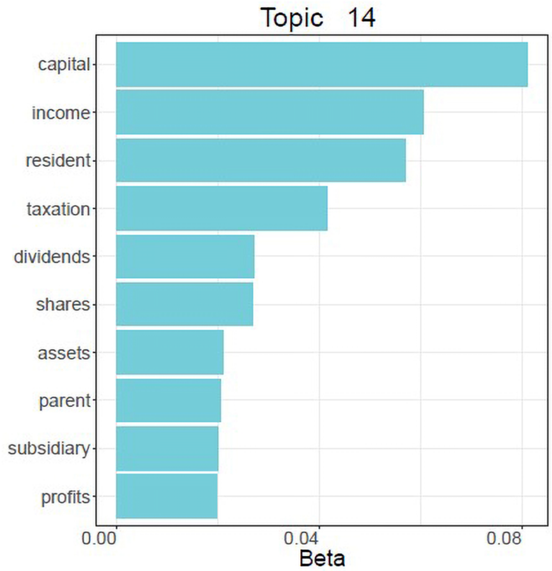
In topic modelling, topics are modelled as probability over words and documents as probability over topics. To generate the topics, the algorithm tries to find which probabilities are most likely to have generated the observed documents.

Figure 6 illustrates the output of a topic model of preliminary rulings (approximately 8,000 rulings). The number of topics was set at 25. What Figure 6 displays is one of these topics represented by its 10 most distinctive words (note that the higher the beta value, the more characteristic of the topic the word is). Looking at these 'keywords' – which, it is essential to understand, are not chosen by the researcher but emerge from the analysis – we may plausibly summarise this topic as corporate taxation.

In topic modelling, documents are conceptualised as mixtures of topics and, in addition to generating topics, a topic model tells you what proportion of what topic documents are likely to contain. So to check that our interpretation of topic 14 is correct we can inspect the decision that, according to the model, has the highest proportion of this topic. In that case, it turns out to be *Test Claimants in the FII Group Litigation v Commissioner of Inland Revenue*[15], a 2012 Grand Chamber ruling,

15. 12 December 2012, C-446/04.

Figure 6 Topic from topic model of preliminary rulings (1961-2016)



which according to the model is 99% about topic 14. Here is a quote from the first ruling:

> The High Court of Justice of England and Wales, Chancery Division, seeks, first, to obtain clarification regarding paragraph 56 of the judgment in Test Claimants in the FII Group Litigation and point 1 of its operative part. It recalls that the Court of Justice held, in paragraphs 48 to 53, 57 and 60 of that judgment, that national legislation which applies the exemption method to nationally-sourced dividends and the imputation method to foreign-sourced dividends is not contrary to Articles 49 TFEU and 63 TFEU, provided that the tax rate applied to foreign-sourced dividends is not higher than the rate applied to nationally-sourced dividends and that the tax credit is at least equal to the amount paid in the Member State of the company making the distribution, up to the limit of the tax charged in the Member State of the company receiving the dividends.

So it does really look like corporate taxation after all.
Topics can be visualised in various ways. In Figure 7, they are represented as a network in which node size represents overall topic proportion in the overall document collection while edge thickness corresponds to the weighted number of shared words. This way we can see themes emerging from the topics.

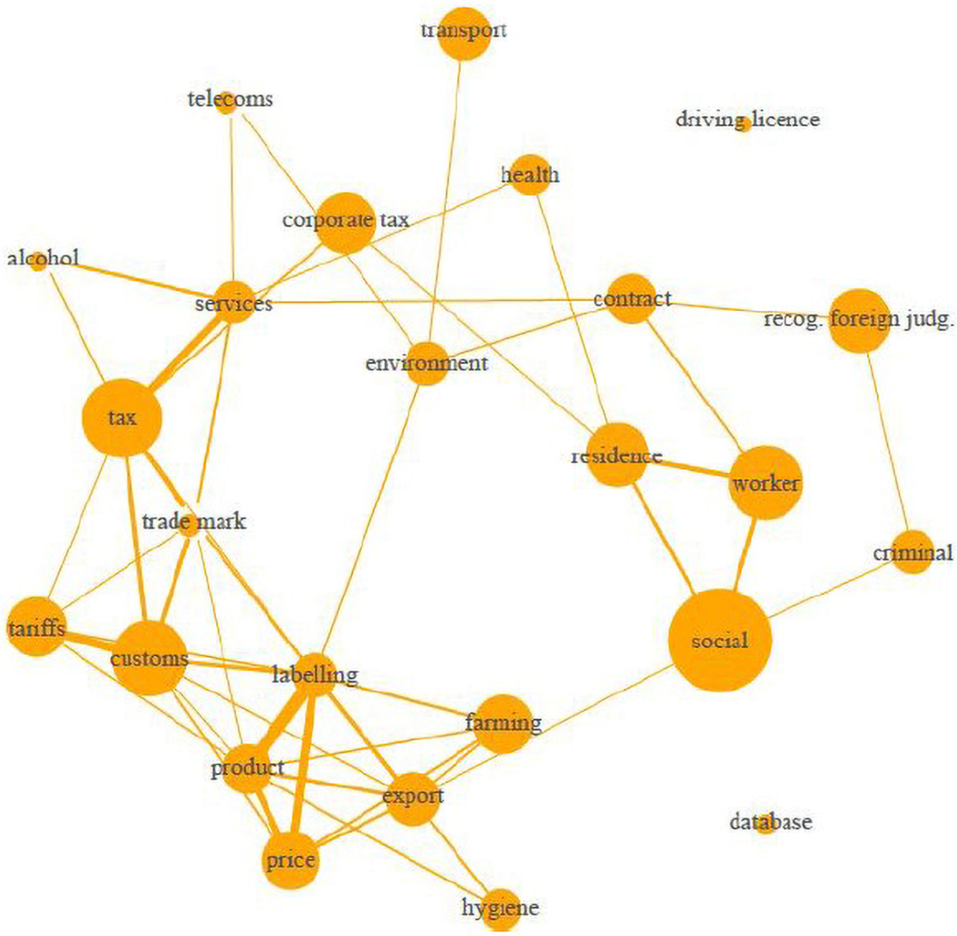
Among other things, Figure 7 suggests that internal market and tax issues represent a big chunk of what the Court of Justice of the European Union (CJEU) does. However, social rights, residence rights and the recognition of foreign judgments (private international law) also make for a substantive share of the cases on which the Luxembourg judges sit.

If you think that 25 categories is too few to get a good sense of issue prevalence in the Court's case law, how about a topic model with 100 categories? In Figure 5 we see that such a topic provides a more detailed picture, although we find the same themes (internal market in the lower-right region, social and immigration issues in the left region).

It is also possible to construct dynamic topic models to study the evolution of case law over time or 'litigant' topic models to study how issues vary across litigant types.
Recent work has applied topic modelling to study relative issue emphasis across infringement, annulment and preliminary rulings, highlighting how the CJEU's case law is influenced by the litigation agenda of case initiators (like the European Commission);[16] to compare topic salience in European Union legislation, CJEU rulings and contributions to the *Common Market Law Review*;[17] to explore Dutch Supreme Court decisions;[18] and to demonstrate the lingering centrality of market regulation in European Union law-making in the twenty-first century.[19] While significant efforts have been expended on manually classifying the legal areas addressed by US Supreme Court rulings, some authors have proposed topic modelling as a more efficient and more accurate alternative.[20] Work by Peter Grazl and Peter Murrel further illustrates how topic modelling can assist in exploring large collections of old legal texts. They apply topic modelling to reports of cases heard by English

16. A. Dyevre and N. Lampach, 'Issue Attention on International Courts: Evidence from the European Court of Justice', *Review of International Organizations* 1-23 (2020).
17. A. Dyevre, M. Ovadek and M. Glavina, 'The Voices of European Law: Legislators, Judges and Law Professors', forthcoming *German Law Journal* (2021).
18. Y. Remmits, *Finding the Topics of Case Law: Latent Dirichlet Allocation on Supreme Court Decisions* (2017).
19. N. Lampach, W. Wijtvliet and A. Dyevre, 'Merchant Hubs and Spatial Disparities in the Private Enforcement of International Trade Regimes', *International Review of Law and Economics* 105946 (2020).
20. D. Rice, 'Measuring the Issue Content of Supreme Court Opinions', 7 *Journal of Law and Courts* 107-27 (2019).

Figure 7 Topic model of CJEU preliminary rulings represented as network



between the fourteenth and eighteenth centuries (N = 52,949).[21]

4.4 Word Embeddings

Tools like LSA, PCA and topic modelling are typical of the BOW paradigm. Word embeddings, by contrast, are part of a new text-mining paradigm inspired by the defining principle of distributed linguistics – 'a word is defined by the company it keeps'.[22]

To explain how word embeddings work, the best is, again, to start with an example. Suppose you want to investigate variations in attention to a particular phenomenon, e.g. politics and populism in posts on a major legal blog. To measure attention to this concept, we might first try to come up with a list of keywords (e.g. 'politics', 'party', 'populism'…) capturing attention to this phenomenon and then determine the extent to which our keywords are actually matched in the document collection. However, this approach often delivers poor results because the same phenomenon can be characterised in many different ways, leading exact matches to either over- or underestimate the true number of relevant instances of attention to the phenomenon in question. (The frustrating feeling is surely one that many jurists have experienced when trying to retrieve documents via a keyword search in some legal database.)

Word embeddings help deal with this problem by representing words not as frequencies – the BOW approach – but as sequences (i.e. a vectors) of occurrence similarities, generated via a shallow neural network. For example, Table 1 displays the first 40 items in the vector of occurrence similarities yielded by a word embeddings model trained on the German-language contributions to the *Verfassungsblog* (a leading constitutional law blog) using the Word2Vec algorithm. The vector corresponds to the words *Politik* (politics), *Parteien* (parties) and *Populismus* (populism).[23]

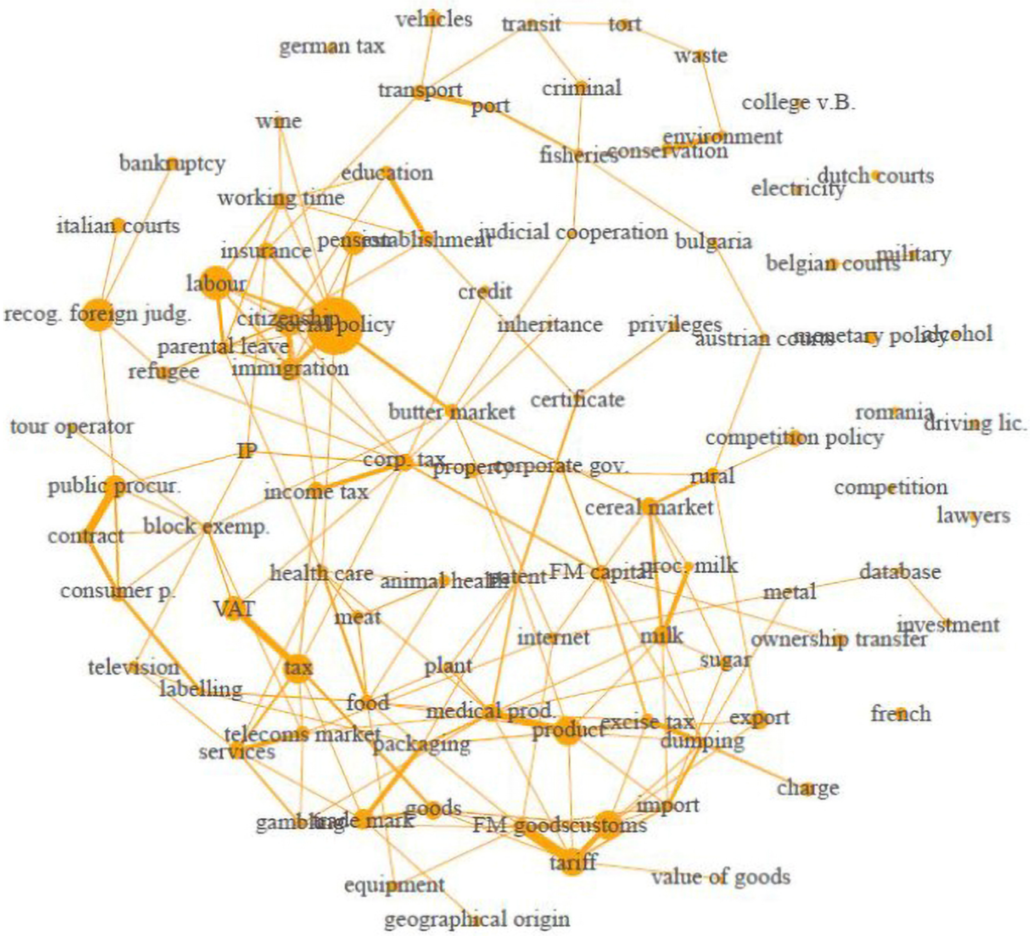
Numbers next to the words in Table 1 indicate the cosine occurrence similarity. The closer it is to 1, the more similar is the word's context of occurrence to that of *Politik* + *Parteien* + *Populismus*. Here the word exhibiting the highest cosine similarity score is *Eliten* (elites), which makes sense since elite-bashing is a defining feature of populist discourse. Other terms, including *Presse* (press), *Medien* (media) and *Bürger* (citizen), frequently

21. P. Grajzl and P. Murrell, 'A Machine-Learning History of English Caselaw and Legal Ideas Prior to the Industrial Revolution I: Generating and Interpreting the Estimates', 17 *Journal of Institutional Economics* 1-19 (2021).

22. T. Mikolov et al., 'Distributed Representations of Words and Phrases and Their Compositionality', in C.J.C. Burges and L. Bottou and M. Welling and Z. Ghahramani and K.Q. Weinberger, *Advances in Neural Information Processing Systems* 3 (111-3119 (2013).

23. Several word embedding algorithms exist, including Word2Vec, Fasttext and Glove. Here we relied on the Word2Vec approach.

Figure 8 Topic model of CJEU preliminary rulings with 100 topics



come in populist rhetoric, too. *Verwaltung* (administration) and *Instrumente* (instruments), though, are less intuitively associated with politics, populism or partisan organisations.

Co-occurrence similarity here refers to the words that tend to appear around the target word. How many words before and after the target word should be considered – the window size – is one of the parameters that have to be set by the researcher before training an embedding model. A window size of 5 means that two words and two after the target word will be considered; a window size of 9 four words before and four words after; and so on. The neural network is then trained to predict either the target word from the surrounding words (continuous bag-of-words method) or the surrounding words given the target word (skip-gram method).

As with machine learning and neural networks in general, the more data (texts) the better. This is why instead of training embeddings from scratch on a relatively small collection of blog posts, it may be preferable to use a pre-trained model built on a much larger corpus. Table 2 shows the words associated with *Politik* + *Parteien* + *Populismus* from a pre-trained embeddings model constructed from all German Wikipedia pages, with a vocabulary of nearly five million words.[24] Pre-trained embeddings constructed from legal documents also exist.[25]

The cosine similarity scores are generally higher in Table 2 than in Table 1, which suggests that the pre-trained model better captures contextual similarity. In fact, it assigns a high cosine similarity scores to typos like 'poltk' (cosine = 0.846). This is because typos appear in the same context as the word with the correct spelling. The similarity scores assigned to typos highlight how word embedding models handle synonymy, which represents a major advance for legal information retrieval tasks.

That pre-trained embeddings can deliver better results than locally trained embeddings (i.e. embeddings trained on the corpus one actually wants to investigate) illustrates the notion of transfer learning. What a model learns about language use from a very large corpus is often transferable to smaller text collections.

24. A wide range of word embedding models spanning multiple languages can be downloaded from a repository made available by the Language Technology Group at the University of Oslo; *see*http://vectors.nlpl.eu/repository (last visited 12 November 2020).
25. I. Chalkidis and D. Kampas, 'Deep Learning in Law: Early Adaptation and Legal Word Embeddings Trained on Large Corpora', 27 *Artificial Intelligence and Law* 171-98 (2019).

Table 1 Top 40 occurrence similarity scores for vector Politik + Parteien + Populismus from embeddings (Word2Vec) trained on German-language contributions to the Verfassungsblog

eliten 0.8215773105621338	veränderungen 0.7256141901016235
vorteile 0.7712808847427368	gerechtigkeit 0.7244186401367188
öffentlichkeit 0.7557699084281921	medien 0.7238696813583374
bürger 0.7556987404823303	bevölkerung 0.7218047976493835
institutionen 0.7544448971748352	instrumente 0.7199758887290955
positionen 0.7507109642028809	kultur 0.7188452482223511
denjenigen 0.7492086291313171	verfassungen 0.7175022959709167
presse 0.7491012215614319	verteidiger 0.7158944606781006
vielfalt 0.7490779757499695	schmieden 0.7158849239349365
minderheiten 0.7462370991706848	verhältnisse 0.7151006460189819
nationalstaaten 0.7429373264312744	etablierten 0.7142930626869202
demokratien 0.7417148351669312	wissenschaft 0.7125871777534485
arena 0.7414146065711975	ideen 0.711867094039917
repräsentanten 0.7355824112892151	unionsbürger 0.7089967131614685
elite 0.7347079515457153	chancen 0.7073072791099548
gesellschaft 0.7327207922935486	debatten 0.7052429914474487
solidarität 0.7316428422927856	staatssekretäre 0.704701840877533
verwaltung 0.7315931916236877	justiz 0.7037882804870605
vernunft 0.7314342260360718	kommunikation 0.7020408511161804
wirtschaft 0.7265527248382568	minderheit 0.701974093914032

One powerful application of word embeddings is to generate weighted lexicons, which can be utilised to detect attention to a particular phenomenon or concept of interest. Figure 9 plots the variation in attention to politics, parties and populism in German-language contributions to the *Verfassungsblog* using the words contained in the vector *Politik* + *Parteien* + *Populismus* to measure the average attention to the underlying phenomenon.
Over time, the *Verfassungsblog* has been posting a growing number of English-language contributions. Figure 10 charts attention to the same phenomenon in English-language posts using the vector *politics* + *parties* + *populism* generated by Google's pre-trained word embedding model (Word2Vec) for English – which boasts a vocabulary of three billion words trained on Google News data.

These are potentially interesting results for scholars interested in the evolution of European constitutional law scholarship and a possible shift from a legalistic, narrowly doctrinal conception of legal scholarship to one that pays greater heed to political behaviours and social dynamics.[26]
To further illustrate the potential of word embeddings for attention detection and document retrieval, note that we can vary the specification of vectors to improve results or to capture conceptual nuances. The vector generated for *politics* alone will be different from the vector generated for *politics* + *parties* + *populism*. But if we wanted to generate a vector for terms associated with politics and political parties but not with populism, we could specify a vector like *politics* + *parties* – *populism*. Remarkably, in the Google pre-trained model, specifying *king* – *man* generates a vector in which the word with the highest cosine similarity score is *queen*.
So, by comparison with document search engines based on exact keyword matching, word embeddings provide a considerably more powerful tool to capture attention to concepts.
Another application of word embeddings is to compare change in the connotations of words across time, in

26. B. Caiepo and F. Benetti, 'How Political Turmoil is Changing European Constitutional Law: Evidence from the Verfassungsblog', *Verfassungsblog* (2020), https://verfassungsblog.de/how-political-turmoil-is-changing-european-constitutional-law-evidence-from-the-verfassungsblog/ (last visited 9 November 2020).

