Special Issue: Should Data Drive Private Law

Technology and Regulation

Should Data Drive Private Law?

Christoph Busch*, Catalina Goanta**, Katarzyna Kryla-Cudna***, Monika Leszczynska**, Vanessa Mak*****

data
private law,
data governance

christoph.busch@uos.de

e.c.goanta@uu.nl

katarzyna.kryla-cudna@bristol.ac.uk

monika.leszczynska@maastrichtuniversity.nl

v.mak@law.leidenuniv.nl

This special issue tackles the question of whether and how data shapes private law. The development of new technologies enabled the generation, collection and processing of both personal and non-personal data on an unprecedented scale. The implications of this phenomenon for private law are threefold. One, how does data affect our understanding of technology regulation in private law relationships? Two, how does data affect the way in which private law is applied? Three, what is the role of data in the design of law from a public policy perspective that transcends doctrinal considerations relating to private law?

1. Introduction

This special issue tackles the question of whether and how data shapes private law. The development of new technologies enabled the generation, collection and processing of both personal and non-personal data on an unprecedented scale. The implications of this phenomenon for private law are threefold.

Firstly, we can think about how data affects our understanding of technology regulation in private law relationships: What sanctions should be imposed on a facial recognition company when it uses data obtained from consumers in a misleading way?[1] Should platforms be held liable for harmful content posted by their users?[2] Are contracts concluded between social media platforms and individuals 'for free' or do consumers render an equivalent to payment by providing their personal data?[3] These are only a few examples of the wide array of questions that arise in the application of private law to data, and that private law scholarship has taken a growing interest in during the past years. Data can drive private law, in this perspective, to the extent that it shapes the way in which current norms are interpreted. It also contributes to a long-standing question whether new legal issues merit the creation of new legal rules,[4] and opens the door for reforms to update existing interpretational frameworks. In this perspective, data is a *subject* of private law rules.

Secondly, we can reflect upon how data affects the way in which private law is applied. As to the application of private law, some of the resulting questions are: How should judges relate to scientific evidence relevant for determining more characteristics and preferences of the average consumer in a given industry?[5] How can public authorities such as consumer protection agencies use public interest technology to measure legal compliance? How can recommender systems be used for information disclosures?[6] What all these examples have in common is that data may be involved in the infrastructure and/or be the subject of quantitative analysis providing evidence that makes the application and enforcement of law possible. In this meaning, data is a *tool* for the application and enforcement of private law rules.[7]

1 Evan Selinger and Woodrow Hartzog, 'The Inconsentability of Facial Surveillance' (2020) 66 *Loy L Rev* 33.

2 Rory Van Loo, 'The New Gatekeepers: Private Firms as Public Enforcers' (2020) 106 *Va L Rev*.

3 Natali Helberger, Frederik Zuiderveen Borgesius & Agustin Reyna, 'The Perfect Match? A Closer Look at the Relationship Between EU Consumer Law and Data Protection Law' (2017) 54 *Common Market Law Review* 1427.

4 Lawrence Lessig, 'The Law of the Horse: What Cyberlaw Might Teach' (1999) 113 *Harvard Law Review* 501.

5 Hanna Schebesta and Kai Purnhagen, 'An Average Consumer Concept of Bits and Pieces - Empirical Evidence on the Court of Justice of the European Union's Concept of the Average Consumer in The UCPD', in Lucila de Almeida, Marta Cantero Gamito, Mateja Djurovic, Kai Peter Purnhagen (eds.) *The Transformation of Economic Law: Essays in Honour of Hans-W. Micklitz* (Bloomsbury 2019).

6 Madalena Narciso, 'The Regulation of Online Reviews in European Consumer Law' (2019) 27 *European Review of Private Law* 557.

7 See for instance the European Commission's endorsement of public authorities deploying so-called 'AI tools' for law enforcement, EC White Paper on AI, February 2020. https://ec.europa.eu/info/sites/default/files/commission-white-paper-artificial-intelligence-feb2020_en.pdf.

* Christoph Busch is Professor of European Private Law, European Business Law and Private International Law at Universität Osnabrück, Germany.

