



Universiteit
Leiden

The Netherlands

Conclusions on data science and law

Mak, V.; Tjong Tjin Tai, T.F.E.; Berlee, A.

Citation

Mak, V., Tjong Tjin Tai, T. F. E., & Berlee, A. (2024). Conclusions on data science and law. In *Research handbook in data science and law* (pp. 388-400). Cheltenham: Edward Elgar.
doi:10.4337/9781035316458.00026

Version: Publisher's Version

License: [Leiden University Non-exclusive license](#)

Downloaded from: <https://hdl.handle.net/1887/4175580>

Note: To cite this publication please use the final published version (if applicable).

19. Conclusions on data science and law

Vanessa Mak, Eric Tjong Tjin Tai and Anna Berlee

1. INTRODUCTION

At the outset of this book the question was put forth: do data-driven technologies require regulation, and *vice versa*, how does data science advance legal scholarship? While there is no resounding answer one way or the other to the first question, we can deduce from the analyses put forward by our authors that the rise of the so-called data economy does pose challenges to regulators, which have only intensified, or are at least more prominently on display, with the advent of generative AI. The challenges are diverse and the answers to the—many—questions put forward in the previous chapters will likely be manifold. We nevertheless perceive some common issues that regulators are likely to encounter in each of the areas of law that were examined. We summarize them in section 2 of this conclusion, and elaborate some thoughts on the direction in which future research on the regulatory aspects of data-driven technologies may be headed.

The second part of the book considered the increasing use of data science and the rise of generative AI in legal scholarship and legal practice. Here also, challenging questions for future research have been identified by our authors. While the replacement of lawyers and judges by robots may still be a science-fiction dream (or nightmare), the use of data analysis and (generative) AI in law is changing the way in which we approach legal (research) questions. We summarize the tentative findings in this field in section 3 of this conclusion.

We round up the book with a final question: with data science and law, are we witnessing the emergence of a new discipline?

2. REGULATING DATA SCIENCE

In the introductory chapter we presented three aspects that can serve as pointers when we ask whether data-driven technologies require regulation. These three are justice, fundamental rights, and legal classification. They resurface in each of the chapters on data and regulation. A common theme that emerges in most of the chapters is, for example, the lack of transparency in veritable ‘black boxes’ of data science usage, which raises questions of justice and the protection of fundamental rights.¹ Other

¹ See Chapter 7 for example, or Chapter 18 in which the need for transparency for AI in relation to the judiciary and justice is stressed.

chapters highlight aspects of legal classification, such as the difficulties of the legal classification of ‘data’ and ‘who owns the data’,² which informs (intellectual) property law as well as liability law and to some extent contract law.³

The rich and elaborate discussion of these aspects in each of the chapters makes it impossible to fully summarize them. We therefore put forward some aspects that we believe stand out as important areas for further research.

2.1 Legal Classification, Legal Responsiveness

A specific issue that emerges from the chapters on regulating data-driven technologies is the question of legal classification or, broader, of legal responsiveness. If regulation is deemed necessary, how can laws integrate new rules and concepts into existing legal frameworks? Can existing laws, without upsetting established rules and doctrines, be responsive to the challenges of data science? Particular concerns that were raised are, for example, how data can be *quantified* if damages need to be calculated.⁴ Another frequently mentioned concern is how data *quality* should be measured. Issues raised extend not only to the actual quality of the data used for data science, but also the way in which it is used, and the skills that researchers and practitioners need to have to assess the quality of data. The importance of these questions prominently comes to the fore in relation to court proceedings.

We will elaborate each of these points a bit further, noting as our authors did that here we find questions that require further research. We will therefore say a few words about data and doctrinal challenges, data valuation, data quality, data transparency, and data skills.