Table 2 Top 40 occurrences for vector Politik + Parteien + Populismus from embeddings trained on German Wikipedia pages

parteitaktik 0.8728734850883484	europapolitik 0.8352898359298706
parteipolitik 0.8678461909294128	hinterzimmerpolitik 0.8350450992584229
parteienpolitik 0.8615270853042603	demokratiedebatte 0.8343005180358887
parteienoligarchie 0.8604286313056946	demokratieverachtung 0.833306074142456
klientel- 0.8596892952919006	demokratieverlust 0.833281397819519
einheitsparteien 0.8587565422058105	eu-kritischer 0.8332792520523071
demokratie 0.8570675849914551	nationalpopulistischen 0.8332092761993408
populisten 0.8544521331787109	systemopposition 0.8329799771308899
euro-kritik 0.8523470163345337	linkspopulisten 0.832958459854126
klüngelei 0.8509451746940613	stimmungskanzlerin 0.8328656554222107
poltik 0.8463006615638733	schröder-ära 0.8327784538269043
stimmungsdemokratie 0.8435574173927307	politischen 0.8327664136886597
eu-zentralismus 0.8430708050727844	politikeliten 0.8319368362426758
parteienkartell 0.8392568826675415	europafeindlichkeit 0.8310469388961792
regierungspolitik 0.839039146900177	wirtschaftslobbyismus 0.8308700323104858
partikularinteressen 0.83860182762146	eurorettungspolitik 0.8304780125617981
parteiengezänk 0.8367708325386047	konzernspenden 0.8304095268249512
troika-politik 0.83570396900177	protestparteien 0.8301451206207275
linkspopulismus 0.8356888890266418	euroskepsis 0.830102801322937
wirtschaftslobbyisten 0.8353110551834106	parteitagsbeschlüsse 0.8300416469573975

Figure 9 Relative incidence of words relating to 'Politik', 'Partei' and 'Populismus' in German-language contributions to the Verfassungsblog, 2009-2019

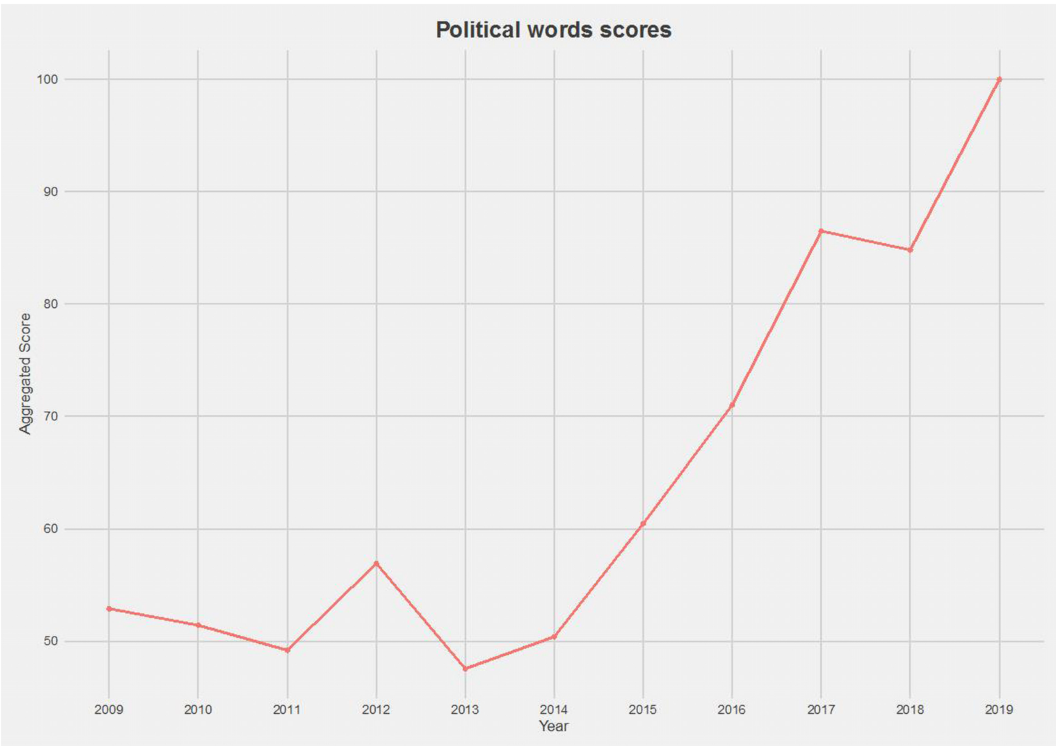
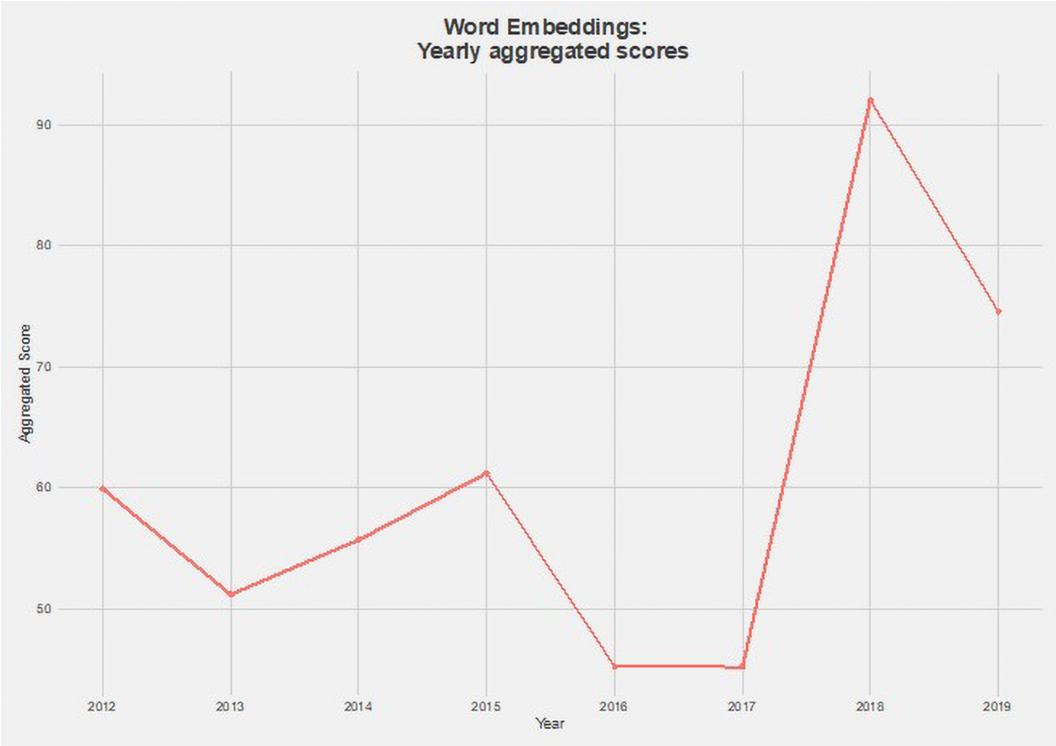


Figure 10 Relative incidence of words relating to 'politics', 'parties' and 'populism' in English-language contributions to the Verfassungsblog, 2012-2019



which case separate embeddings models are trained on subsets of the corpus corresponding to distinct periods.[27] A noted study by Elliott Ash, Daniel Chen and Arianna Ornaghi has applied a similar approach to compare gender stereotypes across the opinions of federal judges in the United States.[28] 380,000 judicial opinions were grouped by authorship, and separate embedding models were trained for each judicial author. The distance between the vector *male – female* and *career – family* was then used to construct a gender slant indicator. The authors find that judges for whom these two vectors are closer – meaning that they more closely associate men and women with traditional gender roles – vote more conservatively on women's rights' issues such as reproductive rights, sexual harassment and gender discrimination. Moreover, they are less likely to assign opinions to female judges but are more likely to reverse lower-court decisions if the lower-court judge is a woman, and they cite fewer female-authored opinions.

A related study by Douglas Rice, Jesse Rhodes and Tatishe Nteta has examined racial biases in a corpus comprising over 1 million state and federal court opinions. The authors find stereotypically African-American names to be systematically associated with more negative words compared with stereotypically European-American names.[29]

4.5 Document Clustering with Word Embeddings: Doc2Vec

Closely related to the word embedding approach just described is a document clustering technique known as Doc2Vec. It relies on the same representation of words, but instead of training the neural network to predict only the target word or the surrounding terms, it is also trained to predict the documents in which they occur. Documents thus become associated with word vectors.

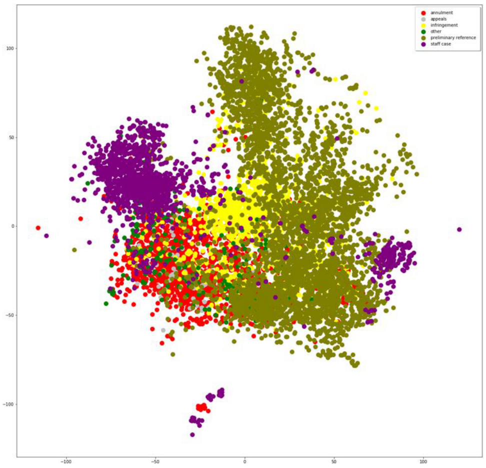
Doc2Vec is similar to PCA/LSA in that it simultaneously relates words and documents. The principal difference, however, is that Doc2Vec draws on a much more sophisticated word representation.

A Doc2Vec model can be visualised by means of a t-SNE (shorthand for 't-distributed stochastic neighbour embedding') plot. A t-SNE plot brings the high-dimensional vector representations of documents into a format where similarities among documents are easier to appreciate.

Figure 11 shows a t-SNE plot of a Doc2Vec model of European Court of Justice rulings, with colours denoting the procedure. The horizontal and vertical axes of a t-SNE plot are not amenable to substantive interpretation. But spatial proximity reflects similarity in word usage. Here the plot suggests some degree of overlap across procedures but greater heterogeneity in rulings originating in preliminary references.

27. Studies adopting this approach have revealed the evolution of gender and ethnic stereotypes or the changing connotations of the word 'gay'; *see* W.L. Hamilton, J. Leskovec and D. Jurafsky, 'Diachronic Word Embeddings Reveal Statistical Laws of Semantic Change', ArXiv Prepr. ArXiv160509096 (2016); N. Garg et al., 'Word Embeddings Quantify 100 Years of Gender and Ethnic Stereotypes', 115 *Proceedings of the National Academy of Sciences* E3635-E3644 (2018).
28. E. Ash, D.L. Chen and A. Ornaghi, *Stereotypes in High-Stakes Decisions: Evidence from US Circuit Courts* (2020).
29. D. Rice, J.H. Rhodes and T. Nteta, 'Racial Bias in Legal Language', 6 *Research & Politics* (2019), doi: 10.1177/2053168019848930.

Figure 11 T-SNE plot of Doc2Vec model of European Court of Justice rulings (colour denotes procedure)



Looking at a large corpus of US Court of Appeals rulings, Daniel Chen and Elliott Ash have explored a variety of possible uses of Doc2Vec for the analysis of judicial opinions.[30]

Because precedents play an important role in legal argumentation, several studies have proposed Doc2Vec as a methodology to identify and measure case similarity.[31]

5 Supervised Classification Methods

Unsupervised approaches produce models and output without human input, which may seem to be a great advantage. However, the models and output generated by unsupervised methods always require ex post human interpretation. There is no absolute guarantee that the topics generated by a topic model will make sense or that the dimensions produced by an LSA model will be interpretable. This is not necessarily a problem if unsupervised techniques are primarily used for exploratory purposes. However, if one purports to rest an empirical assertion on the results of unsupervised methods, some human validation of at least a subset of these results may be required in order to demonstrate intersubjective validity.

Supervised methods, by contrast, do not require ex post validation because they seek to 'emulate' what humans do by discovering patterns in documents labelled by human annotators prior to training.

5.1 Obtaining Labelled Documents

Supervised approaches all require labelled documents. There are only two ways of obtaining labelled data. The first is to rely on documents that other researchers have already annotated. To measure the ideological direction of US federal court opinions, Carina Hausladen, Marcel Schubert and Elliott Ash were able to leverage an existing database (the Songer Database) where ideological direction had been hand-coded for a subset (5%) of federal cases. These annotated opinions were then used to train and test a range of algorithms.[32] Using labelled data sets from outside the legal domain can be tempting. But results may then have to be interpreted with caution. One study, for example, has sought to leverage academic articles from moral philosophy that had been labelled either as 'deontological' or 'consequentialist' to train machine learning classifiers to detect modes of moral reasoning in US federal opinions.[33] However, given the risk of low domain adaptation (the language of academic articles and judicial opinions may diverge too much), the results of studies adopting this strategy should be taken with a grain of salt.

When no labelled data set exists, the only way to obtain labelled data is to build it from scratch. In many areas, supervised machine learning projects rely on crowdsourcing platforms such as Amazon Mechanical Turk, where annotators recruited online tag documents for a modest compensation (and so at a low cost for the researcher).[34] However, crowdsourcing works best when a task is simple, quick and straightforward. So the specificity, technicality and complexity of legal language means that the crowdsourcing approach is not well suited to legal projects.

30. E. Ash and D.L. Chen, 'Case Vectors: Spatial Representations of the Law Using Document Embeddings', *Law as Data, Santa Fe Institute Press, ed. M. Livermore and D. Rockmore* (2019).
31. T. Novotná, 'Document Similarity of Czech Supreme Court Decisions', 14 *Masaryk University Journal of Law and Technology* 105-22 (2020); L.T.B. Ranera, G.A. Solano and N. Oco, 'Retrieval of Semantically Similar Philippine Supreme Court Case Decisions using Doc2Vec', in 2019 *International Symposium on Multimedia and Communication Technology (ISMAC)* 1-6 (2019); P. Bhattacharya et al., 'Methods for Computing Legal Document Similarity: A Comparative Study', ArXiv Prepr. ArXiv200412307 (2020).
32. C.I. Hausladen, M.H. Schubert and E. Ash, 'Text Classification of Ideological Direction in Judicial Opinions', 62 *International Review of Law and Economics* 105903 (2020).
33. N. Mainali et al., 'Automated Classification of Modes of Moral Reasoning in Judicial Decisions', in R. Whalen (ed.) *Computational Legal Studies*, 77–94 (2020).
34. C. Grady and M. Lease, 'Crowdsourcing Document Relevance Assessment with Mechanical Turk', in *Proceedings of the NAACL HLT 2010 Workshop on Creating Speech and Language Data with Amazon's Mechanical Turk* 172-9 (2010).

Law students potentially offer a solution to the document annotation challenge. Without being accomplished legal experts, they tend to be more comfortable with legalistic language and better at parsing judicial prose or statutory provision. It is possible to integrate legal annotation tasks in tutorials and interactive classes. In fact, annotating legal documents can be viewed as an excellent exercise for students to practise and perfect legal analytic skills.[35]

At the KU Leuven Law School, EUTHORITY Project researchers recently conducted an annotation experiment with 52 third-year law students. Two hours per week, the annotation team assembled to annotate Belgian constitutional and Supreme Court (Court of Cassation + Council of State) rulings for explicit references to EU law. The rulings of Belgian top courts vary in both length (from a little more than thousand to several hundred thousand words) and complexity (from relatively simple asylum cases to more arcane regulatory issues). Eventually, the team was able to annotate 519 rulings, which is a reasonable number of documents but obviously pales in comparison with the millions of annotated pictures[36] computer scientists and engineers are able to tap into to train image-recognition algorithms.

Conducting annotation tasks with large groups of students demands a good work flow. Annotators must first be trained to recognise the concepts and information that the labels have been designed to capture. Documents must then be distributed to annotators and annotated documents collected. To produce high-quality annotations, it is recommended to have two or more annotators independently annotating the same document. Discrepancies then have to be identified to compute inter-annotator agreement metrics. To bolster quality, a reconciliation procedure can also be put in place. Software and online platforms have been developed to facilitate the completion of document annotation tasks by teams of annotators. The aforementioned annotation project conducted in Leuven, for example, relied on the cloud version of the TagTog[37] platform (which the project was allowed to use free of charge in exchange for making the labelled data public). Some software solutions, such as WebAnno,[38] are open source and can thus be used free of charge but require local server installation – which can be technically involved, unless technical support is provided.

5.2 Bag-of-Words Methods

As with unsupervised techniques, supervised techniques initially relied on the BOW paradigm. Supervised BOW methods involved the same data preparation steps, including converting texts into document-term matrix format. In a supervised set-up, the document matrix will look very similar, except that it will contain one or more additional columns for the labels produced by human annotators.

Before trying out some classification algorithms, the next stage will be to divide the data into train and test data. As its name suggests, the train set will serve to train many versions of the algorithm, whereas the test set will serve to measure their performance and select the best one. Dividing the data into train and test sets is called the 'holdout' procedure and is only one of many sampling procedures. When the number of annotated documents is small (less than 1,000), it is recommended to use some 'cross-validation' procedure. Cross-validation procedures begin by dividing the annotated documents into several folds (e.g. 10). One of the folds then serves as a test set while the algorithms are trained on the remaining folds. This process is then iterated with a different fold until every fold has served as a training set. Performance is evaluated by looking at the average across test folds. This way cross-validation ensures that as much data as possible is used for training.

When we say that the train set serves to train 'many versions' of an algorithm, we mean many combinations of words correlated with the labels. How many versions of the algorithm are fitted to the train data is for the researcher to decide in light of time and computational constraints (fitting a broader range of possible combinations obviously takes more time).

All these competing versions of the algorithm are then tested against the test data. The version that best predicts the human annotations in that set is then selected as the winner.

By way of illustration, we trained several algorithms to predict the labels 'EU law' and 'no EU law' in the aforementioned student-annotated corpus of Belgian high court rulings. Because this data set is relatively small (519 documents), we employed a cross-validation procedure. We then fitted thousands of versions of a handful of popular algorithms: logistic regression, support vector machine (SVM), random forest and sequential neural network.[39] While explaining the technical specifications of these algorithms is beyond the scope of the present article, Table 3 reports the performance of the 'best version' of each of these algorithms.

The metrics reported in Table 3 are the ones typically used in supervised text-mining classification tasks. Precision indicates the proportion of documents predicted to contain references to EU law that truly do so. On this metric, logistic regression and sequential neural network did best, achieving a precision of 95%. Recall measures the proportion of documents human annotators labelled as featuring EU law that the algorithm was able to

35. Conducting legal AI projects also helps bring greater awareness of the potential of new technologies for legal research and legal practice while contributing to the modernisation of legal education; *see* A. Dyevre, 'Fixing European Law Schools', 35 *European Review of Private Law* 151-168 (2017).
36. *See*www.image-net.org/ (last visited 9 November 2020).
37. *See*www.tagtog.net/ (last visited 3 March 2021).
38. https://webanno.github.io/webanno/ (last visited 4 March 2021).
39. For a concise explanation of these algorithms I refer the reader to S. Yıldırım, '11 Most Common Machine Learning Algorithms Explained in a Nutshell', *Medium* (2020), https://towardsdatascience.com/11-most-common-machine-learning-algorithms-explained-in-a-nutshell-cc6e98df93be (last visited 9 November 2020). For a survey from the perspective of econometrics *see* M. Gentzkow, B. Kelly and M. Taddy, 'Text as Data', 57 *Journal of Economic Literature* 535-74 (2019).

Table 3 Performance metrics of algorithms trained to predict the presence of references to EU law in Belgian high court rulings

	Precision	Recall	F1	MCC
Logistic regression	0.95	0.75	0.84	0.70
SVM	0.83	0.83	0.83	0.63
Random forest	0.91	0.88	0.89	0.77
Sequential neural net	0.95	0.75	0.84	0.70

retrieve. Here random forest did best, retrieving 88% of the documents thus labelled. F1 is a metric that combines precision and recall into a single number. The Matthews correlation coefficient (MCC) is yet another performance metric. It is recognised as the most reliable metric to evaluate a binary classifier because it takes into account the proportion of true negatives (documents predicted to feature no EU law and do not), false negatives (documents predicted to feature no EU law but that actually do), true positives (documents predicted to feature that really do) and false positives (documents predicted to feature EU law but that do not). Here random forest performs best with MCC = 0.77.

Similar BOW supervised approaches have variously been used to predict the outcome of ECHR cases;[40] the ideological direction of US federal opinions;[41] and to detect unfair clauses in online terms of service.[42]

5.3 Transfer Learning and Transformers

The new state of the art in supervised document classification draws its strength from several advances. The first is a revolutionary self-attention mechanism, known as 'transformer', which supports rich, contextualised representations of lexical and sentence meaning.[43] The second are new training methods. Models are trained to predict target words and whether two sentences appear next to each other. The third is greater leverage of transfer learning. Models are pre-trained, without human supervision, on vast repositories of texts. This knowledge can then be transferred to 'local' supervised tasks with additional fine-tuning steps.

These advances are embodied in BERT, the path-breaking natural language processing model developed by Google researchers.[44]

Based on a deep neural network architecture, BERT is able to focus attention on a given word in a sentence while simultaneously identifying the context of all the other words in relation to that word. The 'static', type-based, word embeddings discussed in the previous section represent a word as a vector of co-occurrences with cosine similarity scores reflecting co-occurrence frequencies. This permits static word embedding to handle synonymy (if, for instance, 'car' and 'vehicle' are used to mean the same thing they will have high cosine similarity score) but not polysemy or co-reference resolution (to determine what a pronoun refers to). The vector representing the word 'party', for instance, will not differentiate between party as in 'political party' and the party to a legal case. In a large and relatively diverse corpus, the vector is thus liable to assign high cosine similarity to words associated with both usages (e.g. 'political' and 'court'). By contrast, transformer models like BERT go beyond generalising across contexts. They represent words as dynamic, token-based vector embeddings, thereby coming much closer to capturing the particular, sentence-specific context of occurrence of a word. This, in turn, enables BERT to handle polysemy and co-reference resolution much better than previous language models.

The original BERT was trained on a giant corpus of GoogleBooks (800 million words) and Wikipedia pages (2.5 billion words) without human supervision by simply feeding it raw texts. Yet the power of BERT for supervised classification lies in the possibility to further fine-tune the pre-trained BERT on a 'local' data set. What has been learned from the giant corpus can thus be transferred to the local, smaller data set of direct interest to the researcher. Obviously, there are many linguistic patterns that no algorithm will be able to learn from a small data set. But a small data set may also instantiate specific patterns absent in the giant data set. In short, transfer learning helps combine the strengths of both data sets. Technically, local fine-tuning adds an additional layer of neurons to the neural network, thereby incorporating the local knowledge into the larger model.

BERT has been shown to outperform other algorithms on a wide range of natural language processing tasks.[45] One study has shown BERT to perform well at predicting the issue area codes of EU legislative acts.[46]

Table 4 reports the confusion matrix and performance metrics of a BERT model trained to predict whether EU legislative acts will be litigated. The data set was built by matching EU legislative acts in the EUR-Lex

40. M. Medvedeva, M. Vols and M. Wieling, 'Using Machine Learning to Predict Decisions of the European Court of Human Rights', 28 *Artificial Intelligence and Law* 237-66 (2020).
41. Hausladen, Schubert, and Ash, above n. 31.
42. Ranera, Solano, and Oco, above n. 30.
43. A. Vaswani et al., 'Attention is All You Need', in I. Guyon, U.V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan and R. Garnett, *Advances in Neural Information Processing Systems* 5998-6008 (2017).
44. J. Devlin et al., 'Bert: Pre-training of Deep Bidirectional Transformers for Language Understanding', ArXiv Prepr. ArXiv181004805 (2018).
45. *Id.*
46. I. Chalkidis et al., 'Large-scale Multi-label Text Classification on EU Legislation', ArXiv Prepr. ArXiv190602192 (2019).

Table 4 Confusion matrix and performance metrics of a BERT model trained to predict whether an EU legislative act will be litigated before the CJEU

		Predicted		
		Positive	**Negative**	**Total**
Actual	**Positive**	1,220	280	1,500
	Negative	664	4,836	5,500
Precision (positive): 0.6476				
Recall (positive): 0.1833				
F1 Score: 0.721				
Matthews correlation coefficient: 0.6408				

database to CJEU rulings. Only 3% of all EU legislative acts are ever litigated, and the probability that a given piece of legislation will be litigated in a particular case is very low.

When an outcome is a rare event, as with our example, it is important to think carefully about what the bar for a good model should be for the task at hand. Indeed, inexperienced lawyers and laypeople are often impressed by headline metrics like '90% accuracy'. However, achieving 90% correct classifications may, in many settings, be indicative of a poor performance. In fact, it all depends on the task and data set. With our EU litigation data set, it would have been easy to achieve 97% accuracy, since a model predicting that EU legislative acts are never litigated would be right 97% of the time. So here accuracy is a misleading metric, and precision and recall for the rare outcome provide a better gauge of performance.

Table 4 reports results for a subset of the data, where CJEU decisions featuring EU legislative acts have been deliberately oversampled. Oversampling the rare outcome is important to ensure that the algorithm has enough information to learn the patterns associated with this outcome.

While it is certainly possible to improve on these results through further local fine-tuning, a precision of 0.65 (i.e. out of 1,884 predicted to be litigated, 1,220 actually are) and a recall of 0.81 (i.e. out of 1,500 litigated, 1,220 were predicted to be so) are encouraging results.

Since the release of the first BERT, new variants of BERT have appeared, pre-trained on a wide range of general (RoBERTa) or domain-specific corpora (BioBERT, sciBERT…) in a variety of languages (e.g. robBert in Dutch, flauBERT in French, etc.). Multilingual BERT models, simultaneously pre-trained on multiple languages, have been shown to support transfer learning across languages. BERT models pre-trained on large collections of legal documents have also been released to assist with legal classification and prediction tasks.[47]

The arrival of BERT has triggered an AI race where research teams at big tech firms are vying to attain ever-higher performance with increasingly complex transformer language models: RoBERTa (Facebook), XLNET (Google), GPT-2 (OpenAI), Turing NLG (Microsoft) … The last such model to outperform its rivals, GPT-3 from OpenAI, boasts 175 billion parameters (by comparison, BERT has only 110 million parameters). The pace of technological development holds out great promise for the future of legal text-mining research and natural legal language understanding.

