ML in Law

An Algorithm Must be Seen to be Believed

Right to an Explanation of Automated Decision-Making in the GDPR

May 2019

Name of student: Friðbert Þór Ólafsson

Kennitala: 020494-2399

Supervisor: Alma Tryggvadóttir

Abstract

An Algorithm Must be Seen to be Believed: Right to an Explanation of Automated Decision-Making in the GDPR

Rapid technological developments in the world of artificial intelligence and automated decision-making have brought new challenges for the protection of personal data. Just like their human counterparts, algorithmic decision-makers can be biased and discriminatory which can have serious consequences data subjects. News of automated decision-makers making crucial errors due to these biases are being reported regularly. Errors like systematically classifying some ethnicities as gorillas or charging higher interest rates to minorities compared to non-minorities. To counter these biases and errors, scholars and researchers have stressed the importance of increased transparency and explainability of artificial intelligence. To this end, the General Data Protection Regulation includes a *right to an explanation*, which grants the data subject a right to an explanation for decisions made by artificial intelligent systems. This thesis analyses the scope and the extent of this right. More specifically, it addresses in what situations data controllers are required to provide explanations, as well as requirements to the content, form and timing of the information. Further this thesis discusses how data controllers can comply with this obligation.

In summary, data controllers are required to provide an explanation whenever a decision is based solely on automated processing and produces legal effects concerning the data subject or similarly significantly affects the data subject. The explanation must be provided in a meaningful and understandable way. Therefore, the explanation must include a simple (yet comprehensive) and generic (yet complete) overview of the relevant factors of the underlying system's functionality that are of importance. In conclusion, data controllers shall provide an *ex ante* explanation of the *system functionality*. Lastly, the thesis introduces two methods, subject-centric explanations and counterfactual explanations, which data controllers can utilize to comply with the requirements of the GDPR.

Úrdráttur

Algrím, sjón er sögu ríkari: Réttur til útskýringar á sjálfvirkri ákvörðunartöku í almennu persónuverndarreglugerðinni

Mikil þróun á sviði gervigreindar og sjálfvirkni hefur skapað nýjar áskoranir þegar kemur að persónuvernd. Rétt eins og við mennirnir, þá geta algrímar og gervigreindarforrit einnig verið hlutdræg og óréttlát með alvarlegum afleiðingum fyrir hinn skráða. Fréttir af sjálfvirkum ákvörðunartökum sem hafa farið úrskeiðis vegna hlutdrægni eru nú daglegt brauð. Mistök eins og að flokka fólk af ákveðnum kynþáttum sem górillur eða það að rukka minnihlutahópa um hærri vexti en þá sem tilheyra meirihlutanum. Til að koma í veg fyrir þessa hlutdrægni og mistök hafa fræðimenn ítrekað mikilvægi aukins gagnsæi og útskýranleika gervigreindar. Til að stuðla að slíku, inniheldur almenna persónuverndarreglugerðin (GDPR) svokallaðan *rétt til útskýringar*, sem gefur hinum skráða rétt til útskýringar á sjálfvirkum ákvörðunum teknum af gervigreindarkerfum. Þessi ritgerð greinir umfang réttarins. Nánar tiltekið, þá greinir hún frá því í hvaða tilvikum ábyrgðaraðilar skulu veita umræddar útskýringar, sem og hvaða skilyrði gilda um innihald, form og tímamörk slíkra útskýringa. Þá fjallar ritgerðin einnig um hvernig ábyrgðaraðilar hlíta kröfum reglugerðarinnar.

Í meginatriðum er ábyrgðaraðilum skylt að veita útskýringar þegar ákvörðun byggist eingöngu á grundvelli sjálfvirkrar gagnavinnslu sem hefur réttaráhrif eða sambærileg áhrif á hinn skráða. Útskýringin skal innihalda marktækar upplýsingar og lögð fram á skiljanlegan hátt. Þannig skal útskýringin innihalda einfalt (en samt yfirgripsmikið) og almennt (en samt tæmandi) yfirlit yfir þá undirliggjandi þætti sem eru mikilvægir í ákvörðunartökuferlinu. Skal ábyrgðaraðili því veita fyrirfram útskýringu um virkni hins undirliggjandi kerfis. Að lokum, þá kynnir ritgerð þessi tvær aðferðir til leiks, þ.e. *einstaklingsmiðaðar útskýringar (*e. subject-centric explanations og staðleysu útskýringar (e. counterfactual explanations), sem ábyrgðaraðilar geta nýtt sér til að standast kröfur reglugerðarinnar.

Table of Contents

ABSTRACT ..... **I**

TABLE OF CONTENTS ..... **III**

TABLE OF FIGURES ..... **V**

LIST OF TABLES ..... **VI**

LIST OF PRIMARY SOURCES ..... **VII**

1 INTRODUCTION **1**

2 ALGORITHMIC DECISION MAKING **3**

2.1 ALGORITHMS 3

2.2 ARTIFICIAL INTELLIGENCE 6

2.2.1 Rule-Based Systems *8*

2.2.2 Machine Learning *8*

3 THE GENERAL DATA PROTECTION REGULATION **10**

3.1 PRINCIPLES OF DATA PROTECTION 11

3.1.1 Lawfulness, Fairness and Transparency *12*

3.1.2 Purpose Limitation *14*

3.1.3 Data Minimization *15*

3.1.4 Accuracy *16*

3.1.5 Storage Limitation *17*

3.1.6 Security *19*

3.1.7 Accountability *22*

3.2 DATA PROTECTION AND ARTIFICIAL INTELLIGENCE 22

3.3 RIGHTS OF THE DATA SUBJECT 25

3.3.1 Right to Information (Article 13-14) *25*

3.3.2 Elements of Transparency *25*

3.3.3 Content *27*

3.3.4 Timing *28*

3.3.5 Article 13 Exceptions *30*

3.3.6 Article 14 Exceptions *30*

3.4 THE RIGHT TO ACCESS (ARTICLE 15) 33

3.5 AUTOMATED INDIVIDUAL DECISION MAKING (ARTICLE 22) 34

3.5.1 Solely Automated Processing *35*
3.5.2 Legal Effects or Similarly Significantly Affects *36*
4 THE RIGHT TO AN EXPLANATION 37
4.1 WHY DO INDIVIDUALS WANT A RIGHT TO AN EXPLANATION? 37
4.2 LEGAL BASIS FOR THE RIGHT TO AN EXPLANATION 42
4.2.1 Safeguards of Article 22 and Recital 71 ..... *43*
4.2.2 Right to Information (Article 13-14) *46*
4.2.2.1 Timing and Scope 47
4.2.2.2 Meaningful Information 48
4.2.2.3 The Logic Involved and Envisaged Consequence 50
4.2.3 Right to access (Article 15) *52*
4.3 ARGUMENTS FOR THE EXISTENCE OF AN EX POST RIGHT TO AN EXPLANATION 55
5 POSSIBLE SOLUTIONS 57
5.1 SUBJECT-CENTRIC EXPLANATIONS 58
5.2 COUNTERFACTUALS 60
5.3 DO THE PROPOSED SOLUTIONS MEET THE REQUIREMENTS OF THE GDPR? 62
5.3.1 Meaningful Information About the Logic Involved *62*
5.3.2 Envisaged Consequences *62*
5.3.3 Easily Understandable *63*
5.4 ETHICAL REQUIREMENTS TO EXPLAINABILITY 64
6 CONCLUSION 65
BIBLIOGRAPHY 67

Table of Figures

FIGURE 1 3
FIGURE 2 6
FIGURE 3 21
FIGURE 4 59

List of Tables

TABLE 1 17
TABLE 2 22
TABLE 3 28
TABLE 4 33
TABLE 5 41
TABLE 6 42
TABLE 7 54

List of Primary Sources

C-131/12 Google Spain SL and Google Inc. v Agencia Española de Protección de Datos (AEPD) and Mario Costeja González [2014] EU:C:2014:317

Charter of Fundamental Rights of the European Union [2012] OJ C 326/391 (The Charter)

Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data [1995] OJ L 281/31

Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC, OJ 2016 L 119/1 (GDPR)

SCHUFA, BGH, VI ZR 156/13, 28 January 2014

Treaty on the Functioning of the European Union [2012] OJ C 326/47

1 Introduction

For generations, scientists, engineers and philosophers have discussed and theorized the possibility of intelligent machines. In the 1950's paper: Computing Machinery and Intelligence[1] the mathematician and computer scientist Alan Turing considered the question '*can machines think?*' and introduced the Turing Test, a method which attempted to define a machine's intelligent.[2] This has often been considered the first steps in theorizing and developing what is now called *artificial intelligent*. The term *artificial intelligence* itself however wasn't introduced until in the year 1956 by John McCarthy in his proposal for the Dartmouth Summer Research Project on Artificial Intelligence.[3] In the study an attempt was made 'to find how to make machines use language, form abstractions and concepts, solve kinds of problems now reserved for humans, and improve themselves.'[4]

Even though the idea of artificial intelligence is more than sixty years old, the focus and goals of artificial intelligence research and development have changed drastically in the same amount of time.[5] Due to the rapid technological advances in the last sixty years[6], artificial intelligence systems are getting more powerful. This is mainly due to increased availability of data and increased computing power.[7] Further, corporations and governments are heavily incentivized to advance further in the field of artificial intelligence as increased automation generally provides more efficiency and lower costs for these parties.[8] Therefore, it is expected that there will be a continuing increase in utilization of artificial intelligence in most sectors of communities worldwide.[9] Consequently, this increased automation will lead to a decrease in

---

[1] A.M. Turing, 'Computing Machinery and Intelligence' (1950) 49 Mind 433.
[2] A.M. Turing, 'Computing Machinery and Intelligence' (1950) 49 Mind 448.
[3] J. McCarthy and others 'A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence' (31 August 1955) <http://jmc.stanford.edu/articles/dartmouth/dartmouth.pdf> accessed 1 May 2019.
[4] J. McCarthy and others 'A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence' (31 August 1955) <http://jmc.stanford.edu/articles/dartmouth/dartmouth.pdf> accessed 1 May 2019.
[5] Deyi Li and Yi Du, *Artificial Intelligence With Uncertainty* (2nd edn, CRC Press, Taylor & Francis Group 2017) 8.
[6] More specifically, advances in the hardware and other supplemental technologies required to power artificial intelligence systems in accordance with Moore's law; *Moore* (n 7).
[7] Gordon E. Moore, 'Cramming more components onto integrated circuits' (1965) 38 Electronics; Dean Takahasi, 'Forty years of Moore's Law' *(The Seattle Times,* 18 April 2005) <https://www.seattletimes.com/business/forty-years-of-moores-law/> accessed 11 May 2019.
[8] Infosys, 'Amplifying Human Potential – Toward purposeful Artificial Intelligence: A Perspective for CIOs' (*Infosys,* 2017) <https://www.infosys.com/aimaturity/Documents/amplifying-human-potential-CIO-report.pdf> accessed 11 May 2019; Infosys, 'Amplifying Human Potential – Toward purposeful Artificial Intelligence' (*Infosys,* 2017) <https://www.infosys.com/aimaturity/Documents/amplifying-human-potential-CEO-report.pdf> accessed 11 May 2019.
[9] Kasey Panetta, '5 Trends Emerge in the Gartner Hype Cycle for Emerging Technologies, 2018' (*Smarter with Gartner*, 16 August, 2018) <https://www.gartner.com/smarterwithgartner/5-trends-emerge-in-gartner-hype-cycle-for-emerging-technologies-2018/> accessed 8 December 2018.

human involvement in decision-makings. Then again, this increased automation also has a dark side,[10] due to the nature of artificial intelligence systems and some of their inherent technical features.

Examples of these issues of artificial intelligence systems are that they can (a) be very opaque and difficult to explain; (b) involve methods which neither the developer nor the data subject considered at the time of collection;[11] (c) incentivize the collection of vast amount of data; and (d) incentivize the storage of that data for long times,[12] resulting in the developers and users of such systems to collect and use as much data as they can get their hands on. These characteristics and incentives (a) - (d) are in many ways incompatible with the fundamental principles of the General Data Protection Regulation ("**GDPR"**)[13] which stipulates how personal data can be processed in a legitimate way.[14]

In this thesis I will mainly focus on the issue of opaqueness of artificial intelligent systems and the GDPR's *right to explanation* which, in certain situations, grants the data subject a right to an explanation of automated decisions, including decisions made by artificial intelligent system. The scope of the GDPR's *right to explanation* is the subject of much debate and some scholars even argue that this right does not exist in the final and implemented version of the GDPR.[15] Therefore, it is necessary to analyze the scope and the extent of this right. More specifically I will analyze in what situations data controllers are required to provide explanations, as well as requirements to the content, form and timing of the information. This exercise will be done in Chapter 4 of this thesis. Lastly, I will assess how data controllers can comply with this obligation, which can be a difficult task due to the opaqueness and complexity of artificial intelligent systems. This will be performed in Chapter 5 of this thesis by introducing and analyzing possible solutions found in scholarly works in the field.

---

[10] Frida Polli, 'The Dark Side Of Artificial Intelligence' (*Forbes,* 5 December 2017) <https://www.forbes.com/sites/fridapolli/2017/12/05/the-dark-side-of-artificial-intelligence/#3fe68e4a1261> accessed 11 May 2019.

[11] Tal Z. Zarsky, 'Incompatible: The GDPR in the Age of Big Data' (2017) 47 Seton Hall Law Review 995, 1006.

[12] Tal Z. Zarsky, 'Incompatible: The GDPR in the Age of Big Data' (2017) 47 Seton Hall Law Review 995, 1011.

[13] Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC, OJ 2016 L 119/1 (GDPR).

[14] The principles of the GDPR will be discussed in Chapter 3.1.

[15] Sandra Wachter, Brent Mittelstadt and Luciano Floridi, 'Why a Right to Explanation of Automated Decision-Making Does Not Exist in the General Data Protection Regulation' (2017) 7 International Data Privacy Law 76, 78.