2.1.1 Data and doctrinal challenges

It appears that the private law chapters in particular seem to struggle with the question of how to integrate, within existing doctrinal laws, the concept of ‘data’. The legal qualification of data, for example, is a necessity in order to allocate data as a resource, but also who is liable should the data be lost, or if the data is incorrect. Who is the owner of the dataset, or the individual pieces of data, and what if the data concerns personal data? Does this change the qualification? Mak, in relation to contract law, refers to this as an ethical challenge, formulating an answer to the question ‘whether it would be unethical for data subjects to be able to trade their

² See Chapter 8 regarding property law and the legal qualification issues discussed in Chapter 9 regarding data and intellectual property law, but also Chapter 5 in relation to the liability for data breaches and loss.

³ See Chapter 4 on digital content and consumer law, whereby digital content is examined. In this chapter a closer look is taken at the legal classification of data. Moreover, in this chapter also the question of whether personal data can amount to ‘consideration’ is touched upon, and the conceptual difficulties that arise when discussing (personal) data in relation to the remedies that consumers may have in contract law.

⁴ Chapter 5.

data in an economic transaction'. Is personal data to be commodified, or monetized at all?⁵ Op Heij in a new chapter in this edition, also looks at the question of classification of (personal) data but specifically in relation to consumer law. What are the difficulties that arise when we accept personal data as a counter-performance, specifically in relation to non-conformity and remedies in general?⁶ A discussion of data as a counter-performance is also part of Kryla-Cudna's contribution regarding the personal data generated by the use of smart devices.⁷

The qualification of data is of importance for the question of property law and intellectual property law. As Mattioli shows in the chapter on intellectual property law the question of how to classify data and datasets, for the US at least, has been a difficult one. Protecting valuable data by way of trade secrecy is of limited use, he argues, because its liability is limited to 'only those who misappropriate', and using trade secrecy protection may 'frustrate some socially desirable uses of data'.⁸ Copyright protection in turn is too 'thin', and provides difficulties as to classification, as data is typically not eligible for copyright protection, on account of its factual nature.⁹ However, the overview Mattioli provides shows that copyright protection may nevertheless be afforded to types of predictions or estimates that involve 'a degree of judgment or expertise', as well as to 'original selections and arrangements of data'.¹⁰ The latter category is however of limited use, as the originality requirement is problematic. Many companies that make use of electronic databases have structured these in 'unoriginal arrangements' and 'as a practical matter, little originality may often be put into data selection', and as such no copyright is afforded. Nevertheless, should there be a sufficient degree of originality present, then the protection afforded is weak, as 'a copyist could vary the dataset in a trial way to avoid infringement'.¹¹ Patent law in turn offers some support, but its impact is debated and requires more empirical study to inform policy, due to concerns about data access and the varying effects across different technological and industrial contexts.¹² The *sui generis* database protection right within the EU has been more successful,¹³ but there is no equivalent in the US, and hence the data holders rely heavily on contracts and technological measures such as encryption to protect their data. This offers protection to those that hold the data, but makes this data unavailable for researchers who might be able to put it to good use. Moreover, the manner in which the data is made

⁵ See also Chapter 2, section 3.1.

⁶ Chapter 4.

⁷ Chapter 3, section 2.2.

⁸ Chapter 9, regarding Intellectual Property Law, section 2.

⁹ Chapter 8, section 3 regarding Copyright Protection for Data.

¹⁰ Chapter 8, section 3 regarding Copyright Protection for Data.

¹¹ Chapter 8, section 3 regarding Copyright Protection for Data.

¹² Chapter 8, section 6.

¹³ See also Chapter 7 on property law, whereby the *sui generis* database right is also discussed as a possible alternative to pledge and attachment.

available for research which could benefit society, as such the data transparency issue looms large here as well.¹⁴

An important discussion at present is whether data can be ‘owned’. Although data does not conform to the traditional characteristics of tangible objects, there are in practice various interests of both consumers and businesses which stand in need of more effective protection, and could benefit from rules analogous to those applicable to tangible objects. Some of those areas are discussed in the chapter by Swinnen, who considers this issue from a comparative law perspective, looking at the Netherlands and Belgium, discussing the need for and ways in which pledge or attachment of data files is made possible.¹⁵ The new Belgian Civil Code which has come into force since the first edition of this book, has expanded the scope of property law in Belgium, and has had an influence on ownership, pledge, and attachment of data as well. Specifically, by adhering to a broader concept of a ‘good’ under the new Belgian Civil Code, the legislator has according to Swinnen perhaps opened the possibility of e.g. pledging data to a creditor.