While transformers have just come along and applications to the legal domain are only starting to appear in publications and conference proceedings, an article by Evan Gretok, David Langerman and Wesley Oliver provides an interesting illustration of the application of transformer models to the study of legal doctrines. The authors trained transformer-based algorithms to classify rulings pertaining to the Fourth Amendment of the US Constitution depending on whether they applied a bright-line or a totality-of-the-circumstances rule. The best model (based on BERT) achieves an accuracy of 92%.[48]

As researchers begin to realise the potential of natural language processing for large-scale doctrinal analysis, we should expect to see many studies along these lines in the near future.

In the multilingual context of continental Europe, researchers may further seek to leverage the power of multilingual transformers to develop legal documents classifiers or predictors that can be deployed across multiple jurisdictions.

47. *Id.*

48. E. Gretok, D. Langerman and W.M. Oliver, 'Transformers for Classifying Fourth Amendment Elements and Factors Tests', *Legal Knowledge and Information Systems* JURIX 63-72 (2020).

6 Learning Text-Mining Methods

How can lawyers with no prior training in machine learning or data science get started?
One answer (at least for the motivated reader) is to learn a programming language like Python either by following one of the many free online tutorials or by taking a class at a nearby university campus. Python[49] is the language of choice for machine learning, text-mining and data harvesting tasks and the most popular among researchers and developers. Its ecosystem of libraries support the latest models and algorithms. While some lawyers may find the mention of 'programming' off-putting, Python is actually a very intuitive programming language. Moreover, the libraries provide many shortcuts that make it possible to complete a task with very few lines of code.
An alternative to Python is R, another popular programming language with many libraries designed for text-mining tasks, from data-harvesting to topic modelling and LSA/PCA. R was primarily developed for statistical analysis and does not support more advanced embeddings and transformer methods.
Both Python and R, along with their libraries, are entirely open source. They all can be downloaded and installed from the internet. The same goes for the pre-trained embeddings and transformers mentioned in this article (except for GPT-3).
Finally, for those who would prefer to avoid any kind of programming, RapidMiner comes with a graphical user interface to carry out end-to-end text-mining tasks without writing code.[50] Unlike Python and R, RapidMiner is a commercial platform. Yet its free version supports a wide range of supervised as well as unsupervised methods for data sets with up to 10,000 rows.

7 Conclusion

Text-mining and natural language understanding have been making great strides in recent years. Some of these techniques are at the heart of the hyper-hyped 'AI revolution' and are fuelling the development of legaltech.
To be sure, anyone who has actually attempted to use the techniques surveyed here will have realised that algorithms do not process language the way humans do. All techniques, even the most advanced ones, have limitations. Yet, thanks to their scalability, they open up a new possibility for legal research to explore and canvass vast repositories of legal documents.
There exist many variants of the techniques reviewed in this article and many more tasks to which they either already have been or may potentially be applied. However, I hope that the illustrations provided here and the techniques surveyed give the reader a sense of the potential that these techniques offer for academic legal research.

49. www.python.org (last visited 9 November 2020).
50. https://rapidminer.com (last visited 9 November 2020).

Big Data Ethics: A Life Cycle Perspective

Simon Vydra, Andrei Poama, Sarah Giest, Alex Ingrams & Bram Klievink*

Abstract

The adoption of big data analysis in the legal domain is a recent but growing trend that highlights ethical concerns not just with big data analysis, as such, but also with its deployment in the legal domain. This article systematically analyses five big data use cases from the legal domain utilising a pluralistic and pragmatic mode of ethical reasoning. In each case we analyse what happens with data from its creation to its eventual archival or deletion, for which we utilise the concept of 'data life cycle'. Despite the exploratory nature of this article and some limitations of our approach, the systematic summary we deliver depicts the five cases in detail, reinforces the idea that ethically significant issues exist across the entire big data life cycle, and facilitates understanding of how various ethical considerations interact with one another throughout the big data life cycle. Furthermore, owing to its pragmatic and pluralist nature, the approach is potentially useful for practitioners aiming to interrogate big data use cases.

Keywords: big data, big data analysis, data life cycle, ethics, AI

1 Introduction

The transformative potential of big data has attracted considerable academic attention in the last two decades, focusing not only on developments in the private sector, but also on big data's impact on various aspects of governance and public policy. As this special issue demonstrates, big data analytics are also becoming more widely used in the legal domain – a development that raises new ethical questions and concerns. In the legal context, key concerns have to do with judicial and legal principles and can be a lot more controversial than in contexts where model performance is the key (or perhaps the only) relevant metric.

This article explores how the use of big data analytics in the legal domain raises moral questions, looking closely at five illustrative cases and doing so in a structured systematic way. Our approach here is informed by the fact that many ethically significant decisions arise either before analysis, when collecting, storing, and aggregating data, or after analysis, when results are communicated, decisions are reached and lessons are learned. Thus, examining the morality of big data use in a systematic way requires that we include and consider the moral dimensions of all stages of the process, for which we utilise the 'data life cycle' concept. To interrogate the morality of big data use along the various stages of a data life cycle we adopt a 'lawyerly' mode of ethical reasoning that is ethically pluralistic and pragmatic in the sense of arriving at a convincing ethical argument for or against a given practice. The contribution of the article is thus twofold: First, it summarises five big data uses cases from the legal domain in a structured way, pointing to details that might not be obvious otherwise. Secondly, it proposes an approach of morally interrogating big data use cases that combines the logic of following the data's life cycle with a 'lawyerly' perspective, making it potentially valuable to those aiming to interrogate (and change) big data systems in practice.

This article is structured as follows. In Section 1, we articulate a framework for examining big data ethics as applied to the legal domain. We do this by first defining 'big data' (Section 1.1) and specifying the big data life cycle model on which we rely to articulate a systematic view of big data ethics (Section 1.2). Also, the article advances a working conception of the type of considerations deemed 'ethical' in the context of big data ethics (Section 1.3). Section 2, which makes up the bulk of the article, first introduces the five illustrative cases and then proceeds to briefly describe each case and illustrate its ethical significance for the stages of the big data life cycle. In Section 3 we bring together the insights gained through the different cases to offer a systematic overview of some of the key ethical concerns that might arise along the big data life cycle. Section 4 notes some limitations and concludes the discussion.

1.1 Big Data

The term 'big data' is conceptually fuzzy. The industry-standard definition of 'big data' uses a set of 'Vs': attributes of a data set that all begin with 'V', most commonly volume, variety, velocity and veracity.[1] In academic writing, volume, variety and velocity are used most commonly.[2] Some authors expand on this list by includ-

* Simon Vydra is a Researcher at the Institute for Public Administration, Leiden University, the Netherlands. Andrei Poama is Assistant Professor at the Institute for Public Administration, Leiden University, the Netherlands. Sarah Giest is Assistant Professor at the Institute for Public Administration, Leiden University, the Netherlands. Alex Ingrams is Assistant Professor at the Institute for Public Administration, Leiden University, the Netherlands. Bram Klievink is Professor of Digitization and Public Policy at the Institute for Public Administration, Leiden University, the Netherlands.

1. IBM, '4 Vs', www.ibmbigdatahub.com/tag/587 (2012); J.S. Ward and A. Barker, 'Undefined By Data: A Survey of Big Data Definitions', http://arxiv.org/abs/1309.5821 (2013).
2. O. Ylijoki and J. Porras, 'Perspectives to Definition of Big Data: A Mapping Study and Discussion', 4(1) *Journal of Innovation Management* 69, at 79 (2016).

ing veracity, variability, visualisation and value.[3] Defining big data using a set of 'Vs' allows for simple categorisation but lacks a threshold on these 'Vs' when 'data' become 'big data'. These thresholds are not only somewhat subjective, but also constantly changing as a result of technological advancements.

An alternative approach adopted in this article is to partially avoid these issues by focusing on the overall process and analytics necessary to use this data.[4] Using this approach enables us to focus on the processes linked to utilising data instead of focusing on differences in the data itself. For such a definition this article uses the work of Klievink et al. (2017), who distil a set of five criteria from the available literature:

1. Use and combining of multiple, large datasets, from various sources, both *external and internal* to the organization;
2. Use and combining of *structured* (traditional) and *less structured or unstructured* (nontraditional) data in analysis activities;
3. Use of incoming data streams in *real time* or near real time;
4. Development and application of *advanced analytics and algorithms*, distributed computing and/or advanced technology to handle very large and complex computing tasks;
5. *Innovative use* of existing datasets and/or data sources for new and radically different applications than the data were gathered for or spring from.[5]

This definition remains fuzzy because defining 'advanced analytics' or 'innovative use' remains subjective; however, it captures an important aspect of big data crucial for this article: the fact that big data is often 're-purposed' data that was not originally intended for the analysis it is being used for. It also aligns with the data life cycle perspective adopted by this article.

1.2 Big Data Life Cycle

The ambition of the data life cycle concept is to 'present a structure for organising the tasks and activities related to the management of data within a project or an organization'.[6] The concept is operationalised into various data life cycle models that cover the entire life of (big) data from generation to archiving/deletion and views the entire process as feeding into the next iteration of the same process, making it a cycle. What makes such a concept crucial for this article is that it 'provide[s] a structure for considering the many operations that will need to be performed on a data record throughout its life'.[7] Ethically significant decisions are not limited to the stage of generating insight and using it. Operations performed on/with data that precede and follow the decision-making step are no less ethically significant, and to conduct a thorough review of the ethical aspects of big data use every step of the cycle should be considered.[8]

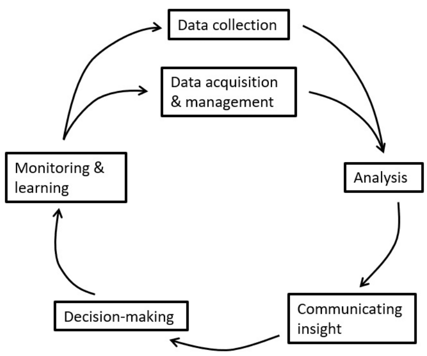
One particular challenge of a life cycle approach is that there is no unified (big) data life cycle model as data life cycles are very different per domain, field and even organisation.[9] Although there have been attempts to develop a scenario-agnostic data life cycle model with a broad application,[10] they do not capture big data use in the legal domain well enough. Approaches that adapt the data life cycle to big data make useful changes but are based on a 'Vs' definition of big data and are not specific to the legal domain.[11] This forces us to select a particular life cycle model, and in doing so we face an important trade-off between generality and complexity: the more complex a data life cycle model is, the better it describes an individual case but lacks generality for describing other big data use cases. This is important because in the legal domain different uses of data can correspond to different life cycles: data can just be archived for record-keeping, can be used for a one-off lawmaking decision or can be continuously processed in a decision-support system. When interrogating a singular big data use case it is, of course, reasonable to follow a data life cycle model that fits that case as much as possible – we recognise that generality is more a feature of social scientific inquiry than a feature of the legal process. In this article we aim to maintain a level of generality to illustrate the merit of our selected approach across multiple cases and to be able to articulate some more general conclusions about ethical concerns with big data use cases in the legal domain.

The data life cycle model we adopt is described in Figure 1, and it aims to be simple enough to generalise across many different use cases in the legal domain but also specific enough to capture meaningful and distinct 'stages'. In this data life cycle model we include six distinct stages: the collection of data, which can involve both actively seeking and storing information and more passive collection of information of no obvious analytical value at the time of collection. The acquisition of that data, which entails purchasing or otherwise obtaining data that is already collected by another actor to either

3. *Ibid.*
4. Approach that is arguably similar to that of G.H. Kim, S. Trimi, and J.H. Chung, 'Big-Data Applications in the Government Sector', 57(3) *Communications of the ACM* 78 (2014).
5. B. Klievink, B.J. Romijn, S. Cunningham, and H. de Bruijn, 'Big Data in the Public Sector: Uncertainties and Readiness', 19(2) *Information Systems Frontiers* 267 ,at 269 (2017).
6. L. Pouchard, 'Revisiting the Data Lifecycle with Big Data Curation', 10(2) *International Journal of Digital Curation* 176, at 180 (2016).
7. A. Ball, 'Review of Data Management Lifecycle Models', https://researchportal.bath.ac.uk/en/publications/review-of-data-management-lifecycle-models (2012), at 4.
8. J.M. Wing, 'The Data Life Cycle', 1(1) *Harvard Data Science Review* 1, at 4 (2019).
9. Ball, above n. 7; M. El Arass, I. Tikito, and N. Souissi, 'Data Lifecycles Analysis: Towards Intelligent Cycle', *Intelligent Systems and Computer Vision ISCV 2017* (2017); Pouchard, above n. 6; A. Sinaeepourfard, J. Garcia, X. Masip-Bruin, and E. Marín-Torder, 'Towards a Comprehensive Data LifeCycle Model for Big Data Environments', in *Proceedings of the 3rd IEEE/ACM International Conference on Big Data Computing, Applications and Technologies* (2016).
10. Sinaeepourfard et al., above n. 9.
11. Pouchard, above n. 6; Sinaeepourfard et al., above n. 9.

further utilise the data on its own or to join it with already collected/acquired data. The analysis of that data, which includes cleaning and processing of data to extract conclusions that are valuable for various types of decision-making. The communication of these conclusions, which involves the selection of 'useful' conclusions and techniques (such as data visualisation) aimed at reducing complexity and information overload associated with communicating these conclusions to (technically non-expert and time constrained) decision-makers. And, finally, monitoring and learning, which includes interrogating (internally or externally) the outcomes and taking corrective or optimising steps for the next iteration of the cycle.[12] Not all of these stages have to be carried out by a single actor (e.g. since big data is often considered to be repurposed, the data is often collected by a different actor) in order to justify the moral relevance of utilising the date.

Figure 1 *Big data life cycle*[13]



This structure not only allows us to conduct a structured review, but also outlines a potential starting point for practitioners and legal researchers to interrogate real-world big data use cases in terms of ethical concerns. As mentioned previously, for assessing individual cases the big data life cycle model can be much more specific, but even for other studies striving for a degree of generality the individual stages can (and should) be different from those we select for this article. In general, data life cycle models mention very similar stages but name them differently and attach them to or detach them from each other differently,[14] meaning that the version of it we adopt here is not the only justifiable one. This includes how decision-making phases or management activities are included in the model (if at all) – in our case those phases are included and considered important. Such decisions are made in an attempt to balance the generality and complexity of the resulting model, but decisions on this trade-off are inherently subjective and fit for purpose.

1.3 Ethical Significance

In each stage of the data life cycle we aim to describe relevant ethical considerations, requiring a definition of ethical considerations that distinguishes them from any other class of considerations. To do so, we propose and adopt a 'lawyerly' perspective of ethical reasoning. In doing this, we draw on an analogy philosophers sometimes use to distinguish between (currently two) different perspectives that might inform and explain how we reason ethically: the lawmaker's perspective and the judge's perspective.[15] The lawmaker's or politician's perspective focuses on devising laws and policies to increase their constituents' aggregate well-being. This is the practical perspective usually at work in consequentialist ethical theories: thinking like an ideal lawmaker means thinking about how to maximise a particular moral value – typically, utility, but also solidarity, community or care. The judge's perspective, on the other hand, focuses on solving a conflict in a specific case according to a fixed set of rules. This is the practical perspective usually at work in deontological ethical theories: thinking like an ideal judge means thinking in terms of respecting the constraints and prohibitions posited by a specific rule (or set of rules).

Drawing on this role analogy, we advance a third practical perspective: the lawyer's perspective. Unlike the lawmaker's or the judge's perspective, the lawyer's practical perspective focuses on convincingly contesting or defending an action or practice on the basis of a given ethical consideration. The focus is thus not on the structure or substance of the ethical consideration that guides the lawyer's contestation or defence, but rather on the consideration that the 'ethical case' they make is persuasive and sound. Construed from the lawyer's perspective, the point of ethical reasoning is to produce winning ethical arguments – i.e. arguments that have a concrete practical bearing and contribute to making a change in individuals' lives. The lawyerly mode of reasoning can thus be considered ethically pluralistic and opportunistic,[16] in the sense that it remains pragmatically open to a plurality of ethical considerations, and focuses on the values and principles that are most fitting for making a successful argument in a specific case. Although this is merely a cursory description of how the lawyerly mode of ethical reasoning might work, many lawyers might admit that partiality, adversariality and pragmatism are

12. More complexity can be added to this model by including tasks that iteratively happen in every stage, such as the 'describe' and 'assure' steps proposed by Pouchard, above n. 6 to guarantee data quality. Such additional steps support the overall ambition of this article to interrogate big data use at every stage.
13. Authors' own diagram.
14. M. El Arass & N. Souissi, 'Data Lifecycle: From Big Data to SmartData', 2018 *Colloquium in Information Science and Technology* 80, at 80-81 (2018).
15. R. Hare, *Moral Thinking: Its Levels, Method, and Point* (1981); J. Rawls, 'Two Concepts of Rules', 64(1) *The Philosophical Review* 1, at 6 (1955). Here Rawls famously suggests that 'different sorts of arguments are suited to different offices. One way of taking the differences between ethical theories is to regard them as accounts of the reasons expected in different offices'.
16. On moral opportunism, *see* K. Winston, 'Moral Opportunism: A Case Study', 40 *NOMOS: American Society for Political and Legal Philosophy* 154 (1998).

central to how they reason and argue qua lawyers.[17] It is also a mode of ethical reasoning compatible with moral psychology theories that argue that when supporting an ethical position, human reason is more analogous to a lawyer defending a partial position than to an impartial judge applying general rules (Haidt 2012).

Assuming this perspective, we require a definition of 'ethical concerns' that is not committed to a specific ethical theory or a moral outlook and one that is based on guidelines that are more likely to influence big data policies and use cases in the real world (as compared with ethical theories that do not go beyond the confines of academia). Consequently, we distinguish between ethical concerns and non-ethical concerns on the basis of widely shared legal and policy documents. Ethical considerations contained in those documents are considered 'ethical considerations' for the purposes of this article. This allows us, as our lawyerly perspective requires, to avoid stipulating or engaging in normative arguments about what ultimately counts as an ethically significant consideration. Our choice here is pragmatic, in that we defer to those organisations whose recommendations are either binding or widely accepted (in the European context) to determine what an ethically significant consideration is without endorsing any particular ethical viewpoint ourselves. This allows us to avoid the daunting task of finding a common denominator to radically opposed ethical theories and also to rely on a rough-and-ready 'consensus' and a 'practice-informed' account of the considerations that matter ethically.

This definition naturally begs two questions: which documents are considered and how are relevant ethical considerations extracted from them? In terms of the documents considered, we select for documents that are a) either legal prescriptions or government-endorsed recommendations that have practical traction in the sense that big data use case should be reasonably expected to take them into account b) applicable in the European context, and c) concerned either with general research conduct or with advanced analytical methods often used for big data analysis (such as artificial intelligence (AI) or machine learning). The documents meeting these criteria range from EU-level legislation such as the General Data Protection Regulation (GDPR) to more global recommendations endorsed by a majority of European countries (such as recommendations of the UN or the OECD). The sampling of these documents is purposive, providing a good breadth of various types of documents but not assuring that our selection of documents is exhaustive.

In terms of extracting ethical considerations from these documents, the method is conventional content analysis (in the sense that the coding scheme is inferred from the documents themselves) and is focused on explicit declarations of ethical principles within the selected documents. In other words, either the selected documents have a statement of ethical principles that guide the recommendations or the entire document is a list of ethical principles to be followed. Given that these are often listed as individual points or principles, their extraction from the text is very straightforward. Our treatment of the individual ethical considerations extracted from these documents is also rather simple: we list them in a table together with the documents that mention them and whether these documents are about general research conduct, big data use, or both (Annex A). In compiling this table our commitment to avoid ethical theorising means that we refrain from merging various concerns into overarching categories. This renders the table in Annex A rather long and filled with functional overlap between the individual ethical considerations. However, the overall approach is methodologically straightforward, easily adjustable to changes in ethical standards and pragmatic in multiple ways: The ethical concerns it works with are directly relatable to prescriptive guidelines and, when combined with a data life cycle approach, it outlines a great number of potential points of contention relevant for the 'ethical case' for or against a practice (some of which may be difficult to identify in a more holistic approach) and provides a basis for coherently contesting big data practices that are structurally similar.

17. A. Applbaum, *Ethics for Adversaries: The Morality of Roles in Public and Professional Life* (2000); D. Markovits, *A Modern Legal Ethics: Adversary Advocacy in a Democratic Age* (2010).

2 Cases

In the remainder of this article we draw on five different big data use cases from the legal domain. We select those cases primarily to cover a range of applications in the legal process, target users, target problems, system managers and countries of implementation. We summarise these relevant features of our cases in Table 1 for the five cases we select: A decision-support system for judges (COMPAS), a predictive policing system (CAS), a crowdsourcing system for lawmaking (vTaiwan) and two cases of welfare fraud detection systems (US welfare fraud detection and SyRI, which is a similar case from the Netherlands). We include the two fraud detection cases to also capture a degree of similarity and illustrate that even if two systems share similarities, they will not necessarily raise the same ethical considerations.

We now proceed case by case in the following five subsections (Sections 2.1-2.5), each offering a brief description of the case, our summary of what ethical concerns at what big data life cycle stages we can identify for that given case, and a narrative description of what makes these stages ethically consequential. That said, our ambition here is primarily exploratory, and our analysis, although systematic, is by no means exhaustive. We do not cover all six life-cycle stages for each case, primarily because we lack perfect information about those cases: for some life-cycle stages of some cases the available information is very scant, and arguing for specific ethical concerns for those stages would be over-interpreting

Table 1 Characteristics of selected cases

	COMPAS	CAS	vTaiwan	US welfare fraud detection	SyRI
Legal process	Adjudication	Enforcement	Lawmaking	Enforcement	Enforcement
Target users	Judges making sentencing and bail decisions	Police officers on patrol	Legislators proposing regulation	Institutions providing welfare benefits	Institutions providing welfare benefits
Target problem	Lack of information and potential bias of judges	'Excessive' occurrence of specific crimes	Lack of public participation in lawmaking	Welfare fraud	Welfare fraud
System managements and ownership	Privately owned (Equivant)	Dutch national police	Civic community of citizens (g0v - gov zero)	Multiple public institutions	the Ministry of Social Affairs and Employment
Country of implementation	United States of America	Netherlands	Taiwan	United States of America	Netherlands

the available information and engaging in arguments for ethical positions that can be subject to reasonable disagreement, which does not align with our normatively non-committal account of ethical concerns. Furthermore, we do not want to assume that there have to be ethical concerns at every stage of the data life cycle, but we also cannot claim an absence of ethical concerns simply for want of evidence. Which stages are addressed and which are omitted is not an a priori decision (all stages are considered for each case), but for some stages we cannot argue for an ethical concern (without engaging in hypotheticals). Given the exploratory nature of this article, we choose to omit from our analysis some of these stages for some cases rather than convey a false sense of exhaustiveness that cannot be supported with the available information. This can be problematic for arriving at generalised conclusions in an article such as this, but it is less problematic in practical application: Systematising ethical concerns in this way can be done continuously, and the assessment of various stages can be 'filled in' as sufficient information becomes available. In fact, this approach can aid in identifying knowledge gaps about a case that needs to be filled in order to conduct an exhaustive analysis.