** Catalina Goanta is Associate Professor in Private Law and Technology at Utrecht University, the Netherlands.

*** Katarzyna Kryla-Cudna is lecturer in law at the Bristol Law School, Bristol University, United Kingdom.

**** Monika Leszczynska is Assistant Professor of Empirical Legal Research at Maastricht University, the Netherlands.

***** Vanessa Mak is professor of Civil law at Leiden University, the Netherlands.

Received 13 June 2022, Accepted 13 June 2022, Published: 17 June 2022.

Thirdly, from a public policy perspective that transcends doctrinal considerations relating to private law, there is also the possibility to consider the role of data in the design of law. The same evidence which could be used for the better understanding of how (private) law ought to be applied in practice is equally relevant when reconsidering the normative scope of that law in the light of its fitness. An illustration of this prospect is the concept of 'granular law', which could entail the redesign of law's core principles, such as fairness or certainty. Granular law, sometimes also referred to as personalized law, is 'precision law characterized by two primary features: individualization and machine-sorted information'.[8] In other words, granular law reflects a completely new paradigm of tailoring legal rules for individual use. From this perspective, data is a *parameter* of public policy design.

The contributions to this special issue discuss specific manifestations of data as a subject of, or as a tool for the application of private law rules. The focus is therefore on the first two perspectives set out above. Our introduction aims to provide a brief overview of the context within which the papers should be read, setting out also the questions that are bound to shape debates on data and private law in the coming years. In doing so, the introduction also embraces the third consideration relating to the design of private law, particularly with a view to addressing the changing nature of private law. If we perceive (private) law to be an operating system which relies on principles inherited from legal ages of either no or less data, questions arise as to the future of such a system. We therefore want to propose this background for a broader research agenda, guided by the following defining questions:

1. What architectural elements of our current legal system are concessions to an operating environment where data is not available widely and we have constraints to data processing?

2. How these architectural elements could be redesigned given an increasing availability of data and a rapid development of data analysis methods?

3. What do we gain and what do we lose if private law adopts, or reverts back to, a more granular approach to individuals' rights and interests?

The structure of this introduction is as follows. Section 2 examines the ways in which data challenges the current architecture of private law, connecting that question to existing scholarship on complexity and law, granularity, and certainty versus fairness. Section 3 sets out a research agenda in which we identify which questions, we think, will shape the discussion on data and private law in the coming years. We refer to issues identified by the authors of this special issue as well to broader research questions that we identify in the field of data and private law. Section 4 provides an overview of the contributions contained in this special issue. Section 5 concludes.

2. Complexity & cohesion (theoretical frameworks)

A transition to data-driven private law could make it possible to formulate and implement more "granular" legal norms tailored to the needs, preferences and capacities of individuals based on their data profiles.[9] This shift from an impersonal and generalized statement of legal rules to a more context-sensitive law could lead to a readjustment of the balance between individual fairness and legal certainty. In the past, as a general rule, a higher degree of individual fairness could only be achieved at the price of less legal certainty or, in economic terms, higher complexity costs. In the near future, advances in information technology could change this equation and redefine the "optimal complexity"[10] of legal rules. This new granular law would no longer be based on typifications and standardised normative models, but on data-rich profiles. The datafication of the legal process and the resulting "granularization" of the law will affect the relationship between law and individuality on several levels.

In this sense, data-driven law not only changes how the law "views" individual human beings, but also inanimate objects. The law could take inspiration from developments in other areas where an abundance of data and new data-processing technology enables innovations such as personalized healthcare or precision farming. Another example could be "predictive maintenance". Here, real-time data profiles about machinery (sometimes referred to as "digital twins" existing in a "mirror world"[11]) make it possible to treat elevators and escalators as "individuals". In a legal context, such a change of perspective could have a disruptive effect for example on traditional concepts such as the numerus clausus principle in property law and give rise to a new "granular property law".

One key for understanding the recalibration in the relationship between individuality and the law is provided by recent sociological research on the impact of the digital transformation.[12] According to some observers the pervasive trend towards personalisation – some would even say "hyper-individualisation" – results in a fundamental change in the relationship between the individual and society. The ubiquitous quantification and "datafication" of individuals and their social relations leads to a dissolution of collective categories (e.g. citizen, consumer) and shifts the focus towards finely chiselled differences between individuals. In the emerging "granular society"[13] (or "society of singularities"[14]) the individual is no longer considered as a representative of a certain social group defined by general criteria based on an average model, but rather as a singular and solitary being defined by a cloud of data points. According to some observers this shift from generality to singularity is a symptom of a more fundamental "crisis of the general" caused by the advent of big data.[15]

Legal and societal categorisations, therefore, are losing their significance as data-driven technologies push towards more individualised approaches to the qualification of legal subjects and the rights and duties attached to those qualifications. Data enables us to view the "consumer" or the "citizen" as an individual, rather than a representative of a broader category. That change of perspective gives

8 Omri Ben-Shahar and Ariel Porat, *Personalized Law: Different Rules for Different People* (Oxford University Press 2021) at 19.