The qualification of data in private law is furthermore of importance for determining liability. In particular, when data is lost or incorrect, as is the topic of Chapter 5. In order to award damages for data losses, liability will first have to be established, which immediately brings up the question of who is responsible for the data. As such Zeno-Zencovich notes that in the area of liability legal qualification issues arise as well in relation to determining ‘ownership’ of data.¹⁶

While data may as such be the subject of economic transactions and legal debates, data also has significant impact on society at large. Data analysis increasingly drives business and government decisions. Deep learning on the basis of big data is the core of generative AI, robots, self-driving vehicles, search algorithms, and so on. In private law this leads to the question of liability for algorithms and other semi-autonomous systems, which is the topic of two chapters in this volume. In Chapter 7 de Graaf discusses the topic of liability for AI from a black letter law perspective, and discusses current and upcoming EU law in this area. Tjong Tjin Tai takes a broader view and provides a comprehensive overview of various liability theories and how these (may) apply in relation to AI in Chapter 6. *Who* should be liable, and for what exactly, are age old questions, but prove especially difficult to answer when applied to AI systems that give rise to liability. Tjong Tjin Tai concludes that at present the tort of negligence (and its civil law equivalent) provides a basis for liability for AI, but only with limited value. While the question of ‘is this enough’ liability is a matter of policy, Tjong Tjin Tai considers the sector-specific approach to be preferable above a general regime applicable to all forms of AI.¹⁷ This approach is further elaborated on in the contribution by de Graaf who discusses the AI safety and liability proposals,

¹⁴ See further section 2.1.4 below.

¹⁵ See extensively Chapter 8.

¹⁶ Chapter 5, section 2.

¹⁷ Chapter 6, section 10.

which adhere to such a sector-specific liability regime.¹⁸ Tjong Tjin Tai further suggests that perhaps a distinction between commercial and private activity may also be contemplated, as well as a distinction between finished products and intermediate parts, with each of these distinct categories having their own liability regime. Tjong Tjin Tai furthermore signals three issues that may require legal change: (1) strict liability which would have to be adopted for certain kinds of AI, (2) product liability could be extended to algorithms, and (3) damages for pure economic loss should not be completely excluded, although some restrictions may be defended.¹⁹

De Graaf, in a new chapter since the first edition, has taken a closer look at the proposals for AI safety and liability that have been presented by the EU Commission in the last couple of years,²⁰ in particular in the way in which these proposals²¹ touch upon the topic of non-contractual liability for manufacturers of AI systems. His overview of the various instruments and proposals for legislation, considering how these relate to one another as well as how they differentiate from current liability regimes, is concluded with a succinct and practical enumeration of the options that are available to the party who suffered damage based on these EU instruments.

The possibilities of data analysis and AI also have farther-reaching consequences for law in practice, which may be positive and negative. Busch and De Franceschi discuss an intriguing possibility, namely the use of data analysis and AI to allow more fine-grained rulemaking, ideally even tailored to individuals.²² As an alternative to the traditional use of general legal typifications such as the general ‘reasonable person’, granular legal norms could match the idiosyncracies of the individual person and without loss of certainty assess what the individual should do to act reasonably. As such it could strengthen the effectiveness of regulation. Next to questions regarding the practical costs associated with implementing such a system, Busch and De Franceschi also discuss more fundamental objections which relate to the immense amount of personal data and profiling which is required. This processing of large quantities of personal data to achieve personalized law not only raises significant privacy concerns, but also raises the question whether it is compatible with the fundamental principle of equal treatment under the law.