2.1 COMPAS

COMPAS (Correctional Offender Management Profiling for Alternative Sanctions) is a need-assessment and risk-assessment tool used in the US as decision support for judges in bail and sentencing decisions. It provides an assessment of individuals' criminogenic needs that aims to aid with case planning and an assessment of risk,[18] the latter being the focus of this article. The risk assessment consists of three measures of risk calculated for a defendant – pre-trial release risk, general recidivism risk and violent recidivism risk. The purpose of these risk scores is to predict recidivism: 'The purpose of the risk scales is prediction – the ability to discriminate between offenders who will and will not recidivate.'[19] The models used to arrive at these scores are not publicly disclosed as COMPAS is a commercial product owned by Equivant (formerly known as Northpointe), and disclosing details of the system would be against their commercial interests.

The data used by COMPAS to generate the risk scores is collected by the institutions utilising it and combined with publicly available data. The data can be collected using a self-reported questionnaire or by conducting an interview during which answers to these questions are recorded.[20] This survey has 137 questions, whose answers are fed into COMPAS. Since the nature of the model utilised by COMPAS is unclear, many researchers have since studied the outcomes of COMPAS utilising data about more than 7,000 defendants in a county of Florida, including demographic details of defendants, the risk scores assigned to them by COMPAS and whether they eventually reoffend. In the case of these defendants COMPAS was approximately 65% accurate.[21] The scores themselves have been approximated by various surrogate models, but the same accuracy can also be achieved by models as simple as logistic regression utilising only two variables – age and number of prior offences.[22] Non-expert human annotators are slightly less accurate than COMPAS (62.8%) individually but more accurate than COMPAS when aggregat-

18. Equivant, 'Practitioner's Guide to COMPAS Core', www.equivant.com/practitioners-guide-to-compas-core (2019).

19. *Ibid.*, at 7.

20. Northpointe, 'COMPAS Risk & Need Assessment System Selected Questions Posed by Inquiring Agencies Ease of Use', www.northpointeinc.com/files/technical_documents/Selected_Compas_Questions_Posed_by_Inquiring_Agencies.pdf (2010).

21. J. Angwin, J. Larson, S. Mattu, and L. Kirchner, 'Machine Bias: There's Software Used Across the Country to Predict Future Criminals. And It's Biased Against Blacks', www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing (2016).

22. J. Dressel & H. Farid, 'The Accuracy, Fairness, and Limits of Predicting Recidivism', 4(1) *Science Advances* 1, at 3 (2018).

Table 2 Ethical concerns with COMPAS

Data collection	Data acquisition & management	Analysis	Communicating insight	Decision-making	Monitoring& learning
		Independence from bias, transparency, accountability	Independence from bias, fairness, causing adverse effects for individuals	Accountability, fairness, transparency, independence from bias, respect	Independence from bias, causing adverse effects for individuals, obeying the law

ing multiple annotators together (67%). In a review of multiple instruments predicting recidivism risk in the US (including COMPAS), the authors state that 'no one instrument stood out as producing more accurate assessments than the others, with validity varying with the indicator reported'.[23]
For this case we highlight ethical concerns in four stages of the data life cycle – analysis, communicating insight, decision-making, and monitoring and learning – focusing on 'independence from bias' at every stage and illustrating the various forms of this concern, especially as related to 'accountability' and 'transparency'. These stages and the concerns we address in them are summarised in Table 2.

2.1.1 Analysis

In terms of analysis, COMPAS has been critiqued as a racially biased tool and has been shown to exhibit racial bias in the errors of the model's predictions: 44.9% of blacks labelled as 'higher risk' did not actually reoffend compared with 23.5% whites. 28% of blacks labelled lower risk did reoffend, compared with 47.7% whites.[24] Even though a defendant's race is not a feature provided to the model, other features associated with race are enough for race to arguably constitute a latent feature. This can be considered ethically significant in and of itself for the model's **independence from bias**, but the question of racial bias in COMPAS is not as straightforward and points to another ethically significant aspect – who interprets issues of justice and fairness and how it gets done: the allegation of racial bias itself is not the focus here as Equivant[25] as well as academic researchers[26] have issued rebuttals exposing serious methodological errors in the critique. What we focus on here is the ensuing debate, which, despite its largely technical character, exposed a fundamental disagreement about the meaning of 'fairness' and how it can be operationalised mathematically. Fairness can refer to accurate calibration between groups (a risk score translates to identical recidivism rate across population subgroups) or to a correct balancing of the negative and the positive classes (average risk scores for reoffenders are identical across population subgroups).[27] In other words '[t]here is no single mathematical definition of fairness. The people developing a "fair" algorithm must decide on the uniformity or variation that is necessary for a functioning system'.[28] Furthermore, this issue cannot be resolved by adjusting the algorithm, as the definition of fairness adopted by the critics and the one adopted by Equivant cannot mathematically be satisfied simultaneously unless our predictions are flawless or the base-rate of the predicted variable (reoffending) is identical for different population subgroups.[29] This means that in designing such a system one has to make the choice about what 'fairness' means (mathematically), which in this case is a decision made by technical experts without any political **accountability**, one hidden in a completely **non-transparent** algorithmic 'black box' and one that is of crucial ethical significance.

2.1.2 Communicating Insight

There are potentially ethically significant features of how COMPAS scores get communicated to judges (and other stakeholders in the legal process). Both recidivism risk scores are communicated as the score itself accompanied by a label of low risk (1-4), medium risk (5-7) or high risk (8-10).[30] The scores themselves are interpretable only with reference to a norm group and represent a specific decile of the scores of everyone in the norm group ranked in ascending order (e.g. score 1 refers to the least likely to recidivate 10% in the norm group). This means that individuals can be assigned a high risk score but not actually be highly likely to reoffend (if their norm group is generally unlikely to recidivate) and

23. S.L. Desmarais, K.L. Johnson, and J.P. Singh, 'Performance of Recidivism Risk Assessment Instruments in U.S. Correctional Settings', 13(3) *Psychological Services* 206, at 213 (2016).
24. Angwin et al., above n. 21.
25. W. Dieterich, C. Mendoza, and M.T. Brennan, 'COMPAS Risk Scales: Demonstrating Accuracy Equity and Predictive Parity', http://go.volarisgroup.com/rs/430-MBX-989/images/ProPublica_Commentary_Final_070616.pdf (2016).
26. A. Flores, K. Bechtel, and C. Lowenkamp, 'False Positives, False Negatives, and False Analyses: A Rejoinder to "Machine Bias: There's Software Used Across the Country to Predict Future Criminals. And It's Biased Against Blacks"', 80(2) *Federal Probation* 38 (2016).
27. A. Chouldechova, 'Fair Prediction with Disparate Impact: A Study of Bias in Recidivism Prediction Instruments', 5(2) *Big Data* 1, at 2-3 (2017); J. Kleinberg, S. Mullainathan, & M. Raghavan, 'Inherent Trade-Offs in the Fair Determination of Risk Scores', https://arxiv.org/pdf/1609.05807.pdf (2016).
28. A.L. Washington, 'How to Argue with an Algorithm: Lessons from the COMPAS ProPublica Debate', 17(1) *The Colorado Technology Law Journal* 131, at 151 (2019).
29. Kleinberg et al., above n. 27.
30. Equivant, above n. 18.

vice versa.[31] The ethical significance of this is dependent on judges' understanding of how to accurately interpret these scores, but the fact that a norm group conditions the actual score a defendant receives is a potential issue for **independence from bias** and **fairness**. Furthermore, the fact that these scores are interpretable as decile scores brings into question the labelling of low, medium or high risk since judges themselves can make the decision on what particular part of the distribution of risk scores is 'high' or 'low' risk for a particular norm group and case. The system providing this label could be a source of bias in and of itself.

Another issue is communicating when exactly a risk assessment tool such as COMPAS should be utilised in the legal process. The tool itself as well as general guidelines for using risk assessments state that risk assessment 'should not be used as an aggravating or mitigating factor in determining the severity of an offender's sanction'.[32] However, the original critique levied against COMPAS mentions cases where the risk score has seemingly influenced severity of punishment when reviewed by the judge.[33] If the interpretation and use of these risk scores are not fully understood by judges it can amount to causing **adverse unjustified effects to individuals**.

2.1.3 Decision-making

COMPAS is ethically significant in three distinct ways at the decision-making stage. First, by using COMPAS the assumptions about fairness (addressed in Section 2.1.1) are made part of the legal process. Judges are the ones with the authority to interpret laws (and the conception of 'fairness' and 'justice' they capture), which appears paradoxical since they are precisely the users of the system and (knowingly or not) adopt assumptions about **fairness** made by technical experts. It is true that human judgment is often flawed and suffers from the same racial bias that COMPAS is criticised for.[34] But even if COMPAS could in some ways be less biased than judges it obscures the **accountability** for this bias: the decisions judges make and the reasoning underlying them are generally a matter of public record, and any potential bias in their decisions can be scrutinised, and they can ultimately be held accountable for it. COMPAS removes a portion of this responsibility by providing an 'impartial' and 'technical' tool whose bias is much more difficult to interrogate as the algorithm itself is a trade secret (and thus not **transparent**).

Secondly, COMPAS scores are not the only piece of information judges consider when making a decision. This is related to the issue of inappropriate interpretation or overusing the tool itself (as mentioned in Section 2.1.2), but also to a more complex interaction between a risk score and other information a judge considers in a decision. For example, the socio-economic status of an individual is another important factor and one that is associated positively with risk of recidivism but negatively with the blameworthiness of an individual for the crime they have already committed. As such, providing a judge with risk assessment information can reduce the likelihood of incarceration for relatively wealthy individuals and increase this risk for relatively poor individuals (the information being identical).[35] The very inclusion of a risk assessment score can thus violate **fairness** principles and **introduce bias** to the decision-making process independently of any bias of the risk score itself.

Thirdly, introducing a tool like COMPAS into deciding individual legal cases needs to be reconciled with the individualistic nature of the legal process: any algorithm basing predictions on existing data judges individual behaviour on the basis of group characteristics. In the case of COMPAS, '[t]he moral issues involve political unease when decisions are based on immutable characteristics over which individuals have no personal control or that may serve directly or by proxy to replicate discriminatory practices'.[36] Some of the 137 features derived from a compass questionnaire are not problematic in this respect, but some are (directly or by proxy) about individuals' immutable characteristics or about their environment (e.g. criminal history of their friends and family). In general, the 'use of group tendencies as a proxy for individual characteristics'[37] is rejected in multiple pieces of US case law,[38] and the moral implications of accepting algorithmic output relying on precisely this type of inference are significant in terms of **independence from bias** and **respect** for individuals.

2.1.4 Monitoring and Learning

In terms of monitoring and learning the case of COMPAS is complicated by the lack of transparency of its inner workings, making the inspection of the algorithm itself impossible. However, the use of COMPAS can still be evaluated on the basis of its predictive outcomes and adherence to legal principles. In terms of legal principles, COMPAS has actually been legally challenged in *State vs. Loomis,* a case that ultimately reached the Wisconsin Supreme Court. In this case the defendant argued that the use of COMPAS in a sentencing decision **violates two of his legal rights** that are also ethically significant: the right to due process and the right to individualised sentence.[39] Since COMPAS is a trade secret its output cannot be scrutinised by the defence (which is a part of due process) but undeniably plays a

31. *Ibid.*
32. J. Elek, R. Warren, & P. Casey, 'Using Risk and Needs Assessment Information at Sentencing: Observations from Ten Jurisdictions', nicic.gov/using-risk-and-needs-assessment-information-sentencing-observations-ten-jurisdictions (2015), at 5.
33. Angwin et al., above n. 21.
34. Dressel & Farid, above n. 22.
35. J. Skeem, N. Scurich, and J. Monahan, 'Impact of Risk Assessment on Judges' Fairness in Sentencing Relatively Poor Defendants', 44(1) *Law and Human Behavior* 51 (2020).
36. M. Hamilton, 'Risk-Needs Assessment: Constitutional and Ethical Challenges', 52 *American Criminal Law Review* 231, at 242 (2014).
37. S.B. Starr, 'Evidence-Based Sentencing and the Scientific Rationalization of Discrimination', 66(4) *Stanford Law Review* 803, at 827 (2014).
38. *Ibid.*
39. *Case 881 N.W.2d State* v. *Loomis*, www.courts.ca.gov/documents/BTB24-2L-3.pdf (2016).

role in sentencing decisions. COMPAS also partially bases its output on aggregate data about recidivism for groups similar to the defendant but not on data for the defendant themselves (including variables like gender), potentially not delivering an individualised sentence, which would make it **biased** and cause **unjustified effects for individual defendants**. The claim of violation of due process was rejected by the court, and in commenting on the right to individualised sentence the Court upheld the use of COMPAS in sentencing decisions because it is not the determining factor for a sentence and 'is helpful in providing the sentencing court with as much information as possible in order to arrive at an individualised sentence'.[40]

In terms of monitoring the predictive outcomes of COMPAS, it is certainly possible (as evidenced by the criticism of its racial bias and the ensuing response), but there is no evidence indicating that such monitoring, which is identified as one of the guiding principles for using risk assessment tools,[41] is being done by the institutions utilising COMPAS. In terms of learning and adjustment the Wisconsin Supreme Court issued a cautionary statement to lower courts to be aware of the limitations of the COMPAS tool and the fact that it predicts group behaviour and not individual behaviour, meaning that judges need to explain factors other than the risk score that ultimately determine their decision in order to avoid undue bias and unfairness. In sum, the monitoring of the performance and legality of COMPAS was somewhat extensive, but not initiated by the institutions utilising it, and the adjustment stemming from this monitoring is very limited.

2.2 CAS

CAS (Crime Anticipation System) is a data-mining software used by most of the police forces in the Amsterdam area, the Netherlands. The software was piloted in 2014 and is currently managed by the Dutch National Police. The officially stated aim of the software is to predict the location and time of 'high impact crimes' (HIC). HIC are narrowly defined to include four offence categories, namely robbery, nuisance by youth, street robbery and bicycle and scooter theft. The software takes the form of a map where different crime categories appear as coloured squares (red for burglary, green for street robbery, blue for youth nuisance and pink for bicycle and scooter theft). Brighter intensities of the same colour indicate higher risk levels for incidents within each category (*e.g.* brighter red means a higher risk of burglary in the designated area), and each square corresponds to an area of 125 × 125 metres within the city. The combined data sources provide 78 data points for each of these squares, and the city is divided into a total of 11,500 squares. The model itself relies on a neural network, i.e. an algorithm that gradually learns to recognise and update its recognition of patterns in the data that it receives.[42]

Crime risk levels are calculated on the basis of three data sources linked together in this system. The first database is provided by the BVI (*Central Crime Database*) and includes the distance to the address of the suspect of an incident registered within the previous 6 months, the number of suspects of incidents registered within the previous 6 months that live within 500 metres, and the number of suspects of incidents registered within the previous 6 months that live within 1 kilometre from that area. The second database is provided by the CBS (*Central Statistics Office*) and currently includes 15 indicators such as the number of inhabitants, their gender, the number and average size of households, the average property and average income level, as well as the number of social benefits recipients within an area.[43] The data provided by the CBS used to include an indicator eliminated in 2017, namely the number of 'non-Western allochthones' living within an area. The third database is the BRP (*Municipal Administration*), which is used to identify streets and specific addresses (for instance, the address of a shop that might be the target of a burglary).

For this case we highlight ethical concerns in four stages of the data life cycle – data collection, data acquisition and management, communicating insight and decision-making. The key concerns appearing throughout these stages are related to 'independence from bias' and 'causing adverse effects for groups', but they do vary from stage to stage and connect to other ethical principles like 'fairness', 'respect' or 'transparency' and 'proportionality'. These stages and the concerns we address in them are summarised in Table 3.

2.2.1 Data Collection

At the stage of data collection, CAS relies on data from multiple agencies, the most ethically significant being data from the BVI. This data relies exclusively on information about the addresses of individuals who are considered crime suspects, raising a series of moral concerns. First, there are reasons to think that the list of 'suspects' might be overinclusive (and thus unreliable) because of the ethnic and classist **biases** that influence who tends to be identified as a suspect for any specific crime incident.[44] The influence of these biases might be accentuated by the fact that identifying someone as a suspect requires a relatively low evidentiary threshold and, as a result, meets little to no epistemic constraints.[45] The influence of these biases can also be com-

40. *Ibid.*, at 764.
41. Elek et al., above n. 32.
42. P. Mutsaers & N. Tom, 'Predictively Policed: The Dutch CAS Case and Its Forerunners', www.researchgate.net/publication/346593158_Predictively_policed_The_Dutch_CAS_case_and_its_forerunners/citation/download (2020).
43. S. Oosterloo & G. van Schie, 'The Politics and Biases of the "Crime Anticipation System" of the Dutch Police', in *Proceedings of the International Workshop on Bias in Information, Algorithms, and Systems 2103* (2018).
44. S. Çankaya, *De controle van marsmannetjes en ander schorriemorrie: het beslissingsproces tijdens proactief politiewerk* (2012).
45. A. Das & M. Schuilenburg, '"Garbage In, Garbage Out": Over Predictive Policing and Vuile Data', 47(3) *Beleid en Maatschappij* 254 (2020).

Table 3 Ethical concerns with CAS

Data collection	Data acquisition & management	Analysis	Communicating insight	Decision-making	Monitoring & learning
Independence from bias, causing adverse effects for groups, fairness, respect, honesty, integrity	Consent, fairness, causing adverse effects for groups, respect, transparency, data minimisation		Transparency, explainability	Causing adverse effects for groups, fairness, proportionality, respect, professionalism	

pounded (and arguably entrenched) further down the road: because of the socio-economic biases that might be incorporated into CAS *via* BVI data, police officers who follow the software's advice might gradually be brought to over-control certain areas and categories of the population and under-control others and thus gradually form or confirm a distorted image about typical offence suspects **causing adverse effects for those groups**. Furthermore, when such stereotypes about suspects influence the outputs of the system, the moral costs that come with data collection – for instance, being surveilled, stopped and interrogated – are also spread in unequal ways, which is ethically significant with regard to **fairness** and causing adverse effects to groups. Furthermore, whenever stereotypes work implicitly, the unfairly distributed costs of coping with police interference remain largely hidden in the inner workings of an algorithm.

Secondly, there is a more diffuse moral concern about the type of information that is deemed relevant for predicting future offences. By focusing on persons who are suspects in past crimes as proxies for the kinds of persons who might commit similar crimes in the future, CAS might be perceived as promoting that idea that 'no one is a suspect innocently', and thus undermine the 'innocent until proven guilty' rule that is constitutive of fair criminal justice practices. Relatedly, it might promote an objectionably stigmatising image of those thus selected as unredeemable 'criminals' or 'villains'. This militates against basic respect and equal treatment norms that should inform both police activities and the research practices on which these activities rely.

Thirdly, how CAS is presented is arguably at odds with the principles of **honesty and integrity,** given that there is a mismatch between its public image as a fine-grained predictive tool and the accuracy of the data it works with. For instance, since many burglaries and thefts happen in the absence of the victim, police officers cannot estimate the exact moment when they were committed. To deal with this problem, police officers usually choose a mid-point between the moment the victim left the house or parked the stolen bicycle and the moment when the burglary or theft was noticed.[46] The inaccurate nature of these estimates might mean that CAS is ultimately a very rough tool when it comes to calculating the timing of certain offences. Presenting it as a fine-grained tool might be empirically dishonest, at least until its effectiveness can be transparently shown.

2.2.2 *Data Acquisition and Management*

The data acquisition practices involved in setting up and using CAS are ethically significant in at least four distinct ways. First, it is not clear whether the acquisition of information from the CBS – in particular, data that makes it possible to geographically locate individuals living in non-traditional family settings or who are social benefits recipients – is submitted to the **consent** or oversight of the relevant human rights protection organisations.

Second, when CAS acquires data about the spatial distribution of socio-economic disadvantage as well as about the location of 'non-Western allochthones' (a practice that was terminated in 2017), for the specific goal of predicting crime, there is a risk of **unfairly** reinforcing existing stigma. The ethical significance of including this data is emphasised because of the absence of conclusive evidence that socio-economic disadvantage or ethnic difference are causally linked to higher crime rates.[47] The very inclusion of such data shows a willingness to single out individuals with a specific social and ethnic background as potential criminals. This can constitute a violation of basic norms of **respect** and principles of equal treatment and not **causing unjustified adverse effects for groups or individuals**.

Third, the Dutch National Police provide no publicly accessible information about the list of indicators and data they acquire from other sources, such as the CBS or BRP. This is arguably an infringement of **transparency**, as it leaves citizens in the dark about the considerations that guide police surveillance and other forms of interference that affect their and their co-citizens' lives. One reply here might be that the precise indicators and data included in CAS cannot be publicly advertised because doing so would affect its predictive effectiveness – for instance, by allowing some future offenders to foresee where and when police forces will be deployed. This rejoinder, however, does little to alleviate transparency concerns, especially as we currently lack evidence about the effectiveness of CAS.

46. Oosterloo & van Schie, above n.43.

47. T. Newburn, 'Social Disadvantage: Crime and Punishment', in D. Hartley and L. Platt (eds.), *Social Advantage and Disadvantage* (2016).

Fourth, CAS raises concerns about **data minimisation** requirements: when the decision was made in 2017 to exclude information about the number of 'non-Western allochthones' living within an area, the administrators of the software argued that the variable did not add to its predictive power.[48] This means that, for approximately 3 years following its introduction, CAS was substantively violating **data minimisation** requirements. This decision raises a more general question about the extent to which the specific indicators included in CAS are needed to ensure that it performs well in predicting the timing and location of crime incidents.