9 See Christoph Busch and Alberto De Franceschi, 'Granular Legal Norms: Big Data and the Personalization of Private Law', in: Vanessa Mak, Eric Tjong Tjin Tai, and Anna Berlee (eds), *Research Handbook on Data Science and Law* (Elgar 2018) 408-424.

10 See Louis Kaplow, A Model of the Optimal Complexity of Legal Rules, 11 *Journal of Law, Econ. & Organization* 150 (1995).

11 See David Gelernter, *Mirror Worlds* (Oxford University Press 1993).

12 See Christoph Busch, 'Algorithmic Regulation and (Im)Perfect Enforcement in the Personalized Economy', in Christoph Busch and Alberto De Franceschi (eds) *Algorithmic Regulation and Personalized Law* (Hart 2021), 279, 280-81.

13 Christoph Kucklick, *Die granulare Gesellschaft: Wie das Digitale unsere Wirklichkeit auflöst* (Ullstein 2014); see also Steffen Mau, *The Metric Society: On the Quantification of the Social* (Polity 2019).

14 Andreas Reckwitz, *The Society of Singularities* (Polity 2020).

15 Ibid at 24.

rise to questions. Can the legal architecture of private law do justice to the new complexities of such individualised approaches? Or may we expect a return to the pre-nineteenth century general private laws and lose the social oil that the categorical protection of consumers and workers has embedded into (European) private laws?[16] Likely, we will end up somewhere in between. There is however a wealth of questions to consider for the development of private law in an age of increasing datafication.

These questions also prompt a growing number of ethical considerations. Certain features of data-driven private law may not only be desirable, but rather necessary, and an example in this respect is the need to apply and enforce consumer protection on digital markets. If a marketplace does not take the necessary steps to prevent unsafe goods, it must be held liable for consumer protection violations; yet the picture of one or two national consumer inspectors manually checking hundreds out of thousands of products to investigate if violations take place in practice is a defeating reflection of the information, power and resource asymmetry between digital market actors and public authorities. To be able to scale operations, digital investigations and enforcement activities need to mirror the scale of digital markets.[17] Yet this example entails the further automation of the administrative state, which raises a lot of concerns for potential harms arising out of algorithmic decision-making.[18] Similar concerns arise not only in the enforcement dimension of the law, but also in its interpretation and design. This is surely not the first time in history when the complexity of codification is observed, and technological solutions proposed.[19] Still, with further reflections on transparency and accountability, it may still be possible to reap some of the benefits of data-driven private law, while avoiding its greatest perils.

3. Structuring a research agenda

We identify three main strands that can provide the basis for a research agenda, reflecting different angles of the current debate on how data shapes private law:

1. Consumer vulnerability and digital asymmetry
2. Beyond consumer contracts
3. The use of data in the design and application of law.

3.1. Consumer vulnerability and digital asymmetry

Data can be instrumental in developing individualized or personalized consumer concepts. The datafication of consumer markets therefore challenges the architecture of consumer law, which is grounded in the idea that consumers as a category deserve protection, as they are generally the weaker, less-informed party in a transaction. That general conceptualization has worked well to increase consumer protection around the world from the mid-twentieth century onwards. At the same time, the reference to a general category of 'consumers' means that consumer protection is sometimes over-inclusive and sometimes under-inclusive. For example, the buyer of a countryside mansion of 2 million euros will benefit from consumer protection, even if they are in fact able to set the terms of renovations to the house and will also be able to financially carry the costs of damage or defective performance should these occur. One may similarly wonder if consumer law should treat a person as a consumer in a credit relationship even if that person is a lawyer and the loan is secured on a building owned by the borrower's law firm.[20] At the same time, the consumer concept, which in the EU is built around the notion of an average consumer who is 'reasonably well-informed, reasonably observant and circumspect', fails to protect consumers who are less literate or less aware of their rights. The use of a 'vulnerable consumer' category only goes so far in addressing this problem.[21]

What can we gain, therefore, by using data to develop personalized approaches to consumer protection? And what do we stand to lose?