2 Algorithmic Decision Making

Before I explore the data protection issues of artificial intelligent systems it is essential to differentiate between the terms regularly used in the discussion of these issues, that is the difference between a) *algorithms*, b) *artificial intelligence*, and c) *machine learning*. The simplest way to differentiate between the terms and define them is to realize that algorithms are a broad umbrella term which the latter terms are all subset of. In fact, all the terms are also subset of each other as is clear from the following Figure 1.

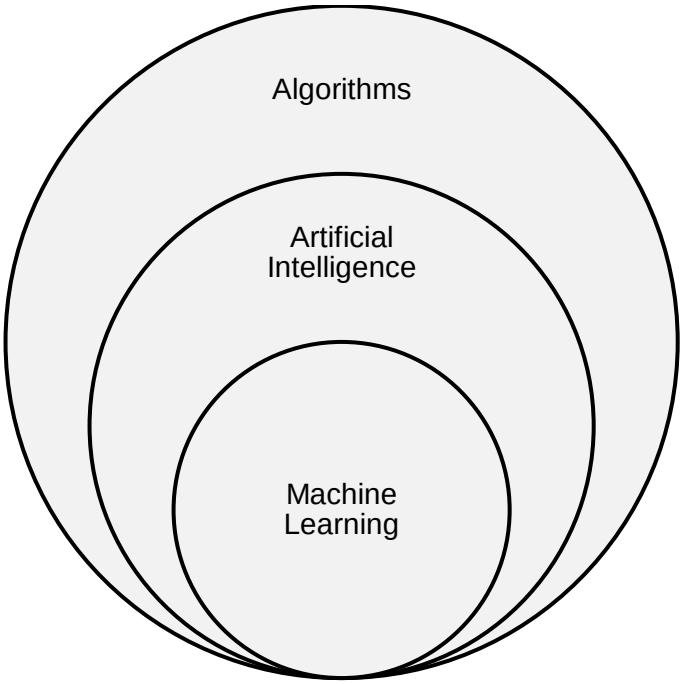

FIGURE 1

As seen, *machine learning* is a type of *artificial intelligence*, which is a type of *algorithm*. This fact is essential when understanding the concept and issues of artificial intelligence. In the following sections each of the terms will be discussed, defined and their characteristics explored.

2.1 Algorithms

In the broadest and most simple sense an algorithm is *a set of instructions for accomplishing a goal.*[16] In other words, an algorithm can be viewed as a recipe to solve a specific problem.

[16] 'Meaning of algorithm in English' (*Cambridge Dictionary*) <https://dictionary.cambridge.org/us/dictionary/english/algorithm> accessed 11 May 2019, 'a list of instructions

However, this definition is too broad for our purposes as this would mean that every instructions or guidelines, including cooking recipe could be viewed as an algorithm, which is clearly not the case.

The American computer scientist and mathematician Donald Knuth has acknowledged these similarities between algorithms and other common methods of providing instructions or guidelines and pinpointed what important features of algorithms differentiate them from other set of instructions:

> The modern meaning for algorithm is quite similar to that of recipe, process, method, technique, procedure, routine, rigmarole, except that the word 'algorithm' connotes something just a little different. Besides merely being a finite set of rules that gives a sequence of operations for solving a specific type of problem, an algorithm has five important features …[17]

These five important features are a) *finiteness*, b) *definiteness*, c) *input*, d) *output* and e) *effectiveness*.[18] *Firstly*, an algorithm must be finite. That is, an algorithm must always terminate after a finite number of steps. For practical use, this not only be finite number 'but a *very* finite number, a reasonable number.'[19] *Secondly*, 'each step of an algorithm must be precisely defined; the actions to be carried out must be rigorously and unambiguously specified for each case.'[20] *Thirdly*, an algorithm must have zero or more inputs. More specifically an algorithm must be fed some set of information (data) before it initiates, or dynamically while it runs.[21] It uses this input to logically reach its goal. *Fourthly*, an algorithm always has at least one output or 'quantities that have a specified relation to the input.'[22] That is, an algorithm must always come up with an answer to the problem. This is an algorithm's main purpose: to solve a problem or to accomplish a goal. The output usually varies, depending on the input given to the algorithm. *Fifthly*, generally it is expected that an algorithm is effective, that is, each step of the algorithm must be basic enough that they could be executed in exactly the same way by

---

for solving a problem'; 'Meaning of algorithm in English' (*Cambridge Dictionary*) <https://dictionary.cambridge.org/us/dictionary/english/algorithm> accessed 11 May 2019,
'a set of mathematical instructions or rules that, especially if given to a computer, will help to calculate an answer to a problem'; 'Definition of *algorithm*' (*Merriam-Webster Dictionary*) <https://www.merriam-webster.com/dictionary/algorithm> accessed 11 May 2019, 'a step-by-step procedure for solving a problem or accomplishing some end'.

[17] Donald Knuth, *The Art of Computer Programming: fundamental algorithms* (3rd edn. 1997) 4.
[18] Donald Knuth, *The Art of Computer Programming: fundamental algorithms* (3rd edn. 1997) 4 – 6.
[19] Donald Knuth, *The Art of Computer Programming: fundamental algorithms* (3rd edn. 1997) 5.
[20] Donald Knuth, *The Art of Computer Programming: fundamental algorithms* (3rd edn. 1997) 5.
[21] Donald Knuth, *The Art of Computer Programming: fundamental algorithms* (3rd edn. 1997) 5.
[22] Donald Knuth, *The Art of Computer Programming: fundamental algorithms* (3rd edn. 1997) 5.

someone using a pen and paper.[23] With these five features in mind, it is clear what differentiates an algorithm from a cooking recipe:

> A recipe presumably has the qualities of finiteness (although it is said that a watched pot never boils), input (eggs, flour, etc.) and output (TV dinner, etc.), but it notoriously lacks definiteness. There are frequent cases in which a cook´s instructions are indefinite: 'Add a dash of salt.' A 'dash' is defined to be 'less than 1/8 teaspoon,' and salt is perhaps well enough defined; but where should the salt be added – on top? on the side? Instructions like 'toss lightly until mixture is crumbly' or 'warm cognac in a small saucepan' are quite adequate as explanations to a trained chef, but *an algorithm must be specified to such a degree that even a computer can follow the directions*.[24]

The fact that the algorithm must be specific enough for a computer to follow the given instructions leads us to our last consideration. Clearly, in the context of computer science, computer programs and artificial intelligence systems, the definition provided hereabove is not sufficient. A more meaningful and narrower definition is required, a definition which takes Knut's features and conclusions into account.

The Portuguese computer scientist Pedro Domingos has offered a simple definition that is perfect for the purposes of this thesis, 'an algorithm is a sequence of instructions telling a computer what to do.'[25] In this way Domingos has limited the scope of algorithms to set of instructions understandable and executable by a computer.[26] This is similar to the requirement provided by Knuth, that an algorithm must be specific enough so that a computer can follow the directions. This is perhaps the fundamental difference between a simple cooking recipe and a computer algorithm. A cooking recipe could only be viewed as an algorithm if the recipe is so specific and structured in such a way that a computer could follow the recipe step by step and prepare a fully cooked meal.

The example given here above, of an algorithm that can instruct a machine to prepare a fully cooked meal (given that the machine has the required hardware and has the necessary ingredients at its disposal) seems like an intelligence machine. But is such a machine artificially

---

[23] Donald Knuth, *The Art of Computer Programming: fundamental algorithms* (3rd edn. 1997) 6.
[24] Donald Knuth, *The Art of Computer Programming: fundamental algorithms* (3rd edn. 1997) 6. (emphasis added)
[25] Pedro Domingos, *The Master Algorithm: How the Quest for the Ultimate Learning Machine Will Remake Our Worlds* (Basic Books 2015) 5.
[26] Jean-Luc Chabert and others, *A history of Algorithms: From the Pebble to the Microchip* (Springer 1999) 2, defines the algorithm as being 'any process that can be carried out automatically' which adds to the previous definition the requirement of automation.

intelligent? In the following Chapter 2.2 the concept of artificial intelligence will be discussed, and the term defined.

2.2 Artificial Intelligence

Artificial intelligence has been the subject of many research fields which 'currently encompasses a huge variety of subfields'[27] and is in fact a universal field. When striving to define the concept of *artificial intelligence* the professional literature refers to four different approaches to artificial intelligence, historically followed scholars. The four groups are artificial intelligence being a system that (a) thinks like a human; (b) thinks rationally; (c) acts like a human; and (d) acts rationally as seen in the following Figure 2[28]:

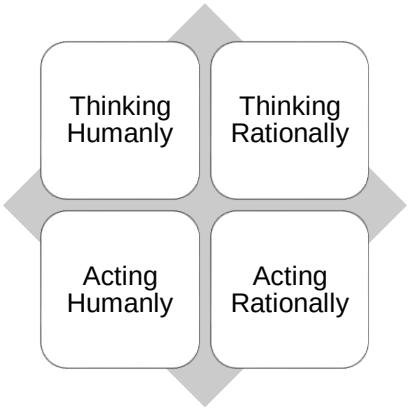

FIGURE 2

The four different approaches can be split into two categories on how an artificial intelligence system is viewed. The first category focuses on how the system works (or thinks), the two approaches in the upper row of Figure 2. The second category however focuses on how the system acts, that is the approaches in the lower row of Figure 2. For the purpose of this thesis, in my analysis of the *right to an explanation,* I will mainly focus on the functionality of artificial systems in decision making, that is the rationality of the system. Therefore, I will approach artificial intelligence being a system acting rationally. A rational system is a system 'that acts so as to archive the best outcome or, when there is uncertainty, the best expected outcome'[29] This is in line with the idea of the Turing Test, but all of the skills needed for the Turing Test also allow a system to act rationally.[30] In fact, from Turing's perspective, artificial

[27] Stuart Russel and Peter Norvig, *Artificial Intelligence: A Modern Approach* (3rd edn, Prentice Hall 2010) 1.
[28] Stuart Russel and Peter Norvig, *Artificial Intelligence: A Modern Approach* (3rd edn, Prentice Hall 2010) 1.
[29] Stuart Russel and Peter Norvig, *Artificial Intelligence: A Modern Approach* (3rd edn, Prentice Hall 2010) 4.
[30] Stuart Russel and Peter Norvig, *Artificial Intelligence: A Modern Approach* (3rd edn, Prentice Hall 2010) 4.

intelligence, was the idea of machines or computers acting rationally, a skill previously only attributable to humans. [31]

Having considered this approach, a definition of the concept is required that explains artificial intelligence and its functionality in a clear way. The High-Level Expert Group on Artificial Intelligence (**"AI HLEG"**), an expert group tasked by the European Commission to implement the European Union's strategy on AI[32] has proposed the following definition of artificial intelligence:

> Artificial intelligence (AI) refers to systems designed by humans that, given a complex goal, act in the physical or digital world by perceiving their environment, interpreting the collected structured or unstructured data, reasoning on the knowledge derived from this data and deciding the best action(s) to take (according to pre-defined parameters) to achieve the given goal. *AI systems can also be designed to learn to adapt their behaviour by analysing how the environment is affected by their previous actions.*[33]

AI HLEG's definition, does not only describe the goal of an artificial system but also its functions.[34] By this definition, artificial intelligence has a lot of the same important features as algorithms as identified by Donald Knuth[35] and referred to earlier. By this definition artificial intelligences systems are designed to achieve complex goals by using data (input) to find the best way to achieve the given goal (output). Lastly, the definition includes the possibility of machine learning systems by including the fact that an artificial intelligent system could be designed in a way that it can adapt and learn from its previous actions.[36] In the following chapter these self-learning capabilities of artificial intelligence, which are often called, machine learning, will be explored. But first, rule-based systems will be introduced.

---

[31] A.M. Turing, 'Computing Machinery and Intelligence' (1950) 49 Mind 448.
[32] European Commission 'Call for a High-Level Expert Group on Artificial Intelligence' (*European Commission,* 9 March 2018) <https://ec.europa.eu/digital-single-market/en/news/call-high-level-expert-group-artificial-intelligence> accessed 11 May 2019.
[33] Independent High-Level Expert Group on Artificial Intelligence - Set Up By The European Commission, 'A Definition of AI: Main Capabilities And Disciplines' (European Commission 2018) (emphasis added).
[34] Andreas Kaplan and Michael Haenlein, 'Siri, Siri, in my hand: Who's the fairest in the land? On the interpretations, illustrations, and implications of artificial intelligence' 62 (2019) Business Horizons 15, defines artificial intelligence being a 'system's ability to interpret external data correctly, to learn from such data, and to use those learnings to achieve specific goals and tasks through flexible adaptation.'
[35] Donald Knuth, *The Art of Computer Programming: fundamental algorithms* (3rd edn. 1997) 4 – 6
[36] Emphasized in AI HLEG's definition above.

2.2.1 Rule-Based Systems

The simplest artificial intelligence systems are *rule-based* which means that their algorithm is unchanging, and its logic coded by its creator. Rule based systems can be described as follows:

> A rule-based system (e.g., production system, expert system) uses rules as the knowledge representation. These rules are coded into the system in the form of if-then-else statements. The main idea of a rule-based system is to capture the knowledge of a human expert in a specialized domain and embody it within a computer system. That's it. No more, no less. Hence, knowledge is encoded as rules.[37]

As the logic or rules are *hardcoded* into the system, they are very inflexible and its tough, and sometimes impossible, to 'add rules to an already large knowledge base without introducing contradicting rules.' Here it is worth pointing that, in order to be considered artificial intelligence, within the definition given in the foregoing chapter, the *rule-based system* needs to be able to achieve a complex goal that would be called intelligent if a human would have achieved it. In other words, it needs to be able to *act rationally*.

2.2.2 Machine Learning

In contrast to rule-based systems, machine learning systems have a very ambitious goal.[38] Machine learning can be described as the 'science of getting computers to act without being explicitly programmed.'[39] Its intention is to enable the system to 'learn' on its own by supplying the system with data to make predictions: [40]

> The ability to learn causes adaptive intelligence, and adaptive intelligence means that existing knowledge can be changed or discarded, and new knowledge can be acquired. Hence, these systems build the rules on the fly. That is what makes learning systems so different from rule-based testing.'[41]

---

[37] Tricentis, 'AI Approaches Compared: Rule-Based Testing vs. Learning' (*Tricentis*, undated) <https://www.tricentis.com/artificial-intelligence-software-testing/ai-approaches-rule-based-testing-vs-learning/> accessed 11 May 2019.