2.2.3 Communicating Insight

In terms of communicating insight, CAS is a closed system,[49] which means that users have access to information only about the risk levels of a particular crime category in a given area based on a fixed number of variables. Consequently, police officers do not have the option of zooming in on an area that is designated as high risk for a crime category (e.g. street robbery) to get more information that would allow them to make more empirically informed hypotheses about the context or seriousness of the crime risk level going upwards in that particular area within a specific time interval. This limitation of CAS can be ethically significant for the principle of explainability, if and insofar as this principle applied reflexively to the police officers themselves, and not only (as is usually the case) to the citizens who are policed. In using CAS, police officers might not be able to understand why a particular output was reached by its underlying algorithm, and, by way of consequence, they might not be able to explain why the output was reached to those they police. Also, the fact that CAS remains insensitive to whether an increase in crime risks is driven by any particular stable variable included in the software or by more conjunctural events (e.g. street parties) is significant for **transparency** as it does not allow users to properly grasp why crime patterns are changing in any particular area at a given moment in time.

2.2.4 Decision-making

CAS raises at least two distinct moral problems when it comes to policing decisions that are based on it. First, and insofar as it relies on 'dirty data'[50] that carries forward patterns of discrimination and disadvantage, CAS-based policing could contribute to compounding or entrenching the disproportionate amount of surveillance and interference that some neighbourhoods and categories of the population are submitted to, and thus **causing unjustified negative effects for groups**. This could violate both **fairness** principles (by upsetting the fair distribution of social burdens across persons and groups within the general population) and **proportionality** principles (with the benefits generated by the application of CAS largely unknown, it is difficult to determine whether the risks that come with over-policing are justified). Second, as police officers have noted themselves,[51] focusing too much on the advice given by CAS reduces policing to prevention and thus diverts officers from other obligations they are expected to tend to, such as assisting people with the coordination of various social activities or establishing a rapport with the inhabitants of any particular neighbourhood. At this level, the focus that CAS puts on prediction and prevention might be in tension with the **professionalism and respect** that citizens can also reasonably expect from their police officers.

2.3 vTaiwan

vTaiwan (v stands for vision, voice, vote and virtual) is a legislative crowdsourcing system used since 2015 by the Taiwanese government to give citizens a way to propose and debate new laws with the output of this discussion ideally influencing future legislation. Conceptually, legislative or policy crowdsourcing 'involves giving ordinary citizens, rather than political and bureaucratic elites, the chance to cooperate to come up with innovative new policies'.[52] It can thus be used for both policymaking and statutory lawmaking. In the latter form, it involves collaborative lawmaking between official lawmakers and networks of citizens and civil society organisations that aims to build the quality of legislative documents and increase political legitimacy of new legislation.[53] It also involves a new kind of role for citizens in the legislative process, moving from top-down models of legislative development to approaches that address information asymmetries between professionals and consumers or citizens, giving the latter more influence.[54] The data for the vTaiwan system is contributed by citizens in a range of different forms such as social media comments, discussion forums or online petitions. These contributions are then analysed using big data analysis instead of the traditional approach of being comprehended only through time-intensive human reading of texts.

Many countries use online platforms as a resource for members of the public to track laws proposed by parliament and make comments about them (e.g. regulations.gov in the US or Avoin Ministerio in Finland), and there have also been notable crowdsourcing approaches to specific legislative initiatives such as the

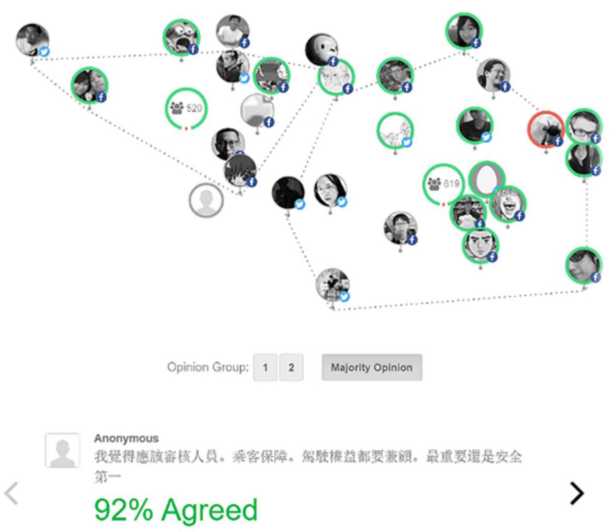
48. Oosterloo & van Schie, above n. 43.
49. *Ibid.*
50. Das & Schuilenburg, above n. 45; P. Mutsaers, 'A Public Anthropology of Policing. Law Enforcement and Migrants in the Netherlands' (dissertation at University of Tilburg) (2013).
51. A. Drenth & R. van Steden, 'Ervaringen van straatagenten met het Criminaliteits Anticipatie Systeem', 79(3) *Het tijdschrift voor de Politie* 6 (2017).
52. H.S. Christensen, M. Karjalainen, and L. Nurminen, 'Does Crowdsourcing Legislation Increase Political Legitimacy? The Case of Avoin Ministeriö in Finland', 7(1) *Policy & Internet* 25 (2015).
53. V. Burov, E. Patarakin, and B. Yarmakhov, 'A Crowdsourcing Model for Public Consultations on Draft Laws', in *Proceedings of the 6th International Conference on Theory and Practice of Electronic Governance* (2012).
54. T. Heikka, 'The Rise of the Mediating Citizen: Time, Space, and Citizenship in the Crowdsourcing of Finnish Legislation', 7(3) *Policy & Internet* 286, at 287 (2015); S. Ranchordás and W. Voermans, 'Crowdsourcing Legislation: New Ways of Engaging the Public', 5(1) *The Theory and Practice of Legislation* (2017).

Table 4 Ethical concerns with vTaiwan

Data collection	Data acquisition & management	Analysis	Communicating insight	Decision-making	Monitoring & learning
		Respect, fairness, independence from bias, transparency, explainability	Honesty, independence from bias	Accountability, honesty, respect, principled performance	

Internet Bill of Rights in Brazil[55] or the new constitutional processes in Ireland and Iceland.[56] vTaiwan is a notable case because of the original way that it analyses and openly presents digital information about its users in real time to facilitate deliberation. It has four main stages in the development of legislation: proposal, opinion, reflection and approval. In the opinion stage stakeholders are identified, and more stakeholders are included using a rolling survey, followed by gathering and analysing a large number of public opinions with the goal of distinguishing important clusters of topics that can be visualised in the form of network diagrams. These models have the explicit purpose of facilitating consensus: users can up-vote or down-vote suggestions (but cannot comment on them) or issue their own suggestions, and the algorithm automatically separates those suggestions into 'opinion groups' and de-emphasises the areas of disagreement between them. This system results in people looking for consensus across various 'opinion groups' and creating new suggestions that even people from disparate 'opinion groups' will up-vote. An example of such a network diagram is provided in Figure 2. Once a satisfactory level of consensus has been reached a given round of opinion mining is concluded. The algorithm itself utilises the pol.is open-source system, which relies mainly on principal components analysis (PCA) and k-means clustering to obtain and visualise 'opinion groups'.[57]

Figure 2 Example of opinion groups in a network diagram[58]



In this case we focus on ethical considerations in three stages of the data life cycle – analysis, communicating insight, and decision-making – mainly highlighting the role of 'honest' and 'fair' summarisation of citizens' opinions but also touching on how this relates to 'independence from bias', 'explainability' or 'principled performance' of the system itself. These stages and the concerns we address in them are summarised in Table 4.

2.3.1 Analysis

In the analysis stage vTaiwan is ethically significant in terms of respect, fairness and the potential for containing biases. The analysis of opinion data is predicated on the programmers' understanding of how 'opinion groups' should be constructed and what consensus (or lack thereof) looks like between these groups. The task of accurately modelling a corpus of texts is in and of itself reliant on subjective assumptions about what content is relevant and how it should be described, which is in this case amplified by reducing this information down to a two-dimensional space to be able to visualise the clustering that defines 'opinion groups'. The two principal components defining this space do not have an

55. D. Arnaudo, 'Computational Propaganda in Brazil: Social Bots During Elections', *University of Oxford Working Paper* 8 (2017).
56. S. Suteu, 'Constitutional Conventions in the Digital Era: Lessons from Iceland and Ireland', 38 *Boston College International and Comparative Law Review* 251 (2015).
57. Y.T. Hsiao, S.Y. Lin, A. Tang, D. Narayanan, and C. Sarahe, 'vTaiwan: An Empirical Study of Open Consultation Process in Taiwan', https://doi.org/10.31235/osf.io/xyhft (2018).
58. Figure reproduced from NESTA, 'vTaiwan' www.nesta.org.uk/feature/six-pioneers-digital-democracy/vtaiwan/ (last visited 16 November 2020).

58. Figure reproduced from NESTA, 'vTaiwan' www.nesta.org.uk/feature/six-pioneers-digital-democracy/vtaiwan/ (last visited 16 November 2020).

inherent meaning and can be constructed in various ways, none of which are strictly 'wrong' or 'right'. This means that how opinions are grouped, what is considered salient for legislative formation and what is consensus between those groups can be expressed in multiple ways. By selecting one of those ways some opinions inevitably get downplayed, others get up-played, and some opinions that do not fit well into large 'opinion groups' might effectively be silenced by being aggregated into these groups (de-emphasising their uniqueness). Summarising opinions in this way can be violative of norms of **respect** for individuals (and their opinion) and **fairness**, and can also introduce a **bias** in terms of what gets highlighted and what gets lost.

The concern with bias is also relevant because crowdsourced legislative commentary is highly diverse in terms of the kinds of populations that may be involved. Crowdsourcing social media data for legislative development can sometimes circumvent this problem if the data is representative of a population at large. However, the problem is most acute if the analysis is of crowdsourced commentary that may be contributed disproportionately by specific interest or demographic groups. This **bias** emerges from the moment the data is collected, but it also affects the roles that different kinds of citizens play in the monitoring of analysis. Even in a technologically advanced country such as Taiwan, digital skills and access inequalities exist among population subgroups such as the elderly or less wealthy. Achieving fairness in such circumstances is vital, but more than being a pervading ethical principle, it must also have safeguards provided by measures to ensure transparency and explainability of algorithms used in the analysis as well as information about the representativeness of those who contribute their opinions. vTaiwan does particularly well with regard to **transparency** principles as it is based on several open source platforms, but **explainability** is much more challenging as not only is computer code understood only by a small section of the population but the decisions in model specifications need to be interrogated in terms of their impact and justified more than simply being made transparent.

2.3.2 Communicating Insight

In terms of communicating insight, many of the concerns previously addressed apply here as well. In some ways, the analysis in this case is, at its core, a communication exercise: the aim of the model itself is to provide a summary of the crowdsourced opinions in a way that is understandable and conducive to consensus seeking. This is particularly problematic given how the delivery of crowdsourced legislation attenuates conflicting interests of analysts and politicians where the former are focused on the technical quality of analysis and the latter are tasked with turning the results of the analysis into actionable results with political consequences. In this case vTaiwan has not yet been used for major legislation that would make those types of challenges sufficiently apparent, but even if its performance is good the ethical significance of visualising an extremely multifaceted data (written opinions) in a two-dimensional space without introducing **bias** remains.

The communicating insight phase is crucially important for this case because democratic systems that rely on citizen input are implicitly (and often explicitly) responsible for reporting back to citizens on what resulted from their contributions. In this respect, communicating insights should, in its general form and processes, be an honest and accurate interpretation of the crowdsourced data, which is very difficult to assess in this case as different analysts and politicians might genuinely see different summaries of the texts as 'accurate'. The **honesty** concern here is particularly pertinent, since the platform aims to de-emphasise disagreement and foster consensus, arguably not summarising the corpus of texts in a fully 'honest' way. The attempt to communicate with citizens in this way should also show democratic values such as equality and understanding of the contextual situation of citizens by making the insights understandable to citizens of different education and technical levels. For citizens with more technical skills, open access to the data and programming steps needed to reproduce the analysis are essential for supporting principles of equality as well as transparency and accountability.

2.3.3 Decision-making

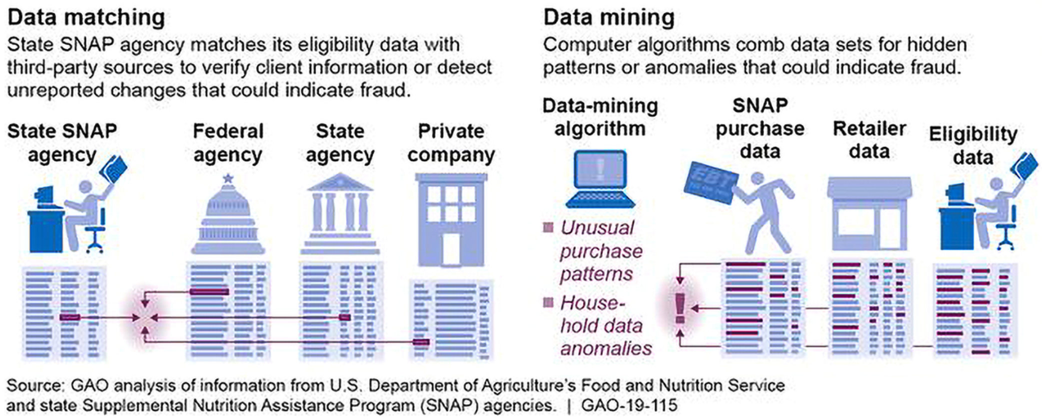
The decision-making stage of this case is particularly contentious as further political steps such as parliamentary debate and voting are required for citizens' opinions to have any effect. This stage actually substantially affects the ethical concerns that are relevant in previous stages: even though the system is generally transparent and open to the public, the government does not have to follow its outputs, making its **principled performance** uncertain. This conjures concerns of 'open washing' by having a system that is transparent and provides the government with legitimacy in decision-making but one that can ultimately be easily ignored when it comes to legislation.[59] This is a concern relevant to **respect** for citizens' opinions and **honesty** but also one that impacts the entire life cycle – if there is no **accountability** of decision makers for simply dismissing the output of this system, the system should be viewed in a different light.

Depending on where the decision-making process falls on the technical-political spectrum, the ethical principles will be different. For technical decisions, transparency of the technical decision-making processes involved in turning raw data into new legislative changes is of primary importance. For political decisions, legal and constitutional compliance in the ways that the political decision-making process are followed are of primary importance. Both technical and political decisions share the ethical problem of respecting citizen rights and dignity and protecting citizens from adverse effects that may result from changes to the law.

59. *Ibid.*

doi: 10.5553/ELR.000190 - ELR 2021 | No. 1

*Figure 3 Data matching and mining**



* Figure retrieved from GAO, above n. 60, at 1.

2.4 US Welfare Fraud Detection

Data surveillance to detect various types of fraud in US welfare programmes is an overarching system of data collection, management, analysis, and decision-making with the purpose of discovering fraud in utilisation of welfare programmes. Exact algorithms to identify fraudulent behaviour vary and are not fully transparent, but they are generally scouring large linked databases to identify patterns indicating potential fraud. These databases are constructed by mining and matching data across program-specific databases such as the Supplemental Nutrition Assistance Program (SNAP), federal and state agencies as well as private company data. This cross-referencing is particularly relevant to the SNAP programme, because it allows recipients to spend their benefits outside their state of residence.[60] Figure 3 demonstrates the extent of the database linking efforts spanning state agencies, federal agencies as well as private companies. Most of this data collection is done at the state level, and states have discretion in how to handle the information. In addition, 'current legal frameworks offer little protection for privacy-related harms experienced by the poor', giving additional leeway to the government for utilising such data.[61]

Figure 3 also touches on the fact that in the era of big data these efforts have gained even more momentum. One example is the replacement of food stamps with the Electronic Benefit Transfer (EBT) Card. The prepaid debit card provides an electronic way to pay in stores without showing stamps but also gives government the opportunity to track purchases. Another new data source being integrated is social media: some services require the use of specific platforms by potential welfare recipients in order to access information and resources. Currently, the US government is planning to use social media profiles to detect welfare fraud with some monitoring already utilised in fraud and abuse investigations.

> In 2014, the SSA's Office of the Inspector General (OIG) utilized social media reviews to help arrest more than 100 people who defrauded Social Security Disability Insurance (SSDI) out of millions of dollars. Investigators found photos on the personal accounts of disability claimants riding on jet skis, performing physical stunts in karate studios and driving motorcycles.[62]

In this case we focus on ethical concerns in four stages of the big data life cycle – data collection, data acquisition & management, decision-making and monitoring and learning. The primary ethical concern here has to do with 'bias' and the resulting undue 'adverse effects for individuals and groups', touching on some additional unequally distributed 'privacy' concerns as well as 'principled performance' concerns raised by automating systems like this. These stages and the concerns we address in them are summarised in Table 5.

2.4.1 Data Collection

The primary ethical concern at the stage of data collection is the over-surveillance of certain population subgroups, in this case welfare recipients. In the US, low-income individuals are more likely than others to experience monitoring by the government, which is relevant to **causing adverse effects to groups**. In fact, 'low-income communities are among the most surveilled communities in America'.[63] This goes back to the 1996 welfare reform bill entitled 'Personal Responsibility and Work Opportunity Reconciliation Act' (PRWORA), which calls for an elaborate system of performance indi-

60. GAO, 'Supplemental Nutrition Assistance Program: Disseminating Information on Successful Use of Data Analytics Could Help States Manage Fraud Risks', www.gao.gov/products/GAO-19-115 (2018).
61. M. Madden, M. Gilman, K. Levy, and A. Marwick, 'Privacy, Poverty, and Big Data: Matrix of Vulnerabilities for Poor Americans', 95(1) *Washington University Law Review* 53, at 113 (2017).
62. M. Miller, 'U.S. Government Weighs Social-Media Snooping to Detect Social Security Fraud', www.reuters.com/article/us-column-miller-socialmedia-idUSKCN1RA12R (2019), at 2.
63. K. Waddell, 'How Big Data Harms Poor Communities', www.theatlantic.com/technology/archive/2016/04/how-big-data-harms-poor-communities/477423/ (2016), at 1.

Table 5 Ethical concerns with US welfare fraud detection

Data collection	Data acquisition & management	Analysis	Communicating insight	Decision-making	Monitoring & learning
Privacy, independence from bias, causing adverse effects for groups	Independence from bias, causing adverse effects for groups, privacy, fairness			Independence from bias, causing adverse effects for groups, principled performance	Accountability, independence from bias, causing adverse effects for groups, transparency

cators with the main goal of 'welfare-to-work' efforts.[64] The establishment of those indicators allows intrusive data collection. For example, it 'empowered state governments to delve into the personal and sexual lives of women of all ages, by requiring single mothers to identify the biological fathers of their offspring' and by capping welfare payments if women had more children while in the programme.[65] Collecting this data is done on top of cross-referencing databases, following up on tips from welfare fraud hotlines, drug-testing and physically surveilling the poor.[66] Digitisation of certain provisions such as the adoption of EBT cards further highlights this issue, as electronic EBT transactions data shows purchase histories and amounts that were spent. Already in 1999, there was a discussion around **privacy** intrusion based on the digitising services, such as EBT, that would apply only to those in need of government support.

> Previously recipients could anonymously cash their checks or spend their food stamps. That is, the transaction did not link the individual to the purchase. With EBT, a permanent record of precisely what the person does with the government benefit often will be created.[67]

Beyond data that is collected at the individual level, there is another dimension of neighbourhood surveillance. Welfare recipients tend to live in poorer neighbourhoods, which are also subjected to more police presence and CCTV monitoring. In addition, people who live in crowded, urban neighbourhoods are more likely to suffer warrantless searches by government agents.[68] 'As a result, they are much more likely than other people in other contexts to become entangled with the criminal justice and child welfare systems, both of which are highly stigmatizing and privacy-stripping'[69] – all of these interactions are ethically relevant with regard to **independence from bias**. These interactions then become embedded in large and linked databases that use this input to assess other things than fraud detection such as housing, employment or educational opportunities.[70]

2.4.2 Data Acquisition and Management

This case is ethically significant in terms of data acquisition owing to its utilisation of social media data and its ambition to extend this practice to frontline workers who work with claimants.[71] This is problematic in two ways: first, social media data makes it possible to assess a person's network as well as create a profile on the basis of online behaviour and preferences. 'Poor Americans have long suffered from guilt by association, meaning they bear the stereotypes and stigma of their social class (and race and gender) in ways that impede their economic progress and well-being',[72] which is relevant to potential violations of both **privacy** principles and **fairness** principles. And second, potential knowledge gaps around privacy can make welfare recipients' 'privacy vulnerable', in particular because of their reliance on mobile connectivity and fewer restrictions they put on the content being posted online.[73] Acquisition and utilisation of this data that has an inherent **bias** could reinforce existing patterns of neglect and socioeconomic disadvantage, resulting in **adverse effects for groups and individuals**. The fact that these conclusions are reached by leveraging individuals' associations within their social network is seemingly not aligned with the individualistic nature of the legal process and raises **fairness** concerns.

2.4.3 Decision-making

There are additional ethically significant concerns linked with the move towards automated decision-making in this case. Many states have started to use data-

64. N. Maréchal, 'First They Came for the Poor: Surveillance of Welfare Recipients as an Uncontested Practice', 3(3) *Media and Communication* 56 (2015); M. Wiseman, 'Welfare Reform in the United States: A Background Paper', 7(4) *Housing Policy Debate* 595 (1996).
65. Maréchal, above n. 65; B. O'Connor, *A Political History of the American Welfare System: When Ideas Have Consequences* (2003).
66. Maréchal, above n. 65.
67. P.P. Swire, 'Financial Privacy and the Theory of High-Tech Government Surveillance', 77(2) *Washington University Law Quarterly* 461, at 505-6 (1999).
68. M. Gilman, 'AI Algorithms Intended to Root Out Welfare Fraud Often End Up Punishing the Poor Instead', https://theconversation.com/ai-algorithms-intended-to-root-out-welfare-fraud-often-end-up-punishing-the-poor-instead-131625 (2020).
69. Madden et al., above n. 61, at 66.
70. *Ibid.*
71. Miller, above n. 63.
72. Madden et al., above n. 61, at 66.
73. Madden et al., above n. 61.

doi: 10.5553/ELR.000190 - ELR 2021 | No. 1

mining techniques for data analysis and automatic identification of fraud in, for example, the food stamp programme or unemployment insurance. In this effort, states are receiving the federal government's support to upgrade their technology and software. Recent examples show that this automation of fraud detection in combination with payouts is not reliable, raising questions about the **principled performance** of these systems:

> In Michigan, a $47 million automated fraud detection system adopted in 2013 made roughly 48,000 fraud accusations against unemployment insurance recipients - a five-fold increase from the prior system. Without any human intervention, the state demanded repayments plus interest and civil penalties of four times the alleged amount owed … As it turns out, a state review later determined that 93% of the fraud determinations were wrong.[74]

Michigan is not the only state experiencing problems with automated decision-making in fraud detection. Similar issue are reported from Indiana, Arkansas, Idaho and Oregon.[75] Reasons for such faulty systems are manifold. States lack funding, skilled analysts as well as data to make automated decision-making systems work.[76] This results in both bias and adverse effects for individuals in the decision-making stage.

