Tilburg Institute for Law, Technology, and Society (TILT)

The right to explanation: means for 'white-boxing' the black-box?

Research into the ability of the 'right to explanation' about decisions based solely on automated decision-making of Articles 13(2)(f), 14(2)(g), 15(1)(h) and 22(3) of the General Data Protection Regulation, as well as of current explanation methods, to solve the legal problems arising from algorithmic decision-making.

Name:	Janneke H.N. Janssen
Student number:	1258019
Program:	LL.M. Law and Technology
Date:	January 2019
Supervisors:	Dr. R.M.R Gellert, C. Quelle

Table of Content

List of abbreviations

AI	Artificial Intelligence
DPIA	Data Protection Impact Assessment
EDPB	European Data Protection Board
GDPR	General Data Protection Regulation
ICO	Information Commissioner’s Office

1. Introduction

1.1. The 'black-box' of algorithmic decision-making

The importance of automated decision-making has grown gigantically in the data driven economy. For example, China is establishing a nationwide social credit rating system with the aim to score the trustworthiness of citizens. The government uses records to calculate individual ratings that determine what services citizens are entitled to. The Chinese government rewards or punishes people according to their scores. To illustrate, already nine million people with low scores have been denied from buying tickets for flights within the country.[1] Like China, most nations rely on credit ratings to quantify financial risks associated with firms and individuals.[2] Software systems and algorithms increasingly make important decisions about people's lives. Credit reporting agencies collect and maintain consumer credit information and resell it to other businesses in the form of a credit report.[3] Hiring companies use algorithms to sort résumés for job applications and advertisers use algorithms to decide who sees certain advertisements.[4]

1.1.1. Algorithmic decision-making

The government and companies use algorithmic decision-making in different ways and in different sectors. Organizations in both the public and the private sector use algorithmic decision-making for non-commercial and/or for commercial purposes.[5] Several problems occur when organizations use algorithmic decision-making.[6] Experiments have repeatedly confirmed that data and algorithms are as biased as society. They reproduce real life inequality.[7] When looked upon sorting résumés for job applications, an algorithm trained to select the best candidates for primary school teachers is likely to develop a preference for female candidates, since more than 75 percent of all current primary school teachers is female.[8] Another problem is the fact

---

1 Josh Chin and Gillian Wong, 'China's New Tool for Social Control: A Credit Rating for Everything' *Wall Street Journal* (28 November 2016).

2 'China's dystopian social credit system is a harbinger of the global age of the algorithm', *The Conversation* (15 January 2018).

3 Latoya Irby, 'Who are the major credit reporting agencies?' (*The balance*, April 2018) <www.thebalance.com/who-are-the-three-major-credit-bureaus-960416> accessed 9 May 2018.

4 Caplan and others, 'Algorithmic accountability: a primer', (*Data & Society*, April 2018), <datasociety.net/wpcontent/uploads/2018/04/Data_Society_Algorithmic_Accountability_Primer_FINAL-4.pdf> accessed 23 October 2018 2.

5 Gianclaudio Malgieri and Giovanni Comandé, 'Why a Right to Legibility of Automated Decision-Making Exists in the General Data Protection Regulation' [2017] 7(4) *International Data Privacy Law* 265.

6 Gerards and others, 'Algoritmes en grondrechten', Universiteit Utrecht (*Officiële bekendmakingen*, 27 August 2018) <zoek.officielebekendmakingen.nl/blg-853458> accessed 28 August 2018 83.

7 The Conversation (n 2).

8 Lokke Moerel, 'Algorithms can reduce discrimination, but only with proper data' (*IAPP*, 16 November 2018) <iapp.org/news/a/algorithms-can-reduce-discrimination-but-only-with-proper-data/> accessed 30 November 2018.

that companies and data subjects involved often do not understand the algorithms.[9] Scholars describe algorithmic decision-making systems as a 'black-box'.[10]

1.1.2. Algorithmic accountability: transparency and explainability

Algorithmic accountability means that companies must be responsible for the results of the algorithms they use and the impact on society. Although algorithms calculate and process data in a way humans are not able to, humans are ultimately the ones providing the input, creating the design of the model, and using the outcomes. Algorithms impose risks such as information asymmetry and discrimination. There should be mechanisms for redress in place when such harm occurs.[11] Algorithmic accountability aims to safeguard the quality of algorithmic decision-making. However, the increasing complexity of algorithms and the speed at which new decision-making tools are developed make it difficult to assure algorithmic quality.[12] There are different ways to enhance accountability. The General Data Protection Regulation [**GDPR**][13] contains several provisions that enhance algorithmic accountability. Examples are the implementation of appropriate technical and organizational measures of Article 25 GDPR, the Data Protection Impact Assessment [**DPIA**] of Article 35 GDPR, approved codes of conduct of Article 40 GDPR, and voluntary certification of Article 42 GDPR. Another way to enhance algorithmic accountability is by creating transparency and explainability. The society demands transparency and explainability because of the impact on data subjects and the public at large.[14] The increased prominence of algorithmic models and the speed at which techniques are developed have led many scholars to call for increased transparency and explainability.[15]

1.1.2.1. Transparency

Algorithmic transparency involves the process of making a decision-making process visible. It requires that companies are open about the purpose and the actions of the algorithms they use.[16] There have been discussions in literature about the benefit of transparency in algorithmic accountability. Some suggest that the publication of datasets

---

9 Article 29 Data Protection Working Party, 'Guidelines on Automated individual decision-making and Profiling for the purposes of Regulation 2016/679 (wp251rev01)' (*EC*, 6 February 2018) <ec.europa.eu/newsroom/Article29/item-detail.cfm?item_id=612053> accessed 1 April 2018 5.

10 See for example: Sandra Wachter and others, 'Counterfactual Explanations Without Opening the Black Box: Automated Decisions and the GDPR' (*SSRN*, 2 November 2017) <papers.ssrn.com/sol3/papers.cfm?abstract_id=3063289> accessed 10 March 2018.

11 Caplan and others (n 4) 10.

12 Jakko Kemper and Daan Kolkman, 'Transparent to whom? No algorithmic accountability without a critical audience' [2018] *Information, Communication & Society* <doi.org/10.1080/1369118X.2018.1477967> accessed 19 October 2018 3.

13 European Parliament and Council of the European Union, 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, repealing Directive 95/46/EC (General Data Protection Regulation), (OJ L 119, 4.5.2016), 27 April 2016.

14 Bryce Goodman and Seth Flaxman, 'EU Regulations on algorithmic decision-making and a right to explanation' (A*rxiv*, 28 June 2016) <arxiv.org/pdf/1606.08813v1.pdf> accessed 1 April 2018 6.

15 Kemper and Kolkman (n 12) 3.

16 Bernhard Waltl and Roland Vogl, 'Explainable artificial intelligence – the new frontier in legal informatics' (*Towards data science*, 2018) <www.matthes.in.tum.de/file/13tkeaid0rhkz/Sebis-Public-Website/-/Explainable-Artificial-Intelligence-the-New-Frontier-in-Legal-Informatics/Wa18a.pdf> accessed 25 October 2018 6.

enhances accountability and fairness. Others suggest that organizations should consider the opacity of algorithms within the context of their use. Initially, many scholars have called for algorithmic transparency through open sourcing. However, this might not have the required effect. Sharing all available documentation might not constitute transparency when the relevant audience does not understand the information.[17] It does not fall within the scope of this research to assess all raised arguments, but it is important to keep in mind that algorithmic operators often have other goals that conflict with transparency. Hence, transparency can only be useful when there is a sufficient motive to disclose information and to reduce information asymmetry on the part of the creator or user of the algorithm.[18]

1.1.2.2. Explainability

Algorithmic explainability aims to explain how and why algorithms work the way they do. Explainability tries to make AI easily understandable for humans.[19] Explainability in algorithmic decision-making refers to the interpretability of the output of the model and the appropriateness of the whole process surrounding the model.[20] Is the goal of the system valid? Do controllers use the correct data? Is the model appropriate for the task?

1.1.3. The right to explanation of the General Data Protection Regulation

Although there are many different ways to solve problems as information asymmetry and discrimination, this research emphasizes on one tool the GDPR provides to enhance transparency and explainability, namely the 'right to explanation'. Articles 13(2)(f), 14(2)(g), 15(1)(h) and 22(3) GDPR establish this right. Data subjects have the right not to be subject to automated decision-making with legal or similar significant effects. When data subjects are subject to automated decision-making, the GDPR provides data subjects with several safeguards. Articles 13(2)(f), 14(2)(g) and 15(1)(h) GDPR primarily try to enhance transparency. Controllers have to provide data subjects with meaningful information about the logic involved, as well as the significance and the envisaged consequences of such processing. Article 22(3) GDPR primarily tries to enhance explainability. Controllers should implement suitable measures to safeguard the data subject's rights, freedom and interests. These measures include at least the right to obtain human intervention, the right to express the data subject's point of view and the right to contest the decision. Recital 71 GDPR adds the right to obtain specific information and an explanation of the decision reached. With regard to Article 22(3) GDPR, this research only emphasizes on the right to obtain specific information and an explanation of the decision, since the other safeguards do not establish transparency and explainability per se. The specific notion of the right to explanation is not clear.[21]

1.2. Research significance

Algorithmic decision-making is a questionable practice for unequal treatment and discrimination, and companies and data subjects do not understand the algorithms that are used. The right to explanation of the GDPR tries to tackle these problems, but the specific notion of the right is not clear. Therefore, this research aims to indicate whether

17 Kemper and Kolkman (n 12) 3.

18 Nicolas Diakopoulos, 'Algorithmic Accountability, Journalistic investigation of computational power structures' [2015] 3(3) *Digital Journalism* 403.

19 Waltl and Vogl (n 16) 5.

20 Ibid 7.

21 Goodman and Flaxman (n 14) 1.

the right to explanation of Articles 13(2)(f), 14(2)(g), 15(1)(h) and 22(3) of the GDPR solves the legal problems arising from algorithmic decision-making. Scholars have conducted much research on the right to explanation, but this research mainly used a 'black letter' methodology of the GDPR or emphasized on a specific explanation method. The author takes a different approach. The author identifies shortcomings of the right to explanation for enhancing transparency and explainability and for solving the legal problems arising from algorithmic decision-making. Furthermore, this research looks upon the way current explanation methods implement the right to explanation and assesses whether these explanation methods are sufficient to establish a right to explanation. It is important to assess the meaningfulness of explanation methods from a data subject's perspective.[22] Therefore, the author takes a quantitative analysis on the opinions of data subjects on several explanation styles into account.[23]

1.3. Research questions

The central research question of this thesis is:

> "Are the 'right to explanation' about decisions based solely on automated decision-making of Articles 13(2)(f), 14(2)(g), 15(1)(h) and 22(3) of the General Data Protection Regulation, as well as current explanation methods, able to solve the legal problems arising from algorithmic decision-making?"

In order to answer this research question, the following sub-questions have been formulated:

1. What legal problems arise from algorithmic decision-making?
2. What is the 'right to explanation' of Articles 13(2)(f), 14(2)(g), 15(1)(h) and 22(3) GDPR?
3. Are the right to explanation and current explanation methods able to solve the legal problems arising from algorithmic decision-making?

1.3.1. Scope and limitations

In order to gain better insight in the scope of this research, some concepts have to be explained.

'Right to explanation'

The GDPR contains multiple articles that aim to establish transparency and explainability.[24] This research only emphasizes on Articles 13(2)(f), 14(2)(g), 15(1)(h) and 22(3) GDPR. According to Articles 13(2)(f), 14(2)(g) and 15(1)(h) GDPR, the controller shall provide data subjects with:

> "information on the existence of automated decision-making, including profiling, referred to in Article 22(1) and (4) and, at least in those cases, meaningful information about the logic involved, as well as the significance and the envisaged consequences of such processing for the data subject".

According to Article 22(3) GDPR:

---

22 Dimitra Kamarinou and others, 'Machine Learning with Personal Data: Profiling, Decisions and the EU General Data Protection Regulation' (*ML and the law*, 2016) <www.mlandthelaw.org/papers/kamarinou.pdf> accessed 5 March 2018 23.

23 Reuben Binns and others, 'It's Reducing a Human Being to a Percentage: Perceptions of Justice in Algorithmic Decisions' (*Arxiv*, 31 January 2018) <arxiv.org/pdf/1801.10408.pdf> accessed 10 March 2018.

24 See for examples Articles 5(1)(a,d), 12 and the other sections of Articles 13, 14 and 15 GDPR.

> "the data controller shall implement suitable measures to safeguard the data subject's rights and freedoms and legitimate interests, at least the right to obtain human intervention on the part of the controller, to express his or her point of view and to contest the decision".

Recital 71 GDPR clarifies that data subjects have the right to obtain specific information and an explanation of the decision. There is a discussion in literature whether Recital 71 GDPR is binding or not. While only the text of the GDPR is legally binding, both Recitals and guidelines of the European Data Protection Board [**EDPB**] play a significant role in interpreting provisions, since they are indicative of what the GDPR's enforcers will do.[25] This research does not emphasize on all safeguards of Article 22(3) and Recital 71 GDPR. This research only emphasizes on the right to obtain specific information and an explanation of the decision, since the other safeguards do not establish transparency and explainability per se. Therefore, the right to explanation in this research refers to Articles 13(2)(f), 14(2)(g) and 15(1)(h) GDPR and the right to obtain specific information and an explanation of the decision of Article 22(3) GDPR.

'Decisions based solely on automated decision-making'

The right to explanation relates to automated decision-making. Article 22(1) GDPR states that:

> "data subjects have the right not to be subject to a decision based solely on automated processing, including profiling, which produces legal effects concerning him or her or similarly significantly affects him or her."

The author uses the definition of automated individual decision-making from the EDPB, formerly the Article 29 Data Protection Working Party, but joins the relative notion of this provision as is mentioned in literature. In short, automated decision-making is the ability to make decisions by technological means without human involvement, with or without profiling.[26] In literary discussions, scholars argue that rather a relative than a strict interpretation is necessary for the provision to be meaningful. According to the relative interpretation, the scope is not limited to 'solely automated decision-making'. It extends to either human or machinery decisions based solely on automated decision-making.[27] The UK Information Commissioner's Office [**ICO**] has released an opinion arguing that the interpretation of the word 'solely' is intended to cover those automated decision-making processes in which humans do not exercise a real influence on the outcome of the decision.[28] In algorithmic decision-making, companies often take a passive human decision based on the outcome of the automated decision-making process.[29] The author would prevent the application of Article 22 GDPR in most cases if the author would join the strict interpretation. This research is not limited to automated decisions with legal or similar significant effects, since Articles 13(2)(f), 14(2)(g) and 15(1)(h) GDPR grant data subjects the right to

---

25 Margot Kaminski, 'The right to explanation, explained' (*SSRN*, 23 July 2018) <papers.ssrn.com/sol3/papers.cfm?abstract_id=3196985> accessed 9 September 2018 13.