Tailoring is also featured in the chapter by Kryla-Cudna. She provides a broad outline of legal issues, specifically regarding contract law, in relation to smart devices. Kryla-Cudna’s contribution identifies areas where current contract law fails to address issues in transactions involving smart devices. Here too, like in the contribution by Op Heij²³ remedies for non-conformity and data as a counter-performance are

¹⁸ See Chapter 7, section 1.

¹⁹ *ibid.*

²⁰ Specifically, the amended Product Liability Directive, the AI Act, and the AI Liability Directive, see Chapter 7.

²¹ The AI Act was adopted on 13 March 2024.

²² Chapter 16.

²³ See also Chapter 2.

discussed but from a different perspective. Furthermore, Kryla-Cudna also discusses servitization and the issues that may arise when the conclusion of a contract is via a smart device. She concludes that further legal consideration at both EU and national level is required due to the evolving nature of smart devices which traditional contract law does not fully address. One example of a proposed change by Kryla-Cudna is allowing direct claims against manufacturers for defects of non-compliance, which under the current law is especially difficult when non-conformity issues arise from digital components or services provided by third parties. She also advances that the unique aspects of smart device transactions in consumer-to-consumer contracts, such as continuous service requirements and data handling, also necessitate tailored legal approaches. An ‘updated’ and more nuanced legal framework is therefore preferred by Kryla-Cudna. A similar call to action arises from the contribution by Mak on consumer contract law and data driven technologies, who states that while ‘to some extent existing regulation may be able to provide adequate answers to the question which rights consumers can enforce against traders or suppliers. Overall, however, a re-evaluation of consumer contract law is needed to determine whether new or revised regulation is required for contracting in the data economy’.²⁴

2.1.2 Data valuation and interacting areas of law

On an abstract level, the value of data is clear. Several of the chapters refer to data as ‘the new oil’.²⁵ However, while we may know what the market prices are for profiles of individuals used for marketing purposes, we have very little evidence of how much a database of clients is worth in a business transaction. And what is the price of data loss when it concerns personal correspondence, or photographs, or addresses? Zeno-Zencovich in his chapter on data loss immediately recognizes that this is not necessarily a ‘new’ question, private law is already familiar with estimations of value, however whether existing approaches are the most suitable manner of dealing with the valuation of data is to be questioned.²⁶

While data protection law lays down rules for the protection of personal data because personal data is deemed valuable, this value is not generally expressed in terms of monetary value. While data protection law therefore does not take as its starting point the economic value of personal data, other areas of the law do.

Mak for example discusses in more depth what Op Heij and Kryla-Cudna touch upon as well, that it is the use of personal data as a counter-performance for a (consumer) contract which necessitates that personal data can be commodified. In doing so she discusses an interesting but complex interplay (and friction) between data protection and (consumer) contract law. On the one hand these fields may complement one another, such as with the EU Directive on Unfair Commercial Practices and the EU Directive on Unfair Contract Terms, which can offer a safety net against

²⁴ Chapter 2, section 2.2.

²⁵ Chapters 2, 5, 8, 9 and 11.

²⁶ Chapter 2, section 4.

unlicensed use of personal data. But these, and other instruments, may also be the source of friction, as the focus on the economic nature of transactions in the directives is not necessarily in line with data protection law. This is highlighted by Mak's examination of the issues put forth by the Dutch data protection authority (DPA) regarding the implementation of EU legislation introducing remedies for consumer contracts where the counter-performance was personal data. The DPA's proposed solutions for bringing back a fairer balance were in part rooted in the application of contractual remedies for issues in data protection law. While innovative, they appear to be difficult if not impossible to implement.²⁷