2.4.4 Monitoring and Learning

In terms of monitoring these data processes, these issues in decision-making as well as in data collection and acquisition are hard to track because welfare programmes are the responsibility of the state and the federal government has little to no authority to 'oversee or assess the adequacy of benefit levels, bureaucratic process, or the return on investment in terms of assuring a decent quality of life for the poorest'.[77] There is thus little control over these automated processes that further entrench existing **bias** and result in **adverse effects for individuals** and vulnerable groups without much human oversight. This raises issues around accountability and transparency when it comes to retracing the steps that were taken and the ability to assess the process.

2.5 SyRI – Dutch Benefit Fraud Detection

SyRI (System Risk Indication) was created by the Dutch government following legislation passed by the Dutch Parliament in 2014. Multiple national governmental bodies can request the minister to use the system, including Dutch municipalities, the Employee Insurance Agency, the Social Security Bank, the Tax Authority, and the Ministry of Social Affairs and Employment. Governmental agencies contributing data to the system were even more numerous, as SyRI was originally allowed (by the legislation allowing its operation) to utilise 17 categories of government data, including taxes, fines, residence, debts and benefits. Because of this breadth of included personal data, the Council of State recommended to install a 'select before you collect' principle, as per the Council's conclusions published in the Staatscourant.[78] The principle requires that parties first determine what data is needed to achieve the objective and then only selectively acquire the data needed, rather than collecting all data accessible to them.

The totality of this data was then fed into an 'artificial intelligence algorithm', the details of which remain secret to this day. To use SyRI one of the agencies authorised to utilise it would request the Ministry of Social Affairs and Employment and identify a neighbourhood they believe to have an elevated risk of benefit fraud. SyRI can then identify specific individuals and addresses in those neighbourhoods that pose an elevated risk of benefits fraud or misuse. Any such risk identification by the system is not a form of evidence of a violation and in and of itself cannot be used for law enforcement[79] – the goal is to identify cases for further inspection and communicate those (excluding false positives the ministry itself can identify) to the agency making the request.

In this case we focus on ethical concerns in three stages of the big data life cycle – data acquisition and management, analysis and monitoring and learning. The primary ethical concerns here have to do with 'privacy', 'data minimisation', and 'proportionality' of the gathered data. These stages and the concerns we address in them are summarised in Table 6.

2.5.1 Data Acquisition and Management

The case of SyRI is ethically significant here because of the relatively unchecked breadth of data sources it links together, potentially violating the principles of **data minimisation** and appropriate balancing of invasiveness with societal benefits of a system. Even as SyRI was being established, the Council of States noted that the list of personal data that may be utilised by SyRI is 'so broad that it is hardly possible to think of any personal data that does not fall under it. The list does not seem intended to be limiting, but to have as much leeway as possible'.[80] The ethical consideration of **proportionality** applies here as well (the risks to individual privacy should be proportional to the societal benefit). In this case the threat to **privacy** is certainly substantial because of both the sheer variety and volume of data SyRI utilised and the fact that individuals simply have to live in the 'wrong' neighbourhood to be potentially analysed by SyRI. The benefit seems to be rather ques-

74. Gilman, above n. 69.
75. *Ibid.*
76. T. Newcombe, 'Aiming Analytics at Our $3.5 Billion Unemployment Insurance Problem', www.govtech.com/data/Aiming-Analytics-at-Our-35-Billion-Unemployment-Insurance-Problem.html (2017).
77. Maréchal, above n. 65, at 63.
78. Staatscourant, 'Raad van State. Ontwerpbesluit houdende regels voor fraudeaanpak door gegevensuitwisselingen en het effectief gebruik van binnen de overheid bekend zijnde gegevens (Besluit SyRI)', *Staatscourant nr. 26306* (2014).
79. Dutch Government, 'Answer to Parliament Questions 2018Z18418', *Buitenweg. Ref. 2018-0000177182* (2018).
80. Staatscourant, above n. 79 (authors' own translation).

Table 6 Ethical concerns with SyRI

Data collection	Data acquisition & management	Analysis	Communicating insight	Decision-making	Monitoring & learning
	Data minimisation, proportionality, privacy	Transparency, explainability, causing adverse effects for groups			Data minimisation, proportionality, transparency, privacy, obeying the law, causing adverse effects for groups

tionable as SyRI analysis was conducted only in four Dutch cities and likely resulted in no new cases of fraud identified.

There is also a secondary and a much more practical concern that is potentially also ethically significant in the data management of the SyRI system linked to data storage: in a reply to parliamentary questions, the government acknowledged that source files were not always destroyed when they should have been but in given cases almost 18 months too late.[81] The data was stored in a secure area that could be accessed only by those authorised to work with SyRI, but this does exacerbate the **proportionality** concerns as personal information was exposed to a greater risk of misuse or potential security breach than necessary with no demonstrable societal benefit.

2.5.2 Analysis

This case is ethically significant for analysis in two distinct ways: first, a lack of clarity in the indicators and risk profiles used may lead to a 'fishing expedition'.[82] Even though the SyRI law originally allowed the practice, the data acquired for this system was in different systems, administered by different organisations and collected for different goals. Inappropriate use of this data may eventually lead to reluctance to share it with the government, ultimately limiting effectiveness. The government opted for not disclosing data sources and methods of analyses to avoid disclosing the modus operandi and thus lowering the risks of those committing fraud and gaming the system.[83] This is a clear instance of **transparency** and **explainability** being sacrificed to maintain the effectiveness of the system (in terms of data acquisition and the difficulty of circumventing it by malicious actors).

Secondly, there are ethical concerns with regard to reinforcing existing stigmatisation and discrimination as SyRI 'benefits from a relatively clear public legal framework' despite its 'alleged discriminatory character'.[84] Especially the targeting of specific areas led to civil advocacy against the system and to accusations of discriminatory or stigmatising effects of the system owing to its use of a broad range of data likely including protected characteristics. This can result in over-surveillance of individuals based on existing stigmatising patterns or the over-surveillance of entire neighbourhoods based on factors largely beyond the control of any individual living within it, both of which are undue **adverse effects for individuals and groups**.

2.5.3 Monitoring and Learning

In the case of SyRI the monitoring and learning process was rather public. In fact, what make this case well known are reports by international news and professional media[85] related to a lawsuit that made it a test case for algorithmic governance. In this case the Court ruled that the law establishing SyRI was in violation of the European Convention on Human Rights because it was too invasive,[86] making this ethically relevant in terms of **obeying the law** but also **privacy**. In its ruling, the Court argued that deployment of new technologies towards these ends can be legitimate but also that the government has a special responsibility to find the right balance between deploying such technologies for the public good and respecting and protecting privacy, referring to **proportionality**. The ruling concluded that using this much data on this level violates private life, does not fit with principles of **transparency** and restraint in data use (**data minimisation**) and creates risks that the system might discriminate and thus cause **adverse effects for individuals and groups**,[87] leading to the immediate termination of SyRI.

3 Discussion

An overview of the case summaries presented in Section 2 is presented in Table 7, which combines the individual one-row tables used for each of the preceding

81. Dutch Government, above n. 80.
82. Staatscourant, above n. 79.
83. Dutch Government, above n. 80.
84. S. Ranchordás and Y. Schuurmans, 'Outsourcing the Welfare State: The Role of Private Actors in Welfare Fraud Investigations', 7 *European Journal of Comparative Law and Governance* 5, at 6 (2020).
85. The Economist, 'Humans Will Add to AI's Limitations', www.economist.com/technology-quarterly/2020/06/11/humans-will-add-to-ais-limitations (2020); T. Simonite, 'Europe Limits Government by Algorithm. The US, Not So Much', www.wired.com/story/europe-limits-government-algorithm-us-not-much/ (2020).
86. Rechtbank Den Haag, Ruling ECLI:NL:RBDHA:2020:865 (2020).
87. *Ibid*.

Table 7 Ethical concerns in selected cases throughout the big data life cycle

	COMPAS	CAS	vTaiwan	US welfare fraud detection	SyRI
Data collection		Independence from bias, causing adverse effects for groups, fairness, respect, honesty, integrity		Privacy, independence from bias, causing adverse effects for groups	
Data acquisition & management		Consent, fairness, causing adverse effects for groups, respect, transparency, data minimisation		Independence from bias, causing adverse effects for groups, privacy, fairness	Data minimisation, proportionality, privacy
Analysis	Independence from bias, transparency, accountability		Respect, fairness, independence from bias, transparency, explainability		Transparency, explainability, causing adverse effects for groups
Communicating insight	Independence from bias, fairness, causing adverse effects for individuals	Transparency, explainability	Honesty, independence from bias		
Decision-making	Accountability, fairness, transparency, independence from bias, respect	Causing adverse effects for groups, fairness, proportionality, respect, professionalism	Accountability, honesty, respect, principled performance	Independence from bias, causing adverse effects for groups, principled performance	
Monitoring & learning	Independence from bias, causing adverse effects for individuals, obeying the law			Accountability, independence from bias, causing adverse effects for groups, transparency,	Data minimisation, proportionality, transparency, privacy, obeying the law, causing adverse effects for groups

cases. Table 7 provides a structured summary of what we found to be important ethical considerations at various stages of the big data life cycle.

Despite the exploratory nature of this systematic overview and the largely illustrative case selection, we can make a few insightful observations. Primarily, we show that relevant ethical concerns can indeed emerge across the entire big data life cycle, substantiating the arguments that claim this to be the case.[88] This is not a surprising finding, as it is generally accepted that this is the case, but this article is innovative in that it operationalises this approach and shows that this intuition applies to the legal domain.

Table 7 can be read in multiple ways. Reading the table as a whole shows that issues of bias and adversely affecting individuals and groups are the most frequent ethical considerations. Other issues such as transparency are also prominent in the table as a whole, but some issues, such as accountability, privacy or obeying the law, can be identified far less often. This observation can be enhanced by reading the table row by row, by which means issues of bias and adversely affecting individuals or groups are shown to be cross-cutting and can be identified in every single data life cycle stage multiple times, even though we focus only on five illustrative

88. Wing, above n. 8.

cases. Other considerations such as transparency remain relatively cross-cutting but clearly over-represented in certain life cycle stages (in the case of transparency this is the 'analysis' stage). Other considerations remain far more circumscribed to a specific life cycle stage, such as obeying existing law, which we can identify only at the stage of monitoring and learning in all five cases. Needless to say, the generality of these observations is limited by analysing only five distinct cases and by the limited available information about these cases, but it is an indication that some ethical considerations tend to apply more to specific big data life cycle stages than to others.

The more interesting observation comes from reading the table column by column (case by case); this shows the interconnectedness of individual stages in any given case and allows us to get a better grasp of how this interconnectedness plays out for ethical concerns. It seems that there are situations where a key concern emerges in multiple stages in a slightly different form. This suggests that concerns from earlier stages of the data cycle can get 'transferred further' or even compounded throughout the data life cycle. The compounding is apparent in, for example, the CAS case, where data collection itself results in 'dirty data' owing to bias in suspect identification and acquisition of specific type of data about ethnicities reflects a further discriminatory assumption, culminating in concerns about discrimination at the level of decision-making. However, it also seems that it is not just a question of issues earlier in the data cycle influencing what happens next – even issues that happen later in the cycle can be significant for ethical concerns at a preceding stage. This is apparent in the case of vTaiwan, where the risk for legislators to ignore the analysis at the stage of 'decision-making' can influence what the relevant ethical concerns are earlier in the cycle. We believe this effect can also be positive, for example, anonymisation of data during the 'data management' stage can alleviate privacy concerns that took place at the 'data collection' stage. This also raises questions for future research, for example, whether bias at the data collection stage carries through to the acquisition and analysis stages or whether new or additional forms of bias are introduced at that point in the life cycle. It also facilitates a discussion around whether rectifying bias at the collection stage will solve bias-related challenges later on in the life cycle or whether they are reintroduced in a different way.

4 Conclusions and Limitations

This article adopts a process-oriented definition of big data and a relatively simple model of the big data life cycle to systematise ethical concerns along the stages of this life cycle. To do so, it adopts an ethically pluralistic and pragmatic perspective on 'ethical' concerns and selects five cases that together capture a broad range of big data uses in the legal domain: a decision-support system for sentencing and bail decisions (COMPAS), a predictive policing system (CAS), a legislative crowdsourcing system (vTaiwan) and two welfare fraud detection systems (one deployed in the US and the other in the Netherlands). Discussing each case in turn, the article provides an overview of these cases, delivering on the intended systematic summary and making a few interesting observations. In particular, the life cycle perspective is capable of highlighting how ethically significant practices and choices may manifest themselves differently in different stages of a use case.

Despite delivering the intended systematic summary, the article has a few limitations worth highlighting: first, the various ethical concerns we refer to throughout the article (and that we list in Annex A) are not mutually exclusive but considerably overlap. This is a direct result of our decision not to merge or re-categorise the identified ethical concerns, as that would necessitate distinctive normative commitments. As a result, some ethical concerns will often incorporate other ethical concerns, and there are multiple ways to label a given issue. For example, anything consequential for the 'independence from bias' is often also related to 'not causing unjustified or adverse effects for individuals or groups' (because that is what biased data tends to result in) or to 'respect' or 'fairness' (as that is often violated by treating individuals as stereotypical examples of a group they belong to). This limitation is important to highlight because it suggests that the number of ethical concerns we include in a stage does not necessarily reflect the ethical 'seriousness' of any particular practice. Put differently, the ethical considerations we list and highlight in the five cases do not necessarily aggregate in the balance of moral concerns. Of course, others attempting to do a similar systematisation, especially when it comes to practitioners implementing or morally interrogating a big data system, can have a more committed and normative definition for ethics to resolve this issue. Second, the stages of the big data life cycle are not as distinct in practice as they are in our model and systematisation. The fact that these stages functionally overlap is apparent from, for example, the case of vTaiwan, where the goal of analysis is to communicate something in a specific way, which makes the stages of 'analysis' and 'communicating insight' inseparable. Sometimes even clearly distinct stages are ethically intertwined. Consider the value of 'due diligence to evaluate data practices of third-party collaborators', which implies that even if data is only obtained and not collected, data collection practices still need to be considered and morally assessed. This makes the stages of 'data collection' and 'data acquisition' conceptually very different but ethically closely connected. Third, our selection of cases also introduces a bias: since many cases of big data use are not fully transparent, we focus on relatively well-described cases in order to have sufficient information for our systematisation. However, this information often comes as a result of investigative journalism, activism or court trials that are more likely prompted by cases that blatantly violate sensitive ethical norms. Thus, it may be that most big data use cases are significant in terms of less inflammatory ethical concerns or are less ethically contentious in general than the cases we address.

Despite these shortcomings, the article does generate some useful insights. First, it supports the claim that ethically significant decisions are made at various stages of a big data life cycle. Although this is not a novel insight, this is the first article (in our estimation) to actually apply this logic so systematically and to do so specifically for the legal domain. Consequently, the ethics of big data practices should look beyond issues that are discretely tied to any one single stage and to scrutinise existing big data use cases along the entire data life cycle. Second, our approach offers a more structured and holistic view than what one would obtain by simply going concern by concern for a given case, potentially missing how some concerns manifest themselves in different stages and connect to one another. It thus allows us to see the prevalence of certain ethical considerations throughout the data life cycle (generalisable to the five analysed cases) and to be more thorough about how ethical concerns get compounded, alleviated or transformed throughout the life cycle.

Third, the approach adopted in this article is, in and of itself, a useful heuristic for 'lawyerly' ethical reasoning about big data use cases, which might be valuable in legal practice or in the development of big data systems. This approach can serve as a useful starting point for examining which ethical concerns tend to appear at a given data life cycle stage or even to highlight structural similarities that might be useful for developing an ethically informed typology of cases of big data practices, which could then be used to examine and address some of these cases collectively, rather than individually. From a scholarly perspective, this approach has benefits in terms of its non-committal attitude towards ethical theorising – not siding with any one particular ethical theory – offering a wider menu of ethical views for examining the morality of big data in the future. Keeping the range of ethical considerations open is arguably more conducive to fostering a discipline of big data ethics that is pluralistic and substantively richer than many of the current attempts focused on one or a limited number of master moral [.[89] This is desirable for a field of study as recent as big data ethics, where favouring any one theory or set of moral values and principles might be normatively and theoretically premature.

Finally, it bears emphasising that, despite the ethical concerns that they raise, the big data tools examined were all deemed compatible with legal norms across a variety of jurisdictions.[90] That law and morality cover distinct normative domains is hardly a surprise to most lawmakers, lawyers and legal practitioners. But here the distinction is worth recalling: since our perspective on ethics is uniquely lawyerly, the question of whether (and how) moral critiques can be recast as legal challenges remains an important area of future work for data scientists, legal scholars, legal practitioners and ethicists. The pragmatism of our approach in this article can potentially aid this recasting (given that the ethical considerations we rely on are drawn from documents that already matter for big data practices), but demonstrating the value of our approach in this respect remains a topic for further research.

89. M. Boeckhout, G.A. Zielhuis, and A.L. Bredenoord, 'The FAIR Guiding Principles for Data Stewardship: Fair Enough?', 26(7) *European Journal of Human Genetics* 931 (2018); J. Collmann & S.A. Matei, *Ethical Reasoning in Big Data: An Exploratory Analysis* (2016); D. Shin & Y.J. Park, 'Role of Fairness, Accountability, and Transparency in Algorithmic Affordance', 98 *Computers in Human Behavior* 277 (2019).

90. A special case here might be SyRI, which got through the legislative level but was finally banned at the judicial one.

Annex A

Ethical considerations[91]

Ethical consideration	Document type	Documents
Reliability	General (research conduct)	ALLEA 2017
Honesty/Integrity (qua honesty and truthfulness)	General (research conduct)	ALLEA 2017, GCC 2019, WHO 2017, UNDP 2017
Respect/mutual respect for human dignity/intrinsic value of people	General (research conduct) & Data-specific	ALLEA 2017, WHO 2017, UNDP 2017, OECD, 2016, European Commission 2019
Accountability	General (research conduct) & Data-specific	ALLEA 2017, UNDP 2017, WHO 2017, OECD 2020
Fairness	General (research conduct) & Data-specific	GCC 2019, OECD 2013, OECD 2020, European Commission 2019, EU Regulation 2016/679, CEPEJ 2018
Care (duty not to harm the subjects of research)	General (research conduct) & Data-specific	GCC 2019, ICC & ESOMAR 2007, European Commission 2019
Independence and impartiality (from bias, discrimination, prejudice and undue influence)	General (research conduct) & Data-specific	WHO 2017, UNDP 2017, UNDG 2017, OECD 2020, CEPEJ 2018
Professional commitment/Professionalism	General (research conduct)	WHO 2017, UNDP 2017
Transparency (of method and application)	General (research conduct) & Data-specific	UNDP 2017, ICC & ESOMAR 2007, UNDG 2017, OECD 2016, OECD 2013, OECD 2020, IHSN 2010, European Commission 2019, EU Regulation 2016/679, CEPEJ 2018
Explainability (as addition to transparency)	Data-specific	OECD 2016, OECD 2013, OECD 2020, IHSN 2010, European Commission 2019, EU Regulation 2016/679, CEPEJ 2018
Principled Performance/Results orientation (demonstrable benefits of the system)	General (research conduct) & Data-specific	UNDP 2017, OECD 2020, European Commission 2019
Obeying the law/Lawfulness	General (research conduct) & Data-specific	UNDP 2017, OECD 2013, OECD 2020, IHSN 2010, European Commission 2019, EU Regulation 2016/679, CEPEJ 2018
Not violating human rights	Data-specific	UNDG 2017, OECD 2020, European Commission 2019, CEPEJ 2018

91. The documents we refer to in this table are the following (ordered alphabetically): ALLEA, 'The European Code of Conduct for Research Integrity', https://allea.org/code-of-conduct/ (2017); CEPEJ, 'European Ethical Charter on the Use of Artificial Intelligence in Judicial Systems and Their Environment', https://rm.coe.int/ethical-charter-en-for-publication-4-december-2018/16808f699c (2018); EU Regulation 2016/679; European Commission, 'Ethics Guidelines for Trustworthy AI', https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai (2019); GCC, 'Global Code of Conduct for Research in Resource-poor Setting', www.globalcodeofconduct.org/ (2019); ICC & ESOMAR, 'International Code on Market and Social Research', https://iccwbo.org/content/uploads/sites/3/2008/01/ESOMAR-INTERNATIONAL-CODE-ON-MARKET-AND-SOCIAL-RESEARCH.pdf (2007); IHSN, 'Dissemination of Microdata Files: Principles, Procedures and Practices', https://ihsn.org/sites/default/files/resources/IHSN-WP005.pdf (2010); OECD, 'The OECD Privacy Framework', www.oecd.org/sti/ieconomy/oecd_privacy_framework.pdf (2013); OECD, 'Research Ethics and New Forms of Data for Social and Economic Research', https://doi.org/10.1787/5jln7vnpxs32-en (2016); OECD, 'Recommendation of the Council on Artificial Intelligence', https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449 (2019); UNDG, 'Data Privacy, Ethics and Protection: Guidance Note on Big Data for Achievement of the 2030 Agenda', https://unsdg.un.org/resources/data-privacy-ethics-and-protection-guidance-note-big-data-achievement-2030-agenda (2017); UNDP, 'Code of Ethics', www.undp.org/content/undp/en/home/accountability/ethics.html (2017); WHO, 'Code of Conduct for Responsible Research', www.who.int/about/ethics/code-of-conduct-for-responsible-research (2017).