26 WP 251 (n 9) 8.

27 Malgieri and Comandé (n 5) 251.

28 ICO, 'Feedback request – profiling and automated decision-making' (*ICO*, June 2017) <ico.org.uk/media/about-the-ico/consultations/2013894/ico-feedback-request-profiling-and-automated-decision-making.pdf> accessed 5 June 2018 19.

29 Malgieri and Comandé (n 5) 251.

receive information about the existence of automated decisions irrespective of whether it is caught by Article 22 provisions.[30]

'Legal problems'

This research emphasizes on two legal problems that arise from algorithmic decision-making, namely information asymmetry because of opacity, and discrimination and unfairness. In this research, discrimination from a legal perspective is the application of different rules or practices to comparable situations or the use of the same rule or practice to different situations.[31]

'Algorithmic decision-making'

In algorithmic decision-making, algorithms are used in the automated decision-making process. Algorithmic decision-making is a very broad concept and there are many different algorithms. Underlying algorithms may for example be either self-learning or not.[32] This means that some algorithmic decision-making models are not as complex or unintelligible for the average data subject as others. This research looks upon self-learning algorithms, since these are challenging to understand and therefore a problem in creating transparency and explainability.

'Current explanation methods'

This research assesses five different explanation methods that controllers currently use in practice. This research firstly distinguishes between model-centric explanations and subject-centered explanations. Apart from model-centric explanations, this research looks upon four subject-centered explanations, namely: (i) sensitivity-based explanations, (ii) input influence-based explanations, (iii) case-based explanations, and (iv) demographic explanations.

'To be able'

With regard to the right to explanation, 'to be able' means that the textual analysis of the GDPR and the analysis of practical complications of algorithmic decision-making do not indicate any shortcomings that may hamper its problem solving capacity. With regard to current explanation methods, 'to be able' means that the specific explanation method fulfills all requirements of the GDPR.

1.3.2. Overview of chapters

In order to answer the research questions, chapter two explains what algorithmic decision-making is and what legal problems arise by its use. Chapter three explains what the right to explanation is. The chapter clarifies the terminology of the GDPR and elaborates on the main discussion within literature, which relates to the tripartite structure of the right to explanation. Chapter four assesses whether the right to explanation solves the legal problems identified in chapter two, whether existing

---

30 WP 251 (n 9) 16, 25.

31 Bruno Lepri and others, 'Fair, Transparent, and Accountable Algorithmic Decision-making Processes' (*Springer*, 15 August 2017) <link.springer.com/Article/10.1007/s13347-017-0279-x> accessed 10 March 2018 4.

32 Gerards and others (n 6) 30.

explanation methods fulfill the requirements of the GDPR, and provides recommendations.

1.4. Methodology

1.4.1. Type of research

This research emphasizes on both transparency and explainability. The concepts of transparency and explainability are intertwined. Scholars mention all four aforementioned articles when they define the right to explanation. Transparency is necessary to explain algorithmic decision-making systems, and explainability requires transparency.[33] This research is based on literature research to answer the research questions described. The research starts by reviewing the functionality of algorithmic decision-making and the type of knowledge decision-makers are able to gain, using technological reports and articles. The author identifies two legal problems that arise by its use. Secondly, this research reviews the right to explanation of the GDPR through a doctrinal legal research. This part is based on a 'black letter methodology', interpreting and explaining the meaning of the right to explanation using the law and academic literature. Lastly, this research combines the legal ruling and the technological context.[34] This part identifies shortcomings of the right to explanation for solving the legal problems arising from algorithmic decision-making. The author analyzes both the text of the GDPR and looks at practical complications. After identifying the shortcomings, the author assesses whether current explanation methods fulfill the requirements of the GDPR. Current explanation methods will not enhance transparency and explainability and will not solve the legal problems when they do not fulfill the requirements of the GDPR. The author includes a quantitative analysis on the opinion of data subjects on several explanation styles to consider the data subjects' perspective.[35] The author assesses those subject-centered explanation methods that are included in the research on the perception of the public. Lastly, the author recommends two other concepts of the right to explanation. The author looks upon the way these concepts can take the current shortcomings of the right to explanation into account, and provides recommendations.

1.4.2. Literature review

The author has conducted a systematic quest through the search engines Google Scholar, SSRN and Hein Online to gather literature. The author has excluded literature not emphasizing on the research objectives. The author uses articles explaining the technology in order to analyze algorithmic decision-making practices. Many articles identify the legal problems that algorithmic decision-making creates. Key articles and different publications are examined in order to analyze the meaning of the right to explanation. Guidelines of the EDPB are reviewed because this gives a better insight in the meaning of the GDPR. Authors disagree on the legal existence and the benefit of the right to explanation in practice, and approach the question from different perspectives.[36]

---

33 See for example Recital 71 GDPR. Recital 71 GDPR elaborates on Article 22 GDPR, which enhances explainability. However, it also mentions transparent processing as a goal.

34 Paul Chynoweth, Legal research. in Andrew Knight and Les Ruddock (eds), *Advanced Research Methods in the Built Environment* (Wiley-Blackwell 2008) 30.

35 Binns and others (n 23) 1.

36 Wachter, Mittelstadt and Floridi claim that a right to explanation of automated decision-making does not exist in the GDPR and propose a 'right to information' to improve transparency and accountability. Malgieri and Comandé try to undermine Wachter and other's opinion and state that a right to legibility of automated decision-making does exist in the GDPR.

2. Algorithmic decision-making

In order to answer the central research question, this chapter firstly explains what the functionality of algorithmic decision-making is. Secondly, this chapter identifies the legal problems that arise by its use.

2.1. The context of algorithmic decision-making: Big Data and Artificial Intelligence

In order to understand what algorithmic decision-making is, this paragraph shortly elaborates on the contexts in which algorithmic decision-making takes place, namely 'Big Data' and 'Artificial Intelligence'. Humans used to make all kind of decisions by themselves in the past. However, several developments in information- and communication technologies have created emerging opportunities to collect and process data.[37] Originally, organizations analyzed datasets with the aim of verifying specific and predefined assumptions. Traditional data analyses were hypothesis-driven; the data were means to get an answer to a specific question or to prove a certain hypothesis that humans had predetermined. 'Big Data' analyses, in which a vast amount of various kinds of data is collected and converted, are on the other hand data-driven. Data analysists use algorithms to identify correlations in datasets.[38] Algorithms test large amounts of connections and try to filter relevant information from the data. Algorithms can discover unexpected connections because the knowledge that they obtain is no longer limited to predetermined hypotheses by humans.[39] As humans no longer make decisions completely by themselves, intelligence is no longer limited to humans. Artificial Intelligence [**AI**] relates to the intelligence of artefacts. Significant developments in technology have produced systems that challenge or even exceed the human ability at highly skilled tasks.[40] For example, the question-answering computer system Watson was already capable of answering questions posed in natural language by using automated reasoning technologies and won against human beings back in 2011.[41] An important characteristic of AI is the high degree of autonomy. Artefacts are able to act, learn, understand and respond autonomously in their environment and are able to adapt to changes to reach the best outcome.[42] This autonomy creates several problems. AI-systems may have such degree of autonomy that human control is no longer possible. Moreover, actions of AI-systems may be unpredictable because it does not correspond to the human thinking-process. Consequently, AI-systems may derive unexpected correlations from data.[43] There are many different applications of AI. In the context of this research, the most important application is the technology that enables machines to

---

37 Gerards and others (n 6) 5.

38 Ibid 8.

39 Ibid.

40 Informatics Europe and EUACM, 'When Computers Decide: European Recommendations on Machine-Learned Automated Decision Making' (*ACM*, 2018) <www.acm.org/binaries/content/assets/public-policy/ie-euacm-adm-report-2018.pdf> accessed 1 March 2018 3.

41 Jo Best, 'IBM Watson: The inside story of how the Jeopardy-winning supercomputer was born, and what it wants to do next' (*Tech republic*, 2012) <www.techrepublic.com/Article/ibm-watson-the-inside-story-of-how-the-jeopardy-winning-supercomputer-was-born-and-what-it-wants-to-do-next/> accessed 5 June 2018.

42 Gerards and others (n 6) 25.

43 Ibid 6.

learn without that humans have programmed them to do so. This is called Machine Learning. In Machine Learning, algorithms are able to learn on their own based on experiences.

2.2. Algorithmic decision-making

This paragraph elaborates on the functionality of algorithmic decision-making and the type of knowledge that it produces.

2.2.1. What is an algorithm?

An algorithm is a set of instructions for how a computer should accomplish a particular task. It calculates an answer to a problem by taking a set of values as input and producing some values as output.[44] Algorithms test large amounts of connections and try to find relevant information. There are many different types of algorithms. The most commonly used type in AI is the self-learning algorithm.[45] Self-learning algorithms are capable of changing itself or its set of instructions based on accumulated data. The algorithms learn to emphasize on the things they should be looking for to solve the problem statement.[46] These models are also called 'artificial neural networks'.[47] The most advanced application is Deep Learning. In 'deep' architectures, the model looks likes the structure of a brain with biological networks and neurons. Similar to the human brain that organizes neurons in layers, the deep architecture is organized in layers. Figure 1 presents a neural network.

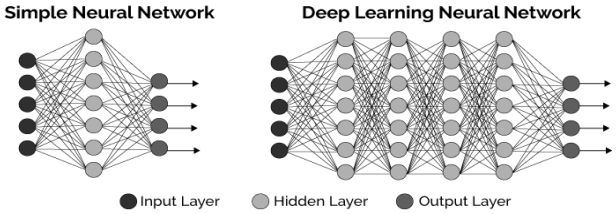

Figure 1 Neural networks[48]

Neurons in the input layer of the brain receive signals from the inputs and neurons in the output layers of the brain provide the answer. The selection of the neurons that connect with other neurons in the next layer(s) and details like the exact number of hidden layers come from experience.[49] Metaphorically, some areas of the brain are sensitive to certain stimuli and other areas are not. Algorithms are able to distinguish the different categories of stimuli.[50]

---

44 Mikella Hurley and Julius Adebayo, 'Credit scoring in the era of big data' [2016] 18(1) *Yale Journal of Law and Technology* 159.

45 Gerards and others (n 6) 26-27.

46 Argyro Karanasiou and Dimitris Pinotsis, 'A study into the layers of automated decision-making: emergent normative and legal aspects of deep learning' [2017] 31(2) *International Review of Law, Computers & Technology* 172.

47 Ibid 174.

48 Anna Gomez, 'Deep learning in digital pathology' *(Global-engage,* 2 February 2018) <www.global-engage.com/life-science/deep-learning-in-digital-pathology/> accessed 23 June 2018.

49 Karanasiou and Pinotsis (n 46) 172.

50 Ibid 174.

2.2.2. The use of algorithms in automated decision-making

Humans use algorithms to make all different kind of decisions, for example to allocate social services.[51] How do organizations use algorithms to make those decisions? To explain this, scholars divide the Big Data process into three different steps: (i) the collection of data and the aggregation of datasets, (ii) the analysis of the data and (iii) the actual use of the model.[52] At first, all different kind of data originating from all different kind of sources have to be collected and datasets have to be created. Secondly, algorithms must derive relevant information from the datasets. There are many different techniques to analyze the collected data, referred to as 'Big Data Analytics'.[53] Paragraph 2.2.3 identifies some of these techniques. Lastly, organizations use the relevant information for policy purposes or for decision-making purposes. Algorithms can make the decision by themselves, or algorithms can assist humans in making the decision.[54]

2.2.3. Data analysis: what knowledge are we able to gain?

There are many different techniques to analyze data. These techniques find their basis in Machine Learning. One of the main technologies for Big Data Analytics is data mining. By the use of data mining, algorithms derive patterns from large datasets.[55] There are four techniques that can be distinguished, namely classification-, clustering-, regression- and association techniques. Classification techniques aim to locate data in categories. Programmers have created those categories on beforehand. The algorithms 'learn' from examples that are already classified by systematically comparing the different categories. The algorithms are capable of distilling rules and applying them to new cases. An example is the classification of patients leaving the hospital in predefined categories. Each category reflects a different risk of re-entering the hospital.[56] Clustering techniques aim to group data that are very similar to each other. An example is a customer base of a shop that divides different types of customers based on their purchasing behavior.[57] The difference between classification and clustering is that classification contains pre-defined classes, whereas clustering aims to create such classes based on the data analysis.[58] Regression techniques aim to formulate numerical predictions based on identified correlations derived from the dataset. For example, a bank is able to predict how likely it is that a loan will not be repaid based on data obtained.[59] Association techniques aim to search for correlations between data and aim to formulate rules based on these correlations. An example is a Netflix recommendation based on previous watched movies.[60] Another technology in Big Data Analytics is profiling. Profiling techniques use algorithms to create profiles of individuals or groups

---

51 Caplan and others (n 4) 2.

52 Gerards and others (n 6) 8-13.

53 Ibid 9.

54 Paul de Laat, 'Algorithmic Decision-Making Based on Machine Learning from Big Data: Can Transparency Restore Accountability?' (*Springer*, 12 November 2017) <link.springer.com/Article/10.1007/s13347-017-0293-z> accessed 14 March 2018 9.

55 Gerards and others (n 6) 9.

56 Ibid 9-10.

57 Ibid 10.

58 Ibid.

59 Ibid.

60 Ibid.

of individuals. Profiling strongly relates to data mining since profiling is often based on data mining techniques.[61]

2.3. Legal problems arising from algorithmic decision-making

Several legal problems arise from algorithmic decision-making. This paragraph elaborates on two of these problems, namely information asymmetry because of opacity, and discrimination and unfairness.