In the area of competition law, we can also note an interesting interplay with data protection law. Podszun and Kreifels discuss extensively the development of competition law in the area of data, and in doing so also discuss the CJEU case in which data protection was considered a 'significant parameter of competition'.²⁸ Specifically, when determining whether a finding of an abuse of dominant position could be substantiated, the competition authority was also allowed to examine a breach of data protection law. Here competition law is impacted by legal norms outside of competition law. Whereas the former impact can be seen as an expansion of competition law's reach, recent EU legislation, in particular the Digital Markets Act (DMA), seems to have the somewhat opposite effect. Podszun and Kreifels make clear that doctrinally the DMA 'is a deliberate and clear turn away from competition law', which changes the regulatory landscape in the digital sphere. However its limited application in scope also means that for many situations 'competition law will remain the definitory area law, setting the outer limits of market activities'.²⁹

2.1.3 Data quality

Another recurring theme in the chapters is the question how a high quality of data can be guaranteed. Here the legal protection afforded to the data itself can act as an incentive to maintain a high-quality dataset, but also other legal requirements of data quality may come into play.

When data is (made) available,³⁰ it is not always useful to researchers or practitioners. The quality of the data released itself as well as the meta-data, i.e. how it was collected and what it can describe, is necessary for researchers to assess the use of the data itself.³¹ This is all the more prevalent when the information is used in the context of legal proceedings.³² However, as Mattioli noted in Chapter 9, the lack of adequate

²⁷ Chapter 2, section 4.

²⁸ Chapter 11, section 1.

²⁹ Chapter 11, section 6

³⁰ Or the methods used are made available, see link with section 2.1.4. below.

³¹ Chapter 8, section 6.

³² See also section 2.1.5 below and Chapter 18 extensively.

protection provides those that collect the data with ‘weak economic incentives’ to maintain its quality for other purposes such as research.³³

An economic incentive may nevertheless exist in the form of a preventive risk assessment. In order to limit one’s liabilities, be it tax liabilities,³⁴ liability for AI systems,³⁵ or otherwise, a high(er) degree of data quality may be warranted. In this regard it is relevant that, where personal data is involved, current EU legislation demands a high degree of personal data quality and forthcoming EU legislation in relation to the use of AI does something similar.³⁶

2.1.4 Data transparency

In order to be able to assess the quality of the data used however, that data has to be made available. In practice this does not always happen, as the data itself, as well as the methods used to process this data, the algorithms, are considered ‘proprietary’ and remain hidden. That can be problematic. As mentioned above, where there is no transparency, the ‘science’ aspect of data science comes under pressure. If there is no way to validate the source material itself or the methods used to work with this data, then the research becomes impossible to verify and falsify. The users of data—researchers, but also others, such as lawyers, legislators, governments, and the judiciary—will then be unable to assess the reliability of data sets.

Existing laws already address this problem to some extent. For example, Russo puts forth in his chapter on data science and tax law that tax authorities ‘should assess new techniques in the light of the boundaries of the rights they have’. These boundaries are determined by the different fundamental rights of the tax payer, such as privacy, what information is off limits to the tax authorities, but also principles such as equality—and Fair Play.³⁷ Equality is of great importance, as it requires that tax payers in similar circumstances should be treated the same and in different circumstances differently.³⁸ However, it is exactly this kind of information which is not made public in many states. Hence, the lack of transparency makes it more difficult to audit the auditors themselves.

In the area of administrative law we observe something similar. Ranchordás and Klop note that in employing, or becoming, a ‘smart city’, governments are ‘forced to contract with private companies that often develop and manage the analysis of data’. The problem lies in the ‘risk of corporate capture of public power’ when these companies ‘will shape the way in which public bodies employ technology to pursue the public good’.³⁹ The actual manner in which the information is collected and used, is

³³ Chapter 8, section 6.

³⁴ Chapter 14, section 2.2.

³⁵ ‘[...] the operation of AI is determined by the quality and content of the data with which it has been fed’. Chapter 6, section 2.

³⁶ See Chapter 7, section 1.3.

³⁷ On Fair Play, mainly a Dutch affair, see Chapter 14, section 3.2.3.

³⁸ Chapter 14, section 3.2.2.