Research in this area can focus on whether it is possible and desirable to develop a contextualized approach to consumer protection, connected to a more diversified palette of consumer concepts. One concept that deserves attention is the 'prosumer'. The emergence of non-professional traders who offer products and services, in other words consumers acting as producers or 'prosumers', has been observed since 1980.[22] It has taken a flight since online platforms started facilitating transactions between consumers in the last decade.[23] While this creates new opportunities for consumers to become active in markets for goods and services, it also creates new challenges. For the prosumer, the question is whether or when they become subject to rules governing professional traders. Should a prosumer be expected to provide the same standard of quality or service as a professional trader? Should they be subject to tax duties for small businesses? Which criteria determine when a prosumer changes from a natural person not acting in the course of a business – a hobbyist – into a professional trader? The European Court of Justice has given some guidance in the Kamenova case as to which criteria should be considered to determine whether we are dealing with a professional trader.[24] However, the list is not exhaustive and outcomes in individual cases will depend on applications by national courts. The European legislator has also acknowledged that specific rules should apply to prosumers in the platform economy, introducing new information duties in its 2019 Modernisation Directive, amending the Consumer Rights Directive and the Unfair Commercial Practices Directive.[25] However, these rules also do not provide objective factors for determining when a prosumer should be considered a professional trader. Nor do the rules specify whether the same legal obligations should be imposed upon prosumers as on professional traders. The analysis of prosumer data, that is the activity of prosumers on online platforms, could provide us with information on their characteristics. That could be a step towards determining in which ways the existing legal framework is sufficiently tailored to their interests, and in which respects

16 Hans-W Micklitz, *Brauchen Konsumenten und Unternehmen eine neue Architektur des Verbraucherrechts?* Gutachten A zum 69. Deutschen Juristentag (CH Beck 2012) 11-12, 26, with reference to Otto von Gierke, *Die soziale Aufgabe des Privatrechts* (Springer 1889).

17 See for instance, Commission Regulation 2017/2394 of the European Parliament and of the Council of 12 December 2017 on Cooperation Between National Authorities Responsible for the Enforcement of Consumer Protection Laws and Repealing Regulation (EC) No 2006/2004 [2017] OJ L 345.

18 See for instance, Reuben Binns, 'Algorithmic Decision-Making: A Guide for Lawyers' (2020) 25(1) Judicial Review 2; Therese Enarsson, Lena Enqvist and Markus Naarttijarvi, 'Approaching the Human in the Loop - Legal Perspectives on Hybrid Human/ Algorithmic Decision-Making in Three Contexts' (2022) 31 Info & Comm Tech L 123.

19 See the concept of codification as a structural science building on natural language processing, Cary G. DeBessonet, 'A Proposal for Developing the Structural Science of Codification' (1980) 8 Rutgers Computer & Tech LJ 47.

20 CJEU, case C-110/14 Costea, ECLI:EU:C:2015:538.

21 See e.g. Directive 2005/29/EC concerning unfair business-to-consumer practices in the internal market [2005] OJ L149/22, Art. 5(3).

22 Alvin Toffler, The Third Wave (Bantam Books 1980).

23 Ian Brown and Chris T Marsden, Regulating Code: Good Governance and Better Regulation in the Information Age (MIT Press 2013).

24 CJEU, case C-105/17 Kamenova, ECLI:EU:C:2018:808.

25 Directive (EU) 2019/2161 as regards the better enforcement and modernisation of Union consumer protection rules [2021] OJ L328/7.

improvements may be made. For example, should the platform bear liability alongside the prosumer for non-conformity of goods or product liability?