2.3.1. Opacity and information asymmetry

As mentioned in paragraphs 2.1.2 and 2.2.1, self-learning algorithms have the ability to act autonomously. Their actions and outcomes are not always foreseeable and their 'thinking-process' differs from the human's thinking-process. Self-learning algorithms are therefore 'opaque'.[62] The opacity concern arises in the hidden layers of the neural network. People do not know what is going on in the middle phase of the network, the 'black-box'. It is not clear what bits of data algorithms select and how algorithms use these bits to provide output.[63] Opacity in algorithmic decision-making creates several problems. There are fears that systems are not accurate and unfairly target certain populations. Individuals have the right to understand why algorithmic decision-making models adversely affect them, but opacity prevents the public from figuring that out.[64] Information asymmetry occurs when it is not clear to individuals how and on what basis algorithms make decisions. A powerful few have access and use resources and tools that the majority does not have access to. This leads to an asymmetry in power between the state and the big companies on one side, and the majority of the people on the other side. The Dutch Administrative Jurisdiction Division of the Council State identifies this problem.[65] When interested parties want to use legal remedies against decisions based on an algorithmic decision-making system, it may result in an unequal procedural position of parties. Interested parties cannot check on what basis algorithms have made a particular decision.[66] Scholars identify three different kinds of opacity.

2.3.1.1. Intrinsic opacity

Intrinsic opacity refers to the opaque nature of algorithms.[67] Programmers cannot provide an explanation why algorithms recommend a specific decision, or at least not in understandable terms. This has several causes. Programmers can code algorithms in such way that their logic is comprehensible, but the rules that the algorithms use to generate the output alters as they train themselves.[68] In the case of neural networks, the weight of input variables may change when the process repeats. The final model displays the weights, but programmers cannot interpret the contribution of the different

---

61 Ibid.

62 Karanasiou and Pinotsis (n 46) 174.

63 Ibid.

64 Lilian Edwards and Michael Veale, 'Slave to the algorithm? Why a right to an explanation is probably not the remedy you are looking for'(*SSRN*, 6 December 2017) <papers.ssrn.com/sol3/papers.cfm?abstract_id=2972855> accessed 20 September 2018 41.

65 The Dutch Administrative Jurisdiction Division of the Council State, *Stichting Werkgroep Behoud de Peel v. het college van gedeputeerde staten van Noord-Brabant,* 17 May 2017, ECLI:NL:RVS:2017:1259.

66 Ibid, paragraphs 14.3 and 14.4.

67 De Laat (n 54) 12

68 Ibid.

input variables in determining the final weights.[69] Algorithms find the exact number of parts or layers in the automated decision-making system by experience. This is not determined in advance. The faster the machine learns, the more difficult it is to understand the reasons behind the decisions.[70]

2.3.1.2. Illiterate opacity

Algorithms are technically very complex. Most people lack the technical skills to understand algorithms and Machine Learning models. The majority of the people does not know the basic principles on which algorithms operate and does not know how to read or write a code. The output of Machine Learning is therefore very difficult to interpret.[71] Scholars refer to this as 'illiterate opacity'.[72]

2.3.1.3. Intentional opacity

A lot of algorithmic opacity is deliberate. Organizations simply do not want others to know how their systems work and decide to withhold information about the way they make a decision.[73] There are two main rationales for intentional opacity. Firstly, organizations try to prevent that interested parties 'game the system'. This means that interested parties may be able to detect a way to evade undesirable results for them. The value of the model diminishes in that case and the accuracy of the algorithm might be undermined because false data will be included in the decision-making system.[74] Secondly, organizations may consider their algorithms as their trade secret or intellectual property, since algorithms distinguish the organization from their competitors.[75] The organization will weaken its market position if it discloses this information. Facilitating transparency also generates costs, which might be a reason not to disclose information.[76]

2.3.2. Discrimination and unfairness

Apart from the problems regarding opacity and information asymmetry, algorithmic decision-making creates discrimination and unfairness. As stated by the White House:

> "Big data techniques have the potential to enhance our ability to prevent discriminatory harm. But, if these technologies are not implemented with care, they can also perpetuate, exacerbate, or mask harmful discrimination."[77]

Algorithmic decision-making often uses personal data that relates to protected characteristics, such as gender and race.[78] As laid down in Article 21 of the EU Charter of fundamental rights and Article 14 of the European Convention of Human Rights, any discrimination based on sex, race, color, language, religion, political or other opinions, national or social origin, association with a national minority, property, birth or other status is prohibited. Not all correlations that arise in Machine Learning systems relate to

---

69 Ibid 13.

70 Maja Brkan, 'AI-supported decision-making under the General Data Protection Regulation' (*ACM*, June 2017) <dl.acm.org/citation.cfm?id=3086513> accessed 27 April 2018 6.

71 Jenna Burrell, 'How the machine 'thinks': Understanding opacity in machine learning algorithms' [2016] *Big Data & Society* 4.

72 Ibid.

73 Lepri and others (n 31) 9.

74 De Laat (n 54) 11.

75 Ibid 12.

76 Tal Z. Zarsky, 'Transparent predictions' [2013] *University of Illinois Law Review* 1553.

77 Gerards and others (n 6) 94.

78 Edwards and Veale 2017 (n 64) 28.

characteristics protected by law. The use of these correlations might lead to unfairness instead of discrimination.[79] Algorithmic discrimination and unfairness have several causes. Firstly, discrimination and unfairness can arise from the decision to use an algorithm.[80] Organizations use data mining techniques to distinguish individuals.[81] Scholars consider such categorization as a form of direct discrimination. The use of data driven decision-making processes may result in individuals being denied based on the actions of others with whom they share characteristics, instead of their own actions.[82] Apart from the decision to use an algorithm, discrimination and unfairness may arise in all steps of the modelling process. There are many different ways to build algorithmic decision-making models, but all designs have three general steps in common. These three steps are: (i) define the problem that has to be solved and describe a target variable that represents the outcome, (ii) gather training data and transform it into a useable format and (iii) develop and refine the model through exposure to training data.[83] At first, the model developer has to describe a target variable. When developers do not correctly define target variables, discrimination may arise. Certain classes would happen to be subject to less favorable determinations.[84] Developers assign labels to classification attributes in either an objective or subjective way. Subjective labelling involves human interpretation. Objective labelling does not involve human interpretation. When developers define attributes to a classification in a subjective way, the human judgment could result in bias.[85] Once the developer has identified the target variable, he must gather information about individuals for which various outcomes are already known. This information is the training data. If the set of examples in the training data does not fairly represent the data on which the algorithm runs, the model may disadvantage misrepresented groups.[86] For example, when the police surveilles in neighborhoods where mostly people from ethnic minorities live, databases are in a large extent filled with information about these minorities.[87] Moerel indicates that it is important to make biases in the training data transparent to prevent the bias from influencing future outcomes.[88] Once the training data is collected, the developer must translate the data into a format that a computer is able to process. After this, algorithms analyze the training data and identify the most significant input variables. They assign the appropriate weights to all variables. The developer will combine the most predictive variables in the eventual model. In this stage, the model may produce implicit forms of bias since factors that are neutral at first sight may be correlated with sensitive characteristics. Scholars call this the 'proxy-problem'.[89] The proxy-problem makes it difficult to distinguish between the discriminatory and the non-discriminatory parts of

79 Ibid 30.
80 Ibid 28.
81 Gerards and others (n 6) 94.
82 Hurley and Adebayo (n 44) 183.
83 Ibid 168.
84 Ibid 173.
85 Ibid.
86 Eva Thelisson and others, 'Regulatory mechanisms and algorithms towards trust in AI/MI (*Earth link*, 2017) <earthlink.net/~dwaha/research/meetings/ijcai17xai/9.%20(Thelisson,%20Padh,%20&%20Celis%20XAI-17)%20Regulatory%20Mechanisms%20and%20Algorithms%20towards%20 Trust%20in%20AIML.pdf> accessed 11 September 2018 2.
87 Gerards and others (n 6) 97.
88 Moerel (n 8).
89 De Laat (n 54) 7.

the dataset.[90] An example is a zip code. A zip code is not a protected characteristic at first sight, but it may serve as an indicator for race.[91] When there are numerous data points to work with, the Machine Learning process may use sensitive characteristics, even when the model does not directly assign these as input values.[92]

2.4. Conclusion

Organizations use algorithms in algorithmic-decision making to make all different kind of decisions that have an impact on human lives. Actions and outcomes of self-learning algorithms are not always foreseeable because they have the ability to act autonomously. Scholars consider algorithmic decision-making as a 'black-box', since people generally do not know how algorithms make certain decisions. Different legal problems arise from algorithmic decision-making, such as information asymmetry because of opacity, and discrimination and unfairness.

90 Ibid.
91 Ibid.
92 Hurley and Adebayo (n 44) 182.

3. The right to explanation

In order to answer the central research question, this chapter explains what the right to explanation is. Firstly, this chapter elaborates on Articles 13(2)(f), 14(2)(g), 15(1)(h) and 22(3) GDPR. Secondly, this chapter clarifies the terminology of the GDPR. Thirdly, this chapter elaborates on the main discussion in literature regarding the tripartite structure of the right to explanation.

3.1. Articles establishing the right to explanation

Different provisions of the GDPR address transparency and explainability. Articles 13(2)(f), 14(2)(g), 15(1)(h) and 22(3) GDPR create the legal basis for the right to explanation.

3.1.1. Notification to data subjects

Articles 13 and 14 GDPR require controllers to notify data subjects when they obtain personal data. Article 13 GDPR relates to information that controllers must provide when personal data have been obtained from the data subject. Article 14 relates to information that controllers must provide where personal data have not been obtained from the data subject. According to these articles, the controller shall provide the data subject at the time when he obtains personal data with the following information necessary to ensure fair and transparent processing:

> "the existence of automated decision-making, including profiling, referred to in Article 22(1) and (4) and, at least in those cases, meaningful information about the logic involved, as well as the significance and the envisaged consequences of such processing for the data subject."[93]

Controllers must ensure that they provide data subjects with information about how automated decision-making processes work in a clear and easily understandable way.[94] The information should inform data subjects and help them to understand why the automated decision-making system has reached a particular decision.[95] The EDPB mentions that Articles 13 and 14 GDPR set out the information that controllers must provide at the beginning of the processing cycle.[96]

3.1.2. Duty to provide access

According to Article 15(1)(h) GDPR, the data subject shall have the right to obtain confirmation as to whether or not personal data of him or her is being processed, and has the right to access information about:

> "the existence of automated decision-making, including profiling, referred to in Article 22(1) and (4) and, at least in those cases, meaningful information about the logic involved, as well as the significance and the envisaged consequences of such processing for the data subject".

Article 15 GDPR gives data subjects the possibility to obtain details about their personal data that the controller uses for automated decision-making. Article 15(1)(h) GDPR entitles data subjects to obtain the same information as described in Articles 13(2)(f) and 14(2)(g) GDPR. The intention of Article 15(1)(h) GDPR seems to be to provide a control mechanism for data subjects to request more or less the same information as

---

93 Articles 13(2)(f) and 14(2)(g) GDPR.

94 WP 251 (n 9) 16.

95 Ibid 25.

96 Article 29 Data Protection Working Party, 'Guidelines on transparency under Regulation 2016/679 (WP260)' (*EC*, 24 January 2018) <ec.europa.eu/newsroom/Article29/item-detail.cfm?item_id=615250> accessed 1 April 2018 13-15.

provided under Articles 13 and 14 GDPR at any time.[97] Recital 63 GDPR states that data subjects should have the right to access to obtain 'communication'. This way, data subjects can become aware of a decision made concerning him or her. The EDPB mentions that the controller has to make the input data available.[98] Controllers should provide general information about the factors that are taken into account in the decision-making process and their 'weight' on an aggregate level.[99]

3.1.3. General safeguards

Article 22(1) GDPR contains a general prohibition on fully automated decision-making that has legal or similar significant effects. There are several exceptions to this rule, as laid down in Article 22(2) GDPR. When one of the exceptions applies, there must be measures in place to safeguard data subjects' rights and freedoms, which is mentioned in Article 22(3) GDPR. The measures that relate to the right to explanation are mentioned in Recital 71 GDPR, namely the right to obtain specific information and the right to obtain an explanation of the decision reached. The EDPB mentions that data subjects will only be able to challenge a decision or express their view if they fully understand how a decision has been made and on what basis.[100] However, the EDPB does not make clear what specific information controllers must provide. The EDPB refers to the transparency requirements of Articles 13(2)(f), 14(2)(g) and 15(1)(h) GDPR in this respect.

3.2. Terminology of the GDPR

This paragraph explains several terms of the GDPR to understand what the right to explanation means.

3.2.1. 'Automated decision-making'

Articles 13(2)(f), 14(2)(g) and 15(1)(h) GDPR mention 'automated decision-making'. To explain this definition, the EDPB refers to automated decisions as described in Article 22(1) GDPR.[101] Article 22(1) GDPR states:

> "The data subject shall have the right not to be subject to a decision based solely on automated processing, including profiling, which produces legal effects concerning him or her or similarly significantly affects him or her."

The first element to elaborate on is the term 'automated processing'. Is 'automated processing' limited to certain automated decision-making models or technologies? Does it include any human involvement? According to the EDPB, 'automated decision-making' is the ability to make decisions by technological means without human involvement.[102] Controllers can base automated decisions on any type of data, such as data provided directly by data subjects, data observed about individuals, and derived or inferred data such as a profile that a controller already has created. Controllers can make automated decisions with or without profiling.[103] The EDPB mentions the use of algorithms several times in its guidance, but does not provide specific guidance on the

97 Wachter and others (n 10) 30.
98 WP 251 (n 9) 17.
99 Ibid 27.
100 Ibid.
101 Ibid 25.
102 Ibid 8.
103 Ibid.

technological scope of automated decision-making processes.[104] Scholars provide additional guidance on certain aspects. With regard to the different automated decision-making models, Brkan notes that automated decision-making in the GDPR seems to encompass a multitude of decision types. The notion of automated decision-making is not a unitary concept, covering only a specific type of decision.[105] Following the rationale of the GDPR, it is for example clear that predictive models are included in the scope of Article 22 GDPR.[106] As explained in paragraph 1.3.1, the relative notion with regard to human involvement in the automated decision-making process prevails. The term 'automated decision-making' extends to either human or machinery decisions based solely on automated decision-making.[107] The second element to elaborate on is the term 'legal effect'. This requires that the decision affects someone's legal rights, such as the right to take legal actions. It may also affect someone's legal status.[108] The threshold for the term 'significant' is similar to that of a decision producing legal effects, for example the exclusion or discrimination of individuals that affects someone's financial circumstances.[109] To summarize, it is unclear what specific models and phases are included in the definition of 'automated decision-making'. The EDPB could give additional guidance to clarify the scope of the definition. With regard to this research, a passive 'human' decision based solely on algorithmic decision-making falls under the scope of Article 22 GDPR.