³⁹ Chapter 13, section 5.

kept in ‘proprietary black boxes’, which ‘obfuscates who retains control and responsibility for outcomes’ the authors argue. The lack of knowledge and clarity about ‘widespread algorithmic governance’ is an area in which one can wonder whether the law is prepared to offer a response for. Moreover, this ‘undermines the trust in good and transparent policymaking’.⁴⁰

Similarly, van Etteken and Prins, while discussing data science in judicial practice, point out the importance of transparency,⁴¹ as does Mak when discussing it in relation to contract.⁴²

2.1.5 Data skills

When the data and the methods employed by the data scientists are made available, different authors note another challenge, specifically the lack of skill with lawyers, traditional legal scholars, and judges to assess the methods used to collect and process the data, as well as the quality of the source material itself.

For example, Ranchordás and Klop state that the use of data science ‘invites us to rethink the skills that civil servants and citizens of the future should have in order to collect data and understand what data is being collected about them, and how that data is used to better predict the outcomes of the decision-making process, and tailor public services to their needs’. Moreover, it ‘challenges lawyers to reflect upon the principles of good administration that have traditionally guided public bodies and question their meaning in a world where algorithms and visual data collected by IoT might be difficult to explain to citizens’.⁴³

A similar point is made by van Etteken and Prins, who discuss the opportunities and risks involved with the introduction of data science and AI techniques in judicial practice. Judges and law clerks will have to gain data skills to profit from the possibilities opened up by data science, which offer new information and can help courts to fulfill the ideals they aim to uphold, such as consistency and equal treatment. Simultaneously they will need these skills in order to critically assess the results of such technology, as undoubtedly parties will introduce algorithmic results in court submissions. Furthermore, as judicial data such as case law becomes widely available to the public, this may open privacy risks, hence the judiciary will need to devise methods to temper such risks.

Risks also exist in relation to corporate governance, as discussed by Lafarre and Van der Elst. They find that ‘[w]hile technology holds promise in reducing agency costs and decentralizing governance, the practicality of entirely agent-less entities remains uncertain’. Their chapter explores Decentralized Autonomous Organizations (DAOs) and the potential of blockchain and smart contracts to modernize AGMs and shareholder voting. Such technologies offer huge potential for the reduction of voting

⁴⁰ Chapter 13, section 5.

⁴¹ Chapter 18.

⁴² Chapter 2, section 3.

⁴³ Chapter 13, section 5.

costs and could also enhance corporate decision-making. In the five years that have passed since the publication of the first edition of this volume, however, the authors note that little progress has been made with the use of such technologies. Challenges to a successful implementation are (a lack of) legislative readiness, accountability concerns, and the unsuitability of existing infrastructures.

In relation to criminal law Porcedda and Wall discuss the rise in data crimes, where they specifically look at ransomware and the role of AI. They examine the causes and contributing factors, and discuss potential future developments. They point out that techniques such as artificial intelligence can assist practitioners in identifying patterns of offending and ‘also (possibly) the path of offending which could lead law enforcement to data criminals’.⁴⁴ Using such techniques however raises a number of questions about ‘admissibility of evidence in court, and also the direction of the burden of proof, both of which require a suitable legal basis’.⁴⁵ Similar questions arise in the chapter by Strikwerda, Mensink and Timmers, in which the use of data science is discussed in three stages: police surveillance, criminal investigation, and sentencing. Providing illuminating examples of application of such techniques by Dutch law enforcement. The authors emphasize the need for a balanced approach in applying data science in criminal law, acknowledging the benefits for community safety while being cognizant of the legal and moral values it may strain.⁴⁶ Furthermore, and this is seen in other chapters as well, the question of skill of the lawyers, public prosecutors, and judges to assess the value and use of the data in a particular process is important.

Custers, in his chapter on methods,⁴⁷ furthermore considers from a methodological perspective the drawbacks of the use of data research methods for law include ‘the lack of human intuition, an abundance of results (that are not always relevant), limited insights in underlying causality, issues with repurposing, self-confirmation, self-fulfilling prophecies and reliability issues’.⁴⁸ While he discusses these drawbacks in relation to legal big tech in particular for researchers, the possible drawbacks hold true for legal practice as well.