This is one example of a changing perspective on consumer categories. Another question, bringing us back to the core question of this contribution on data and private law, concerns the concept of the digital consumer. It has been suggested that in digital markets all users are vulnerable, as they are subject to a structural imbalance. Businesses use data-driven technologies to learn how consumers behave and to influence their purchasing decisions, often without consumers even realizing that their behaviour is being manipulated.[26] This can be dubbed a "digital asymmetry".[27] It raises questions as to what consumer protection should look like in a context where information duties no longer work, as the ways in which tech providers use, process and monetize data are becoming increasingly oblique. Consumer law and data protection law could go hand in hand in seeking to increase the protection of consumers in digital markets.[28] The Unfair Commercial Practices Directive (UCPD) and the Unfair Contract Terms Directive (UCTD) provide frameworks for information, transparency, and liability or the setting aside of unfair terms.[29] These directives can provide the basis for a reassessment of consumer protection in digital markets. Questions for research can be: Should the existing categories of 'average' and 'vulnerable' consumers be complemented by a new framework for digital consumers? What should be the baseline of such a framework: should consumers and prosumers be free to decide, on the basis of consent,[30] to exchange their data for goods or services or should restrictions be imposed in the light of the pervasive digital asymmetry through which big tech companies can exploit consumers? Can the development of a new private law architecture benefit from data-driven research to design "granular" protection to consumers in digital markets?

3.2. Beyond consumer contracts

Our examples in the foregoing paragraphs focused on consumer protection. The significance of data as a driving force to a reform, however, is not limited to rules governing consumer contracts but concerns also the general rules of private law. Contributions to this special issue offer examples of how the general contract and property law rules could tackle the problems arisen in the context of data-driven technologies. Sjef van Erp and Koen Swinnen examine the traditional principles of property law in the context of the use of and access to co-generated data. The authors claim that even though data as a legal object differ considerably from tangibles and intangibles that are traditionally recognized as legal objects, an analogy to the well-established property law doctrines and its rationales can help to solve the problem of distribution of economic benefits regarding co-generated data. Antonio Davola examines the existing rules on defects of consent and misrepresentation in order to evaluate whether these rules can accommodate issues arising from the use of personalized practices. He argues that the current rules on defective consent can serve as an instrument to promote social justice and fairness where tailored commercial techniques are used. The rules on defective consent are meant to embrace diverse factors which may affect the parties' decision to enter into a contract and to assent to the specific content of a contract. Thus, where actions undertaken by one party artificially affect the understanding of contractual terms by the other party, the rules on defects of consent can offer a tool to eliminate the negative consequences of this conduct.

For future research, arguably other private law doctrines also offer openings for addressing the consequences arising from the use of data driven technologies. In particular, the duty to act in good faith recognized in civil law jurisdictions appears to be flexible enough to tackle some problems arising from the use of personalized practices. In civil law jurisdictions, the parties are required to act in good faith at all stages of the life of a contract, including the precontractual stage and the process of performance of a contract.[31] In general terms, this means that a party should take the interest of the other party into account. Consequently, a party to a contract should not take actions that could harm the legitimate interests of the other party. Personalized commercial practices can potentially constitute an example of such actions.

Furthermore, in some jurisdictions, the exploitation of a party's limitations, as an act against the principle of good faith or against public policy, may also qualify as a tort (delict). For example, section 826 of the German Civil Code refers to the liability for damage caused by an individual who, in a manner contrary to the public policy, intentionally inflicts damage on another person. German courts have applied this provision in the area of capital markets to sanction issuers who take advantage of the lack of financial literacy of investors in order to sell economically disadvantageous financial products.[32] It has been argued that, analogously, this provision could also be invoked in the context of tailored commercial practices where businesses rely on informational asymmetry to exploit weaknesses of the other party to the contract.[33]

Based on these examples, it seems that the existing private law rules are adaptable and have the potential to respond to, at least, some problems arising from the increasing use of data-driven technologies.

3.3. The use of data in the design and application of law

Fast generation and easy access to the vast amounts of data opened up new possibilities of its use not only by companies but also by legislators, courts and law enforcement agencies. The mere accessibility of data does not yet render the data useable for these purposes.

26 Shoshana Zuboff, The Age of Surveillance Capitalism (PublicAffairs 2019).

27 Agnieszka Jablonowska, Maciej Kuziemski, Anna Maria Nowak, Hans-W Micklitz, Prezmyslaw Palka and Giovanni Sartor, 'Consumer Law and Artificial Intelligence. Challenges to the EU Consumer Law and Policy Stemming from the Business' Use of Artificial Intelligence', final report of the ARTSY project, EUI Working Paper LAW 2018/11.

28 Catalina Goanta, 'European Consumer Law: The Hero of Our Times' (2021) 10 Journal of European Consumer and Market Law 177.