3.2.2. 'Data subjects'

A data subject is "an identifiable natural person who can be identified, directly or indirectly, in particular by reference to an identifier".[110] With regard to algorithmic decision-making, data subjects are those individuals whose personal data controllers obtain for algorithmic decision-making processes. This research elaborates more extensively on the data subjects in algorithmic decision-making processes in paragraph 4.2.

3.2.3. 'Meaningful information about the logic involved'

According to the EDPB, Articles 13(2)(f), 14(2)(g), 15(1)(h) and 22(3) GDPR try to meaningfully position data subjects so that they can vindicate their rights and hold controllers accountable for the processing of their personal data.[111] Scholars argue that 'meaningful' means that controllers must consider the data subjects' perspective.[112] According to Selbst and Powles, information must be meaningful to the data subject without particular technical expertise.[113] Kuner and others note that a high level, non-technical description of the decision-making process is likely to be meaningful.[114]

---

104 Ibid 25.

105 Brkan (n 70) 3.

106 WP 251 (n 9) 7. Recital 71 GDPR provides an example of predictive modelling, such as the automatic refusal of an online credit application.

107 WP 251 (n 9) 7.

108 Ibid 21.

109 Ibid.

110 See: Article 4(1) GDPR.

111 WP 260 (n 96) 26.

112 Kamarinou and others (n 22) 23.

113 Andrew Selbst and Julia Powles, 'Meaningful information and the right to explanation' [2017] 7(4) *International Data Privacy Law* 236.

114 Christopher Kuner and others, 'Machine Learning with personal data: is data protection law smart enough to meet the challenge?' [2017] 7(1) *International Data Privacy Law* 2.

Malgieri and Comandé argue that 'meaningful' means that information is relevant, significant, important, and intends to show the meaning of algorithmic decision-making. The explanation should be both complete and comprehensible.[115] The complexity of Machine Learning makes it difficult to understand how an automated decision-making process functions. With regard to the explanation of the 'logic involved', controllers should find simple ways to inform the data subject about the rationale behind the decision and the criteria relied on in reaching the decision. The EDPB clarifies that it is not necessary to provide a complex explanation of the algorithms. The controller does not have to disclose the full algorithm. It is sufficient to provide comprehensive information in order to make data subjects understand the reasons behind the decision.[116] The EDPB notes that comprehensive information contains the following information: (i) the data categories that controllers use in the decision-making process, (ii) the reasons why these data categories are relevant, (iii) the way controllers build a profile including statistics used in the analysis, (iv) the reasons why the profile is relevant in the process, and (v) the way controllers use the profile for a decision.[117]

3.2.4. 'Significance' and 'envisaged consequences'

With regard to the significance and the envisaged consequences, the information that controllers must provide relates to intended or future processing and must inform about the way automated decision-making might affect data subjects. Controllers should give real and tangible examples of possible effects. Controllers can use visual and interactive techniques to explain how they have made past decisions and the consequences thereof.[118]

3.2.5. 'Fair and transparent processing'

According to Recitals 60 and 71 GDPR, controllers should provide data subjects with information necessary to ensure fair and transparent processing. The principles of fair and transparent processing require controllers to inform data subjects about the existence of the processing operation and its purposes.[119] It should be transparent to natural persons that controllers collect, use, consult or otherwise process their personal data and to what extent the controller will process their personal data in the future.[120] With regard to 'fair' processing, the GDPR does not explicitly provide further guidance. The EDPB states that profiling may be unfair and may create discrimination, for example by denying people access to employment opportunities.[121] 'Fair processing' therefore relates to preventing discrimination.

3.3. Textual analysis: the scope of the different information requirements

When looked upon Articles 13(2)(f), 14(2)(g) and 15(1)(h) GDPR, controllers have to provide a different kind of information to a different scope of data subjects. All data subjects involved have a right to receive information about the existence of automated

115 Malgieri and Comandé (n 5) 257.
116 WP 251 (n 9) 25.
117 Ibid 31.
118 Ibid 26.
119 Recital 60 GDPR.
120 Recital 39 GDPR.
121 WP 251 (n 9) 10.

processing by the controller.[122] It does not matter whether automated decision-making meets the provisions of Article 22(1) GDPR. The controller only has to provide meaningful information about the logic involved, as well as the significance and the envisaged consequences of such processing when automated decision-making meets the definition of Article 22(1) GDPR.[123] This means that only data subjects that are subject to automated decisions that produce legal or similar significant effects have a right to receive this kind of information. With regard to Article 22(3) GDPR, only data subjects that have been subject to a decision based solely on automated processing which produces legal or similar significant effects have a right to obtain information and an explanation about the decision. To summarize, data subjects receive information on three different bases:

1. Data subjects that are involved in automated processing but are not subject to an automated decision that meets the definition of Article 22(1) GDPR receive information about the existence of automated decision-making, based on Articles 13(2)(f), 14(2)(g) and 15(1)(h) GDPR.
2. Data subjects that are subject to an automated decision that meets the definition of Article 22(1) GDPR receive meaningful information about the logic involved, as well as the significance and the envisaged consequences of such data processing, based on 13(2)(f), 14(2)(g) and 15(1)(h) GDPR.
3. Data subjects that are subject to an automated decision that meets the definition of Article 22(1) GDPR have a right to receive specific information and an explanation based on Article 22(3) GDPR.

3.4. The tripartite structure of the right to explanation

This paragraph elaborates on the main discussion in literature regarding the existence of the right to explanation, namely whether the GDPR requires an ex ante and/or an ex post explanation. This discussion is relevant for this research. As paragraph 4.3 will explain, transparency requires both an ex ante and ex post explanation.

3.4.1. The tripartite structure

Scholars identify two different criteria for categorizing explanations. These two criteria are content and timing. With regard to the content, Wachter and others distinguish the explanation of the 'system functionality' and the 'specific decisions'. The 'system functionality' contains the logic, significance, envisaged consequences and the general functionality of the automated decision-making system.[124] The 'specific decision' contains the rationale of, reasons for and the individual circumstances of a specific decision, such as the weighting of features, specific decision rules and information about profile groups.[125] In terms of timing, they distinguish 'ex ante' and 'ex post' explanations. An ex ante explanation occurs prior to the decision and an ex post explanation occurs after the decision. They argue that ex ante explanations can only address the system functionality and ex post explanations can address both the system

122 Ibid 16.
123 Ibid 25.
124 Sandra Wachter and others, 'Why a Right to Explanation of Automated Decision-Making Does Not Exist in the General Data Protection Regulation' [2017] 7(2) *International Data Privacy Law* 78.
125 Ibid.

functionality and the specific decisions.[126] Hence, there are three different kinds of explanations with respect to algorithmic decisions: (i) an ex ante explanation about the system functionality, (ii) an ex post explanation about the system functionality and (iii) an ex post explanation about a specific decision.[127] Other scholars recognize this tripartite structure as well.[128]

3.4.2. GDPR requirements

Scholars have different views on what the GDPR exactly requires. According to Wachter and others, the GDPR is generally forward-looking. They state that the articles establishing the right to explanation primarily aim to regulate controllers to act before the time of the collection of personal data.[129] Articles 13 and 14 GDPR namely mandate controllers to inform data subjects before they collect personal data. They argue that Article 15(1)(h) GDPR is future oriented as well, because controllers should provide information about the 'envisaged consequences'.[130] Malgieri and Comandé have also conducted a textual analysis of the GDPR. According to them, the use of the same sentence in Articles 13(2)(f) and 14(2)(g) on the one hand and Article 15(1)(h) GDPR on the other hand does not mean that their scope is the same. Data subjects can exercise the right to access personal data at any moment. When data subjects exercise the right after the controller has taken the decision, it may also lead to the disclosure of information about the specific decision. Therefore, Article 15(1)(h) GDPR does not provide a mere ex ante right to be informed.[131] The EDPB does not clarify whether the right to explanation requires an ex post explanation. The EDPB mentions that Article 15(1)(h) GDPR entitles data subjects to have the same information as required under Articles 13(2)(f) and 14(2)(g) GDPR. The EDPB also mentions that Article 15(1)(h) GDPR mandates the controller to provide the data subject with information about the envisaged consequences of the processing, rather than an explanation of a particular decision. At first sight, the EDPB seems to argue that Articles 13(2)(f), 14(2)(g) and 15(1)(h) GDPR only require an ex ante explanation. The purpose of this research is not to assess whose legal interpretation of Articles 13(2)(f), 14(2)(g) and 15(1)(h) GDPR is valid. The fact that there is an elaborate discussion in literature indicates that the legislator could define the GDPR more clearly. Nevertheless, it is worth mentioning that the whole discussion only seems to emphasize on Articles 13(2)(f), 14(2)(g) and 15(1)(h) GDPR, not taking Article 22(3) GDPR into account. According to Recital 71 GDPR, suitable measures include among others the right for data subjects to receive specific information and the right to obtain an explanation of the *decision reached*. Data subjects are only able to challenge a decision if they fully understand how a controller has made a decision and on what basis.[132] Article 22(3) and Recital 71 GDPR would therefore require an ex post explanation.

---

126 Ibid.

127 Ibid.

128 Tae Wan Kim and Bryan Routledge, 'Informational privacy, a right to explanation, and interpretable AI' (*Research gate*, August 2018) <www.researchgate.net/profile/Tae_Wan_Kim3/publication/327160853_Informational_Privacy_A_Right_to_Explanation_and_Interpretable_AI/links/5b7d7f4092851c1e1227b22e/Informational-Privacy-A-Right-to-Explanation-and-Interpretable-AI.pdf> accessed 11 September 2018 2.

129 Wachter and others (n 124) 82.

130 Ibid 83-84.

131 Malgieri and Comandé (n 5) 255-256.

132 WP 251 (n 9) 27.

3.5. General elements of transparency

The controller must take general elements of transparency into account when he provides data subjects with information. Article 12(7) GDPR clarifies that Articles 13 and 14 GDPR aim to provide data subjects with a meaningful overview of the intended processing in an easily visible, intelligible and clearly legible manner. This means that controllers must present the information efficiently and briefly in order to avoid information fatigue.[133] The EDPB recommends that controllers should use layered notices to link various categories of information rather than displaying all information in a single notice.[134] The average member of the intended audience must understand the information that the controller provides.[135] Furthermore, the controller must immediately make clear where the data subject can access information. The controller must provide the information in a manner as simple as possible, which means that the information should not contain overly technical language. The controller must write in an active rather than in a passive way. The controller must structure paragraphs and sentences by using bullets for example, and the controller should not only rely on predictable examples.[136] Controllers can trial different modalities and ask feedback on how understandable the proposed measure is for data subjects.[137] Controllers may consider publicizing a part of the DPIA in order to create trust.[138]

3.6. Intellectual property rights and trade secrets

As mentioned in paragraph 2.3.1.3, controllers might consider their algorithms as their trade secret or intellectual property. This paragraph explains how the GDPR balances the right to explanation with controllers' trade secret rights and intellectual property rights. Recital 63 GDPR asserts that the right to access should not adversely affect the rights and freedoms of others, including trade secrets and intellectual property rights. Intellectual property rights are less relevant in relation to the protection of algorithms than trade secrets in many cases. In order to obtain patent protection, the claims of the application must contain a detailed description of the invention, which will be published.[139] Therefore, patent protection is not desirable in the case controllers want to prevent that interested parties game the system. The Computer Programs Directive may protect the expression of a computer program.[140] However, this protection does not prevent disclosure of information that grants interested parties the possibility to game the system. Controllers are able to prevent disclosure when they use trade secrets, since the algorithm would remain unknown to the public in that case. Scholars have discussed the conflict between algorithmic transparency and trade secret protection. Malgieri and Comandé have conducted a textual analysis of the Directive on Trade Secrets[141] and the

133 WP 260 (n 96) 7.
134 Ibid.
135 Ibid.
136 Ibid 8-9.
137 Ibid 14.
138 Ibid 22.
139 Recital 149 of the Convention on the Grant of European Patents (European Patent Convention, as amended).
140 Recital 11 and Article 1 of Directive 2009/24/EC on the legal protection of computer programs (Computer Programs Directive), (OJ L 111/16, 5.5.2009), 23 April 2009.
141 European Parliament and the Council, Directive (EU) 2016/943 on the protection of undisclosed know-how and business information (trade secrets) against their unlawful acquisition, use and disclosure (OJ L 157/1, 15.6.2016), 8 June 2016.

GDPR. They infer a legal preference for data protection rights, even if there is a 'non-prevalence' principle between data protection law and trade secret law.[142] The EDPB clarifies that companies cannot rely on trade secret protection as an excuse to deny access or refuse to provide information, and agrees with a case-by-case approach when balancing both rights.[143] The trade secret restriction only relates to Article 15(1)(h) GDPR.[144] This means that the GDPR does not limit the right to receive ex ante meaningful information about algorithmic functionalities of Articles 13(2)(f), 14(2)(g) and 22(3) GDPR. Malgieri and Comandé note that the disclosure of rationales of specific decisions and information about auditing cannot be considered as adversely affecting trade secrets or intellectual property.[145] They argue that data controllers should at least provide specific information about the rationales of decisions while they can avoid disclosing all details about the technological functionality.[146]

3.7. Conclusion

Articles 13(2)(f), 14(2)(g), 15(1)(h) and the right to obtain specific information and an explanation about the decision of Article 22(3) GDPR establish the right to explanation. The right to explanation requires controllers to provide specific and easily accessible information about automated decision-making based solely on automated processing. Data subjects must receive meaningful information about the logic involved, as well as the significance and the envisaged consequences of such processing for the data subject when they are subject to a decision based solely on automated processing which produces legal or similar significant effects. The articles establishing the right to explanation relate to a different scope of information and a different scope of data subjects. Data subjects that are involved in automated processing but are not subject to an automated decision that meets the definition of Article 22(1) GDPR receive information about the existence of automated decision-making, based on Articles 13(2)(f), 14(2)(g) and 15(1)(h) GDPR. Data subjects that are subject to an automated decision that meets the definition of Article 22(1) GDPR receive meaningful information about the logic involved, as well as the significance and the envisaged consequences of such data processing, based on 13(2)(f), 14(2)(g) and 15(1)(h) GDPR. Data subjects that are subject to an automated decision that meets the definition of Article 22(1) GDPR have a right to receive specific information and an explanation based on Article 22(3) GDPR. It is not necessary to provide a complex explanation of the algorithm, but controllers should find simple ways to inform the data subjects about the rationale behind the decision, taking the general elements of transparency into account. The right to explanation has a tripartite structure. Scholars have different views on whether the GDPR requires an ex post explanation or not.