[38] Tricentis, 'AI Approaches Compared: Rule-Based Testing vs. Learning' (*Tricentis*, undated) <https://www.tricentis.com/artificial-intelligence-software-testing/ai-approaches-rule-based-testing-vs-learning/> accessed 11 May 2019.

[39] Stanford University, 'Machine Learning' (*Coursera*, undated) <https://www.coursera.org/learn/machine-learning/home/info> accessed 11 May 2019, 'machine learning is the science of getting computers to act without being explicitly programmed').

[40] Tricentis, 'AI Approaches Compared: Rule-Based Testing vs. Learning' (*Tricentis*, undated) <https://www.tricentis.com/artificial-intelligence-software-testing/ai-approaches-rule-based-testing-vs-learning/> accessed 11 May 2019.

[41] Tricentis, 'AI Approaches Compared: Rule-Based Testing vs. Learning' (*Tricentis*, undated) <https://www.tricentis.com/artificial-intelligence-software-testing/ai-approaches-rule-based-testing-vs-learning/> accessed 11 May 2019.

Machine learning can be split into two categories: a) supervised learning, and b) unsupervised learning. In supervised learning, algorithms are built on labelled datasets. In this sense, the algorithms have been guided or trained 'how to map from input to output by the provision of data with "correct" values already assigned to them.' After this training phase the algorithms have generated models which they can use for predictions in later phases.[42] On the other hand, unsupervised leaning is 'when the algorithms are not trained and are instead left to find regularities in input data without any instructions as to what to look for.'[43]

Neural networks, one of the most popular and effective forms of machine learning systems[44] are an example of machine learning that are prone to very fast learning. Neural networks are essentially developed to replicate the activity of the human brain whereas different *nodes* connect with each other in a network in the similar way as the activity in networks of brain cells called neurons.[45] 'Due to this ability of interconnection and due to the fact that it mirrors human brain, neural network has an ability to learn while processing data.'[46]

Lastly, it is worth noting that the self-learning capacity of machine learning algorithms makes it so that the developer often has little or no control over *how* the system achieves the given goal. All the system *cares* about is programmed to do is find the best and most efficient way to achieve the given goal, which in some cases is not the way the developer had foreseen and can often be a complete mystery to the developer.[47] This can be especially problematic when the system is used to process personal data which has to be processed in a lawful, fair and transparent manner in accordance with the GDPR. In the following Chapter 3 I will discuss the GDPR, its purpose and relevant provisions in more detail.

---

[42] Information Commissioner's Office 'Big data, artificial intelligence, machine learning and data protection' (*Information Commissioner's Office,* 2017) 7 <https://ico.org.uk/media/for-organisations/documents/2013559/big-data-ai-ml-and-data-protection.pdf> accessed 11 May 2019.
[43] Information Commissioner's Office 'Big data, artificial intelligence, machine learning and data protection' (*Information Commissioner's Office,* 2017) 7 <https://ico.org.uk/media/for-organisations/documents/2013559/big-data-ai-ml-and-data-protection.pdf> accessed 11 May 2019.
[44] Stuart Russel and Peter Norvig, *Artificial Intelligence: A Modern Approach* (3rd edn, Prentice Hall 2010) 728.
[45] Stuart Russel and Peter Norvig, *Artificial Intelligence: A Modern Approach* (3rd edn, Prentice Hall 2010) 727.
[46] Dr. Maja Brkan, 'Do Algorithms Rule the World? Algorithmic Decision-Making In The Framework of the GDPR and Beyond' ('Terminator or the Jetsons? The Economics and Policy Implications of Artificial Intelligence' conference, Washington February 2018) 21.
[47] Andrew Griffin, 'Facebook's Artificial Intelligence Robots Shut Down After They Start Talking To Each Other In Their Own Language' (*Independent*, 31 July 2017) <https://www.independent.co.uk/life-style/gadgets-and-tech/news/facebook-artificial-intelligence-ai-chatbot-new-language-research-openai-google-a7869706.html> accessed 11 May 2019.

3 The General Data Protection Regulation

The protection of persons personal data is a fundamental right. According to Article 8(1) of the Charter of Fundamental Rights of the European Union[48] and Article 16(1) of the Treaty on the Functioning of the European Union[49] everyone has the right to the protection of personal data concerning him or her.[50] The principles and rules of the GDPR are intended to protect these fundamental rights[51] and provide a framework where the processing of personal data is designed to serve mankind.[52] This however, is not an absolute right and must be 'balanced against other fundamental rights, in accordance with the principle of proportionality.'[53]

Additionally, '[r]apid technological developments and globalisation have brought new challenges for the protection of personal data.'[54] The scale of processing of personal data has increased significantly. These technological developments allow both private companies and public authorities to make use of personal data on an unprecedented scale in their day to day activities. Individuals increasingly make personal data available publicly and globally. Further, these technological developments have transformed both the economy and social life. The GDPR acknowledges this fact and aims to 'further facilitate the free flow of personal data within the Union and the transfer to third countries and international organizations, while ensuring a high level of the protection of personal data.'[55] 'In order to ensure a consistent and high level of protection of natural persons and to remove the obstacles to flows of personal data within the Union',[56] the GDPR seeks to harmonize the level of data protection within each of its member states.

[48] Charter of Fundamental Rights of the European Union [2012] OJ C 326/391 (The Charter).
[49] Treaty on the Functioning of the European Union [2012] OJ C 326/47 (TFEU).
[50] GDPR, recital 1.
[51] GDPR, recital 2; GDPR, art 1.
[52] GDPR, recital 4.
[53] GDPR, recital 4.
[54] GDPR, recital 6.
[55] GDPR, recital 6.
[56] GDPR, recital 10

In order to achieve its goals, the GDPR regulates the processing of personal data. Its scope is limited to situations where a *controller*[57] *processes*[58] the *personal data*[59] of *data subjects*[60] who are in the European Union ('**EU**' or the '**Union'**) where the processing activities are related to one of the following:

a) 'the offering of goods or services, irrespective of whether a payment of the data subject is required, to such data subjects in the Union; or'[61]
b) 'the monitoring of their behavior as far as their behaviour takes place within the Union.'[62]

In addition, the GDPR also applies to the processing of personal data by data controllers or processors established in the EU whether the processing takes place in the EU or not.[63] Lastly, the GDPR also applies 'to the processing of personal data by a controller not established in the Union, but in a place where Member State law applies by virtue of public international law.'[64]

In the following chapters I will explore some of the rights of the data subject regarding information disclosure and transparency. But first, I will explore the seven principles of data protection found in the GDPR and how they interplay with artificial intelligence systems.

3.1 Principles of Data Protection

The GDPR sets out seven fundamental principles which lie at the heart of the European Data Protection framework. For processing of personal data to be lawful it must follow these principles. They don't give hard and fast rules, but rather embody the spirit of the general data

---

[57] GDPR, Article 4(7), ' "controller" means the natural or legal person, public authority, agency or other body which, alone or jointly with others, determines the purposes and means of the processing of personal data; where the purposes and means of such processing are determined by Union or Member State law, the controller or the specific criteria for its nomination may be provided for by Union or Member State law'.
[58] GDPR, Article 4(2), ' "Processing" means any operation or set of operations which is performed on personal data or on sets of personal data, whether or not by automated means, such as collection, recording, organisation, structuring, storage, adaptation or alteration, retrieval, consultation, use, disclosure by transmission, dissemination or otherwise making available, alignment or combination, restriction, erasure or destruction'.
[59] Article 4(1), ' "Personal data" means any information relating to an identified or identifiable natural person ('data subject'); an identifiable natural person is one who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, an online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person'.
[60] Information Commissioner's Office, 'Guide to the General Data Protection Regulation (GDPR): The principles' (*Information Commissioner's Office,* December 2018) <https://ico.org.uk/for-organisations/guide-to-the-general-data-protection-regulation-gdpr/principles/> accessed 11 May 2019.
[61] GDPR, art 3(2)(a).
[62] GDPR, art 3(2)(b).
[63] GDPR, art 1.
[64] GDPR, art 3(3).

protection regime – and there are only very limited exceptions from the principles. Compliance with these fundamental principles is considered an essential building block for good data protection practice.[65] These principles are found in Article 5 of the GDPR and are as follows:

a) The principle of lawfulness, fairness and transparency; [66]
b) The principle of purpose limitation;[67]
c) The principle of data minimization;[68]
d) The principle of accuracy;[69]
e) The principle of integrity and confidentiality;[70] and
f) The principle of accountability;[71]

In the following Chapters 3.1.1 - 3.1.7 these principles will be discussed. Thereafter, I will highlight how the principles affect data controllers who utilize artificial intelligence systems.

3.1.1 Lawfulness, Fairness and Transparency

According to the *lawfulness, fairness and transparency principle* any processing of personal data should be both lawful and fair. Additionally, it should always be clear and transparent to the data subject that their personal data is being processed and to what extent their data is being or will be processed. Furthermore, the it is required that any information and communication relating to the processing is both easily accessible and easily understandable. Data controllers must provide the information on the identity of the data controller, the purposes of the processing and any further information required to ensure fair and transparent processing in respect of the data subject and his right to obtain confirmation and communication of data concerning them which is being processed. Consequently, the data subjects should be made aware of any risks, rules, safeguards and rights in relation to the processing of their data and how they can exercise their rights in relation to the processing. [72]

Lawfulness

[65] Information Commissioner's Office, 'Guide to the General Data Protection Regulation (GDPR): The principles' (*Information Commissioner's Office,* December 2018) <https://ico.org.uk/for-organisations/guide-to-the-general-data-protection-regulation-gdpr/principles/> accessed 11 May 2019.
[66] GDPR, art 5(1)(a).
[67] GDPR, art 5(1)(b).
[68] GDPR, art 5(1)(c).
[69] GDPR, art 5(1)(d).
[70] GDPR, art 5(1)(f).
[71] GDPR, art 5(2).
[72] GDPR, recital 39.

For processing to be lawful, data controllers are required to identify and specify the grounds of the processing. Article 6 of the GDPR provides six different grounds of lawful processing. Additionally, if the data belongs to one of the special categories of data,[73] some additional conditions are required for the processing to be lawful.[74] The principle also provides that data controllers cannot process unlawful data in a more general sense. Therefore, for an example, the processing may be unlawful and in breach of this principle if it results in one of the following 'an infringement of copyright; a breach of an enforceable contractual agreement; a breach of industry-specific legislation or regulations.'[75]

Fairness

Processing must also be fair. Meaning that data controllers should handle personal data in the way expected by the data subject. Additionally, the data should not be processed in a way that has unjustified adverse effects on the data subject. How the data controller obtained the information is of upmost importance when assessing fairness. If the data controller deceived or mislead the data subject when obtaining the data, the processing is most likely unfair. [76]

Transparency

Lastly, processing must be transparent. This principle is closely linked with the idea of fairness. The data controller must be clear, open and honest about who he is and how and why he processes the personal data. Transparency is especially important in situations where the data subject has a choice to enter into an agreement or relationship with the data controller. The transparency principle provides that the data subject is informed about the scale and purpose of the processing and enables the data subject to a take an informed decision about whether to

[73] Article 9 of the GDPR identifies the following as special categories of data: personal data revealing racial or ethnic origin, political opinions, religious or philosophical beliefs, or trade union membership, and the processing of genetic data, biometric data for the purpose of uniquely identifying a natural person, data concerning health or data concerning a natural person's sex life or sexual orientation.

[74] GDPR, art 8.

[75] Information Commissioner's Office, 'Guide to the General Data Protection Regulation (GDPR): Principle (a): Lawfulness, fairness and transparency' (*Information Commissioner's Office,* December 2018) <https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/principles/lawfulness-fairness-and-transparency/> accessed 11 May 2019.

[76] Information Commissioner's Office, 'Guide to the General Data Protection Regulation (GDPR): Principle (a): Lawfulness, fairness and transparency' <https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/principles/lawfulness-fairness-and-transparency/> (*Information Commissioner's Office,* December 2018) accessed 11 May 2019.

sign the agreement or enter into the relationship. Further, the data subject could use this knowledge to renegotiate the terms of the agreement or relationship. [77]

Transparency is also important when the data controller collects the data from a third party and has no direct relationship with the data subject. One could even argue that in such cases it is more important to be transparent about the processing as the individual has no knowledge that the data controller has collected or processed their personal. Furthermore, not having this knowledge greatly affects the data subjects' ability to exercise their rights granted by the GDPR.[78]

3.1.2 Purpose Limitation

In accordance with the *purpose limitation principle,* the specific purposes for which personal data is processed should be explicit and legitimate and determined beforehand, that is, at the time of collection of data. The data should therefore be adequate, relevant and limited to only what is necessary for the predetermined purposes of processing.[79] The principle aims to ensure that the data controller is clear and honest about the purpose of collection and processing and what he does with the data is in line with the reasonable expectations of the data subject.[80] Data controllers should specify the purpose in a way easily accessible for the data subject in accordance with Articles 13(1)(c) and 14(1)(c).[81]

The principle does not completely prohibit processing for other purposes than the data was originally collected for. Data controllers can process the data for new purposes which were not originally anticipated, only if:

a) the new purpose is compatible with the original purpose;
b) the data subject has given its consent for the new purpose; or

---

[77] Information Commissioner's Office, 'Guide to the General Data Protection Regulation (GDPR): Principle (a): Lawfulness, fairness and transparency' (*Information Commissioner's Office,* December 2018) <https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/principles/lawfulness-fairness-and-transparency/> accessed 11 May 2019.
[78] Information Commissioner's Office, 'Guide to the General Data Protection Regulation (GDPR): Principle (a): Lawfulness, fairness and transparency' (*Information Commissioner's Office,* December 2018) <https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/principles/lawfulness-fairness-and-transparency/> accessed 11 May 2019.
[79] GDPR, recital 39.
[80] Information Commissioner's Office, 'Guide to the General Data Protection Regulation (GDPR): Principle (b): Purpose limitation' (*Information Commissioner's Office,* December 2018) <https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/principles/purpose-limitation/> accessed 11 May 2019.
[81] See Chapter 3.3.1.

c) there is a clear 'legal provision allowing the new processing in the public interest – for example, a new function for a public authority.'[82]

When assessing whether a purpose of further processing is compatible with the purpose for which the personal data are originally collected, the controller, should take into account a range of factors,[83] including the following, *inter alia:*

- Is there a link of any kind between the original purposes and the purposes of the intended further processing?
- What is the context in which the personal data have been collected, in particular the reasonable expectations of data subjects based on their relationship with the controller as to their further use?
- What is the nature of the personal data?
- What are the consequences of the intended further processing for data subjects?
- Are there appropriate safeguards in place for both the original and intended further processing operations? [84]

To sum up, generally, 'if the new purpose is either very different from the original purpose, would be unexpected, or would have an unjustified impact on the individual, it is likely to be incompatible with [the data controller's] original purpose.'[85]

3.1.3 Data Minimization

The data minimization principle requires the data controller to ensure that personal data is *adequate*, *relevant* and *limited* to what is necessary in relation to the purposes for which they are processed.[86] Personal data should only be processed if the data controller could not reasonably fulfill its purpose by any other means.[87] The principle is closely linked to the data subject´s right to rectification which allows the data subject to have incomplete personal data

---

[82] Information Commissioner's Office, 'Guide to the General Data Protection Regulation (GDPR): Principle (b): Purpose limitation' (*Information Commissioner's Office,* December 2018) <https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/principles/purpose-limitation/> accessed 11 May 2019.