29 Directive 2005/29/EC concerning unfair business to consumer practices in the internal market [2005] OJ L149/22; Council Directive 1993/13/EEC on unfair terms in consumer contracts [1993] OJ L95/29.

30 Art 6 and Art 4 sub 11 of Regulation (EU) 2016/679 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation) [2016] OJ L119/1 (GDPR).

31 For a broader analysis of the principle of good faith in contract law see Reinhard Zimmermann, Simon Whittaker, and Mauro Bussani (eds.), Good Faith in European Contract Law (Cambridge University Press, 2000). See also Christian Eckl, Treu und Glauben im spanischen Vertragsrecht (Mohr Siebeck, 2007); Karl Riesenhuber, System und Prinzipien des Europäischen Vertragsrechts (De Gruyter Recht, 2003), p. 398 ff; Ole Lando, 'Is Good Faith an Over-Arching General Clause in the Principles of European Contract Law?', 15 European Review of Private Law 841 (2007); Martijn Hesselink, 'The Concept of Good Faith', in Arthur Hartkamp et al. (eds), Towards a European Civil Code (Kluwer Law International, 2011), 619.

32 See, e.g., BGH, Case XI ZR 170/07, NJW 2008, 1734, 1737 para. 29.

33 Philipp Hacker, 'Personal Data, Exploitative Contracts, and Algorithmic Fairness: Autonomous Vehicles Meet the Internet of Things', 7 International Data Privacy Law 266, 280 (2017).

Instead we need to rely on the techniques of data processing and analysis that have been developed in order to extract meaningful and helpful information out of large data sets and different types of data. For instance, more recent advances in the natural language processing (based on e.g., deep learning)have enabled automatic processing on unprecedented scale and new uses a of text as data.[34] Large datasets can now be employed to train machine learning algorithms to perform different types of tasks such as predictions, classifications or clustering.[35] Although very promising, these new developments also have limitations that need to be taken into account when relying on data in application and enforcement of law or design of public policy. This section highlights some of these limitations and formulates questions that still need to be addressed in order to make sure that the use of data in private law is corresponding to the conclusions that can be drawn based on its analysis.

As introduced above, automatic text analysis is one of the data analysis developments that opens up promising avenues for the use of data in private law. The article by Fabiana di Porto, Tatjana Grote, Gabriele Volpi and Riccardo Invernizzi in this special issue provides an example of such a use. The authors analyzed submissions by stakeholders in a consultation process over Digital Markets Act and Digital Services Act. They employed word embeddings to understand the context in which stakeholders use various terms related to transparency duties. The authors suggest that the methods they developed together with other text analysis techniques could complement the analysis of stakeholders' inputs conducted by public officials. Furthermore, it could provide important insights to legal scholars investigating the rationale of legislative acts. CLAUDETTE is another example of the application of automated text analysis that can be helpful in monitoring potentially unfair clauses in terms of use and provisions of privacy policies that may be incompatible with GDPR.[36] The tool could be used not only to empower consumers[37] but also by law enforcement agencies.[38]

Despite the clear advantages, one should also realize the limitations of these analysis. First, as Di Porto and co-authors rightly notice, interpretation of the results of their analysis will always involve a certain dose of subjectivity, since the output of the analysis provide merely a quantitative measure of a difference in the use of same words by different stakeholders. The analysis also provides information about the words usually occurring in close proximity to the terms of interest. The interpretation of these results will need to be conducted by a human, and thus, will necessarily involve a subjective element. A similar issue applies to the tools developed within CLAUDETTE project. CLAUDETTE informs its users whether the term is potentially unfair and provides a confidence score of this prediction. This means that the final determination of the term's unfairness will need to be conducted by a human. The tool, however, speeds up the screening process identifying suspicious clauses.

Given that the results of automatic text analysis still require human input to interpret them or make final conclusions, it is necessary to reflect upon a number of questions such as:

- How should these results be communicated in order to be interpretable and understandable by their addressees?
- What preparation does this interpretation require? For instance, what kind of training in data science will public officials relying on the results of automatic analysis of submissions in the consultation process need in order to interpret these results?
- How does information provided by the automatic text analysis affect people's behavior? Will they trust it or will they rather be suspicious? Will they be willing to rely on it?