142 Malgieri and Comandé (n 5) 264.
143 WP 251 (n 9) 17.
144 See Recital 63 GDPR. This Recital relates to Article 15.
145 Malgieri and Comandé (n 5) 264.
146 Ibid.

4. The right to explanation in algorithmic decision-making

In order to answer the central research question, this chapter combines chapters two and three. This chapter assesses whether the right to explanations solves the legal problems arising from algorithmic decision-making. The author analyzes the text of the GDPR and looks at practical complications to identify shortcomings of the right to explanation. Furthermore, this chapter assesses whether current explanation methods fulfill the requirements of the GDPR to enhance transparency and explainability. Lastly, this chapter recommends two other concepts of the right to explanation. The author looks upon the way these concepts can take the shortcomings of the right to explanation into account.

4.1. The purpose of the right to explanation in algorithmic decision-making

This paragraph elaborates on the purposes of the right to explanation in algorithmic decision-making. Paragraph 2.3 of this research has already identified the legal problems that arise by the use of algorithmic decision-making. It creates information asymmetry because of opacity, and discrimination and unfairness. Does the right to explanation also try to solve these problems? This paragraph derives the purpose of the right to explanation from the text of the GDPR.

4.1.1. Textual analysis of the GDPR

Articles 13(2)(f), 14(2)(g) and 15(1)(h) GDPR state that a controller has to provide data subjects with information about the existence of automated decision-making to ensure *fair and transparent processing*. Transparent processing requires that controllers are transparent to data subjects about the fact that they process personal data.[147] By demanding fair processing, Articles 13(2)(f), 14(2)(g) and 15(1)(h) GDPR try to prevent discrimination and unfairness.[148] With regard to Article 22(3) GDPR, Recital 71 notes that the measures must ensure *fair and transparent processing*. Article 22(3) GDPR thus also aims to improve transparency and aims to prevent discrimination and unfairness.

4.1.2. The right to explanation: means to solve the legal problems?

Articles 13(2)(f), 14(2)(g), 15(1)(h) and 22(3) GDPR aim to position data subjects to vindicate their rights and to hold controllers accountable.[149] The right to explanation has emerged as attractive remedy, since it intuitively promises to open the algorithmic black-box and heightens accountability.[150] However, there are different reasons why the right to explanation will probably not solve the legal problems algorithmic decision-making creates, as elaborated on in the upcoming paragraphs.

4.2. Transparency: who has a right to explanation?

This paragraph answers the question who specifically has a right to explanation according to the GDPR. Does the right to explanation improve transparency for all parties involved? As elaborated on in paragraph 3.3, data subjects receive information on three different bases. What exactly does the different scope of data subjects mean in

147 Recital 39 GDPR.
148 See paragraph 3.2.5.
149 See paragraph 3.2.3.
150 Edwards and Veale 2017 (n 64) 1.

algorithmic decision-making practices? To answer this question, this paragraph analyzes the practical context of algorithmic decision-making.

4.2.1. Data subjects in practice

Kim and Routledge explore which parties are involved in an algorithmic decision-making process.[151] They break the second segment of the decision-making process, the data analysis, into three different components. Figure 2 presents a schematic diagram of this.

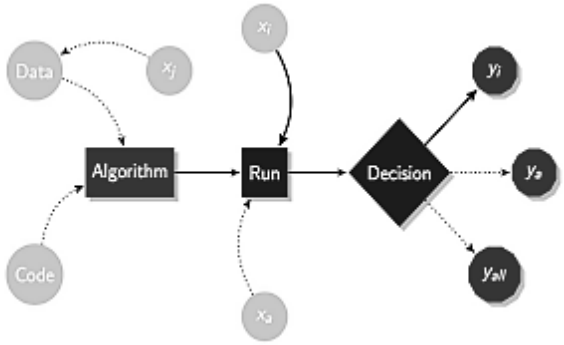

Figure 2 Schematic diagram of the data analysis[152]

In this figure, the algorithm contains a code and training data. The algorithm runs on the input data and produces a decision. Different parties are involved in the decision-making process.

4.2.1.1. The individual providing input that causes a decision (i)

Firstly, there are individuals that directly interact with the algorithm by providing their own personal data as input data. This causes the algorithm to produce a decision that has an impact on him or her. These individuals are subject to a decision based solely on automated processing, which produces legal or similar significant effects. With regard to the text of the GDPR, these individuals have a right to receive information about the existence of automated processing, meaningful information about the logic involved, as well as the significance and the envisaged consequences based on Articles 13(2)(f), 14(2)(g) and 15(1)(h) GDPR. These individuals also have a right to receive specific information and an explanation based on Article 22(3) GDPR.

4.2.1.2. The individual contributing to the training data (j)

Large-scale decision-making systems receive input from many individuals that contribute to the training data. Decision-makers collect training data from a sample of individuals. These individuals are involved in algorithmic decision-making, but are not directly subject to a decision based solely on automated processing with legal or similar significant effects.[153] This means that these individuals only receive information about the existence of algorithmic decision-making, based on Articles 13(2)(f), 14(2)(g) and 15(1)(h) GDPR.

151 Kim and Routledge (n 128) 2.
152 Ibid.
153 Ibid.

4.2.1.3. The company providing input (a)

Companies will provide input data about the products or services they provide and personal data they have on person 'i'. The data of the company may alter the decision for person 'i'.[154] According to Kim and Routledge, companies have to receive an explanation about the way their input has influenced the decision. They state that it is not clear whether the right to explanation requires that the explanation to the company and to individual 'i' are identical or equivalent.[155] This argumentation has shortcomings. Only natural persons have a right to explanation when looked upon the text of the GDPR.[156] A company (i.e. a legal person) is not a natural person, and would therefore not have a right to explanation. Since it does not fall in the scope of data protection law, the right to receive an explanation will most likely be part of the commercial relationship between the owner of the algorithm and the company.[157]

4.2.1.4. The public (all)

Decision-making algorithms are an important component of economic life. An informed public policy debate requires an explanation of algorithmic decision-making to the public at large. Policymakers are only able to make the right policy decisions when they understand how algorithms make decisions that have an impact on their citizen's lives.[158] Does the public at large have a right to explanation about the way algorithms make decisions? As noted in literature, a right to explanation for the public at large might stretch the concept of informed consent regarding automated decision-making.[159] The public at large does not need to consent to automated decision-making, since controllers do not process their personal data. Information about the decision-making process is thus not required. Based on the GDPR, the public at large does not have a right to receive information based on Articles 13(2)(f), 14(2)(g) and 15(1)(h) GDPR and does not have a right to an explanation based on Article 22(3) GDPR.

4.2.2. No transparency for all parties involved

The right to explanation does not improve transparency for all parties involved in algorithmic decision-making. As mentioned in paragraph 4.1, Articles 13(2)(f), 14(2)(g), 15(1)(h) and 22(3) GDPR have an important role in creating transparency. However, there are difficulties that prevent improving transparency with regard to the individuals contributing to the training data, the companies providing input and the public at large. Individuals that contribute to the training data only have a right to receive information about the existence of automated processing based on Articles 13(2)(f), 14(2)(g) and 15(1)(h) GDPR. They do not have the right to receive meaningful information about the logic involved, the significance and the envisaged consequences. They neither have a right to obtain an explanation based on Article 22(3) GDPR. The EDPB notes that controllers should provide this information as good practice, even though it is not required.[160] Companies (i.e. legal persons) that provide input do not fall under the scope of data protection law since they are no natural persons. Because of

154 Ibid 4.
155 Ibid.
156 See Article 4(1) GDPR: personal data means any information relating to an identified or identifiable natural person ('data subject').
157 Kim and Routledge (n 128) 4.
158 Ibid.
159 Ibid.
160 WP 251 (n 9) 25.

this, they do not have a right to explanation about how their input influences the decision-making process. With regard to the public at large, it is important to have an explanation on algorithmic decision-making systems to make informed public policy decisions. Informing the public at large about algorithmic decision-making practices might reduce public concerns, since it will become clear how controllers use personal data. The public at large does not have a right to receive information under the current data protection regime. Of course, many questions need to be asked before regulators provide a right to explanation to these parties. How should controllers provide such information? Does it really have added value? Is it even feasible? For example, there is a difficulty in providing an explanation to individuals contributing to training data because of the prevalence of the data use. Platforms may share personal data for a vast number of studies and decision-making algorithms. Twitter has provided for example personal data for studies and decision-making algorithms for speech recognition, political polls, sarcasm in language and even earthquake detection.[161] The future data uses are hard to predict. That makes a right to an explanation about how controllers use data in the future infeasible.[162] This research only intends to show that certain parties that might need a right to explanation do not have this right under the current data protection regime. It requires further research to answer the questions raised.

4.3. Transparency: ex ante and ex post explanations

As elaborated on in paragraph 3.4 of this research, there is a discussion in literature whether the right to explanation requires an ex ante and ex post explanation. The current draft of the GDPR lacks a clearly supporting expression for an ex post explanation.[163] This paragraph explains why both explanations are necessary to establish transparency. The right to explanation undermines transparency when it does not clearly support both ex ante and ex post explanations.

4.3.1. Ex ante explanations

An ex ante explanation should offer an explanation about the system functionality to reasonably well inform data subjects about the nature of algorithmic processing. Data subjects need sufficient information to consent to the data processing.[164] As elaborated on in chapter 2, there are difficulties in explaining algorithmic decision-making practices. A complete enumeration of the model is not possible because of intrinsic opacity. A complete enumeration is neither informative because of illiterate opacity. The incomplete nature of an ex ante explanation poses a serious problem. An ex ante explanation may often hardly be specific or complete about the ways the algorithmic decision-making system uses personal data and what kind of insights it makes in the end. Data subjects cannot consent to the whole process in a well-informed manner when controllers only provide an ex ante generic explanation about the system functionality.[165]

---

161 Kim and Routledge (n 128) 4.
162 Ibid.
163 Ibid 1.
164 Ibid 5.
165 Ibid.

4.3.2. Ex post explanations

Ex post explanations are meaningful in two different ways. Firstly, ex post explanations offer harmed individuals an explanation about how the algorithmic system has created the impact on them, including specific features used in the processing. The ex post explanation is tailored to the data subject that is subject to the automated decision.[166] The explanation might contain the rationale of, reasons for and the individual circumstances of a specific decision, such as the weighting of features and specific decision rules.[167] This enables harmed individuals to challenge the decision. Secondly, data subjects have a so called 'right to an updating explanation'.[168] An informative ex post explanation on the system functionality explains the actual factors that the algorithm has used to make a decision instead of the assumed factors.[169] Algorithmic decision-making may involve risks or uncertainties that controllers do not foresee before the processing, because self-learning algorithms are capable of changing itself or its set of instructions based on accumulated data. Ex post explanations about the system functionality might therefore differ from ex ante explanations about the system functionality. This ex post explanation provides a possibility to opt-out.

4.3.3. Ex ante and ex post explanations required

Data subjects cannot consent to the whole process in a well-informed manner when controllers only provide an ex ante generic explanation about the system functionality. An ex post explanation is also necessary, since these explanations offer harmed individuals an explanation about how the algorithmic system has created the impact on them, which enables them to challenge the decision. The 'right to an updating explanation' is necessary because of the unforeseen risks and uncertainties that arise by the use of self-learning algorithms. Companies must assure data subjects about their readiness to offer ex post information that enables them to redress in the case of harm, or to opt-out in the case of unexpected risks.[170] The current draft of the right to explanation lacks a clearly supporting expression for ex post information. Controllers will not provide such information to data subjects when it is not specifically required. Therefore, transparency lacks as long as the GDPR does not clearly support an expression for an ex post explanation.

4.4. Discrimination and unfairness: lack of specific guidance

4.4.1. Textual analysis of the GDPR

Articles 13(2)(f), 14(2)(g) and 15(1)(h) GDPR demand controllers to provide information about the existence of automated decision-making to prevent unfair processing and discrimination.[171] However, neither the text of the GDPR nor the EDPB clarifies what specific information controllers must provide with regard to discrimination and unfairness. Does it mean that controllers must provide information about whether they use discriminating and unfair factors or not? Does it mean that controllers must provide information about the specific factors and discriminatory and/or unfair correlations? Article 22(3) GDPR tries to prevent discrimination as well.

166 Ibid 5.
167 Ibid 1.
168 Ibid 5.
169 Ibid.
170 Ibid 9.
171 WP 251 (n 9) 10.

'Suitable measures' must prevent discriminatory effects on natural persons.[172] The EDPB does not clarify what information controllers must provide to data subjects or what information the explanation should contain. The EDPB only elaborates on other means to prevent discrimination, such as algorithmic auditing and quality assurance checks.[173] The right to explanation hinders the prevention of discrimination and unfairness as long as the GDPR or the EDPB does not specifically indicate what information controllers have to provide and what information explanations must contain.

4.4.2. Practical complications

Apart from the lack of guidance, there are also some other practical complications that have to be taken into account. Firstly, controllers must carefully consider how they provide information to data subjects. The public might misunderstand the meaning of correlations that algorithms use in the decision-making process, or the public might in general wrongfully interpret correlations as characteristics about individuals.[174] Correlations might cause the mistreatment of individuals by other people based on the personal traits they share with the pattern.[175] Secondly, controllers might not detect certain discriminating or unfair correlations because of intrinsic algorithmic opacity.[176] Factors that are neutral at first sight may be correlated with sensitive characteristics without controllers being aware of this. This means that data subjects cannot become aware of all discriminating or unfair factors that algorithms use in the decision-making process.

4.5. Transparency: is the right to explanation a remedy for opacity?

There are not only specific characteristics of the right to explanation that hinder the ability to solve the legal problems. It is important to look at a more general question as well. Is the right to explanation even able to solve opacity?