3. APPLICATION OF DATA SCIENCE TO LAW

The second part of the book focused on the use of data in legal scholarship and legal practice. The methods and techniques of data science provide an important contribution to legal work in these fields. With that message Custers ends his chapter on methods of data science.⁴⁹ The message is echoed in other chapters as well, where

⁴⁴ Chapter 17, section 5.

⁴⁵ Chapter 17, section 5.

⁴⁶ Chapter 12.

⁴⁷ Chapter 15.

⁴⁸ Chapter 15, section 6.

⁴⁹ Chapter 15, section 6.

examples are discussed of how data science can benefit legal practice and legal scholarship.⁵⁰

In the area of tax law for example, we can note the significance and possibilities of data science, as exemplified by the chapter of Russo.⁵¹ In particular, when focusing on tax authorities, whose primary role it is to ‘ensure (through their supervisory function) that taxpayers fulfill their legal obligations: file correct tax returns, do so on time and pay the resulting taxes. They also have another role, closely linked to the primary one, and that is to promote (voluntary) compliance’.⁵² He shows that tax authorities can use data science in many different ways,⁵³ and are collecting ever more information, gathering data from an increasing number of sources; the traditional tax returns and audits, but also from other tax authorities as a result of the implementation of the OECD BEPS plans (transfer pricing, rulings, country-by-country-reporting, tax advices).⁵⁴ This changes the paradigm from how to collect the information, to how to use data in an efficient way. What Russo notes, is that those tax authorities that are ‘not burdened by past investments in IT have taken the lead in this development’,⁵⁵ for example, Brazil, Portugal, and Poland.⁵⁶ Thus, being an early adopter of a particular method of collecting and processing data can lead to path dependency, and stifle development later on.

4. DEVELOPING A NEW DISCIPLINE

Virtually all authors have noted opportunities and challenges for regulating data science and using data science and AI in regulation and law in general. What the chapters furthermore have in common is that while they provide an overview of the current body of research in a particular area, they also show that many questions remain unanswered. What is notably different from the previous edition is that the EU legislator has been very busy in its attempt to adopt new legislation in this area, trying to tackle many of the issues put forth in the previous edition, it is nevertheless clear that many questions remain. Moreover, the complex legal landscape of EU law in relation to data science has provided the authors with new questions of regulatory alignment between the different instruments.⁵⁷

⁵⁰ For example, Chapter 18.

⁵¹ Chapter 14.

⁵² Chapter 14, section 3.1.

⁵³ See for examples, Chapter 14, section 3.2.

⁵⁴ Chapter 14, section 3.1.

⁵⁵ Chapter 14, section 3.1.

⁵⁶ See M Rabenort, ‘Digitale Fiscale Autoriteiten veroorzaken een revolutie in de fiscale praktijk’ (2017) *Weekblad Fiscaal Recht* 116.

⁵⁷ See for example Chapter 7, and section 2.1.1 above.

What is clear is that these remaining questions are still urgent, as the collection and use of data continues unabated. Further research is therefore not only desirable, but indispensable to reach a society where data operates mostly for good.

Part of what makes research on and with data science hard is that it requires a solid grasp of the actual uses and possibilities of data, which cannot solely be obtained with reasoning by analogy with the physical world. Most of the preceding chapters show that data and use of data has unexpected opportunities, pitfalls and limitations which are nearly impossible to foresee from the outset. While other disciplines are busy investigating data as well, the legal perspective becomes pressing as the issues that nowadays confront society cannot solely be dealt with through social means: legal intervention is unavoidable. Thus the combination of two disciplines is part of the future of both: law and data science point the way to a new area that is in equal measure worrisome and exciting. A few provisional threads may sketch the outlines of what awaits us when we consider data law, the law and data science.