Ethical consideration	Document type	Documents
Consent (including consent for reuse where feasible)	Data-specific	ICC & ESOMAR 2007, UNDG 2017, OECD, 2016, OECD 2013, IHSN 2010, EU Regulation 2016/679
Data minimisation (limit the collection of data to what is relevant for research)	Data-specific	ICC & ESOMAR 2007, UNDG 2017, OECD, 2016, OECD 2013, EU Regulation 2016/679
Privacy	Data-specific	UNDG 2017, OECD 2016, OECD 2013, OECD 2020, IHSN 2010, EU Regulation 2016/679
Confidentiality	Data-specific	UNDG 2017, OECD 2013, IHSN 2010
Not causing unjustified or adverse effects for individuals or groups	Data-specific	UNDG 2017, OECD 2016, European Commission 2019, CEPEJ 2018
Proportionality – Risks of harm need to be proportional to the benefits of data use	Data-specific	UNDG 2017, European Commission 2019, OECD 2016
Sensitivity to context – including focus on vulnerable population groups	Data-specific	UNDG 2017, European Commission 2019
Due diligence to evaluate data practices of third-party collaborators	Data-specific	UNDG 2017, CEPEJ 2018
Data and analysis quality assessments (general and to prevent biases)	Data-specific	UNDG 2017, OECD 2016, European Commission 2019, EU Regulation 2016/679, CEPEJ 2018
Sharing data (to the extent it does not violate other principles)	Data-specific	OECD 2016
Responsibility to maintain adequate security of data	Data-specific	OECD 2013, OECD 2020, EU Regulation 2016/679, CEPEJ 2018
Promotion/adherence to democratic values and individual freedom	Data-specific	European Commission 2019, OECD 2020
Not limiting user autonomy	Data-specific	CEPEJ 2018

Teaching Technology to (Future) Lawyers

Mikołaj Barczentewicz*

Abstract

The article offers a reflection on how applications of computer technology (including data analytics) are and may be taught to (future) lawyers and what are the benefits and limitations of the different approaches. There is a growing sense among legal professionals and law teachers that the technological changes in the practice of law are likely to promote the kind of knowledge and skills that law graduates often do not possess today. Teaching computer technology can be done in various ways and at various depths, and those different ways and levels have different cost and benefit considerations. The article discusses four models of teaching technology: (1) teaching basic technological literacy, (2) more advanced but general technology teaching, (3) teaching computer programming and quantitative methods and (4) teaching a particular aspect of technology – other than programming (e.g. cybersecurity). I suggest that there are strong reasons for all current and future lawyers to acquire proficiency in effective uses of office and legal research software and standard means of online communication and basic cybersecurity. This can be combined with teaching of numerical and informational literacy. I also claim that advanced technology topics, like computer programming, should be taught only to the extent that this is justified by the direct need for such skills and knowledge in students' future careers, which I predict to be true for only a minority of current lawyers and law students.

Keywords: legal education, law and technology, legal analytics, technology education, technological literacy

1 Introduction[1]

It is widely accepted and, I think, true that technology – or, more specifically, computers and computer networks – will be playing an even greater role in the future of legal practice than it is today.[2] Perhaps the changes are not going to be very radical in the short term. However, some specialised software tools will likely continue to transform certain aspects of legal practice. Among the obvious examples are e-discovery (and due diligence),[3] basic legal research[4] and contract automation tools,[5] which though sometimes not precise or 'self-driving' enough for many practical contexts today, are likely to keep improving. The same is true with regard to artificial intelligence algorithms aiming to predict how a court (or another authority) would apply the law in a given factual scenario: there is a lot of promise, but the results are still somewhat underwhelming (for instance, a great deal of manual work is still required to create such tools).[6] There is a growing sense among legal professionals and law teachers that the technological changes in the practice of law are likely to promote the kind of knowledge and skills that law graduates often do not possess today.[7] Different kinds of technological

* Mikołaj Barczentewicz is the Research Director, Surrey Law and Technology Hub, as well as Senior Lecturer (Associate Professor) in Law, University of Surrey School of Law. He is also a Research Associate of the University of Oxford Centre for Technology and Global Affairs. I would like to express my thanks to the anonymous referees.

1. This article develops a short piece published as M. Barczentewicz, 'Learn in Code', *Solicitors Journal* (February 2020).
2. J. Armour, R. Parnham & M. Sako, 'Augmented Lawyering', European Corporate Governance Institute (ECGI) – Law Working Paper 5582020, 2020, at http://dx.doi.org/10.2139/ssrn.3688896.
3. E-discovery may mean both 'the process by which computers search a database for keywords that lawyers agree are marks of relevance' and the more advanced 'predictive coding' where 'lawyers look at a sample of the larger set of documents' and '[c]omputer technicians help construct algorithms that predict whether a document is relevant'; J.O. McGinnis and R.G. Pearce, 'The Great Disruption: How Machine Intelligence Will Transform the Role of Lawyers in the Delivery of Legal Services Colloquium: The Legal Profession's Monopoly on the Practice of Law', 82 *Fordham Law Review* 3041, at 3047 (2013).
4. The improvements include, for instance, 'semantic search' (which 'will allow lawyers to input natural language queries to computers, and the computers will respond semantically to those queries with directly relevant information'), machine judgments on strength of precedent, and automatic detection of the most relevant cases based, *e.g.*, on a scan of the text of a court brief; *ibid.*, at 3048-3050.
5. According to McGinnis and Pearce: 'In the future, machine processing will be able to automate a form, tailor it according to the specific facts and legal arguments, and track its effect in future litigation. As hardware and software capacity improves, so too will the generated documents. We predict that within ten to fifteen years, computer-based services will routinely generate the first draft of most transactional documents'; *ibid.*, at 3050.
6. *See e.g. ibid.*, at 3052-3053; D.M. Katz, M.J. Bommarito II & J. Blackman, 'A General Approach for Predicting the Behavior of the Supreme Court of the United States', 12 PLoS ONE (2017), at https://doi.org/10.1371/journal.pone.0174698; M. Medvedeva, M. Vols & M. Wieling, 'Using Machine Learning to Predict Decisions of the European Court of Human Rights', 28 *Artificial Intelligence and Law* 237 (2020); M. Barczentewicz, 'Combining AI and Digitization of Judgments for Access to Justice', in S. Chishti and others (eds.), *The LegalTech Book: The Legal Technology Handbook for Investors, Entrepreneurs and FinTech Visionaries* (2020).
7. *See, e.g.*, M. Pistone and J.J. Hoeffner, 'No Path But One: Law School Survival in an Age of Disruptive Technology', 59 *Wayne Law Review* 193 (2013); R.W. Staudt, 'Introduction, Justice, Lawyering and Legal Education in the Digital Age', 88 *Chicago-Kent Law Review* 687 (2013); T. Rostain, R. Skalbeck & K.G. Mulcahy, 'Thinking Like a Lawyer, Designing Like and Architect: Preparing Students for the 21st Century Practice', 88 *Chicago-Kent Law Review* 743 (2013); D.M. Katz, 'The MIT School of Law? A Perspective on Legal Education in the 21st Century', *University of Illinois Law Review* 1431 (2014); M. Pistone, 'Law Schools and Technology: Where We Are and Where We Are Heading', 64 *Journal of Legal Education* 586 (2015); V. Janeček, R. Williams & E. Keep, 'Education for the Provision of Technologically Enhanced Legal Services' 40 *Computer Law & Security Review* 105519

expertise are cited as examples of what the lawyers of the future will need to know and be able to do. For instance, Daniel Katz argued that a law school of the future (the MIT School of Law) should have a 'curriculum that strongly emphasize[s] science, technology, process engineering, predictive analytics, and mathematical and computational modelling'.[8] In this article, I reflect on how technology is and may be taught to (future) lawyers and what benefits and limitations are associated with the different approaches. This article is not intended as a comprehensive survey of technology education in law schools and law faculties. The examples I discuss are meant to be illustrative.

I am concerned here solely with computer technology, not with technology in the broader sense. Also, I do not consider the issue of teaching technology law and regulation but only that of acquiring skills and knowledge in technology as such. Furthermore, I do not dispute that the growing importance of computer technology may call for lawyers to acquire business, design or process management skills, which do not fit into the narrower conception of technology skills and knowledge on which I focus in this article.[9]

The crucial preliminary question for any discussion of teaching technology to lawyers and law students is the extent to which lawyers today and in the near future will benefit from such knowledge and skills. In a survey of English solicitors reported by Janeček, Williams and Keep, '90 per cent of English solicitors indicated that they would need some training concerning AI and digital technology in the next three years'.[10] What is more, 71 per cent of respondents expected data analytics to be a training need.[11] It must be noted that virtually all of the lawyers surveyed most likely have but a very vague idea of what data analytics is. Hence, for many, the expectation that they will need training in data analytics is probably an expression of the willingness to jump on the popular bandwagon, not an informed judgment. Armour, Parhnham and Sako are clearly right in predicting that some lawyers will in the near future be consumers of technology, with their jobs augmented by technology but not in a way that requires any detailed knowledge of it.[12] True, some lawyers or law graduates will be involved in the development of legal technology and will need various levels of technological expertise (some – though probably not many – may even benefit from full interdisciplinarity, for instance, as both lawyers and software engineers).[13] Moreover, as today, some lawyers will benefit from technological expertise in their legal practice (especially in areas like intellectual property or, more broadly, technology regulation). The big question is the proportions among those types of future lawyering.

1.1 Three Kinds of Lawyers

In other words, there are three levels of need for technological literacy for current and future lawyers:

1. those who benefit significantly only from basic proficiency (but greater than the current norm),
2. those who benefit significantly from more advanced knowledge about the behind-the-scenes workings of technology (just like planning and construction lawyers benefit from knowledge about architecture and construction technology),
3. those who benefit significantly from even more advanced technological proficiency, including advanced *practical* skills such as computer programming.

My prediction is that the vast majority of future lawyers will stay at levels (1) and (2). Lawyers remain consumers of computer technology and need to know only as much about it as their field of law requires (more in intellectual property, much less in family law). In other words, I do not expect the current situation to change dramatically. Software, unlike traditional in-person legal advice, is easily scalable, and hence a relatively small number of producers can service a very large number of consumers (in this case lawyers or those who seek legal advice). Despite attempts in a growing number of law firms today,[14] I do not expect that in-house software development will be as significant for legal practice as simply licensing software from comparatively few providers (of course, some law firms may become software providers themselves or spin off such companies). Moreover, optimists about technological improvement who expect software to be increasingly better at automating some legal tasks should not forget that software development itself is also likely to be positively affected by this trend.[15] It would be curiously myopic to think that automation means that 'everyone' should learn computer programming as it is practised now. The same (or very similar) technological improvements as the ones that spearhead automation in legal practice are likely to do the same in software development, and to do so even faster.

Even if thinking only about the most immediate future, before any further significant technological improvements, there are good reasons for lawyers and law students to be cautious about investing too much effort in their technology education. For instance, computer programming (or more broadly, software engineering) is easy to do poorly but difficult to do well. The risks asso-

(2021); R. Williams, V. Janeček & E. Keep, 'What Is the Role of Law Schools in the 21st Century?', 2020; A. Smith and N. Spencer, 'Do Lawyers Need to Learn to Code? A Practitioner Perspective on the "Polytechnic" Future of Legal Education', in C. Denvir (ed.), *Modernising Legal Education* (2020); Armour, Parnham & Sako, above n. 2; Barczentewicz, above n. 1.

8. Katz, above n. 7, at 1465.
9. *See, e.g.*, Smith and Spencer, above n. 7; Williams, Janeček & Keep, above n. 7.
10. Janeček, Williams & Keep, above n. 7, at 4.
11. *Ibid*.
12. Armour, Parnham & Sako, above n. 2, at 56-7.
13. *Ibid.*, at 57-8.
14. *Ibid.*, at 35-41.
15. *See, e.g.*, B.W. Sorte, P.P. Joshi & V. Jagtap, 'Use of Artificial Intelligence in Software Development Life Cycle: A State of the Art Review', 3 *International Journal of Advanced Engineering and Global Technology* 398 (2015); M. Barenkamp, J. Rebstadt & O. Thomas, 'Applications of AI in Classical Software Engineering', 2 *AI Perspectives* 1 (2020).

ciated with bad code are very significant, especially in terms of reliability and security. The benefits of division of labour, i.e. letting the coders do the coding, are overwhelming in most circumstances. This is not to deny that relatively few lawyers can make significant contributions to the development of legal technology as fully fledged technology experts, even as computer programmers. It is only important not to lose a sense of proportion over how many people will be in that group and thus on how much of that kind of training needs to be provided.

However, some degree of more advanced understanding of technology will remain crucial (just as it is now) to those lawyers, classified in my level (2) above, who need it in legal practice (intellectual property, technology law and regulation, and so on). Moreover, a small minority of lawyers (my level (3)) will benefit from even greater understanding and from advanced *practical* skills such as computer programming. Such lawyers may work as 'legal engineers' or 'legal technologists' participating in developing legal technology on a par with engineers or as members of multidisciplinary teams. They can also work as empirical legal researchers, particularly in academia or government (although it is likely that empirical research will require less programming in the future owing to the development of appropriate software).

It is a separate question of what proportion of lawyers would benefit from more basic training in issues such as cybersecurity, numerical literacy or even greater proficiency in using office and legal research software. Here the answer, I think, is that most lawyers (as with most professionals in other fields) should gain such fundamental proficiency. And this is true also for the youngest and future members of the profession. There is a popular myth that the young people today are 'digital natives' and that they are more tech-savvy than previous generations.[16] This claim is usually made in a very vague way, making it difficult to verify empirically. However, serious attempts to do so show that it is indeed a myth.[17] Those who grew up in the age of the internet may be more adept at clicking through interfaces of some software applications that they use daily, but their understanding of the behind-the-scenes mechanisms of computer technology is not at all impressive.[18] Characteristically, computer security habits of the younger people are just as bad as those of their elders.[19]

Having broadly reflected on the scope of the need for technology I now turn to a discussion of four models of teaching about technology already present in law faculties. I will then return to the question of what kind of technology education is suitable depending on different career paths of lawyers and law graduates.

2 Models of Teaching about Technology

A number of universities and other education providers offer some form of technology education for undergraduate or postgraduate law students. In this part of the article, I discuss several examples of such courses and classify them in four groups, or 'models'. In the first model basic technological and numerical literacy are taught. The remaining three models are concerned with more advanced technological proficiency. In the second model computer technology in general is taught in a more advanced way. The third model concerns teaching computer programming, and the fourth focuses on one specific aspect of computer technology (for instance, cybersecurity). My discussion is not meant as a comprehensive survey, and I do not claim that the examples chosen are the best in the world (because this would require a comprehensive comparison that I did not undertake), but I do consider them to be well designed.

2.1 Teaching Basic Technological Literacy

The first model of teaching technological proficiency is the most basic one and includes the teaching of effective uses of office and legal research software and standard means of online communication, cybersecurity. This may be paired with training in business skills, service and product design and process management, but I leave that issue aside as it is beyond the scope of this article.[20] Teaching basic technology skills may also be fruitfully connected to assisting (future) lawyers in gaining two other kinds of literacy: informational literacy and numerical literacy, both of which need improvement.

Law students and young lawyers may feel confident in their use of technology (to put it colloquially, they can google things very quickly), but this confidence quickly dissipates when faced with some even seemingly basic tasks required in the study and practice of law (at least in the United Kingdom, many court judgments cannot be found through Google, not to mention information such as what was the judgment's subsequent authoritative treatment).[21] Probably all law schools offer students some form of an introduction to legal research and databases available to them, but this is likely to be limited to several hours in the 'welcome week' of the first year or a

16. *See, e.g.*, E.J. Helsper and R. Eynon, 'Digital Natives: Where Is the Evidence?', 36 *British Educational Research Journal* 503 (2010); D. Bates, 'Are "Digital Natives" Equipped to Conquer the Legal Landscape?', 13 *Legal Information Management* 172 (2013); T. Ståhl, 'How ICT Savvy Are Digital Natives Actually?', 12 *Nordic Journal of Digital Literacy* 89 (2017); T. Judd, 'The Rise and Fall (?) Of the Digital Natives', 34 *Australasian Journal of Educational Technology* 99 (2018).
17. Helsper and Eynon, above n. 16; Bates, above n. 16; Judd, above n. 16, at 99.
18. *See, e.g.*, J. Fraillon and others, *IEA International Computer and Information Literacy Study 2018 Assessment Framework* (2019).
19. *See, e.g.*, S.S. Tirumala, A. Sarrafzadeh & P. Pang, 'A Survey on Internet Usage and Cybersecurity Awareness in Students', *2016 14th Annual Conference on Privacy, Security and Trust (PST)* (2016); J.D. Thompson, G.L. Herman, T. Scheponik, L. Oliva, A. Sherman, E. Golaszewski, D. Phatak, & K. Patsourakos, 'Student Misconceptions about Cybersecurity Concepts: Analysis of Think-Aloud Interviews', 2018 *Journal of Cybersecurity Education, Research and Practice* 5 (2018).
20. Janeček, Williams & Keep, above n. 7; Smith and Spencer, above n. 7.
21. *See also* Bates, above n. 16, at 176.

short online course. Experiences from Cambridge University described by Bates,[22] as well as my own observations, show worrying deficiencies in information literacy even among students in their later years. Similarly, students struggle with effective use of office software to complete tasks like formatting a document in Microsoft Word or using footnotes. This is not a reason to scoff as I am sure that proficiency in, for example, using Microsoft Word's 'styles' or numbering of paragraphs eludes many practising lawyers and academics.

One way this can be addressed is through entry-level modules on legal skills (legal research and writing) designed in a way that does not assume that young people require minimum (if any) instruction in effective use of software. An example of such successful effort in teaching proficiency in Microsoft Word at the University of Buffalo was described by Detweiler.[23] Another possible way is to include such instruction in core law subjects. For example, in the early weeks of a first-semester law module students could be provided with detailed narrated video tutorials (screen recordings) of how to research answers for the questions they are asked to prepare for those weeks.

Regarding numerical literacy, it is already a skill beneficial to all professionals and a necessity in some areas of law (financial regulation, tax law). With the propagation of artificial intelligence (machine learning) tools, there is a growing need for basic numerical literacy to include at least the fundamentals of statistics needed for informed use of products of machine learning. My own experience, which I am sure is widely shared, is that law schools cannot rely on their students' prior education in that respect. Hence, including numerical literacy in legal research modules or as stand-alone modules, perhaps delivered online, may be advisable. While discussing the other teaching models further on, I provide examples of how more advanced numeracy can be taught in 'legal analytics' or 'computational law' courses, often together with computer programming. I emphasise that the need for basic numeracy skills is broader than for programming skills, so it is likely suboptimal if the only teaching (or at least encouragement for independent learning) that a law student receives in this respect is at this advanced level, which is by and large unnecessary.

Finally, with regard to cybersecurity education, this is a particularly difficult issue, because unlike good research or writing skills, cybersecurity literacy for most people lacks immediate, easily perceived rewards. Cybersecurity skills, if practised, reduce – but never eliminate – one's risk of being a victim of a cyberattack. Moreover, cybersecurity can get very technical very quickly – as I show in Section 2.4 while discussing Yale Law School's 'Cybersecurity' course – but it is also not an obvious question to say how much cybersecurity training is too little and how much is too much (for most people). On the other hand, lawyers, bound to protect the confidentiality of information about their clients or employers, constitute very attractive targets for attackers. For a lawyer to be a victim of a cyberattack, especially because of a kind of contributory negligence (e.g. in a phishing attack[24]), could (and often should) have very grave consequences. I am sceptical that a quick online cybersecurity module (especially in the initial weeks of one's studies or work) may have a meaningful effect on most people's security hygiene. One alternative way to approach cybersecurity education for lawyers and law students is through something like a system of fire drills. For instance, students could be regularly targeted with emails containing links that they should not click on (e.g. because of a suspicious originating domain), and if they do they would receive immediate feedback that they failed a cybersecurity drill with an invitation to online training showing them how to be more secure in the future. This kind of education can to largely be automated and delivered on a university (organisation) level.

I now turn to the three models of teaching more advanced technology topics.

2.2 More Advanced Teaching about Computer Technology in General

One way of providing more advanced teaching about computer technology is through a broad survey of salient topics, without singling out any particular topic. This has been the strategy of Oxford University's 'Law and Computer Science'[25] and Harvard University's 'CS50 for Lawyers'.[26] The two courses differ in their methods of learning and teaching. The Harvard course was designed to be delivered online to large numbers of students, whereas the Oxford course emphasises group work and instructor supervision (and thus enrolled only twelve postgraduate computer students in law and an equal number from computer science). The Oxford course is much more law oriented than the Harvard one and seems to cover more technology law than technology as such. Neither of the courses aims to teach students to program computer software on their own.

As reported by members of the teaching team, in its first year the Oxford course was structured in the following way:

> The first half of the course focused on AI and digital technology in legal practice (the sphere primarily relevant to this paper); the second half of the course on

22. *Ibid.*, at 174-5.
23. B. Detweiler, 'A Quick Word About Technology Competence: The University at Buffalo School of Law's Microsoft Word Training Program', 25 *Perspectives: Teaching Legal Research and Writing* 97 (2017).
24. *See, e.g.*, R. Dhamija, J.D. Tygar & M. Hearst, 'Why Phishing Works', *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Association for Computing Machinery 2006), at https://doi.org/10.1145/1124772.1124861.
25. Janeček, Williams & Keep, above n. 7. *See also* 'Law and Computer Science: 2019-2020', available at: www.cs.ox.ac.uk/teaching/courses/2019-2020/LawandCS/, archived at: https://perma.cc/CZE5-RCN9 (last visited 30 January 2021).
26. 'CS50 for Lawyers 2019', available at: https://cs50.harvard.edu/law/2019/, archived at: https://perma.cc/ZBW7-D6TL (last visited 30 January 2021).

questions of substantive law brought about by such technology (which as noted above, are beyond the scope of this paper). The experiential part of the course was based on a group project (each group containing three students from each discipline) that resulted in a pilot product which was demonstrated in a pitch-like session to experts from the profession.[27]

The key idea behind the course design was 'to explore how computer scientists and lawyers of the future will need to work together'.[28] This fits well with the adopted strategy of limited explicit instruction about technology combined with the focus on interdisciplinary group work and exchange of knowledge between students. The law students participating in the Oxford course may have ended up with a less comprehensive picture of computer technology than those who took the Harvard course, but it is possible that they gained skills that are likely to be more practically useful for them as lawyers who interact with engineers. However, purely on the basis of the synopsis of the course content and the relative lack of introductions to basic aspects of computer technology there, it would not be surprising if law students struggled with technological concepts and had to do a considerable amount of independent study. Perhaps supplementing the Oxford course with the kind of contents that are covered in the Harvard course might have made it more effective, at least for some students.

In contrast to the Oxford course, Harvard's 'CS50 for Lawyers' explicitly covered issues like the basics of programming, algorithms, data structures, databases and cybersecurity.[29] The Harvard course even included some very basic programming tasks as at-home assignments. However, those tasks can be seen more as a familiarisation with the idea of programming than as teaching programming as a skill.

Simplifying matters a fair bit, one may be tempted to say that the Oxford course explicitly teaches lawyers to interact with engineers while leaving learning about technology more implicit (as a side effect of the interdisciplinary interactions and of explicit teaching about technology law), whereas the Harvard course teaches about technology explicitly and about interacting with technology specialists more implicitly. Neither of the courses turns lawyers into technology specialists and, in particular, equips them with sufficient knowledge and skills to develop software on their own, but given how unlikely it is that many lawyers today or in the near future will benefit from such in-depth grasp of technology, this is hardly a significant drawback. However, even the level of technology-related skills and knowledge that the two courses aim to provide is unnecessary for the vast majority of lawyers, while being very useful if not crucial for some small minority. Hence, there is a strong case for such technology courses to be offered as optional for law students, but not as core modules in general legal education.