Another type of information that can be generated using large data sets are machine learning based predictions. These types of analysis are common in B2C relationships where companies rely on them to predict consumer behavior such as willingness to pay for a specific product. Prediction is therefore a cornerstone of all personalized business practices such as targeted advertising or personalized pricing.[39] Outside of private law, it is also increasingly used by law enforcement agencies and courts (see, for instance, COMPAS,[40] OxRec,[41] SyRi[42] or "Sensing Project"[43]). As suggested above, predictions could also be used in private law to design 'granular' norms or to personalize application of consumer protection law. Although the use of predictive analysis by agencies and courts have been subject to a heavy criticism elsewhere[44], here we would like to point to one issue that has been rarely mentioned in legal scholarship, i.e., that predictions generated based on the analysis of large data sets do not yet tell us anything about the causal relationship between the outcome and various factors that the algorithm identified as being predictive of the outcome.[45]

34 Justin Grimmer and Brandon M. Stewart, 'Text as Data: The Promise and Pitfalls of Automatic Content Analysis Methods for Political Texts' (2017) 21 *Political Analysis* 267

35 See for instance Susan Athey and Guido W. Imbens, 'Machine Learning Methods That Economists Should Know About' (2019) 11 *Annual Review of Economics* 685

36 Marco Lippi and others, 'CLAUDETTE: an automated detector of potentially unfair clauses in online terms of service' (2019) 27 *Artificial Intelligence and Law* 117; Giuseppe Contissa and others, 'CLAUDETTE meets GDPR: Automating the evaluation of privacy policies using artificial intelligence' https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3208596.

37 Marco Lippi and others, 'Consumer protection requires artificial intelligence' (2019) 1 *Nature Machine Intelligence* 168; Ruta Liepina and others, Explaining potentially unfair clauses to the consumer with the CLAUDETTE tool (CEUR-WS 2020).

38 Further examples of how machine learning based tools can empower both consumers and consumer protection agencies in the context of algorithmic price discrimination see the article in this special issue by Mateusz Grochowski and co-authors, https://doi.org/10.26116/techreg.2022.004.

39 For more details on these practices see articles in this special issue by Mateusz Grochowski and co-authors, and Antonio Davola, https://doi.org/10.26116/techreg.2021.007.

40 William Dieterich, Christina Mendoza and Tim Brennan, 'COMPAS risk scales: Demonstrating accuracy equity and predictive parity' (Northpointe 2016).

41 Seena Fazel and others, 'Prediction of violent reoffending in prisoners and individuals on probation: a Dutch validation study (OxRec)' (2019) 9 Scientific Reports 841.

42 Koen Vervloesem, 'How Dutch activists got an invasive fraud detection algorithm banned' (AlgorithmWatch, 6 April 2020) https://algorithmwatch.org/en/syri-netherlands-algorithm.

43 Amnesty International, 'We sense trouble: Automated discrimination and mass surveillance in predictive policing in the Netherlands' (Amnesty International, 29 September 2020) https://www.amnesty.org/en/documents/eur35/2971/2020/en.

44 See, for instance: J. Dressel and H. Farid, 'The accuracy, fairness, and limits of predicting recidivism' (2018) 4 *Science Advances*; Frank Pasquale and Glyn Cashwell, 'Prediction, persuasion, and the jurisprudence of behaviourism' (2018) 68 *University of Toronto Law Journal* 63 Christopher E. Church and Amanda J. Fairchild, 'In Search of a Silver Bullet: Child Welfare's Embrace of Predictive Analytics' (2017) 68 *Juvenile and Family Court Journal* 67.

45 For further discussion and potential solutions to this issue in policy making and law, see: Susan Athey and Guido W. Imbens, 'The State of Applied Econometrics: Causality and Policy Evaluation' (2017) 31 *Journal of Economic Perspectives* 3; J. Kleinberg and others, 'Prediction Policy Problems' (2015) 105 *Am Econ Rev* 491; Susan Athey, 'Beyond prediction: Using big data for policy problems' (2017) 355 *Science* 483; Sendhil Mullainathan and Jann Spiess, 'Machine Learning: An Applied Econometric