4.5.1. The crux of 'meaningful' information

Data subjects have a right to receive *meaningful* information. The terminology of 'meaningful' prevents organizations from providing complex explanations of the algorithms and decision-making systems that they use.[177] This lowers the burden for organizations to provide explanations to data subjects, which makes it easier for controllers to be GDPR-compliant. On the other hand, this limits the amount of specific information data subjects will receive. Controllers do not have to be completely transparent.

4.5.2. Intrinsic, illiterate and intentional opacity

The right to explanation requires that controllers provide information and an explanation about algorithmic decision-making. However, this does not take away the fact that some algorithmic decision-making systems are simply not understandable and

172 Paragraph 4.1.1.
173 WP 251 (n 9) 32.
174 Zarsky 2013 (n 76) 1560-1561.
175 Ibid.
176 Bryce Goodman, 'A Step Towards Accountable Algorithms?: Algorithmic Discrimination and the European Union General Data Protection' (*ML and the law*, 2016) <www.mlandthelaw.org/papers/goodman1.pdf> accessed 15 March 2018 3.
177 Paragraph 3.2.3.

remain opaque. Hence, the right to explanation does not solve intrinsic opacity. Controllers must provide information that is meaningful to the general public with limited technical expertise. This meets the problem regarding illiterate opacity in one respect. Controllers have to provide information in such way that the majority of the people understands how a system generates the decision. On the other hand, people will never understand the principles behind the algorithmic decision-making practices when controllers only have to provide information in such understandable and incomplete way. The right to explanation solves some issues regarding intentional opacity. The right to access should not adversely affect intellectual property and trade secret rights of controllers.[178] Scholars argue that providing the required information cannot be considered as an adverse effect on controllers. The right to explanation does not demand controllers to provide a complex explanation of the algorithm or to disclose all details about the technological functionality. This means that controllers do not have to be afraid that interested parties game the system or that competitors use technological information about the system and the algorithms to improve their position.

4.6. The right to explanation in practice

As noted in the previous paragraphs of this chapter, the right to explanation tries to enhance transparency and tries to prevent discrimination and unfairness. However, several characteristics of the right to explanation hinder its capacity to solve the legal problems. It is important to look at current explanation methods to see whether these are sufficient to establish a right to explanation. The explanation methods will not enhance transparency and explainability and will not solve the legal problems when they do not fulfill the requirements of the GDPR.

4.6.1. Model-centric explanations

Model-centric explanations provide general information about algorithmic decision-making models and do not relate to specific decisions or input-data. The explanations offer insight into the working of the model, for example the intentions behind the modelling process and information about the predictive skill of the model.[179] They provide information about the logic involved in automated processing, but they do not provide information about the significance and the envisaged consequences of such processing for the data subject. Therefore, model-centric explanations do not meet the requirements of the right to explanation on their own.

4.6.2. Subject-centric explanations

Subject-centric explanations relate to input records. Controllers can only provide subject-centric explanations in reference to a given query.[180] There are different types of subject-centric explanations.

4.6.2.1. Sensitivity-based explanations: counterfactual explanations

Wachter, Mittelstadt and Russel suggest that controllers should offer counterfactual explanations. Counterfactual explanations describe the smallest change that would obtain a desirable outcome. An example of such counterfactual explanation is:

178 Recital 63 GDPR.
179 Edwards and Veale 2017 (n 64) 55.
180 Ibid 56.

> "You were denied an insurance premium because your annual income was X. If your income had been Y, you would have been offered an insurance premium."

Counterfactual explanations have several advantages. Counterfactuals avoid the challenge of explaining internal workings of complex systems by describing the dependency of factors, and do not require that data subjects understand the internal logic of the model.[181] There are also some drawbacks. Controllers cannot always alter data because of their nature, for example data subjects' age.[182] Providing one counterfactual explanation would therefore be insufficient. Articles 13(2)(f), 14(2)(g) and 15(1)(h) GDPR require that controllers inform data subjects about the rationale behind the decision.[183] Counterfactuals do not provide information about the rationale of the decision.[184] Furthermore, counterfactuals do not provide the statistical evidence that controllers and data subjects need to assess whether algorithms are fair and free of bias.[185] Because of this, counterfactual explanations are not sufficient to establish a right to explanation.

4.6.2.2. Input influence-based explanation

Ribeiro, Sing and Guestrin argue that controllers must present textual or visual artifacts that provide an understanding of the prediction of the model and the relationship between components.[186] Explanations must treat the original model as a black-box since many classifiers are not interpretable.[187] They use algorithms that enable end-users to see a list of features that have contributed to the output.[188] Input influence-based explanations are not sufficient to provide a meaningful explanation to data subjects. This kind of explanation only clarifies which features have contributed to the output, together with their corresponding weights. It does not clarify why algorithms use the specific categories of data, how the system creates profiles and what the consequences are for data subjects.

4.6.2.3. Case-based explanations

Case-based explanations present a case from the training data that is most similar to the decision that the controller has to explain.[189] According to Nugent and Cunningham, case-based explanations solve the problems regarding interpretability.[190] They argue that explanations based on the feature ranking and the presentation of selected cases provide data subjects with suitable explanations.[191] Knowledge-intensive case-based

---

181 Wachter and others (n 10) 20.

182 Ibid 12.

183 WP 251 (n 9) 25.

184 Wachter and others (n 10) 43.

185 Ibid.

186 Marco Tulio Ribeiro and others, 'Why should I trust you? Explaining the predictions of any classifier' (*Kdd*, 2016) <www.kdd.org/kdd2016/papers/files/rfp0573-ribeiroA.pdf> accessed 11 July 2018 1.

187 Ibid.

188 Binns and others (n 23) 2.

189 Conor Nugent and Pádraig Cunningham, 'A case-based explanation system for 'black-box'systems' (*SCSS*, 2010) <www.scss.tcd.ie/publications/tech-reports/reports.04/TCD-CS-2004-20.pdf> accessed 14 July 2018 2.

190 Ibid.

191 Ibid 3.

explanations include the specific technical feature ranking to generate explanations. Knowledge-light case-based explanations express the case features.[192] An example is:

> 'The system predicts that the outcome will be X because this was the outcome in case Y. Case Y only differs from the current case in the value of feature F which was f1 instead of f2'.[193]

Case-based explanations do not provide a meaningful explanation to data subjects. As well as rule-based explanations, they only clarify the features that have contributed to the output, together with their corresponding weights. They do not indicate the other required elements.

4.6.2.4. Demographic explanations

Demographic explanations present statistics on the outcome classes for people in the same demographic category, such as gender and income level.[194] This is relevant with regard to discriminatory aspects of the decision, but is not sufficient to provide a meaningful explanation to data subjects. It does not indicate the other required elements of the GDPR.

4.6.3. Perceptions of data subjects

It is important to look at the perception of data subjects towards explanation styles because information must be meaningful to the data subjects. Binns and others' research looks upon the effects of explanations on people's perceptions regarding algorithmic decisions.[195] They have conducted a set of experiments to see people's responses when they were faced with aforementioned explanation styles. The research aims to provide an overview of the ways in which explanation styles might make algorithmic decision-making transparent and justified. The subjects perceive the (lack of) human touch, the interpretation of the reasoning of the system, the acceptability of statistical inference, and the degree of actionability important in an explanation.[196] Explanation styles do generally not affect justice perceptions when subjects are presented with one explanation style. Case-based explanations result in significantly lower perceptions of appropriateness, fair process and deservedness when subjects are presented with multiple explanation styles.[197] Binns and others recognize that the experiment has certain shortcomings and that future work is necessary.[198]

4.6.4. We need more: the legibility test and data chain traceability

Current explanation methods do not fulfill the requirements of the GDPR. Model-centric explanations do not provide information about the significance and the envisaged consequences for data subjects. Subject-centric explanations are data subject-centered, but none of the subject-centric explanations provides the required information. Data subjects do not perceive case-based explanations as an appropriate explanation method. Therefore, we need to look at other possible explanation methods.

---

192 Pádraig Cunningham and others, 'An evaluation of the usefulness of case-based explanation' (*Research gate*, 2003) <www.researchgate.net/publication/225070358_An_Evaluation_of_the_Usefulness_ of_ Case-Based _Explanation> accessed 14 July 2018 3.

193 Ibid.

194 Binns and others (n 23) 2.

195 Ibid 1.

196 Ibid 6.

197 Ibid 9.

198 Ibid 10.

4.6.4.1. Legibility test

Malgieri and Comandé introduce the so called 'legibility test'. They distinguish the logic of the decision-making system (the 'Architecture') and the significance and the consequences for data subjects (the 'Implementation'). The right to explanation contains a duty to perform an audit on both aspects.[199] Controllers have to disclose the answers of the composed questionnaire to comply with the duty to provide an explanation. The legibility test has several advantages. It is an indicator for transparency, comprehensibility and non-discrimination for data subjects.[200] It is useful for controllers to comply with the duties of Articles 13(2)(f), 14(2)(g), 15(1)(h) and 22(3) GDPR, it improves the quality of the decision-making system and it helps controllers to demonstrate that no discrimination or unequal treatment occurs.[201] The questionnaire contains all required information. Controllers need to analyze three elements with regard to the Architecture. These elements are the creation of the algorithm, its functioning and use, and its expected outputs.[202] The elements that are relevant with regard to the Implementation are among others the purposes of processing, the level of human intervention, statistical impacts on past customers and the possibility to reconsider a decision.[203] Malgieri and Comandé note that controllers need to explore the actual structure and content of the proposed questionnaire according to the environment of the automated decision-making process.[204]

4.6.4.2. Data chain traceability

Zarsky has set forth a conceptual framework to understand the role of transparency in algorithmic decision-making. Zarsky argues that transparency refers to the variety of phases within the decision-making process. As elaborated on in paragraph 2.2.2, the Big Data process consists of three different phases: (i) the collection of data and the aggregation of datasets, (ii) the data analysis and (iii) the actual use of the model.[205] Thelisson and others also propose a so called 'data chain traceability' in order to explain algorithmic decision-making. The data chain of Thelisson and others only contains the different steps in the second phase of the decision-making process.[206] The author prefers Zarsky's framework because the GDPR also requires controllers to provide information about the first and third phase of the Big Data process. The concept of data chain traceability has the potential to fulfill the legislative requirements of the GDPR. Controllers can include all required information in the relevant decision-making phase. Controllers can include the collected categories of data and the reasons why these categories are relevant in the first phase, since this information relates to the collection of the data. Controllers can include information about the way in which the system creates a profile including the statistics used in the analysis in the second phase. Controllers can include information about the reasons why the profile is relevant for the decision, information about the way the controller uses the profile for a decision, and the significance and the envisaged consequences for data subjects in the third phase. Furthermore, controllers can include aforementioned explanation methods in the

199 Malgieri and Comandé (n 5) 258.
200 Ibid 259.
201 Ibid 260.
202 Ibid 260-261.
203 Ibid 259.
204 Ibid 265.
205 Zarsky 2013 (n 76) 1521.
206 Thelisson and others (n 86) 4.

relevant phase of the data chain. Controllers can use the legibility test of Malgieri and Comandé as a format to provide technical information within the second phase. Input influence-based explanations relate to the second phase as well, since this explanation presents the value of factors that algorithms use in the decision-making process. Counterfactual explanations, demographic-based explanations and case-based explanations indicate when data subjects are subject to a specific decision. These explanation methods relate to the use of the decision-making model and belong in the third phase. Besides the three aforementioned steps, the author suggests to include a 'pre-phase'. The pre-phase contains a general explanation of the decision-making practice, and the reasons why automated decision-making is used. This information is relevant in relation to discrimination and unfairness, since the use of data driven decision-making processes may result in individuals being denied based on the actions of others with whom they share characteristics, instead of their own actions.[207] Figure 3 presents a visual representation of a data chain.

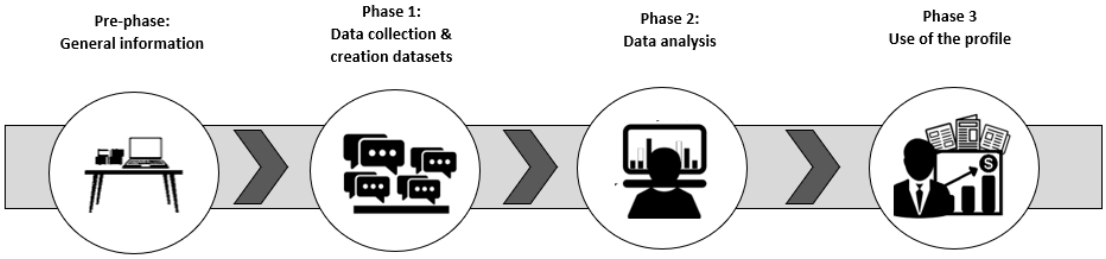

Figure 3 Data chain traceability in algorithmic decision-making

4.7. Proposed explanation methods and identified shortcomings combined

This paragraph elaborates on the possibilities to deal with the identified shortcomings of the right to explanation in the legibility test and data chain traceability. This research has shown that certain parties that might need a right to explanation do not have this right under the current data protection regime, but this research does not assess whether this is also feasible.[208] Therefore, this paragraph only mentions whether controllers can provide all parties involved with explanations in the format of the legibility test or data chain traceability.

4.7.1. Legibility test

4.7.1.1. Data recipients

Malgieri and Comandé do not explicitly clarify which parties receive the outcome of the legibility test. They do note that data subjects can only exercise the right to receive meaningful information about the logic involved, the significance and the consequences when Articles 22(1) and 22(3) GDPR apply.[209] This means implicitly that only the individuals that are subject to a decision have a right to receive the outcome of the legibility test. However, Malgieri and Comandé do not specifically address the questions of the legibility test to these individuals. The legibility test asks general questions with regard to the significance and the envisaged consequences, such as: "What are the possible effects on data subjects? Do they encompass legally recognized

207 Hurley and Adebayo (n 44) 183.
208 See: paragraph 4.2.
209 Malgieri and Comandé (n 5) 265.

rights or freedoms?" These questions do not relate to the specific individual that receives an outcome decision. Controllers could therefore also provide the outcome of the legibility test to other parties involved. Moreover, certain questions are specifically relevant for the individuals that contribute to the training data.[210] Controllers could decide to share the outcome of these questions with these individuals.