[83] Article 29 Data Protection Working Party, 'Opinion 03/2013 on purpose limitation' (WP 203, 2 April 2013) 23.

[84] GDPR, recital 50.

[85] Information Commissioner's Office, 'Guide to the General Data Protection Regulation (GDPR): Principle (b): Purpose limitation' (*Information Commissioner's Office,* December 2018) <https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/principles/purpose-limitation/> accessed 11 May 2019.

[86] GDPR, art 5(1)(c).

[87] GDPR, recital 39.

completed and inaccurate personal data rectified.[88] It is also closely linked to and best enshrined in the data subject's right to erasure (right to be forgotten)[89] which allows the data subject to obtain erasure of its personal data and the data controller is obligated to erase the personal data in the scenarios listed in points (a) – (f) of Article 17(1), including when the 'personal data are no longer necessary in relation to the purposes for which they were collected or otherwise processed.'[90]

Data controllers should not collect, store or otherwise process more data than is necessary for to achieve the purpose of the processing. Nor should the data include any irrelevant details. Therefore, data controllers are prohibited from collecting personal data on the 'off-chance that it might be useful in the future. However, [data controllers] may be able to hold information for a foreseeable event that may never occur if [they] can justify it.'[91]

3.1.4 Accuracy

In accordance with the accuracy principle, a data controller must take every reasonable step to ensure that inaccurate personal data is rectified or even deleted. [92] This principle is enshrined in the data subject's right to rectification, which gives the individual which allows the data subject to have incomplete personal data completed and inaccurate personal data rectified.[93]

The of the data controller to ensure accuracy must always be seen in the context of the purpose of the processing. Consequently, one could imagine scenarios where updating personal data is strictly prohibited because the purpose of the processing is documenting events as a historical 'snap-shot' and updating such information would decrease the validity of the records. On the other hand, one could also imagine scenarios where it is necessary to update the data, 'due to the potential damage which might be caused to the data subject if data were to remain

[88] GDPR, art 16.
[89] GDPR, art 17(1).
[90] GDPR, Article 17(1)(a).
[91] Information Commissioner's Office, 'Guide to the General Data Protection Regulation (GDPR): Principle (c): Data minimisation'(*Information Commissioner's Office,* December 2018) <https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/principles/data-minimisation/> accessed 11 May 2019.
[92] GDPR, recital 39.
[93] GDPR, art 16.

inaccurate.' [94] See the following two examples given in the Handbook on European data protection law,[95] which describe both cases:

Updating data is prohibited	Updating data is necessary
'A medical record of an operation must not be changed, in other words 'updated', even if findings mentioned in the record later on turn out to have been wrong. In such circumstances, only additions to the remarks in the record may be made, as long as they are clearly marked as contributions made at a later stage.'[96]	'If somebody wants to conclude a credit contract with a banking institution, the bank will usually check the creditworthiness of the prospective customer. For this purpose, there are special databases available containing data on the credit history of private individuals. If such a database provides incorrect or outdated data about an individual, this person may suffer negative effects. Controllers of such databases must therefore make special efforts to follow the principle of accuracy.'[97]

TABLE 1

The foregoing examples highlight that there is no *one-size-fits-all* solution when it comes to the accuracy principle. Data controllers must always consider the context of the purpose of processing when deciding whether the data requires to be updated or corrected.

3.1.5 Storage Limitation

In accordance with the storage limitation principle the data controller shall keep personal data in a form which permits identification of data subjects for no longer than is necessary for the purposes for which the personal data are processed. However, personal data may be stored for longer periods insofar as the personal data will be processed solely for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes as long as the storage is subject to implementation of the appropriate technical and organizational measures in order to safeguard the rights and freedoms of the data subject.[98] 'This requires, in particular, ensuring that the period for which the personal data are stored is limited to a strict

[94] Council of Europe and others, *Handbook on European data protection law: 2018 edition* (Publications Office of the European Union 2019).
[95] Council of Europe and others, *Handbook on European data protection law: 2018 edition* (Publications Office of the European Union 2019).
[96] Council of Europe and others, *Handbook on European data protection law: 2018 edition* (Publications Office of the European Union 2019), 128.
[97] Council of Europe and others, *Handbook on European data protection law: 2018 edition* (Publications Office of the European Union 2019), 128.
[98] GDPR, art 5(1)(e).

minimum.'[99] Also, since the GDPR doesn't provide specific time limits, the data controller must set those limits himself and be able to justify how long he keeps the personal data. What is considered a justifiable period depends on the type of data and the purpose of the processing. To ensure compliance, the data controller should establish time limits for erasure or for a periodic review.[100]

This principle is closely linked to the data minimization and accuracy principle as ensuring erasure and anonymization of personal data which is no longer needed will reduce the risk of personal data becoming irrelevant, excessive, inaccurate or outdated. Further, personal data which has been held for too long may become unnecessary which means that the data controller is unlikely to have lawful basis for storing such data. Furthermore, the storage of excessive data is more costly, considering storage and security costs.[101] Lastly, the data controller has a responsibility to respond to data subjects' access request,[102] requests for rectification[103] or erasure.[104] Responding to such queries in relation to old data may be burdensome for the data controller. The foregoing highlights the importance for data controllers to establish retention policies or schedules.

When the data is no longer necessary for processing, it is time for the data controller to delete the data. At that point it is important to for the data controller to remember that the word 'deletion' can mean different things in relation to electronic processing. It is not always possible to delete or erase all traces of electronic data. The key issue here for the data controller is to ensure that the data is put data beyond use. A data controller should make its best efforts to delete the personal data from its systems, as well as any back-ups available, as appropriate.

Alternatively, the data controller could anonymize the data instead of completely deleting it.[105] By anonymizing the data, it is no longer considered personal data within the meaning of Article 4(1) of the GDPR as it is no longer a piece of information which permits identification

---

[99] GDPR, recital 39.
[100] GDPR, recital 39.
[101] Information Commissioner's Office, 'Guide to the General Data Protection Regulation (GDPR): Principle (e): Storage limitation' (*Information Commissioner's Office,* December 2018) <https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/principles/storage-limitation/> accessed 11 May 2019.
[102] GDPR, art 15.
[103] GDPR, art 16.
[104] GDPR, art 17.
[105] Information Commissioner's Office, 'Guide to the General Data Protection Regulation (GDPR): Principle (e): Storage limitation' (*Information Commissioner's Office,* December 2018) <https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/principles/storage-limitation/> accessed 11 May 2019.

of a natural person. Lastly, it's worth noting that pseudonymization[106] usually still permits identification and is therefore generally not sufficient to be in compliance with the storage limitation principle. Nonetheless, 'pseudonymisation can be a useful tool for compliance with other principles such as data minimisation and security.'[107]

3.1.6 Security

The principle of integrity and confidentiality, sometime referred to as the *security principle*, is found in Article 5(f) of the GDPR which states that personal data shall be:

> processed in a manner that ensures appropriate security of the personal data, including protection against unauthorised or unlawful processing and against accidental loss, destruction or damage, using appropriate technical or organisational measures ('integrity and confidentiality').

This entails that personal data shall be processed in such a manner that appropriate security and confidentiality of the personal data is ensured. This also includes that the data controller shall prevent any unauthorized access to or use of the personal data or the equipment used for the processing.[108] The security principle is enshrined in Article 32 of the GDPR which specifies in more detail the security requirements of processing:

> Taking into account the state of the art, the costs of implementation and the nature, scope, context and purposes of processing as well as the risk of varying likelihood and severity for the rights and freedoms of natural persons, the controller and the processor shall implement appropriate technical and organisational measures to ensure a level of security appropriate to the risk …

To adhere to the security requirements of the GDPR, data controllers and processors must evaluate risks inherent in their processing and take measures to mitigate the risks.[109] Such measures include pseudonymization and encryption.[110] Data controllers should also consider the risks presented by the processing of personal data, 'in particular from accidental or unlawful destruction, loss, alteration, unauthorised disclosure of, or access to personal data transmitted,

---

[106] GDPR art. 4(5), ' 'pseudonymisation' means the processing of personal data in such a manner that the personal data can no longer be attributed to a specific data subject without the use of additional information, provided that such additional information is kept separately and is subject to technical and organisational measures to ensure that the personal data are not attributed to an identified or identifiable natural person'.

[107] Information Commissioner's Office, 'Guide to the General Data Protection Regulation (GDPR): Principle (e): Storage limitation' (*Information Commissioner's Office,* December 2018) <https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/principles/storage-limitation/> accessed 11 May 2019.

[108] GDPR, recital 39.

[109] GDPR, recital 83.

[110] GDPR, art 32(1)(a).

stored or otherwise processed'[111] which might 'lead to physical, material or non-material damage.'[112] In other words, as the GDPR does not define specifically which measures should be in place. Rather, the scope of security measures depends on the risk of the processing, purpose of the processing, scope of processing, type of personal data processed, and the size of the data controller or processor. There is no 'one-size-fits-all' solution when it comes to security measures pursuant to the GDPR.

Data controllers who do not ensure proper security to their operations leave personal data at risk of potential harm to the data subjects. Data breaches may result in damage to the data subject. This can include:

> … loss of control over their personal data, limitation of their rights, discrimination, identity theft or fraud, financial loss, unauthorized reversal of pseudonymization, damage to reputation, and loss of confidentiality of personal data protected by professional secrecy. It can also include any other significant economic or social disadvantage to those individuals.[113]

Consequently, the data controller's obligation to implement appropriate security measures is not only a matter of legal compliance but failure to do so can result in harm, embarrassment or inconvenience for both the data subject and the data controller.

As referred to in Article 32(1)(b) of the GDPR the security measures put in place should ensure ongoing confidentiality, integrity, availability of processing systems and services. This means that the measures should ensure:

- that the data can only be accessed, altered, disclosed or delete by people that the data controller has authorized to do so, as well as that those people only act within the scope of authority given to them ('confidentiality');
- that the data hold by the data controller is accurate and complete ('integrity'); and
- that the data remains accessible and usable for the whole period of processing ('availability').[114]

---

[111] GDPR, art 32(2).
[112] GDPR, recital 83.
[113] Article 29 Data Protection Working Party, 'Guidelines on Personal data breach notification under regulation 2016/679' (WP 250rev.01, 6 February 2018), 9; GDPR, recital 75; '';--have i been pwned?' (*have I been pwned*) <https://haveibeenpwned.com/> accessed 1 May 2019.
[114] Information Commissioner's Office, 'Guide to the General Data Protection Regulation (GDPR): Principle (f): Integrity and confidentiality (security)' (*Information Commissioner's Office,* December 2018) <https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/principles/integrity-and-confidentiality-security/> accessed 11 May 2019.

These three points (confidentiality, integrity and availability) are commonly known as the *CIA triad* and are the core principles of information security. All information security controls, safeguards, threats, vulnerabilities and security processes of data controllers are subject to the CIA triad.[115] The main idea behind the CIA triad is that if any of the three elements are compromised, then there is a risk of serious consequences, both for the data controller, and for the data subject. In addition to the CIA triad, Article 31(1)(b) of the GDPR provides that data controllers should also ensure the resilience of their processing systems and services. This refers to how well the systems of the data controller can handle adverse incidents. The systems should be able to continue to operate under adverse conditions. The data controller should also be able to restore the systems to an effective state after the incident.[116] The obligations provided by the CIA triad along with the requirement of the resilience can be seen in the following Figure 3.

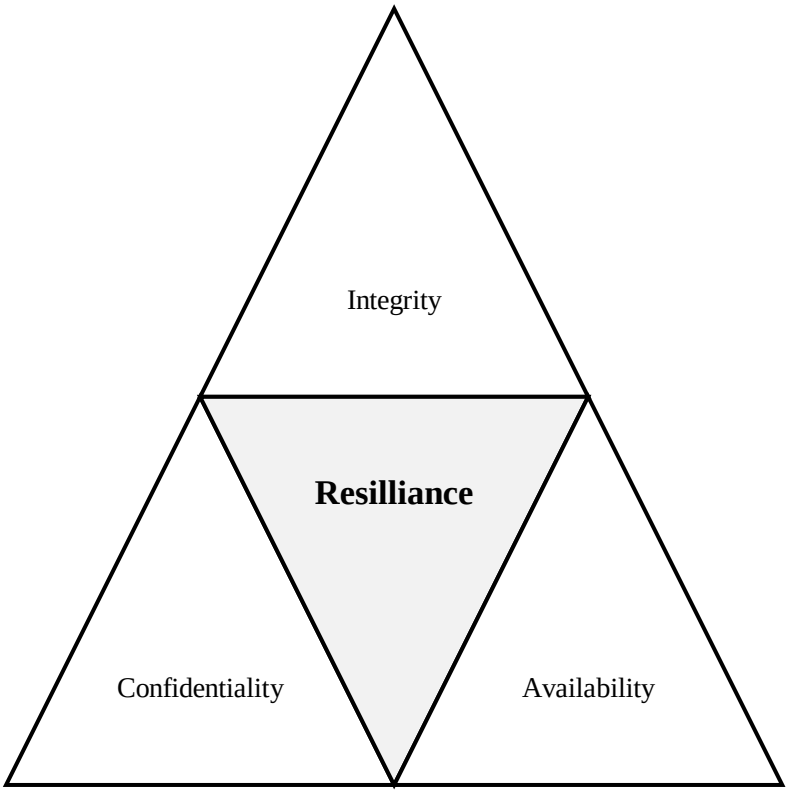

FIGURE 3

[115] Ronald L. Krutz and Russel Dean Vines, *Cloud Security: A Comprehensive Guide to Secure Cloud Computing* (Wiley 2010) 125.
[116] Information Commissioner's Office, 'Guide to the General Data Protection Regulation (GDPR): Principle (f): Integrity and confidentiality (security)' (*Information Commissioner's Office,* December 2018) <https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/principles/integrity-and-confidentiality-security/> accessed 11 May 2019.