This limitation can be best illustrated with an example of businesses' interventions targeted towards customers predicted to be the most likely to churn (i.e., to stop doing business with a given company).[46] In such a situation, businesses often rely on predictions generated by algorithms developed using businesses' past data about customers who left the company. Based on this data, the algorithms learn which customers' features are predictive of churning. Businesses then use such an algorithm to analyze the existing customers' pool and identify those who are most likely to leave. Then, they target various interventions, such as special sales techniques, to those customers in order to retain them. Ascarza studied this issue systematically.[47] She showed that rather than predicting which customers are likely to churn and then target these customers, the businesses should focus on predicting which customers are most likely to positively react to businesses' interventions. The customer pool might be heterogenous with this respect, i.e., some customers may react positively to specific retention practices while for others such a practice might be ineffective or even have the opposite effect. This implies that in order to successfully implement an intervention, one needs to first understand whether and for whom such an intervention is effective. Rather than with predictions, these insights can be generated using experimental methods such as field experiments implemented by Ascarza to illustrate the limitations and challenges of predictions. These limitations raise a number of questions that are relevant for the use of predictions in the design and application of private law, such as:

How to identify and distinguish prediction problems and causality problems in private law (such as, for instance, predicting which consumer is more likely to fall prey to unfair practices or designing/applying norms that will protect them)?

Generating insights about causal relationships is often more time consuming than generating predictions. Given that legal actors applying and enforcing law often need to act quickly and have restricted resources, in which situations should we merely rely on predictions although the issue at hand might require identifying causal relationships?

4. The content of this special issue

The research agenda set out in this Introduction proposes questions going well beyond the scope of this special issue. Our authors, however, provide analyses of the ways in which data should shape private law that can set the stage for debates in the future. The focus of the contributions is on three perspectives: the ways in which private law rules should be applied to data as a subject, and the ways in which data can be used as a tool in research that will shape private law, and the role of data as a parameter of public policy design.

Taking these different perspectives, the papers coming up in this special issue explore the implications of data for private law. The first paper – by Antonio Davola – addresses the question of how the law deals with the use of data by businesses in their interactions with consumers. Davola analyzes existing private law rules on defective consent and argues that these rules could offer a potential protection to consumers when they are targeted by businesses' personalized commercial practices. He juxtaposes this solution with those provided by consumer law and data protection regulation, as well as competition law. Another question of legal application is addressed by Swinnen and van Erp. They explore the treatment of data as legal objects. Relying on existing principles of property law, the authors analyze the legal status of co-generated data (i.e., when more than one party contributed to data creation).

Exploring the second perspective – how data can be used in the development of private law – Fabiana di Porto and her collaborators demonstrate how data can be relied on in the legislative process. Di Porto et al. propose an automated text analysis method to extract information from contributions submitted by stakeholders in the process of public consultation. Specifically, the authors compare the use and understanding of core terms by various stakeholder groups consulted when developing proposals of Digital Markets Act and Digital Services Act.

Finally, Grochowski and co-authors touch upon the third perspective on whether and how data shape private law in a more normative sense. They analyze policy implications of the use of data for setting the prices of products and services in business-to-consumer transactions. The authors conclude that the current regulatory solution in form of information duties may be insufficient to deal with the consequences of algorithmic pricing and suggest that the way forward is to create data-driven consumer empowerment tools that will help consumers detect and avoid the disadvantages of algorithmic pricing.

5. Conclusion

The evolution of private law seems finally to be catching up with the development of data-driven technologies and data analysis methods. In some ways, this leads to a fundamental reassessment of the architecture of private law. When data becomes available on a large scale, and when data analysis methods improve, it seems feasible for private law to provide more individually tailored rules than it currently does. Instead of addressing "consumers" and "businesses" as categories, their legal interests can be addressed through granular law, that is rules tailored to their individual characteristics. In a sense, that is a shift back towards pre-nineteenth century general private laws. If private law indeed moves in that direction, we should consider whether that is the desirable route for the future. What do we lose if we relinquish the categorial protection of consumers? What alternative mechanisms or designs for private law are available?

The contributions published in this special issue provide an overview of the state of play of research on data and private law. We hope that they will be a stepping stone towards a better understanding of private law design and application in the digital age.

Approach' (2017) 31 *Journal of Economic Perspectives* 87; Ryan Copus, Ryan Hübert and Hannah 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 Institute of Science 2019).

46 This example is also referred to by Athey, 'Beyond prediction: Using big data for policy problems'. Athey discusses in more detail the challenge of causality that is faced by policymakers introducing interventions based on predictions generated using big data.

47 Eva Ascarza, 'Retention Futility: Targeting High-Risk Customers Might be Ineffective' (2018) 55 *Journal of Marketing Research* 80.