2.3 Teaching Computer Programming (Coding)

The third and often discussed model of teaching technology is hands-on teaching of computer programming (coding). Teaching of coding to law students may be done without any direct connection with the law – for example through an elective module delivered by a computer scientist and without any adjustments for law students.[30] However, it may also be embedded in a law-specific module, for example on 'legal analytics' or 'Artificial Intelligence in law'.[31] A growing number of law schools offer modules with practical computer programming.[32] Those include modules offered within the Law and Technology Initiative at The University of Manchester,[33] the LLM in LegalTech at the University of Swansea,[34] 'Applied Legal Data Analytics & AI' course at the University of Pittsburgh School of Law,[35] 'Introduction to Quantitative & Computational Legal Reasoning' at the University of Iowa College of Law[36] and 'Computational Law' at the Stanford Law School (previously called 'Legal informatics').[37] Of those examples, only Iowa, Pittsburgh and Swansea state clearly that their teaching involves instruction in computer programming using a general-purpose programming language. In all three cases the language of choice is Python,[38] which is very popular in academic and commercial applications, especially in data science. Hence, the major benefit of learning Python is that it is a highly marketable skill in itself.[39] Naturally, the question remains as to the extent to which it is a marketable skill for law graduates in particular. As I noted in the previ-

27. Janeček, Williams & Keep, above n. 7.
28. 'Law and Computer Science: 2019-2020', above n. 25.
29. 'CS50 for Lawyers 2019', above n. 26.
30. Bocconi University in Milan offers such an elective course that is available to law postgraduates; *see*www.unibocconi.eu/wps/wcm/connect/4d8d627d-9249-4710-887f-585c71a3c861/Scheda+Programming+with+Python.pdf?MOD=AJPERES&CVID=moqkLuG, archived at: https://perma.cc/E5CP-88ZC (last visited 30 January 2021).
31. K.D. Ashley, 'Teaching Law and Digital Age Legal Practice with an AI and Law Seminar', 88 *Chicago-Kent Law* Review 783 (2013).
32. *See* R. Tromans, 'Legal Tech Courses', *Artificial Lawyer*, available at: www.artificiallawyer.com/legal-tech-courses/, archived at: https://perma.cc/B2AU-RFCZ (last visited 30 January 2021).
33. University of Manchester, 'Law and Technology Initiative (LaTI)', available at: www.law.manchester.ac.uk/research/themes/law-money-technology/law-technology-initiative-2/, archived at: https://perma.cc/N26Z-M765 (last visited 30 January 2021).
34. University of Swansea, 'LLM in "LegalTeach"', available at: www.swansea.ac.uk/postgraduate/taught/law/llmlegaltech/, archived at: https://perma.cc/K66D-ZQXW (last visited 30 January 2021).
35. University of Pittsburgh School of Law, 'Applied Legal Data Analytics & AI', available at: www.law.pitt.edu/academics/courses/catalog/5719, archived at: https://perma.cc/ZM2U-33PT; https://luimagroup.github.io/appliedlegalanalytics/, archived at: https://perma.cc/68QT-4ZLG (last visited 30 January 2021).
36. P. Gowder, 'Introduction to Quantitative & Computational Legal Reasoning', available at: https://sociologicalgobbledygook.com/, archived at: https://perma.cc/HNN8-L7GG (last visited 30 January 2021).
37. Stanford Law School, 'Computational Law', available at: https://law.stanford.edu/courses/computational-law/, archived at: https://perma.cc/Q4NQ-JY85; http://complaw.stanford.edu/, archived at: https://perma.cc/8CZP-RRA9 (last visited 30 January 2021).
38. G. Van Rossum and F.L. Drake, *Python 3 Reference Manual* (2009).
39. *See* Stack Overflow, 'Developer Survey Results 2019', available at: https://insights.stackoverflow.com/survey/2019, archived at: https://perma.cc/W9UL-LTL6 (last visited 30 January 2021).

ous section of this article, arguably very few lawyers do and will benefit from being able to code. However, the minority who will benefit from it in the near future will likely benefit the most from knowing general-purpose languages such as Python or JavaScript.

The University of Pittsburgh's 'Applied Legal Data Analytics & AI' shares some of the teaching strategy with the Oxford course discussed in the previous subsection.[40] As in the Oxford course, Pittsburgh students are assessed through a project to be prepared in interdisciplinary groups meant to represent both legal and engineering competencies.[41] The main difference is that the Pittsburgh course requires all participants to complete practical programming and data analysis tasks. However, the course does not include instruction in the basics of programming in Python. To participate in this course, students are expected to either have a background in Python programming or learn the language on their own within the first few weeks of the course. The Pittsburgh course is law specific, but it is open to non-law students. Notably, their approach is focused on data analytics and thus covers an introduction to machine learning (and natural language processing in particular) as well as programmatic extraction and transformation of data originating from legal texts.

'Introduction to Quantitative & Computational Legal Reasoning' at the University of Iowa College of Law has similar teaching aims as the Pittsburgh course.[42] One key difference is in assessment through individually completed problem sets, not through group work. The Iowa course also explicitly covers the basics of programming in Python (which is a prerequisite at Pittsburgh). Similarly, in terms of statistics, the Iowa course focuses on the fundamentals with less time devoted to advanced topics like machine learning. The courses share a focus on practical training in programmatic data analysis of legal texts.

The emphasis on training in statistics and related skills (e.g. extracting data from texts, 'cleaning' the data), clear in both the Iowa and Pittsburgh courses, makes those courses particularly valuable. Students who complete those courses gain not only a capacity to code, which may or may not be of practical use to them in their future careers. They also gain significant numerical literacy, greater than in the first teaching model discussed in Section 2.1.

An additional benefit of learning data analytics and computational modelling, while focusing on case law or on legislation, is that it may help students gain a deeper understanding of the law. As Kevin Ashley argued, developing or reverse engineering computational models of law or legal reasoning forces us to make explicit many issues that lawyers tend to do intuitively and that students are often expected to grasp without having them explained.[43]

The Stanford course took a noticeably different path and focused on teaching the fundamental concepts of computer programming on the example of a logic programming language, Epilog. Epilog is relatively unlikely to be used outside of a teaching context but is arguably well suited for learning of fundamental concepts of computer programming.[44] Epilog can be relatively easy for beginners to learn and use because it follows the syntax of symbolic logic. However, this benefit comes at the opportunity cost of not familiarising students with a different style of general-purpose programming that dominates academic and commercial uses, represented, for example, by Python. To some extent, learning the basic concepts of programming in Epilog should make it easier to start learning a language like Python. However, it is debatable whether choosing a language like Epilog over a language like Python is adequately beneficial even for beginners, especially given that Python skills are much more directly applicable outside of the teaching context. Moreover, the Stanford course does not cover quantitative methods of the kind the Iowa and Pittsburgh courses focus on. This may also contribute to the course being relatively easier for students than the other courses discussed here but, again, at the cost of more direct practical relevance.

A seemingly similar approach, but one that is actually very different from teaching to code, is to enable students to create 'apps' with the use of software tools that do not require programming skills in any of the general-purpose programming languages, or even a language like Epilog. Instead, such tools offer graphical interfaces ('no-code') and simple quasi-programming languages (a kind of 'low-code').[45] Depending on the software platform used, teaching assisted with such tools may potentially have similar benefits to the approach adopted by the Stanford course. That is, it may familiarise students with the basic concepts of programming like data structures and the logic of algorithms. This benefit comes with the limitations discussed in the Stanford case. However, it is also possible that the adopted no-code platform will be so simplified and 'user-friendly' that the students using it will not learn even the basic concepts of programming. At worst, the students may just learn 'how to click' through a particular interface of a particular piece of (soon-to-be-obsolete) software while gaining very limited transferable technology skills. Naturally, a course adopting this approach could deliver other learning outcomes than acquiring hard technology knowledge and skills, so limited value from the technological perspective does not necessarily mean that this is never a worthwhile teaching method. I do not discuss this possibility further as my concern in this article is exclusively on teaching technology.

40. *See also* Ashley, above n. 31.
41. 'Applied Legal Analytics & AI: Spring 2019', available at: https://luimagroup.github.io/appliedlegalanalytics/, archived at: https://perma.cc/68QT-4ZLG (last visited 30 January 2021).
42. Gowder, above n. 36.
43. Ashley, above n. 31, at 787-788.
44. 'Epilog', available at: http://epilog.stanford.edu/, archived at: https://perma.cc/Z5WV-P74U (last visited 30 January 2021).
45. Rostain, Skalbeck & Mulcahy, above n. 7, at 745.

2.4 Teaching a Particular Aspect of Technology – Other Than Programming

The fourth model of teaching about computer technology that I distinguish is advanced teaching of some specific aspect of technology, other than general-purpose programming. This could be done in a more descriptive or in a more practical way. For example, a course devoted to blockchain (decentralised ledger technology) could include exercises in programming smart contracts, but such a course could also be limited to helping students understand the technology at a higher level of abstraction.[46] Both ways may be suitable, depending on the backgrounds and aspirations of the students on the one hand and resources and knowledge of the instructors on the other.

The Cybersecurity course at Yale Law School is an excellent example of the fourth model.[47] It was designed by a law professor (Scott Shapiro) together with a cybersecurity expert (Sean O'Brien). The course content is not law specific and could potentially be offered to students of any discipline (perhaps even as an introductory course for computer science undergraduates). What is distinctive of the teaching method adopted in the Yale course is that it is centred on practical exercises, which include attacks on computer systems. In other words, the course aims to teach cybersecurity (defence and forensics) by teaching how to break into computer systems but does not require any previous knowledge of programming. Given the highly technical content of the course, students who do not already have a knowledge of computer programming, networking and system administration likely have to devote significant amounts of time for independent study to follow the course successfully. What makes the course challenging is also what makes it especially valuable. Completing the course equips students with the level of knowledge and skills in cybersecurity that is very rare among those who are not cybersecurity specialists and should be very helpful in a wide range of career paths. However, this level of expertise is arguably greater than what most law graduates are likely to really benefit from – or at least the course might not offer the best cost-benefit ratio for many students, which may suggest that there is space for a less technical cybersecurity course to be offered alongside this one.

The biggest concern of choosing the fourth model of teaching about technology is that it is possible that the students will not benefit as much from learning in detail about the particular aspect of computer technology as they would from gaining a broader perspective offered by courses that follow the second or even the third model. Hence, it is highly advisable to choose a topic of likely relevance to the students' future careers. From that perspective, a course on blockchain could be seen as too niche to be the only technology elective offered to undergraduate law students, but it may fit very well as one of several courses offered as part of a law and technology pathway or postgraduate degree.[48] On the other hand, cybersecurity is undoubtedly an issue of paramount importance to virtually all career pathways a law graduate may want to pursue. The key question regarding teaching cybersecurity concerns method: the very in-depth technical approach taken by the Yale course may be not only hard for some education providers to implement, but also unsuitable for all students (owing to differences in preparation, predispositions and career aspirations). One potential solution to that is to offer a choice of 'basic cybersecurity' (i.e. the first teaching model discussed in Section 2.1) and 'advanced cybersecurity' as different pathways within a module or separate module (or non-module teaching method like the cybersecurity fire drills I suggested in Section 2.1).

As with the third model (teaching programming), the fourth model requires teachers who are technology experts. The available solutions are the same: from teachers who are interdisciplinary experts (in law and technology) to interdisciplinary teaching teams to non-law specific courses delivered solely by technology experts and potentially offered to non-law students together with law students.

3 How Much Should Lawyers Learn About Technology?

The question of how much technology current and aspiring lawyers should learn is being increasingly debated in academia,[49] on industry blogs and during industry conferences.[50] I suggested earlier in this article (Section 2.1) that, within the profession, it would be very

46. *See* D.M. Katz and N. Rosario, 'Blockchain Law Class', available at: www.blockchainlawclass.com/, archived at: https://perma.cc/XA3L-7B9Z (last visited 30 January 2021).
47. Yale Law School, 'Cybersecurity', available at: https://courses.law.yale.edu/courses/course/2793, archived at: https://perma.cc/27RP-Z3AP, 'Materials for Cybersecurity (LAW 20310) at Yale Law School', available at: https://github.com/seandiggity/yls-cybersec, archived at: https://perma.cc/PEZ9-8R8W (last visited 30 January 2021).
48. This, for example, is the approach taken by the Illinois Tech – Chicago-Kent College of Law, available at: www.thelawlab.com/courses, archived at: https://perma.cc/J3QG-SEEG (last visited 30 January 2021).
49. *See, e.g.*, Ashley, above n. 31; Katz, above n. 7; Williams, Janeček & Keep, above n. 7; Janeček, Williams & Keep, above n. 7; Smith and Spencer, above n. 7.
50. *See, e.g.*, J. Krause, 'Does Learning to Code Make You a Better Lawyer?', *ABA Journal*, 1 September 2016, available at: www.abajournal.com/magazine/article/lawyer_learning_code_zvenyach_ohm/, archived at: https://perma.cc/KTK6-6HBB (last visited 30 January 2021); L. Cheek, 'Lawyers Who Code', available at: www.legalgeek.co/tag/code/, archived at: https://perma.cc/W2FU-5QRA (last visited 30 January 2021); B. Inkster, 'Lawyers and Coding', *The Time Blawg*, 24 February 2018, available at: http://thetimeblawg.com/2018/02/24/lawyers-and-coding/, archived at: https://perma.cc/5R8C-XEKQ (last visited 30 January 2021); Lawtomated, 'To Code or Not to Code: Should Lawyers Learn to Code?', available at: https://lawtomated.com/to-code-or-not-to-code-should-lawyers-learn-to-code-3/, archived at: https://perma.cc/Q59Y-5QPM (last visited 30 January 2021); R. Tromans, 'Should Lawyers Learn To Code? If You Have a Good Use Case, Yes', *Artificial Lawyer*, 30 January 2021, available at: www.artificiallawyer.com/2019/06/14/should-

valuable to spread technological literacy at a relatively basic but still higher level than likely possessed by the vast majority of current lawyers. This should include effective uses of office and legal research software and standard means of online communication and cybersecurity. As I also noted, this may be paired with teaching numerical and information literacy, on the one hand, and training in business skills, service and product design and process management, on the other.[51] I consider this general proposition to be relatively uncontroversial.

The more difficult question pertains to training in more advanced aspects of computer technology, including computer programming or practical data analytics. It will thus be worthwhile to summarise the key reasons for and against it. Beginning with the latter, software engineering is complex and requires significant knowledge and skills to execute at a level needed to deliver software that is being used by consumers. This concern is valid but does not mean that non-engineers should not learn computer programming. The correct lesson to draw from it is that students from other disciplines (like law) face a choice. They could be among the few who are willing to invest very considerable efforts into becoming, so to speak, fully bilingual (on a par with specialist computer programmers). I emphasise that the effort required to achieve that level of competence is, for most, incompatible with full-time practice of law (or even full-time law studies). The alternative is to learn some computer programming up to a level providing a decent measure of understanding of behind-the-scenes workings of computer technology, which may help in working together with experts on legal technology projects or simply in advising clients on technology-related legal issues. However, what is clearly a misconception is that lawyers and law students are able to take a relatively short course and that this will enable them to become fully fledged producers of legal technology.

As in the case with basic programming skills, the concern about learning effort required is not a sufficient argument against learning basic data analytics skills. As illustrated by the Pittsburgh and Iowa courses, discussed in the previous section, it is possible to learn in a semester how to extract and analyse some useful information from legal texts in a programmatic way. It may be enough to perform some research tasks for academic or legal practice purposes. This, however, leads to the second main reason for caution, which is the question of opportunity to use the skills. It is likely that not many lawyers or law graduates will have opportunities to conduct legal research in a programmatic way. For most academic and legal practice purposes, the ready-made tools from the main legal information providers (like WestLaw, LexisNexis) are sufficient. Those tools are gaining new functionalities supporting legal research, making the need for self-programmed solutions obsolete in some circumstances. And even if one has a research question for which the self-programmed way would be more appropriate, they may face the problem of access to legal data (like texts of court judgments, hearing transcripts), which in some countries, like the United Kingdom, are not publicly available for machine processing.[52]

Turning to reasons in favour of learning advanced technology skills, I emphasise that *some* lawyers and law graduates do need them. A relatively small proportion of law graduates will be able to work as 'legal technologists', 'legal engineers' or 'quantitative legal analysts', developing software solutions for legal practice (programming themselves or working in interdisciplinary teams with programmers) or performing advanced legal analytics research (also in academia and in government).[53] Moreover, some kinds of practice of law do benefit from an intimate understanding of technology. However, just as practising construction law may benefit from vastly different non-legal expertise than practising law of patents for chemicals, it may be advisable for relevant computer technology to be taught in a more specialised way depending on the field of law, perhaps as an element of advanced optional law courses in those fields. 'Blockchain Law Class', developed by Katz and Rosario, which includes instruction both in technology and in relevant law, may serve as an example.[54]

Direct need for advanced technology skills, likely applicable to a minority of lawyers, is clearly the strongest argument for teaching such skills. There are also many other potential reasons to teach advanced technology, which by themselves are not strong enough to justify both the cost to law schools (law faculties) of providing such teaching and the significant opportunity cost to students. The opportunity cost is significant because to truly gain advanced skills, such as those taught, e.g. in Yale's Cybersecurity course, while starting from the average level of technological skill, requires at least as much, if not more, effort as mastering a core law subject. However, those reasons are worth considering, especially that in some measure they also count in favour of basic technology education of the sort discussed earlier in this section, where both provision costs and opportunity costs are lower.

One of those weaker reasons is that students may transfer skills from some aspects of computer technology, like computer programming, to legal research and writ-

lawyers-learn-to-code-if-you-have-a-good-use-case-yes/, archived at: https://perma.cc/43AV-9XJT (last visited 30 January 2021).

51. Janeček, Williams & Keep, above n. 7; Smith and Spencer, above n. 7.

52. For instance, British and Irish Legal Information Institute, which operates the www.bailii.org/ website and publishes UK court judgments expressly prohibits 'bulk downloading of documents' from their website; *see*www.bailii.org/bailii/copyright.html, archived at: https://perma.cc/S66Z-JBJP (last visited 31 January 2021).

53. Legal technology investment is growing and reached over 1 billion US dollars in 2018; N. Dolm, '713% Growth: Legal Tech Set an Investment Record in 2018', *Forbes*, 15 January 2019, available at: www.forbes.com/sites/valentinpivovarov/2019/01/15/legaltechinvestment2018/, archived at: https://perma.cc/SF7F-DQF7 (last visited 31 January 2021). *See also* Katz, above n. 7.

54. *See* Katz and Rosario, above n. 46.

ing.[55] For example, computer programming requires rigorous attention to detail, precision in writing and clarity in structuring documents. As Koch noted, both legal writing and computer programming are instances of 'rules-driven writing'.[56] In computer programming, the author usually receives immediate feedback on whether they are complying with the rules of the programming language, which may help instil a habit of meticulous attention to the applicable rules while writing. The acquisition of such transferable skills may be a welcome side effect of otherwise valuable teaching, but it can hardly justify, e.g., teaching of computer programming, where students could devote the same time to the direct study of legal research and writing.

Finally, advanced study of some specific aspects of computer technology – such as learning computer programming – involves learning about many related behind-the-scenes aspects of computer technology.[57] Even though current and future law students may be well-versed as consumers of technology, they rarely have enough understanding of how it works, for example, to make informed decisions regarding how their use of technology affects their privacy, which is a crucial issue given the requirements of client-lawyer confidentiality.[58] This is not a strong reason for learning advanced skills, because there are less costly ways of bringing about the benefit of more general awareness of how technology works (e.g. basic technology education).

4 Conclusions: What Should Legal Education Providers Do?

Both students (including students of continuing professional education) and legal education providers should reflect on what kind of computer technology education suits their particular circumstances. There may be a worry that education providers who decide to teach, for instance, computer programming to law students are merely 'bandwagon-jumping', without serious and systematic consideration of the benefits that it may bring.[59]

In this article, I have emphasised that teaching computer technology can be done in various ways and at various levels of depth and that those different ways and levels have different cost and benefit considerations. I suggested that there are strong reasons for all current and future lawyers to acquire proficiency in effective uses of office and legal research software and standard means of online communication, basic cybersecurity and at fundamentals of quantitative thinking and methods. I also argued that advanced technology topics, like computer programming, should be taught only to the extent that this is justified by the direct need for such skills and knowledge in students' future careers, which I predict to be true for only a minority of current lawyers and law students.

My discussion suggests a number of questions for further study. What are the outcomes of each of the teaching models discussed? Are graduates satisfied with that particular aspect of their education once they have some experience on the labour market? Does the teaching contribute to higher salaries or more satisfactory employment? Are lawyers who learned programming or data analytics any better at some typical legal tasks than others (the questions of transferability of skills)? It would be valuable to observe whether answers to those questions change over time.

What, then, should law schools (law faculties) do? On the one hand, investing in 'teaching to code' may be a successful marketing strategy as long as it remains a way by which law schools can differentiate themselves (i.e. if only some law schools offer it). Also, some law schools may be able to reduce the cost of providing computer technology education by benefiting from the expertise of computer science and engineering faculties within their institutions, e.g., by offering non-law specific technology education to law students (delivered by technology experts), without needing to develop law-specific modules. On the other hand, the question of how many law graduates will really benefit from more advanced technology training should be treated seriously. It may be worthwhile for some (a minority of) law schools to specialise in providing such advanced training. However, since relatively few jobs will benefit sufficiently from it and since those interested in learning advanced technology topics have access to a plethora of excellent online learning options (including free ones), most law schools should think twice about taking this route. What all law schools should do in terms of technology education is either provide training in what I referred to as the basic technological and numerical literacy or at least actively encourage students to learn them from some of the excellent internet resources available.

55. K.L. Koch, 'A Multidisciplinary Comparison of Rules-Driven Writing: Similarities in Legal Writing, Biology Research Articles, and Computer Programming', 55 *Journal of Legal Education* 234 (2005).
56. *Ibid.*, at 237.
57. M. Fenwick, W.A. Kaal & E.P.M. Vermeulen, 'Legal Education in the Blockchain Revolution', 20 *Vanderbilt Journal of Entertainment and Technology Law* 351, at 382 (2017).
58. *See e.g.* Tirumala, Sarrafzadeh & Pang, above n. 19; Thompson and others, above n. 19.
59. *See e.g.* A. Young-Powell, 'More Universities are Teaching Lawtech – But Is It Just a Gimmick?', *The Guardian*, 12 April 2019, available at: www.theguardian.com/law/2019/apr/12/more-universities-are-teaching-lawtech-but-is-it-just-a-gimmick, archived at: https://perma.cc/QJC6-DP26 (last visited 31 January 2021).

doi: 10.5553/ELR.000192 - ELR 2021 | No. 1