At the core is an understanding of what data is. It is not necessary to have a perfect definition, although as seen in the debates about property rights in data, such definitions may matter.⁵⁸ For most purposes, however, the foregoing chapters show that the issue is rather about the various uses and possibilities of data analysis, as well as its limitations. Data is intimately, intrinsically intellectual in nature. The possibilities of data thereby may directly impact that other intellectual discipline, law. While law in the past could be considered mostly as the mind governing the material world, the law now has to find ways to regulate the field of information. And at the same time, law itself may be directly impacted by data. This interactive relationship presents different challenges to those of new technologies in the past.

Algorithms may be able to take over large parts of the traditional field of lawyerly activity, but it is becoming clear also that there are as yet significant obstacles, areas where algorithmic activity falls short of what we ideally expect lawyers and judges to do in their work. Most important here is the problem of transparency.⁵⁹ While we can and do expect human experts to be able to explain, to motivate, what they do, algorithms as yet are unable to give insight by themselves of how they obtain their results. To ensure proper control of algorithms using data therefore requires human expertise; this kind of expertise is still in development. It stands to reason that the lawyers of the future will require at least a basic working knowledge of this area. Such knowledge, furthermore, will be essential if proper regulation is to be devised to ward off undesirable uses or consequences of data use. Justice requires understanding of data science.

In line with the above argument is the question of the responsibility for data. The results and use of data are determined in large part by the *quality* of the data used.

⁵⁸ For law it may be important to take into account the consequences of specific characteristics, such as the multifunctionality of data, the fact that data can be copied and transmitted easily (is not rivalrous).

⁵⁹ See section 2.1.5 above, as well as Chapter 13, section 4 and Chapter 18, section 3.

This is not a simple question, as quality depends also on the purpose for which the data is analysed. The fact that data can be put to a multiplicity of uses is clear from the large variety of use cases discussed in the previous chapters. The multifunctionality of data also makes it impossible to fully foresee future uses, and may require a flexible approach in regulating data use. A further part of responsibility for data is in the *preservation* of data, which is still an uncharted area, as discussed in the chapter on data loss and valuation,⁶⁰ as well as the IP chapter.⁶¹ For practical purposes, the responsibility for data is most important as regards the *use* of data: in developing algorithms,⁶² for market manipulation,⁶³ granular legal norms.⁶⁴ This shows that it is impossible to discuss data without taking into account the use of data, principally through algorithms.

Understanding data science, data use, and regulating responsibility for data seem to be the primary topics that require attention in this new field. Data is not merely a new object for legal thought but influences law as well: it is becoming an intrinsic part of law, and thereby changes the discipline of law, in a manner unprecedented since legal writing formed Roman law. On the other hand, law is not disrupted simply in the same way other areas of society are being disrupted. Law is essentially normative and aimed at humans: it wants to be understood by humans. Mediating between the raw nature and human ideals is what law and morality do, and data as yet does not fully supplant this. Indeed, it may be argued that one of the tasks of law is to make our world more human, by realising our ideal of justice.⁶⁵ In a world driven by data, the task is to realize justice with and through data.

Data has been called the new oil. And indeed, this metaphor is even more apt than may have been envisaged originally. Data is not a single product, but can be put to multiple uses, good as well as bad. Data can be refined to obtain better quality, it can be distributed or kept for proprietary purposes. Data can bring us huge profits and contribute to social welfare, but it can also be used to drive repressive practices. The use of data may also have negative social consequences, as is becoming increasingly clear with the effects of data-driven algorithms employed by social media and the rise of cybercrime. And just like oil is essential for automobiles and airplanes, data has become indispensable for society as we know it. There is only one clear difference: the supply of data will not be exhausted. For now, the emergence of new research questions relating to data science, as this book has shown, also remains unexhausted.

⁶⁰ Chapter 5.

⁶¹ Chapter 8.

⁶² See extensively Chapter 6.

⁶³ See extensively Chapter 11.

⁶⁴ See Chapter 16.

⁶⁵ This is a somewhat Hegelian position.