3.1.7 Accountability

The seventh and last principle relating to the processing of personal data is the principle of accountability. Data controllers shall be responsible and be able to demonstrate compliance with the other principles.[117] The essence of this obligation of the controller is to:

> put in place measures which would – under normal circumstances – guarantee that data protection rules are adhered to in the context of processing operations; and have documentation ready which demonstrates to data subjects and to supervisory authorities the measures that have been taken to achieve compliance with the data protection rules.[118]

Controllers can comply with the requirements of the principle in various ways, mainly by complying with the provisions of the GDPR and documenting their measures. The following table includes ways to facilitate compliance with the accountability principle as highlighted in the Handbook on European data protection law:[119]

Action	Article
Recording processing activities and making them available to the supervisory authority upon request	30
In certain situations, designating a data protection officer who is involved in all issues relating to personal data protection	37-39
Undertaking data protection impact assessments for types of processing likely to result in a high risk to the rights and freedoms of natural persons	35
Ensuring data protection by design and by default;	25
Implementing modalities and procedures for the exercise of the rights of the data subjects	12, 24
Adhering to approved codes of conduct or certification mechanisms	40, 42

TABLE 2

3.2 Data Protection and Artificial Intelligence

As previously referred to, processing personal data with artificial intelligence systems can raise some data protection issues. In this chapter I will highlight some of the incompatibilities of artificial intelligence and the fundamental principles of the GDPR.

---

[117] GDPR, article 5(2).
[118] Council of Europe and others, *Handbook on European data protection law: 2018 edition* (Publications Office of the European Union 2019), 137; Article 29 Data Protection Working Party, 'Opinion 3/2010 on the principle of accountability' (WP 3/2010).
[119] Council of Europe and others, *Handbook on European data protection law: 2018 edition* (Publications Office of the European Union 2019), 135.

Firstly, the processing of personal data by artificial intelligence system is often invisible or done without the knowledge of the data subject. Profiling, machine learning and other big data processing often rely on creating derived or inferred data about the data subject, so called *bastard data*.[120] Obviously, this new *bastard data* has not been provided by the data subject itself, and in some cases, the data subject doesn't even know of its existence. Furthermore, when dealing with artificial intelligence, the data subject may have a difficult time comprehending and understanding the complex methods and techniques involved in the process.[121]

Secondly, processing by artificial intelligence systems can often be biased[122], unfair and discriminating.[123] An example of this is denial of employment opportunities, denial of credit or insurance, or being targeted with excessively risky or costly financial products based on your profile or an automated decision.[124]

Thirdly, the full capabilities of new technologies are not always foreseen at their developing stages. Therefore, data controllers might collect vast amounts of data to train the system before knowing the full capabilities of the system. In other words, they might collect the data before the end-purpose of the system is known. The foregoing can result in the system using personal data which was originally collected for different purposes than its pre-defined purpose. This further use of the data may (or may not) be incompatible with the purposes for which it was collected in the first place.[125] Therefore, developers of any such systems must resist the urge to use datasets collected for purpose [A] for the incompatible purpose [B]. The GDPR encourages that developers carefully determine the purpose beforehand and inform the data subject before the processing commences.[126]

Fourthly, data controllers using artificial intelligence systems are incentivized to collect and use great amount of data. The costs of data storage, transportation, processing large

---

[120] Joe McNamee, 'Is Privacy Still Relevant in a World of Bastard data?' (*Edri,* 9 March 2016) <https://edri.org/enditorial-is-privacy-still-relevant-in-a-world-of-bastard-data> accessed 11 May 2019.
[121] Article 29 Data Protection Working Party, 'Guidelines on Automated individual decision-making and Profiling for the purposes of Regulation 2016/679' (WP 241, 3 October 2017), 7.
[122] Independent High-Level Expert Group on Artificial Intelligence - Set Up By The European Commission, 'Ethic guidelines For Trustworthy AI' (European Commission 2018) 36, biases are defined as 'an inclination of prejudice towards or against a person, object, or position', they can 'be good or bad, intentional or unintentional. In certain cases, bias can result in discriminatory and/or unfair outcomes'.
[123] See further in Chapter 4.1.
[124] Article 29 Data Protection Working Party, 'Guidelines on Automated individual decision-making and Profiling for the purposes of Regulation 2016/679' (WP 241, 3 October 2017), 10.
[125] See Chapter 3.1.2 on the relevant factors to assess whether a new purpose is compatible with the original purpose.
[126] GDPR, art. 14.

amounts of information has decreased in the last few years, also encouraging developers and organizations to collect more data than needed, in the case it might be profitable or useful in the future. Further, most machine learning systems are heavily dependent on having a large amount of training data. Usually, the greater the dataset used for training the algorithm, the more accurate the algorithm becomes. Therefore, the data controller is incentivized to use all the data he can get its hands on to be able to develop a more accurate and effective system.

Lastly, accuracy of the data used by the system is highly important, especially when collecting and analyzing data, building and applying the algorithm. [127]Any inaccuracies, error in the data or the inclusion of outdated data can result in the system being flawed or even biased. The system might provide incorrect or inappropriate decisions in relation to an individual. An example would be inaccurate statements about someone's, health, credit or insurance risk, which provides obvious problems for the individual. [128] In addition to the system or the algorithm being accurately developed, the dataset itself must be accurate as well. Additionally, it possible to think of a dataset built from perfectly accurate and true raw data, but the dataset itself not being fully representative and therefore resulting in an inaccurate system.[129] Furthermore, one can imagine inaccurate or inappropriate methods being used to analyze the accurate raw data resulting in an inaccurate dataset.[130]

In order to counter some of the foregoing risks, biases, errors and issues in relation to decisions taken by artificial intelligence systems, the GDPR provides certain requirements and obligations to data controllers which limit the risks. Firstly, the GDPR requires data controllers to ensure the accuracy of their data on an ongoing basis, including by verifying that the data used and collected is up to date.[131] By minimizing inaccurate data, data controllers greatly limit the risk of errors and biases. Secondly, data controllers are required to minimize the amount of data they use[132] and must be able to clearly explain and justify the necessity of collection and storage of the data. This requirement makes excessive and unnecessary data collection unlawful. Thirdly, the GDPR encourages controllers to anonymize any data that does not

[127] Article 29 Data Protection Working Party, 'Guidelines on Automated individual decision-making and Profiling for the purposes of Regulation 2016/679' (WP 241, 3 October 2017), 12.
[128] Article 29 Data Protection Working Party, 'Guidelines on Automated individual decision-making and Profiling for the purposes of Regulation 2016/679' (WP 241, 3 October 2017), 12.
[129] See the discussion in Chapter 4.1 on uncertainty-bias.
[130] Article 29 Data Protection Working Party, 'Guidelines on Automated individual decision-making and Profiling for the purposes of Regulation 2016/679' (WP 241, 3 October 2017), 12.
[131] GDPR, art 5(2).
[132]GDPR, art 5(1)(c).

necessarily need to be personally identifiable.[133] By completely anonymizing the data, the processing does no longer fall within scope of the GDPR[134] and therefore provides the data controller more flexibility for further processing of the data without risking the privacy of the data subject or being in breach of the GDPR. Lastly, data controllers are required to disclose the data subject meaningful information about the processing and implement safeguards to minimize the risks.[135] These disclosure requirements and safeguards are the subject of the next Chapter 3.3.

3.3 Rights of the Data Subject

3.3.1 Right to Information (Article 13-14)

The principles of fair and transparent processing require that data subjects are informed of the existence of processing operations of the controller and its purposes.[136] The controller should also 'provide the data subject with any further information necessary to ensure fair and transparent processing taking into account the specific circumstances and context in which the personal data are processed.'[137] This duty to notify and inform the data subject about the processing is found in Articles 13-14 of the GDPR.

3.3.2 Elements of Transparency

Article 12 of the GDPR sets out general rules which apply to the provision of information or communication by the data controller, including information provision pursuant to Article 13 and 14. More specifically, the communication or provision of information must at least comply with the following rules:

a) it must be concise, transparent, intelligible and easily accessible;
b) clear and plain language must be used;
c) it must be in writing or by other means,
d) where requested by the data subject it may be provided orally; and
e) it generally must be provided free of charge.[138]

---

[133] GDPR, art 5(1)(c); GDPR, art 25
[134] GDPR, recital 26.
[135] GDPR, art 12-14; GDPR, art 5(1)(a).
[136] GDPR, recital 60.
[137] GDPR, recital 60.
[138] Article 29 Data Protection Working Party, 'Guidelines on transparency under Regulation 2016/679' (WP 260, 11 April 2018), 6.

I will now take a closer look at the first two rules, a) - b) which are highly linked to the right to an explanation.

Concise, transparent, intelligible and easily accessible

The requirement of information provision in 'concise and transparent' manner provides that data controller should present such information efficiently and succinctly. The information provided should always be well differentiated from other non-privacy related information.[139]

The requirement of the information being 'intelligible' provides that the information needs to be understandable by an 'average member of the intended audience'.[140] For this requirement to achieve it purpose, the data controller must be aware of who its main audience is and 'ascertain the average member's level of understanding'.[141] Further, as the intended audience might differ from the actual audience, the data controller should audit his information provisions on regular basis to ensure that the information is still tailored to the right audience.[142]

The requirement of the information being 'easily accessible' provides that it should be obvious for the data subject where they can access this information. The data subject should not have to seek out this information.[143] As the main purpose of Articles 13-14 of the GDPR is providing the data subject an opportunity to determine and understand in advance the scope and consequences of the processing, information should be provided prior to processing, as is clearly required by Article 13-14 of the GDPR and as discussed later on in Chapter 3.3.4.

Clear and plain language

The requirement of having information provided in a 'clear and plain language' means that the information should be provided in a simple manner. The data controller should avoid complex sentences and language structures. The data controller should also avoid using unclear terms

[139] Article 29 Data Protection Working Party, 'Guidelines on transparency under Regulation 2016/679' (WP 260, 11 April 2018), 7.
[140] Article 29 Data Protection Working Party, 'Guidelines on transparency under Regulation 2016/679' (WP 260, 11 April 2018), 7.
[141] Article 29 Data Protection Working Party, 'Guidelines on transparency under Regulation 2016/679' (WP 260, 11 April 2018), 8.
[142] Article 29 Data Protection Working Party, 'Guidelines on transparency under Regulation 2016/679' (WP 260, 11 April 2018), 8.
[143] Article 29 Data Protection Working Party, 'Guidelines on transparency under Regulation 2016/679' (WP 260, 11 April 2018), 8.

which leave room for different interpretation, the information provided should be concrete and definitive. The purpose, and legal basis of the processing should be particularly clear.[144]

3.3.3 Content

Articles 13 and 14 of the GDPR specify in more detail the content of the information disclosure as is listed here below in Table 3:

Information type	Article	Article
The identity and contact details of the controller and where applicable, their representative	13.1(a)	14.1(a)
Contact details for the data protection officer, where applicable	13.1(a)	14.1(b)
The purposes and legal basis for the processing	13.1(a)	14.2(c)
Where processing is based on legitimate interest, information on the legitimate interests pursued by the data controller or a third party	13.1(d)	14.2(b)
Categories of personal data concerned	N/A	14.1(d)
Recipients, or categories of recipients of the personal data	13.1(e)	14.1(e)
Details of transfer to third countries, the fact of same and the details of the relevant safeguards (and the existence or absence of an adequacy decision by the Commission) and the means to obtain a copy of them or where they have been made available	13.1(f)	14.1(f)
Storage period, or if not possible, criteria used to determine that period	13.2(a)	14.2(a)
The rights of the data subject to a) access, b) rectification, c) erasure, d) restriction on processing, e) objection to processing and f) portability	13.2(b)	14.2(c)
When processing is based on consent, the right to withdraw consent at any time	13.2(c)	14.2(d)
The right to lodge a complaint with a supervisory authority	13.2(d)	14.2(e)
The details of whether individuals are under a statutory or contractual obligation to provide the personal data and the possible consequences of failure.	13.2(e)	N/A
The source of the personal data, and if applicable, whether it came from a publicly accessible source	N/A	14.2(f)

[144] Article 29 Data Protection Working Party, 'Guidelines on transparency under Regulation 2016/679' (WP 260, 11 April 2018), 9.

The existence of automated decision-making including profiling and, if applicable, meaningful information about the logic used and the significance and envisaged consequences of such processing for the data subject[145]	13.2(f)	14.2(g)

TABLE 3

Further, in addition to the information provided above and prescribed under Article 13 and 14, the Article 29 Working Party[146] states that data controllers should also, especially in complex, technical or unexpected data processing:

> … separately spell out in unambiguous language what the most important consequences of the processing will be: in other words what kind of effect will the specific processing described in a privacy statement/ notice actually have on a data subject? Such a description of the consequences of the processing should not simply rely on innocuous and predictable 'best case' examples of data processing, but should provide an overview of the types of processing that could have the highest impact on the fundamental rights and freedoms of data subjects in relation to protection of their personal data.[147]

Therefore, in addition to the information provided in Table 3 hereabove, the data controller should inform the data subject about possible consequences or the effects of the processing has on the data subject.