4.7.1.2. Ex ante and ex post

Malgieri and Comandé argue that Articles 13 and 14 GDPR provide a right to receive ex ante information, and Article 15 GDPR provides a right to receive ex post information after specific requests of data subjects.[211] However, the legibility test only seems to provide ex ante information when looked upon the questions of the test. It contains questions emphasizing on the possible effects on data subjects instead of the real effects on data subjects. Moreover, it contains questions considering past data subjects instead of the current data subjects.[212] Companies must assure data subjects about their readiness to offer ex post information that enables them to redress in the case of harm, or to opt-out in the case of unexpected risks.[213] Therefore, the author recommends including the following questions in the legibility test: (i) does the controller provide ex post information in the case data subjects request this, (ii) does the controller provide ex post information in the case of unforeseen risks and consequences, and (iii) what ex post information does the controller provide? Furthermore, controllers can draft an ex post version of the legibility test that contains an updated version of the information about the system functionality, information about the specific decision and the real effects on current data subjects.

4.7.1.3. Discrimination and unfairness

The legibility test contains only one question regarding discrimination, namely

> "considering past data subjects, can the data controller show that the outputs of that decision-making were not illegitimately discriminatory in statistical terms?"[214]

The legibility test does not contain any questions regarding unfair aspects of the decision-making process. Therefore, the author recommends including a question regarding unfair aspects in the decision-making process. Furthermore, it is important that controllers carefully consider how they provide information to data subjects to prevent that data subjects misunderstand the meaning of correlations. Therefore, the author recommends including the following questions: (1) has the controller considered how to provide information to prevent that the public misunderstands the meaning of correlations, and (2) how does the controller try to prevent this?

4.7.2. Data chain traceability

4.7.2.1. Data recipients

Data chain traceability makes it possible to identify the relevant data recipients per phase of the Big Data process.[215] Zarsky has identified the relevant data recipients per phase. He argues that the general public should receive the information about the first

---

210 Ibid 260.
211 Ibid 247.
212 Ibid 261.
213 Kim and Routledge (n 128) 9.
214 Malgieri and Comandé (n 5) 261.
215 Zarsky 2013 (n 76) 1532.

phase. It will namely create interest and uproar among the public when the public considers the collection and aggregation problematic.[216] Compared to paragraph 4.2, this means that all parties involved should receive information about the first phase. Zarksy argues that none of the parties involved should receive information about the second phase because of the technical nature of the information. The potential data recipients in this phase are the internal and external auditors.[217] However, this reasoning has shortcomings. Controllers *can* provide certain information in an understandable way, for example by using the outcome of the legibility test. Zarsky argues that the data recipients in the third phase are both the individuals that receive a decision and the general public. General issues as discrimination might namely be interesting for the public at large.[218] Compared to paragraph 4.2, this means that all parties involved should receive information about the third phase. The author has suggested to include a 'pre-phase'. Following Zarsky's rationale, all parties involved should receive information about the pre-phase because they should know whether and why algorithmic decision-making is taking place.

4.7.2.2. Ex ante and ex post

Data chain traceability offers the possibility to provide ex ante and ex post information. Zarsky notes that the third phase of the Big Data process emphasizes on the actual strategies and practices of the models. Controllers can only measure the effectiveness of the model by assessing them ex ante and ex post, in which the ex post information is assessed through a feedback process.[219] Controllers must regularly inform the public on the state of flaws in the data analysis as part of the ex post disclosure requirements.[220]

4.7.2.3. Discrimination and unfairness

Paragraph 2.3.2 of this research notes that discrimination and unfairness may arise from the decision to use an algorithm. Discrimination and unfairness may also arise in all the phases of the modelling process. Controllers could provide information about the factors that might create discrimination and unfairness in each relevant phase of the data chain. In that case, the pre-phase contains information about the reasons why algorithmic decision-making is taking place. The first phase contains information about the data that is collected. The second phase contains information about the way model developers have defined target variables, about the way training data is gathered, and contains information about sensitive characteristics in correlations. It indicates whether objective or subjective labeling takes place. The third phase indicates whether the model uses any direct or indirect discriminatory or unfair factors.

4.8. Conclusion

The right to explanation of the GDPR has the purpose to solve the legal problems concerning information asymmetry because of opacity, and discrimination and unfairness. However, there are some reasons why the right to explanation will probably not solve these legal problems. Firstly, not all parties involved in algorithmic decision-making have a right to explanation. The terminology of the right to explanation prevents

216 Ibid 1535.
217 Ibid.
218 Ibid 1535 – 1536.
219 Ibid 1521.
220 Ibid 1560.

improving transparency with regard to the individuals contributing to the training data, companies providing input and the public at large. Secondly, the current draft of the right to explanation lacks a clearly supporting expression for ex post information. Ex post explanations are necessary to well inform data subjects. Thirdly, the right to explanation hinders the prevention of discrimination and unfairness as long as it does not specifically indicate what information controllers should provide. The right to explanation does not solve intrinsic opacity. The right to explanation meets the problem regarding illiterate opacity in the sense that controllers must provide information in such way that data subjects understand how the system has generated a decision. The right to explanation solves some issues regarding intentional opacity, since controllers do not have to disclose all details about the technological functionality. This means that they do not have to be afraid that interested parties game the system or that competitors use the information to improve their position. Research into the explanation methods that controllers currently use, shows that these explanation methods do not fulfill the requirements of the GDPR. An explanation in the format of the legibility test or data chain traceability might fulfill the requirements of the GDPR. Controllers can include all required information in the questionnaire of the legibility test or in the relevant decision-making phase of the data chain. Controllers can take the shortcomings of the right to explanation into account when they provide data subjects with an explanation in the format of the legibility test or data chain traceability. With regard to the legibility test, controllers could (i) provide the outcome of the test to all parties involved, (ii) include questions relating to ex post information and draft an ex post version of the test, and (iii) include a question about unfair aspects and make sure that the public does not misunderstand the meaning of correlations. With regard to data chain traceability, controllers could (i) provide the relevant data recipients per phase with meaningful information, (ii) include ex post information in the third phase of the chain, and (iii) include information about the factors that might create discrimination and unfairness in each relevant phase of the data chain.

5. Conclusion

This research aims to answer the following research question: "Are the 'right to explanation' about decisions based solely on automated decision-making of Articles 13(2)(f), 14(2)(g), 15(1)(h) and 22(3) of the General Data Protection Regulation, as well as current explanation methods, able to solve the legal problems arising from algorithmic decision-making?" In order to answer this research question, this research answers the following sub-questions:

1. What legal problems does algorithmic decision-making create?
2. What is the 'right to explanation' of Articles 13(2)(f), 14(2)(g), 15(1)(h) and 22(3) GDPR?
3. Are the right to explanation and current explanation methods able to solve the legal problems arising from algorithmic decision-making?

5.1. The legal problems arising from algorithmic decision-making

In algorithmic-decision making, controllers use algorithms to make all different kind of decisions that have an impact on human lives. Actions and outcomes of self-learning algorithms are not always foreseeable, since they have the ability to act autonomously. People generally do not know what happens between the moment model developers feed input into the algorithm and the moment the algorithm provides the output. Scholars consider algorithmic decision-making systems as a black-box. Different legal problems arise from algorithmic decision-making. Firstly, algorithmic decision-making systems are opaque. There are different kinds of opacity. Intrinsic opacity refers to the opaque nature of algorithms. Illiterate opacity means that most people lack the technical skills to understand algorithms. Intentional opacity means that organizations do not want others to know how their systems work and decide to withhold information about the way they make decisions. Opacity results in informational asymmetry. Secondly, algorithmic decision-making creates discrimination and unfairness. Discrimination and unfairness may arise in every phase of the modelling process.

5.2. The right to explanation of Articles 13(2)(f), 14(2)(g), 15(1)(h) and 22(3) GDPR

The GDPR provides the right to explanation to enhance transparency and explainability. Articles 13(2)(f), 14(2)(g), 15(1)(h) and the right to obtain specific information and an explanation of the decision of Article 22(3) GDPR establish the right to explanation. Articles 13 and 14 GDPR require controllers to notify individuals when they obtain data. Article 15 requires controllers to provide confirmation as to whether or not they process data subject's personal data. Article 22(3) GDPR requires controllers to implement suitable measures to safeguard data subject's rights and freedom. According to Recital 71 GDPR, such measures include at least the provision of specific information and an explanation of the decision to data subjects. The right to explanation requires controllers to provide meaningful information about the logic involved, as well as the significance and the envisaged consequences of such processing for the data subject in the case the decision has legal or similar significant effects. The articles establishing the right to explanation relate to a different scope of information and a different scope of data subjects. Data subjects that are involved in automated processing, but are not subject to an automated decision that meets the definition of Article 22(1) GDPR, receive information about the existence of automated decision-making, based on Articles 13(2)(f), 14(2)(g) and 15(1)(h) GDPR. Data subjects that are

subject to an automated decision that meets the definition of Article 22(1) GDPR receive meaningful information about the logic involved, as well as the significance and the envisaged consequences of such data processing, based on 13(2)(f), 14(2)(g) and 15(1)(h) GDPR. Data subjects that are subject to an automated decision that meets the definition of Article 22(1) GDPR have a right to receive specific information and an explanation based on Article 22(3) GDPR. It is not necessary to provide a complex explanation of the algorithm. The controller should find simple ways to tell the data subject about the rationale behind the decision, taking the general elements of transparency of the GDPR into account. The right to explanation has a tripartite structure, but scholars have different views on whether the GDPR requires an ex post explanation or not.

5.3. The right to explanation: insufficient means for 'white-boxing' the black-box

The right to explanation of Articles 13(2)(f), 14(2)(g) and 15(1)(h) and 22(3) GDPR has the purpose to solve the legal problems concerning information asymmetry because of opacity, and discrimination and unfairness. However, there are some reasons why the right to explanation will probably not solve these legal problems. Firstly, not all parties involved in algorithmic decision-making have a right to explanation. The terminology of the right to explanation prevents improving transparency with regard to the individuals contributing to the training data, companies providing input and the public at large. Secondly, the current draft of the right to explanation lacks a clearly supporting expression for ex post information. Ex post explanations are necessary to well inform data subjects. Thirdly, the right to explanation hinders the prevention of discrimination and unfairness as long as it does not specifically indicate what information controllers should provide. The right to explanation does not solve intrinsic opacity. The right to explanation meets the problem regarding illiterate opacity in the sense that controllers must provide information in such way that data subjects understand how the system has generated a decision. The right to explanation solves some issues regarding intentional opacity, since controllers do not have to disclose all details about the technological functionality. This means that they do not have to be afraid that interested parties game the system or that competitors use the information to improve their position. Research into explanation methods that controllers currently use, shows that these explanation methods do not fulfill the requirements of the GDPR. An explanation in the format of the legibility test or data chain traceability might fulfill the requirements of the GDPR. Controllers can include all required information in the questionnaire of the legibility test or in the relevant decision-making phase of the data chain.

5.4. Recommendations

Transparency and explainability can only be useful to enhance accountability when there is sufficient motive on the part of the controller to disclose information. The current draft of the right to explanation has shortcomings that make it difficult to solve the legal problems arising from algorithmic decision-making. The author recommends several adjustments and further research to solve these shortcomings. The EDPB has an important role in this by providing guidance. Firstly, the EDPB could provide additional guidance to clarify the technological scope of 'automated decision-making'. Secondly, the author recommends further research to figure out whether it is feasible to provide a right to explanation to other parties than the data subject that is subject to a decision with legal or similar significant effects. When it is feasible, the GDPR should extend the

right to explanation to these parties to improve transparency. Thirdly, the right to explanation should clearly require ex post explanations. Transparency will lack as long as the GDPR does not clearly require ex post information. Fourthly, the right to explanation should specifically demand which information controllers must provide regarding discrimination and unfairness. Controllers must carefully consider how they provide information to data subjects since the public might misunderstand the meaning of correlations. Controllers have to implement aforementioned recommendations when they use the legibility test or data chain traceability as explanation method in order to improve transparency and to prevent discrimination and unfairness. With regard to the legibility test, controllers could (i) provide the outcome of the test to all parties involved, (ii) include questions relating to ex post information and draft an ex post version of the test, and (iii) include a question about unfair aspects and make sure that the public does not misunderstand the meaning of correlations. With regard to data chain traceability, controllers could (i) provide the relevant data recipients per phase with meaningful information, (ii) include ex post information in the third phase of the chain, and (iii) include information about the factors that might create discrimination and unfairness in each relevant phase of the data chain.