3.3.4 Timing

Both Article 13 and 14 set out an obligation to the data controller to provide information to the data subject in a timely manner. However, the timing of this obligations differs between Article 13 and 14.

Article 13 only applies to situations where the data is collected directly from the data subject, e.g. in scenarios where the data subject consciously submits the data to the data controller, for example via an online form or by observation. On the other hand, Article 14 only applies where the data has not been obtained from the data subject itself, e.g. in scenarios

[145] This type of information will be discussed in detail later.
[146] This Working Party was set up under Article 29 of Directive 95/46/EC. It is an independent European advisory body on data protection and privacy. Its tasks are described in Article 30 of Directive 95/46/EC and Article 15 of Directive 2002/58/EC. The Working Party ceased to exist as of 25 May 2018 and was replaced by the European Data Protection Board (EDPB).
[147] Article 29 Data Protection Working Party, 'Guidelines on transparency under Regulation 2016/679' (WP 260, 11 April 2018), 8.

where the data controller has collected the data from other sources such as third-party data controllers or data brokers. [148]

In both situation it is crucial that the information is provided in a timely manner. In the case of Article 13 the information should be provided 'at the time when personal data is obtained.'[149] On the other hand, in the case of Article 14, the timing is more flexible as is set out in Article 14.3(a) – c). The general rule is that such information must be provided 'within a reasonable period after obtaining the personal data, but at the latest within one month.'[150]

The GDPR further stipulates that if the data controller plans to use the personal data to communicate with the data subject or to disclose to a third party, the data controller should provide the information at the latest when he first communicates with the data subject[151] or when he discloses[152] the data to the third-party.[153] However, 'in any case, the maximum time limit within which Article 14 information must be provided to a data subject is one month.'[154] Therefore, the one month time limit still applies in these situations.

Lastly, data controllers must take into consideration that the previous mentioned limits are maximum time limits. 'Accountability requires [data] controllers to demonstrate the rationale for their decision and justify why the information was provided at the time it was'.[155] Additionally, the Article 29 Working Party is of the opinion that 'data controllers should provide the information to data subjects well in advance of the stipulated time limits'[156] For these reasons data controllers providing information at the 'last moment' have a hard time complying with the requirements unless this is done for a valid reason.

---

[148] Article 29 Data Protection Working Party, 'Guidelines on transparency under Regulation 2016/679' (WP 260, 11 April 2018), 14.
[149] GDPR, art 13.1.
[150] GDPR, art 14(3)(a).
[151] GDPR, art 14(3)(b).
[152] GDPR, art 14(3)(c).
[153] Information Commissioner's Office, 'Guide to the General Data Protection Regulation (GDPR): The right to be informed – When should we provide privacy information?' (*Information Commissioner's Office,* December 2018) <https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/the-right-to-be-informed/when-should-we-provide-privacy-information/> accessed 11 May 2019.
[154] Article 29 Data Protection Working Party, 'Guidelines on transparency under Regulation 2016/679' (WP 260, 11 April 2018), 15.
[155] Article 29 Data Protection Working Party, 'Guidelines on transparency under Regulation 2016/679' (WP 260, 11 April 2018), 15.
[156] Article 29 Data Protection Working Party, 'Guidelines on transparency under Regulation 2016/679' (WP 260, 11 April 2018), 15.

3.3.5 Article 13 Exceptions

The GDPR provides certain exceptions from the right to be informed. In scenarios whereas the data controller collects the data directly from the individual (Article 13), the data controller does not need to provide them with information 'where and insofar as, the data subject already has the information'.[157] Therefore, if the data controller knows, or if it is obvious that the data subject has the necessary information, the data controller has no obligation to provide it to the data subject again. However, if the data subject does not have all the information, the data controller is obligated to provide the missing information.

3.3.6 Article 14 Exceptions

In scenarios whereas the data controller obtains the information from a third-party source (Article 14), the data controller has a broader set of exceptions. In such scenarios, the data controller does not need to provide the data subject with information when:

a) the data subject already has the information;[158]
b) providing the information would be impossible;[159]
c) providing the information to the data subject would involve a disproportionate effort;[160]
d) providing the information to the data subject would render impossible or seriously impair the achievement of the objectives of the processing;[161]
e) the data controller is required by law to obtain or disclose the personal data;[162] or
f) the data controller is subject to an obligation of professional secrecy regulated by law that covers the personal data;[163]

The Data Subject Already has the Information.

Pursuant to Article 14(5)(a) data controllers do not have to provide information in scenarios whereas the data subject already has the information. This is the same exception as provided in Article 13(4) and has been discussed hereabove in Chapter 3.3.5.

[157] GDPR, art 13(4).
[158] GDPR, art 14(5)(a).
[159] GDPR, art 14(5)(b).
[160] GDPR, art 14(5)(b).
[161] GDPR, art 14(5)(b).
[162] GDPR, art 14(5)(c).
[163] GDPR, art 14(5)(d).

Impossibility

Data controllers are not obligated to provide information when the provision of such information is impossible.[164] Situations where it is impossible to provide the information to the data subject are few and far between.[165] Impossibility within the meaning of Article 14(5)(b) is an all or nothing situation, either something is impossible or not; there are no degrees of impossibility.[166] The data controller must therefore be able to demonstrate that it is indeed impossible to provide the information to the data subject. Further, if situations change, and the factors that prevented the provision of information no longer exists and it suddenly becomes possible to provide the information, the data controller is under an obligation share the information immediately.

Disproportionate Effort

Data controllers are not obligated to share information when there is a disproportionate effort between the effort for the data controller to provide the data subject with the information and the effect that the processing has on them.[167] The greater effect the processing has on a data subject the, the less likely it is that the data controller can rely on this exception.[168] Data controllers should not rely on this exception to routinely escape their obligations of information provision. They need to be able to justify why contacting the data subject is disproportionate in the scenario at hand by performing a balancing exercise to assess (and document) whether the effort involved in contacting the data subject is proportionate considering the rights of the data subjects. 'In that regard, the number of data subjects, the age of the data and any appropriate safeguards adopted should be taken into consideration.'[169]

---

[164] GDPR, art 14(5)(b).
[165] Information Commissioner's Office, 'Guide to the General Data Protection Regulation (GDPR): The right to be informed – Are there any exceptions?' (*Information Commissioner's Office,* December 2018) <https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/the-right-to-be-informed/are-there-any-exceptions/> accessed 11 May 2019.
[166] Article 29 Data Protection Working Party, 'Guidelines on transparency under Regulation 2016/679' (WP 260, 11 April 2018), 25.
[167] GDPR, art 14(5)(b).
[168] Information Commissioner's Office, 'Guide to the General Data Protection Regulation (GDPR): The right to be informed – Are there any exceptions?' (*Information Commissioner's Office,* December 2018) <https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/the-right-to-be-informed/are-there-any-exceptions/> accessed 11 May 2019.
[169] GDPR, recital 62.

Impairment of Objectives

Data controllers are not obligated to provide the information if the provision would render impossible or seriously impair the achievement of the objectives of the processing.[170] As with the other exemptions, data controllers must be able to demonstrate how and why the information provision would nullify the objective of the processing when relying on this exemption. [171]

Collection or Disclosure is Required by Law

Data controllers are not obligated to provide information when the obtaining or disclosure of the personal data 'is expressly laid down by Union or member state law to which the controller is subject.'[172] 'Such a law must directly address the data controller and the obtaining or disclosure in question should be mandatory upon the data controller.'[173]Further, the 'legal basis or legislative measure of processing should be clear and precise and its application should be foreseeable to persons subject to it, in accordance with the case-law of the Court of Justice of the European Union and the European Court of Human Rights'[174] The data controller is required to 'make it clear to data subjects that it obtains or discloses personal data in accordance with the law in question, unless there is a legal prohibition preventing the data controller from doing so.'[175]

Confidentiality

Lastly, data controllers are not obligated to provide information where the personal data 'must remain confidential subject to an obligation of professional secrecy regulated by Union or Member State law, including a statutory obligation of secrecy. To utilize this exemption the data controller must identify and be able to demonstrate how the secrecy obligation prohibits the information disclosure.

---

[170] GDPR, art 14(5)(b).
[171] Information Commissioner's Office, 'Guide to the General Data Protection Regulation (GDPR): The right to be informed – Are there any exceptions?' (*Information Commissioner's Office*, December 2018) <https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/the-right-to-be-informed/are-there-any-exceptions/> accessed 11 May 2019.
[172] GDPR, art 14(5)(c).
[173] Article 29 Data Protection Working Party, 'Guidelines on transparency under Regulation 2016/679' (WP 260, 11 April 2018), 29.
[174] GDPR, recital 41.
[175] Article 29 Data Protection Working Party, 'Guidelines on transparency under Regulation 2016/679' (WP 260, 11 April 2018), 29.

3.4 The Right to Access (Article 15)

Article 15 of the GDPR provides the data subject a *right to access* of its own personal data. This right is also set out as an essential part of the fundamental right of data protection in Article 8(2) of the EU Charter of Fundamental Rights. The data subject has a 'right to obtain from the controller a confirmation as to as to whether or not personal data concerning him or her are being processed, and, where that is the case, access to the personal data.'[176] Further, the data controller shall provide the data subject with the following information:

Information type	Article
Purpose of the processing	15.1(a)
Categories of personal data concerned	15.1(b)
Recipients, or categories of recipients of the personal data	15.1(c)
Storage period, or if not possible, criteria used to determine that period	15.1(d)
The rights of the data subject to, a) rectification, b) erasure, c) restriction on processing, d) objection to processing	15.1(e)
The right to lodge a complaint with a supervisory authority	15.1(f)
When applicable, any available information on the source of the personal data.	15.1(g)
The existence of automated decision-making including profiling and, if applicable, meaningful information about the logic used and the significance and envisaged consequences of such processing for the data subject.	15.1(h)
When applicable, information on the appropriate safeguards pursuant to Article 46 relating to transfer to a third country or to an international organization.	15.2

TABLE 4

In other words, the *right of access* gives the data subject a right to obtain from the data controller:

a) confirmation that he is processing their data;
b) a copy of its personal data; and
c) the supplementary information listed in Table 4 hereabove.

If the content of the supplementary information pursuant to Article 15 (See Table 4), is compared to the content of notifications pursuant the notification duties[177] (see Table 3) one

[176] GDPR, art 15.
[177] GDPR, art 13 - 14.

can see that data controllers should already be providing much of the same content in their privacy notices pursuant to Article 13-14.

3.5 Automated Individual Decision Making (Article 22)

Article 22 of the GDPR states that:

> 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.

It states that, automated decision-making is generally prohibited. This prohibition however, only applies, in very particular scenarios, that is, only when there is decision *based solely* on automated processing, which produces *legal effects* concerning the data subject or *similarly significantly* affects the data subject. Only if these conditions are met, then the processing is prohibited. This prohibition is however, not without exceptions:

> … decision-making based on such processing, including profiling, should be allowed where expressly authorised by Union or Member State law… or necessary for the entering or performance of a contract … or when the data subject has given his or her explicit consent [178]

Only in these cases, that is, when a) it is specifically allowed authorized by law of a member state, b) it is necessary for a contract with the data subject or c) when the data subject has given explicit consent, such processing is allowed, as long as the safeguards referred to in Article 22(3) are ensured. Further, 'automated decision-making and profiling based on special categories of personal data should be allowed only under specific conditions'[179] that is, only when the processing is necessary for a substantial public interest.[180] To summarize, Article 22 stipulates that:

- as a rule, fully automated individual decision-making, including profiling that has a legal or similarly significant effect is prohibited;
- there are exceptions to the rule;
- where one of these exceptions applies, there must be measures in place to safeguard the data subject's rights and freedoms and legitimate interests.[181]

The purpose of Article 22 is to protect the data subject against the possible harms of automated decisions, including decisions made by artificial intelligence systems. The GDPR's *right to an*

---

[178] GDPR, recital 71; GDPR, art 22(2).
[179] GDPR, recital 71.
[180] GDPR, article 22(4).
[181] Article 29 Data Protection Working Party, 'Guidelines on Automated individual decision-making and Profiling for the purposes of Regulation 2016/679' (WP 241, 3 October 2017), 19.

explanation, which I will analyze in Chapter 4 applies to decisions that fall within the scope of Article 22(1). Therefore, in order to understand what kind of decisions fall within the scope, it is necessary to take a closer look at the conditions of Article 22(1). This will be done in the following Chapters 3.5.1 - 3.5.2.

3.5.1 Solely Automated Processing

The first condition of Article 21(1) is that the decision must be *solely* on automated processing. Therefore, for the provision to apply there must be no human involvement in the decision-making process. The scope of the Article was narrowed from prior drafts. In the European Parliament's proposed amendments to the European Council's draft of the GDPR[182] the European Parliament proposed the following amended Article 20(5): [183]

> Profiling which leads to measures producing legal effects concerning the data subject or does similarly significantly affect the interests, rights or freedoms of the concerned data subject shall not be based *solely or predominantly on automated processing* and shall include human assessment, including an explanation of the decision reached after such an assessment. The suitable measures to safeguard the data subject's legitimate interests referred to in paragraph 2 shall include the right to obtain human assessment and an explanation of the decision reached after such assessment.[184]

With these amendments the European Parliament wanted to prohibit automated decisions that were either *predominantly* or *solely* based on automated processing which is much broader than only referring to *solely*. However, the word *predominantly* was not adopted in the final text of the adopted version of the GDPR. A strict interpretation of the provision would therefore lead to a situation 'whereby even nominal involvement of a human in the decision-making process allows for an otherwise automated mechanism to avoid invoking elements of the right

---

[182] European Commission, 'Regulation of the European Parliament and the Council on the Protection of Individuals with Regard to the Processing of Personal Data and on the Free Movement of Such Data (General Data Protection Regulation)' (European Commission 2012) 2012/0011 (COD) <http://ec.europa.eu/justice/data-protection/document/review2012/com_2012_11_en.pdf> accessed 19 April 2019.

[183] Article 22 in the adopted version of the GDPR.

[184] European Parliament Committee on Civil Liberties, Justice and Home Affairs, 'Report on the Proposal for a Regulation of the European Parliament and of the Council on the Protection of Individuals with Regard to the Processing of Personal Data and on the Free Movement of Such Data (General Data Protection Regulation) - A7-0402/2013' (European Parliament 2013) A7–0402/2013 <http://www.europarl.europa.eu/sides/getDoc.do?pubRef=-//EP//TEXT+REPORT+A7-2013-0402+0+DOC+XML+V0//EN> accessed 19 April 2019 (emphasis added).

of access.'[185] However, this seems not to be the case. Article 29 Working Party opinion that Article 22(1) could still apply in cases of insignificant or minimal human interference takes place:

> The controller cannot avoid the Article 22 provisions by fabricating human involvement. For example, if someone routinely applies automatically generated profiles to individuals without any actual influence on the result, this would still be a decision based solely on automated processing.
>
> To qualify as human involvement, the controller must ensure that any oversight of the decision is meaningful, rather than just a token gesture. It should be carried out by someone who has the authority and competence to change the decision. [186]

However, it is still unclear what the threshold for human involvement is.[187] How much or how meaningful involvement is required for the decision to fall outside the scope of Article 22(1)?

3.5.2 Legal Effects or Similarly Significantly Affects

Secondly, Article 22(1) only applies to automated decisions with *legal effects* or *similarly significant effects*. Recital 71 of the GDPR provides two examples of such effects: automatic refusal of an online credit application and e-recruiting practices.

A *legal effect* refers to the decision affecting the data subject's legal rights, such as the 'the freedom to associate with others, vote in an election, or take legal action. A legal effect may also be something that affects a person's legal status or their rights under a contract.'[188] Examples of this are 'cancellation of a contract … entitlement to or denial of a particular social benefit granted by law, such as child or housing benefit … refused admission to a country or denial of citizenship.'[189]

Article 22(1) also applies in situations whereas the decision produces an effect that has a *similarly significant* impact. That means that even if there are no legal or contractual consequences of the processing the article could still apply. An example of this is 'automatic

---

[185] Sandra Wacther, Brent Mittelstadt and Luciano Floridi, '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, 76, 88.
[186] Article 29 Data Protection Working Party, 'Guidelines on Automated individual decision-making and Profiling for the purposes of Regulation 2016/679' (WP 241, 3 October 2017), 21.
[187] SCHUFA, BGH, VI ZR 156/13, 28 January 2014, in this case the German Federal Court of Justice found that a minimal human involvement in a decision, otherwise automatic, could be enough for a decision not being considered *solely* automated.
[188] Article 29 Data Protection Working Party, 'Guidelines on Automated individual decision-making and Profiling for the purposes of Regulation 2016/679' (WP 241, 3 October 2017), 21.
[189] Article 29 Data Protection Working Party, 'Guidelines on Automated individual decision-making and Profiling for the purposes of Regulation 2016/679' (WP 241, 3 October 2017), 21.

refusal of an online credit application or e-recruiting practices without any human intervention.'[190] The decision must have the potential to significantly:

> influence the circumstances, behaviour or choices of the individuals concerned; have prolonged or permanent impact on the data subject; or at its most extreme, the decision may lead to the exclusion or discrimination of individuals.[191]

Although it is difficult to be precise about what sufficiently meets the foregoing threshold, one can imagine the following kind of decisions meeting the threshold:

- decisions that affect the data subject's financial circumstances;
- decisions that affect the data subject's access to health services or education;
- decisions that deny or hinder the data subject an employment opportunity;[192]

Further, when assessing whether the threshold is met, the particular characteristics of the case should be considered. Consider the following example from the Article 29 Working Party Guidelines: 'Someone known or likely to be in financial difficulties who is regularly targeted with adverts for high interest loans may sign up for these offers and potentially incur further debt.'[193] This is evident of processing that might generally have little or non-significant impact on most data subjects but might have a *significant* impact on certain vulnerable groups, in this case an individual in financial difficulties. Therefore, the characteristics of each case should be taken into consideration.

4 The Right to an Explanation

4.1 Why do Individuals Want a Right to an Explanation?

Humans tend to 'make systematic and predictable mistakes, and our decisions are subject to bias.'[194] due to the fact that human cognitive abilities are only finite and that we only have 'limited computational skills and seriously flawed memories.'[195] This fact is perhaps best enshrined in the following observation of the US Supreme Court of Justice Oliver Wendell:

---

[190] GDPR, recital 71.
[191] Article 29 Data Protection Working Party, 'Guidelines on Automated individual decision-making and Profiling for the purposes of Regulation 2016/679' (WP 241, 3 October 2017), 21.
[192] Article 29 Data Protection Working Party, 'Guidelines on Automated individual decision-making and Profiling for the purposes of Regulation 2016/679' (WP 241, 3 October 2017), 21.
[193] Article 29 Data Protection Working Party, 'Guidelines on Automated individual decision-making and Profiling for the purposes of Regulation 2016/679' (WP 241, 3 October 2017), 22.
[194] Tal Zarsky, 'The Trouble with Algorithmic Decisions: An Analytic Road Map to Examine Efficiency and Fairness in Automated and Opaque Decision Making' (2016) 41[1] Science, Technology & Human Values 118, 122.
[195] Christine Jolls, Cass R. Sunstein and Richard Thaler, 'A Behavioral Approach to Law and Economics' (1988) 50 Stanford Law Review 1471, 1477.

> *The life of the law has not been logic: it has been experience*. The felt necessities of the time, the prevalent moral and political theories, intuitions of public policy, avowed or unconscious, *even the prejudices which judges share with their fellow-men, have had a good deal more to do than the syllogism in determining the rules by which men should be governed*.[196]

Other legal realists have phrased the same observation by saying that justice is 'what the judge ate for breakfast.'[197] Further, they have argued that 'rational application of legal reason does not sufficiently explain judicial decisions and that psychological, political, and social factors influence rulings as well.'[198] This hypothesis was tested in a study where researchers looked at more than 1,000 rulings made in 2009 by eight judges in sequential parole decisions.[199] The result of the study was that 'likelihood of a favorable ruling is greater at the very beginning of the work day or after a food break than later in the sequence of cases.' In other words, judges were more lenient at the beginning of the workday or right after a food break. This confirms that extraneous variables can influence judicial decisions, and points to the susceptibility of experienced judges to psychological biases.[200]

The foregoing is just one example of predictable human biases in decision-makers, but the examples are countless.[201] '[T]hese errors of human judgment and bias might be mitigated in the automated environment.'[202] By replacing human decision-makers with artificial intelligence systems, the human biases can potentially be reduced. However, the fact remains that artificial intelligence systems also have the potential to be both be biased and embody values.[203] For one, there is always a risk of algorithms having an *uncertainty-bias*, which arises when two conditions are met:

a) 'One group is underrepresented in the sample, so there is more uncertainty associated with predictions about that group

---

[196] Oliver Wendell Holmes Jr., *The Common Law* (first published 1881, Paulo J. S. Pereira & Diego M. Beltran 2011) 5. (emphasis added)
[197] Alex Kozinski, 'What I Ate for Breakfast and Other Mysteries of Judicial Decision Making' (1993) 26 Loyola LA L Rev 993.
[198] Shai Danziger, Jonathan Levav and Liora Avnaim-Pesso, 'Extraneous factors in judicial decisions' (2011) 108 PNAS 6889.
[199] Shai Danziger, Jonathan Levav and Liora Avnaim-Pesso, 'Extraneous factors in judicial decisions' (2011) 108 PNAS 6889.
[200] Shai Danziger, Jonathan Levav and Liora Avnaim-Pesso, 'Extraneous factors in judicial decisions' (2011) 108 PNAS 6889, 6892.
[201] William Meadow and Cass R. Sunstein, 'Statistics, Not Experts' (2001) 51 Duke Law Journal 629, 630.
[202] Tal Zarsky, 'The Trouble with Algorithmic Decisions: An Analytic Road Map to Examine Efficiency and Fairness in Automated and Opaque Decision Making' (2016) 41[1] Science, Technology & Human Values 118, 122.
[203] Helen Nissenbaum, 'How computer systems embody values' (2001) 34[3] Computer 119, 120.

b) The algorithm is *risk averse*, so it will *ceteris paribus* prefer to make decisions based on predictions about which they are more confident (i.e. those with smaller confidence intervals)'[204]

This means that the algorithm could, favor the group better represented in the training data, 'since there will be less uncertainty associated with those predictions.'[205]Also, '[i]n the case of knowledge-based systems,[206] the knowledge that is fed into the system, and assumptions that are involved in modelling it, may reflect biases of the system designers and data collection process.'[207] However, in the case of machine learning algorithms, there is also another separate source of discrimination:

> If an algorithm is trained on data that are biased or reflect unjust structural inequalities of gender, race or other sensitive attributes, it may 'learn' to discriminate using those attributes (or proxies for them). In this way, decisions based on machine learning algorithms might end up reinforcing underlying social inequalities. This kind of problem might arise when predictive models are used in areas like insurance, loans, housing and policing. If members of certain groups have historically been more likely to default on their loans, or been more likely to be convicted of a crime, then the model may give a higher risk score to individuals from those groups.[208]

These biases and algorithmic discriminations can have serious consequences for the data subject as is evident from the following real-world examples. According to the investigative journalism organization ProPublica, Compas,[209] an algorithm used by the US courts for risk assessment is supposedly biased against black people. The algorithm 'was particularly likely to falsely flag black defendants as future criminals, wrongly labeling them this way at almost twice the rate as white defendants. White defendants were mislabeled as low risk more often than black defendants.' [210] A 2018 University of California, Berkeley study found significant discrimination by algorithmic fintech lenders. African-Americans and Latinx were found to pay a 5.3 basis points higher interest rate for purchase mortgages than White and Asian

---

[204] Bryce Goodman & Seth Flaxman, 'European Union regulations on algorithmic decision-making and a "right to explanation"' [2017] AI Magazine 50, 54.
[205] Bryce Goodman & Seth Flaxman, 'European Union regulations on algorithmic decision-making and a "right to explanation"' [2017] AI Magazine 50, 54.
[206] Sometimes referred to as *rule-based systems*, see Chapter 2.2.1.
[207] Reuben Binns, 'Algorithmic Accountability and Public Reason' (2017) 31[4] Philosophy & Technology 1, 4.
[208] Reuben Binns, 'Algorithmic Accountability and Public Reason' (2017) 31[4] Philosophy & Technology 1, 4.
[209] Abbreviation for 'Correctional Offender Management Profiling for Alternative Sanctions'.
[210] Julia Angwin and others, 'Machine Bias: There's software used across the country to predict future criminals. And it's biased against blacks' (*ProPublica*, 23 May 2016) <https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing> accessed 11 May 2019.

ethnicity borrowers do.[211] Finally, Google Photos' image recognition algorithms have been reported classifying African-American people as gorillas by mistake. [212]

These examples show that algorithms, just as humans, make errors and can biased. There is an expression in computer science: garbage in, garbage out. This means that 'regardless of how accurate a program's logic is, the results will be incorrect if the input is invalid.'[213] The quality of the output is dependent on the quality of the input. For this reason, any underlying biases in the training data for a machine learning algorithm might result in a biased system. By relying on historical data developers risk the system mimicking the mistakes of the past. This complex 'and multifaceted nature of algorithmic discrimination suggests that appropriate solutions will require an understanding of how it arises in practice.'[214] Further,

> [t]ransparency arguably may correct errors in any algorithmic process, thus promoting efficiency. It allows individuals to correct inaccurate data that have been collected about them. In this way, transparency also brings the scrutiny that will pressure agencies to improve their practices. This leads to the conclusion that transparent processes will prove more accurate and, thus, efficient. [215]

This highlights the need for increased transparency in algorithmic decision making and the value of a right to explanations.[216] However, note that transparency is not a magical solution against discrimination, transparency itself 'cannot prevent non-discrimination or ensure fairness,'[217] rather transparency helps spot these biases in order to minimize harm. Some have even argued that transparency alone, does little for the data subject, something more is required.[218]

---

[211] Robert P. Bartlett and others, 'Consumer Lending Discrimination in the FinTech Era' (2017) UC Berkley Public Law Research Paper.

[212] James Vincent, 'Google "fixed" its racist algorithm by removing gorillas from its image-labeling tech: Nearly three years after the company was called out, it hasn't gone beyond a quick workaround' (*The Verge,* 12 January 2018) <https://www.theverge.com/2018/1/12/16882408/google-racist-gorillas-photo-recognition-algorithm-ai> accessed 11 May 2019.

[213] Techopedia, 'Garbage In, Garbage Out (GIGO)' <https://www.techopedia.com/definition/3801/garbage-in-garbage-out-gigo> accessed 11 May 2019.

[214] Bryce Goodman & Seth Flaxman, 'European Union regulations on algorithmic decision-making and a "right to explanation"' [2017] AI Magazine 50, 55.

[215] Tal Zarsky, 'The Trouble with Algorithmic Decisions: An Analytic Road Map to Examine Efficiency and Fairness in Automated and Opaque Decision Making' (2016) 41[1] Science, Technology & Human Values 118, 120. (emphasis added)

[216] Bryce Goodman & Seth Flaxman, 'European Union regulations on algorithmic decision-making and a "right to explanation"' [2017] AI Magazine 50, 55.

[217] Independent High-Level Expert Group on Artificial Intelligence - Set Up By The European Commission, 'Ethic guidelines For Trustworthy AI' (European Commission 2018), 34.