Bibliography

Legislation

European Patent Convention
Convention on the Grant of European Patents (European Patent Convention, as amended)

Computer Programs Directive
European Parliament and the Council, Directive 2009/24/EC on the legal protection of computer programs (Computer Programs Directive), (OJ L 111/16, 5.5.2009), 23 April 2009

General Data Protection Regulation
European Parliament and Council of the European Union, 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, repealing Directive 95/46/EC (General Data Protection Regulation), (OJ L 119, 4.5.2016), 27 April 2016

Directive on Trade Secrets
European Parliament and the Council, Directive (EU) 2016/943 on the protection of undisclosed know-how and business information (trade secrets) against their unlawful acquisition, use and disclosure (OJ L 157/1, 15.6.2016), 8 June 2016

Guidelines European Data Protection Board

WP 251
Article 29 Data Protection Working Party, 'Guidelines on Automated individual decision-making and Profiling for the purposes of Regulation 2016/679 (wp251rev01)' (*EC*, 6 February 2018) <ec.europa.eu/newsroom/Article29/item detail.cfm?item_id=612053> accessed 1 April 2018

WP 260
Article 29 Data Protection Working Party, 'Guidelines on transparency under Regulation 2016/679 (WP260)' (*EC*, 24 January 2018) <ec.europa.eu/newsroom/Article29/item-detail.cfm?item_id=615250> accessed 1 April 2018

Case law

The Dutch Administrative Jurisdiction Division of the Council State 2017
The Dutch Administrative Jurisdiction Division of the Council State, *Stichting Werkgroep Behoud de Peel v. het college van gedeputeerde staten van Noord-Brabant,* 17 May 2017, ECLI:NL:RVS:2017:1259

Books

Chynoweth
Paul Chynoweth, Legal research. in Andrew Knight and Les Ruddock (eds), *Advanced Research Methods in the Built Environment* (Wiley-Blackwell 2008) 28-38

Gunter
Lacey Gunter, *Variable selection for decision making* (The University of Michigan 2009) 1-91

Articles

Association for payment clearing services and others
Association for payment clearing services and others, 'Guide to credit scoring 2000' (*Experian*, 2000) <www.experian.co.uk/assets/responsibilities/brochures/guideToCreditScoring.pdf> accessed 25 March 2018

Binns and others
Reuben Binns and others, 'It's Reducing a Human Being to a Percentage: Perceptions of Justice in Algorithmic Decisions' (*Arxiv*, 31 January 2018) <arxiv.org/pdf/1801.10408.pdf> accessed 10 March 2018

Brkan
Maja Brkan, 'AI-supported decision-making under the General Data Protection Regulation' (*ACM*, June 2017) <dl.acm.org/citation.cfm?id=3086513> accessed 27 April 2018 3-8

Burrel
Jenna Burrell, 'How the machine 'thinks': Understanding opacity in machine learning algorithms' [2016] *Big Data & Society* 1-12

Bygrave
Bygrave, 'Automated profiling, minding the machine: Article 15 of the EC Data Protection Directive and automated profiling', [2001] 17(1) *Computer Law & Security Review* 17-24

Caplan and others
Caplan and others, 'Algorithmic accountability: a primer', (*Data & Society*, April 2018), <datasociety.net/wpcontent/uploads/2018/04/Data_Society_Algorithmic_Accountability_Primer_FINAL-4.pdf> accessed 23 October 2018

Cunningham and others
Pádraig Cunningham and others, 'An evaluation of the usefulness of case-based explanation' (*Research gate*, 2003) <www.researchgate.net/publication/225070358_An_Evaluation_of_the_Usefulness_of_Case-Based_Explanation> accessed 14 July 2018

Danaher
John Danaher, 'The threat of algocracy: reality, resistance and accommodation' [2016] 29(3) *Philosophy and Technology* 245-268

Deloitte
Deloitte, 'Credit scoring Case study in data analytics' (*Deloitte*, 18 April 2016) <www2.deloitte.com/content/dam/Deloitte/global/Documents/Financial-Services/gx-be-aers-fsi-credit-scoring.pdf> accessed 5 March 2018

Diakopoulos
Nicolas Diakopoulos, 'Algorithmic Accountability, Journalistic investigation of computational power structures' [2015] 3(3) *Digital Journalism* 398-415

Edwards and Veale
Lilian Edwards and Michael Veale, 'Enslaving the Algorithm: From a 'Right to an Explanation' to a 'Right to Better Decisions'?' [2018] Forthcoming IEEE *Security & Privacy* <papers.ssrn.com/sol3/papers.cfm?abstract_id=3052831> accessed 10 March 2018

Edwards and Veale 2017
Lilian Edwards and Michael Veale, 'Slave to the algorithm? Why a right to an explanation is probably not the remedy you are looking for' (*SSRN*, 6 December 2017) <papers.ssrn.com/sol3/papers.cfm?abstract_id=2972855> accessed 20 September 2018 1-67

Eiband and others
Malin Eiband and others, 'Normative vs Pragmatic: Two Perspectives on the Design of Explanations in Intelligent Systems' (*Explainable systems*, 2018) <explainablesystems.comp.nus.edu.sg/wp-content/uploads/2018/02/exss_7_eiband.pdf> accessed 11 March 2018

Gerards and others
Janneke Gerards and others, 'Algoritmes en grondrechten', Universiteit Utrecht (*Officiële bekendmakingen*, 27 August 2018) <zoek.officielebekendmakingen.nl/blg-853458> accessed 28 August 2018

Goodman
Bryce Goodman, 'A Step Towards Accountable Algorithms?: Algorithmic Discrimination and the European Union General Data Protection' (*ML and the law*, 2016) <www.mlandthelaw.org/papers/goodman1.pdf> accessed 15 March 2018

Goodman and Flaxman
Bryce Goodman and Seth Flaxman, 'EU Regulations on algorithmic decision-making and a right to explanation' (*Arxiv*, 28 June 2016) <arxiv.org/pdf/1606.08813v1.pdf> accessed 1 April 2018

Hildebrandt 2012
Mireille Hildebrandt, ‘The Dawn of a Critical Transparency Right for the Profiling Era’ (*Selected works*, 2012) <works.bepress.com/mireille_hildebrandt/40/> accessed 2 April 2018

Hildebrandt 2017
Mireille Hildebrandt, ‘Privacy As Protection of the Incomputable Self: From Agnostic to Agonistic Machine Learning’ (*SSRN*, 1 December 2017) <papers.ssrn.com/sol3/papers.cfm?abstract_id=3081776> accessed 5 March 2018

Hurley and Adebayo
Mikella Hurley and Julius Adebayo, 'Credit scoring in the era of big data' [2016] 18(1) *Yale Journal of Law and Technology* 148-216

Informatics Europe and EUACM
Informatics Europe and EUACM, 'When Computers Decide: European Recommendations on Machine-Learned Automated Decision Making' (*ACM*, 2018) <www.acm.org/binaries/content/assets/public-policy/ie-euacm-adm-report-2018.pdf> accessed 1 March 2018

Kaltheuner and Bietti
Frederike Kaltheuner and Elettra Bietti, 'Data is power: Towards additional guidance on profiling and automated decision-making in the GDPR' [2017] 2(2) *Journal of Information Rights, Policy and Practice* 1-17

Kamarinou and others
Dimitra Kamarinou and others, 'Machine Learning with Personal Data: Profiling, Decisions and the EU General Data Protection Regulation' (*ML and the law*, 2016) <www.mlandthelaw.org/papers/kamarinou.pdf> accessed 5 March 2018

Kaminski
Margot Kaminski, ‘The right to explanation, explained’ (*SSRN*, 23 July 2018) <papers.ssrn.com/sol3/papers.cfm?abstract_id=3196985> accessed 9 September 2018

Karanasiou and Pinotsis
Argyro Karanasiou and Dimitris Pinotsis, 'A study into the layers of automated decision-making: emergent normative and legal aspects of deep learning' [2017] 31(2) *International Review of Law, Computers & Technology* 170-187

Kemper and Kolkman
Jakko Kemper and Daan Kolkman, ‘Transparent to whom? No algorithmic accountability without a critical audience’ [2018] *Information, Communication & Society* <doi.org/10.1080/1369118X.2018.1477967> accessed 19 October 2018

Kim and Routledge
Tae Wan Kim and Bryan Routledge, 'Informational privacy, a right to explanation, and interpretable AI' (*Research gate*, August 2018) <www.researchgate.net/profile/Tae_Wan_Kim3/publication/327160853_Informational_Privacy_A_Right_to_Explanation_and_Interpretable_AI/links/5b7d7f4092851c1e1227b22e/Informational-Privacy-A-Right-to-Explanation-and-Interpretable-AI.pdf> accessed 11 September 2018

Kuner and others
Christopher Kuner and others, 'Machine Learning with personal data: is data protection law smart enough to meet the challenge?' [2017] 7(1) *International Data Privacy Law* 1-2

De Laat
Paul de Laat, 'Algorithmic Decision-Making Based on Machine Learning from Big Data: Can Transparency Restore Accountability?' (*Springer*, 12 November 2017) <link.springer.com/Article/10.1007/s13347-017-0293-z> accessed 14 March 2018

Lepri and others
Bruno Lepri and others, 'Fair, Transparent, and Accountable Algorithmic Decision-making Processes' (*Springer*, 15 August 2017) <link.springer.com/Article/10.1007/s13347-017-0279-x> accessed 10 March 2018

Mai
Jens-Erik Mai, 'Big data privacy: the datafication of personal information' [2014] 32(3) *The Information Society* 192-199

Malgieri and Comandé
Gianclaudio Malgieri and Giovanni Comandé, 'Why a Right to Legibility of Automated Decision-Making Exists in the General Data Protection Regulation' [2017] 7(4) *International Data Privacy Law* 243-265

Moerel
Lokke Moerel, 'Algorithms can reduce discrimination, but only with proper data' (*IAPP*, 16 November 2018) <iapp.org/news/a/algorithms-can-reduce-discrimination-but-only-with-proper-data/> accessed 30 November 2018

MSI-NET
MSI-NET, 'Study on the human rights dimensions of automated data processing techniques (in particular algorithms) and possible regulatory implications' (*Council of Europe*, 2016) <rm.coe.int/study-on-algorithmes-final-version/1680770cbc> accessed 8 March 2018

Nugent and Cunningham
Conor Nugent and Pádraig Cunningham, 'A case-based explanation system for 'black-box'systems' (*SCSS*, 2010) <www.scss.tcd.ie/publications/tech-reports/reports.04/TCD-CS-2004-20.pdf> accessed 14 July 2018

Oxborough and others
Chris Oxborough and others, 'Explainable AI – Driving business value through greater understanding' (*PwC*, 2018) <www.pwc.co.uk/audit-assurance/assets/pdf/explainable-ai-xai.pdf> accessed 26 October 2018

Ribeiro and others
Marco Tulio Ribeiro and others, 'Why should I trust you? Explaining the predictions of any classifier' (*KDD*, 2016) <www.kdd.org/kdd2016/papers/files/rfp0573-ribeiroA.pdf> accessed 11 July 2018

Roig
Antoni Roig, 'Safeguards for the right not to be subject to a decision based solely on automated processing (Article 22 GDPR)' [2017] 8(3) *European Journal of Law and Technology* 1-17

Selbst and Powles
Andrew Selbst and Julia Powles, 'Meaningful information and the right to explanation' [2017] 7(4) *International Data Privacy Law* 233-242

Sheshasaayee and Bhargavi
Ananthi Sheshasaayee and K Bhargavi, 'A Study of Automated Decision Making Systems' [2017] 7(1) *International Journal of Engineering And Science* 28-31

Tene and Polonetsky
Omer Tene and Jules Polonetsky, 'Taming the Golem: Challenges of Ethical Algorithmic Decision-Making' [2017] 19(1) *North Carolina Journal of Law & Technology* 125-173

Thelisson and others
Eva Thelisson and others, 'Regulatory mechanisms and algorithms towards trust in AI/MI (*Earth link*, 2017) <earthlink.net/~dwaha/research/meetings/ijcai17-xai/9.%20(Thelisson,%20Padh,%20&%20Celis%20XAI-17)%20Regulatory%20Mechanisms%20and%20Algorithms%20towards%20Trust%20in%20AIML.pdf> accessed 11 September 2018

Veale and Edwards
Michael Veale and Lilian Edwards, 'Clarity, Surprises, and Further Questions in the Article 29 Working Party Draft Guidance on Automated Decision-Making' (*SSRN*, 17 November 2017) <papers.ssrn.com/sol3/papers.cfm?abstract_id=3071679> accessed 15 March 2018

Wachter and others
Sandra Wachter and others, 'Why a Right to Explanation of Automated Decision-Making Does Not Exist in the General Data Protection Regulation' [2017] 7(2) *International Data Privay Law* 76-99

Wachter and others
Sandra Wachter and others, 'Counterfactual Explanations Without Opening the Black Box: Automated Decisions and the GDPR' (*SSRN*, 2 November 2017) <papers.ssrn.com/sol3/papers.cfm?abstract_id=3063289> accessed 10 March 2018

Waltl and Vogl
Bernhard Waltl and Roland Vogl, 'Explainable artificial intelligence – the new frontier in legal informatics' (*Towards data science*, 2018) <www.matthes.in.tum.de/file/13tkeaid0rhkz/Sebis-Public-Website/-/Explainable-Artificial-Intelligence-the-New-Frontier-in-Legal-Informatics/Wa18a.pdf> accessed 25 October 2018

Zarsky 2013
Tal Z. Zarsky, 'Transparent predictions' [2013] *University of Illinois Law Review* 1504-1568

Zarsky 2017
Tal Z. Zarsky, 'Incompatible: The GDPR in the Age of Big Data ' [2017] 47(995) *Seton Hall Law Review* 995-1018

News articles

Chin and Wong
Josh Chin and Gillian Wong, 'China's New Tool for Social Control: A Credit Rating for Everything' *Wall Street Journal* (28 November 2016)

The Conversation
The Conversation, 'China's dystopian social credit system is a harbinger of the global age of the algorithm' *The Conversation* (15 January 2018)

Websites

Analytics Vidhya
Analytics Vidhya, 'A Complete Tutorial on Tree Based Modeling from Scratch (in R & Python)' (*Analytics Vidhya*, 2016) <www.analyticsvidhya.com/blog/2016/04/complete-tutorial-tree-based-modeling-scratch-in-python/> accessed 5 June 2018

Best
Jo Best, 'IBM Watson: The inside story of how the Jeopardy-winning supercomputer was born, and what it wants to do next' (*Tech republic*, 2012) <www.techrepublic.com/Article/ibm-watson-the-inside-story-of-how-the-jeopardy-winning-supercomputer-was-born-and-what-it-wants-to-do-next/> accessed 5 June 2018

Chulu
Hendrik Chulu, 'Let us end algorithmic discrimination' (*Medium*, 3 August 2018) <medium.com/techfestival-2018/let-us-end-algorithmic-discrimination-98421b1334a3> accessed 26 October 2018

Gomez
Anna Gomez, 'Deep learning in digital pathology' (*Global-engage*, 2 February 2018) <www.global-engage.com/life-science/deep-learning-in-digital-pathology/> accessed 23 June 2018

ICO
ICO, 'Feedback request – profiling and automated decision-making' (*ICO*, June 2017) <ico.org.uk/media/about-the-ico/consultations/2013894/ico-feedback-request-profiling-and-automated-decision-making.pdf> accessed 5 June 2018

McCann
Adam McCann, 'Auto insurance scores: What they are, rang & more' (*Wallet hub*, 9 September 2017) <wallethub.com/edu/auto-insurance-score/39224/> accessed 18 April 2018

Shaikh
Faizan Shaikh 'Simple Beginner's guide to Reinforcement Learning & its implementation' (*Analytics Vidhya*, 19 January 2017) <www.analyticsvidhya.com/blog/2017/01/introduction-to-reinforcement-learning-implementation/> accessed 23 June 2018

Sirota
Dimitri Sirota, 'Why keep records on data processes if you don't know where your records are?' (*IAPP*, 30 March 2017) <iapp.org/news/a/why-keep-records-on-data-processes-if-you-dont-know-where-your-records-are> accessed 15 March 2018