[218] Joshua A. Kroll and others, 'Accountable Algorithms' (2017) 165 University of Pennsylvania Law Review 633, 660 'transparency alone does little to explain either why any particular decision was made or how fairly the system operates across bases of users or classes of queries. With such systems, there is the added risk that the rule

Lastly, '[t]he ever-increasing application of algorithms to decision-making in a range of social contexts has prompted demands for algorithmic accountability'[219]

> 'Accountability in its fundamental sense means being answerable for one's actions to some authority and having to suffer sanctions for those actions: 'A is accountable to B when A is obliged to inform B about A's (past or future) actions and decisions, to justify them, and to suffer punishment in the case of eventual misconduct.' [220]

Therefore, in order to be accountable for their automated decisions, data controllers should be able to provide their decision-subjects with reasons and explanations for the design and operation of their automated decision-making system and for their decisions. [221]

For all of the aforementioned reasons, increased transparency in the form of explanation rights, might be 'useful remedy for taming the algorithm'[222] and has been ported as a prime solution to the algorithmic concerns highlighted hereabove.[223] But what does one mean by an explanation of automated decision making? Before, exploring the legal basis for a right to explanation it is necessary to have in mind what kind of explanations may be in question. Mainly one can envisage two kind of explanations of the decision-making process:

System functionality	**Specific decisions**
'The logic, significance, envisaged consequences, and general functionality of an automated decision-making system, eg the system's requirements specification, decision trees, pre-defined models, criteria, and classification structures.' [224]	'The rationale, reasons, and individual circumstances of a specific automated decision, eg the weighting of features, machine-defined case-specific decision rules, information about reference or profile groups.' [225]

TABLE 5

disclosed is obsolete by the time it can be analyzed. Online machine learning systems update their decision rules after every query, meaning that any disclosure will be obsolete as soon as it is made.'

[219] Reuben Binns, 'Algorithmic Accountability and Public Reason' (2017) 31[4] Philosophy & Technology 1. (emphasis added)

[220] Mark Bovens, Thomas Schillemans, and Robert E. Goodin, *The Oxford Handbook of Public Accountability* (Oxford University Press 2014) 6.

[221] Reuben Binns, 'Algorithmic Accountability and Public Reason' (2017) 31[4] Philosophy & Technology 1, 2

[222] Lilian Edwards, and Michael Veale, 'Slave to the Algorithm? Why a 'Right to an Explanation' Is Probably Not the Remedy You Are Looking For' (2017) 16 Duke Law & Technology Review 18.

[223] Independent High-Level Expert Group on Artificial Intelligence - Set Up By The European Commission, 'Ethic guidelines For Trustworthy AI' (European Commission 2018), 18.

[224] Sandra Wacther, Brent Mittelstadt and Luciano Floridi, '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 76, 78.

[225] Sandra Wacther, Brent Mittelstadt and Luciano Floridi, '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 76, 78.

Also, explanations can be distinguished on their timing:

Ex ante	Ex post
'An *ex ante* explanation occurs prior to an automated decision-making taking place. Note that an *ex ante* explanation can logically address only system functionality, as the rationale of a specific decision cannot be known before the decision is made.'[226]	'An ex post explanation occurs after an automated decision has taken place. Note that an ex post explanation can address both system functionality and the rationale of a specific decision.'[227]

TABLE 6

While explanations of a *system functionality* can be provided both *ex ante* and *ex post*, the nature of an explanation of *specific decision* makes it so that it cannot be provided *ex ante* as it can logically only be provided once a decision has been taken.

With these categories of explanations in mind, I will, in the following chapters, explore the legal basis for a right to explanation. For this purpose, I will perform a systematic interpretation of the relevant provisions of the GDPR.

4.2 Legal Basis for the Right to an Explanation

As referred to in the foregoing Chapter, machine learning algorithms, are increasingly important in our society but have also caused a range of concerns, mainly revolving around unfairness, discrimination and opacity. 'Transparency in the form of a "right to explanation" has emerged as a compellingly attractive remedy since it intuitively promises to open the algorithmic "black box" to promote challenge, redress, and hopefully heightened accountability.'[228] However, using the term *right to an explanation* may be misleading as

> [t]here is no single, neat statutory provision labelled the 'right to an explanation' in Europe's new General Data Protection Regulation (GDPR). But nor is such a right illusory. Articles 13–15 provide rights to 'meaningful information about the logic involved' in automated decisions.[229]

[226] Sandra Wacther, Brent Mittelstadt and Luciano Floridi, '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 76, 78.
[227] Sandra Wacther, Brent Mittelstadt and Luciano Floridi, '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 76, 78.
[228] Lilian Edwards, and Michael Veale, 'Slave to the Algorithm? Why a 'Right to an Explanation' Is Probably Not the Remedy You Are Looking For' (2017) 16 Duke Law & Technology Review 18.
[229] Andrew D. Selbst and Julia Powels, 'Meaningful Information and the Right to Explanation' (2017) 7[4] International Data Privacy Law, 233. (emphasis added)

However, whether one uses the phrase *right to an explanation* or not is irrelevant and 'more attention must be paid to the GDPR's express requirements and how they relate to its background goals, and more thought must be given to determining what the legislative text actually means.'[230]In the following Chapters I will follow this general line of thinking and perform a systematic interpretation of the GDPR's provisions where such a right can (or cannot) be derived from. More specifically I will assess whether the following provisions provide a legal basis for the right to an explanation:

a) right to information (Article 13 and 14);
b) right to access (Article 15); and
c) right not to be subjected to automated decision-making (Article 22). [231]

4.2.1 Safeguards of Article 22 and Recital 71

Turning to the first legal basis for the right to an explanation. Some have argued that right to an explanation can be derived from the safeguards referred to in Article 22 and Recital 71 of the GDPR which provide the data subject a right not to be subject to automated decision-making unless exceptions apply.[232] Article 22(3) provides a non-exhaustive list of safeguards that need to be in place whenever a data controller relies on one of these exceptions and utilizes automated-decision making. These safeguards include:

> … 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.

The data controller should therefore at least allow the data subject to a) obtain human intervention, b) express his or her point of view, and c) contest the decision. Note how this does *not* include a right to an explanation after the decision has been made. The right to an explanation is only mentioned in Recital 71 which stipulates the in relation to automated decision-making:

> … such processing should be subject to suitable safeguards, which should include specific information to the data subject and the right to obtain human intervention, to express his or her point of view, *to obtain an explanation of*

[230] Andrew D. Selbst and Julia Powels, 'Meaningful Information and the Right to Explanation' (2017) 7[4] International Data Privacy Law, 233, 234.
[231] 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, 244.
[232] See Chapter 3.5

> *the decision reached after such assessment* and to challenge the decision. Such measure should not concern a child. (emphasis added)

The safeguards referred to are almost identical to the safeguards referred to in Article 22(3). However, the emphasized part, which refers to a right to an explanation, is missing from Article 22(3). This omission is crucial as recitals are not legally binding. To understand the importance of this omission I will briefly discuss the legal status of recitals in European law.

The purpose and role of recitals in European legislation is to 'explain the background to the legislation and the aims and objectives of the legislation'[233] and as such they are not legally binding or enforceable by themselves:

> In principle the ECJ does not give effect to recitals that are drafted in normative terms. Recitals can help to explain the purpose and intent behind a normative instrument. They can also be taken into account to resolve ambiguities in the legislative provisions to which they relate, but they do not have any autonomous legal effect.[234]

This role of recitals has been crystallized in the jurisprudence of the European Court of Justice: '[w]hilst a recital in the preamble to a regulation may cast light on the interpretation to be given to a legal rule, it cannot in itself constitute such a rule.' In line with this, the role of a recital is of a supplementary nature and should only be used to clarify or dissolve ambiguity in the main text of the legislation. Consequently, that the right to an explanation referred to in Article 71 cannot be legally binding or enforceable on its own as this right is not mentioned anywhere in the operative text of the GDPR. Therefore, Recital 71 can only be used to dissolve any ambiguities in the language of Article 22(3). However, the language of Article 22(3) is unambiguous:

> Article 22(3) lists the minimum requirements that have to be met for lawful automated decision-making. There are no ambiguities in the language that would require further interpretation with regard to the minimum requirements that must be met by data controllers. As long as these requirements are met, automated decision-making is lawful and in compliance with the GDPR.[235]

Consequently, the non-binding Recital 71 reference to a right to an explanation has no impact on the interpretation of the legally binding Article 22(3).

---

[233] EUROPA, 'Guide to the Approximation of EU Environmental Legislation ANNEX I' (*Environment*, 2015) <http://ec.europa.eu/environment/archives/guide/annex1.htm> accessed 19 April 2019.

[234] Roberto Baratta, 'Complexity of EU Law in the Domestic Implementing Process' (2014) 2 The Theory and Practice of Legislation, 293.

[235] Sandra Wachter, Brent Mittelstadt and Luciano Floridi, 'Why a Right to Explanation of Automated Decision-Making Does Not Exist in the General Data Protection Regulation' (2017) 7 International Data Privacy Law 76, 80.

Lastly, this omission of the right to an explanation from Article 22(2) but inclusion in the almost identical Recital 71 is intentional rather than a mistake by the legislators. This can be seen in the previous drafts of the GDPR and commentary from the trialogue negotiations. Initially the European Council proposed a draft text of the GDPR which did not include a right to an explanation.[236] In answer to the draft, the European Parliament proposed the following amendments to Article 20(5):[237]

> Profiling which leads to measures producing legal effects concerning the data subject or does similarly significantly affect the interests, rights or freedoms of the concerned data subject shall not be based solely or predominantly on automated processing and shall include human assessment, including an explanation of the decision reached after such an assessment. The suitable measures to safeguard the data subject's legitimate interests referred to in paragraph 2 shall include the right to obtain human assessment and *an explanation of the decision reached after such assessment.*[238]

As seen in the italic part of the Article, the amendments included a right to obtain an explanation. However, these amendments to the list of safeguards were not adopted into the final version of the GDPR, they were rejected. 'This change suggests that legislators intentionally chose to make the right to explanation non-binding by placing it in Recital 71.'[239]

Lastly, the Article 29 Working Party has issued and adopted guidelines on automated individual decision-making and profiling for the purposes of the GDPR. The Article 29 Working Party states the following in relation to the safeguards pursuant to Article 22(3):

> Recital 71 highlights that *in any case* suitable safeguards should also include:
>
> ... specific information to the data subject and the right … to obtain an explanation of the decision reached after such assessment and to challenge the decision.

---

[236] European Commission, 'Regulation of the European Parliament and the Council on the Protection of Individuals with Regard to the Processing of Personal Data and on the Free Movement of Such Data (General Data Protection Regulation)' (European Commission 2012) 2012/0011 (COD) <http://ec.europa.eu/justice/data-protection/document/review2012/com_2012_11_en.pdf> accessed 19 April 2019.

[237] Article 22 in the adopted version of the GDPR.

[238] European Parliament Committee on Civil Liberties, Justice and Home Affairs, 'Report on the Proposal for a Regulation of the European Parliament and of the Council on the Protection of Individuals with Regard to the Processing of Personal Data and on the Free Movement of Such Data (General Data Protection Regulation) - A7-0402/2013' (European Parliament, 2013) A7–0402/2013 <http://www.europarl.europa.eu/sides/getDoc.do?pubRef=-//EP//TEXT+REPORT+A7-2013-0402+0+DOC+XML+V0//EN> accessed 19 April 2019 (emphasis added).

[239] Sandra Wachter, Brent Mittelstadt and Luciano Floridi, 'Why a Right to Explanation of Automated Decision-Making Does Not Exist in the General Data Protection Regulation' (2017) 7 International Data Privacy Law 76, 81.

> The controller must provide a simple way for the data subject to exercise these rights.
>
> This emphasises the need for transparency about the processing. The data subject will only be able to challenge a decision or express their view if they fully understand how it has been made and on what basis. Transparency requirements are discussed in Chapter IV (section E). (emphasis added.) [240]

In this excerpt the Article 29 Working Party highlights the importance of the safeguards referred to in Recital 71 being established, including the right to an explanation. However, the Working Party refers back to Chapter IV (section E) of the Guidelines. Interestingly, Chapter IV (section E) of the Guidelines only discuss Articles 13-15 (right to information and access) rather than Article 22. This is evident that the legal basis for the right to an explanation is not Article 22, but rather Articles 13-14 or Article 15.

Consequently, a right to an explanation cannot be derived solely from the safeguards of Article 22(3). Note however, that data controllers are welcome to, and even encouraged to provide such a right to the data subjects as a one of the suitable measures required to comply with Article 22 (3). However, I conclude that providing such a right to the data subject is not mandatory. Nonetheless, the text of Article 22 and Recital 71 may nonetheless support the existence of such a right deriving from Article 13-14 or Article 15, as discussed in the following chapters.[241]

4.2.2 Right to Information (Article 13-14)

Concluding that a right to an explanation cannot be derived from Article 22 of the GDPR, I will now explore the second legal basis for the right. Scholars have suggested that the right to information provided in Article 13-14 of the GDPR[242] in combination with the safeguards provided in Article 22(3) and Recital 71,[243] provides an right to explanation.[244] As discussed earlier, Article 13 of 14 of the GDPR stipulate that data controllers shall provide to the data subject certain information to ensure fair and transparent processing. In the case of automated decision-making, Articles 13(2)(f) and 14(2)(g) state that the data controller shall provide the data subject with information on:

---

[240] Article 29 Data Protection Working Party, 'Guidelines on Automated individual decision-making and Profiling for the purposes of Regulation 2016/679' (WP 241, 3 October 2017), 27.
[241] [241] Andrew D. Selbst and Julia Powels, 'Meaningful Information and the Right to Explanation' (2017) 7[4] International Data Privacy Law, 233.
[242] See Chapters 3.3.1 and 3.4 for further discussion on the notification duties.
[243] See Chapter 4.2.1 on why these safeguards cannot solely provide a right to an explanation.
[244] Bryce Goodman & Seth Flaxman, 'European Union regulations on algorithmic decision-making and a "right to explanation"' [2017] AI Magazine 50, 54.

> 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. (emphasis added)

Therefore, in the case of automated decision-making, the controller must, in addition to telling the data subject that they are engaging in this type of activity:

a) 'provide meaningful information about the logic involved; and

b) explain the significance and envisaged consequences of the processing.'[245]

Note that these requirements only apply if the processing is solely automated within the meaning of Article 22(1) and has legal or similar effects to the data subject.[246] However, even if the decision does not meet the criterion of Article 22(1) 'it is nevertheless good practice to provide the above information.'[247]

Could it be that these provisions grant data subjects a right to an explanation? In the following Chapters 4.2.2.1 - 4.2.2.3, I will perform systematic interpretation of the requirements of the provisions in order to find out.

4.2.2.1 Timing and Scope

As referred to in Chapters 3.3.4 the disclosure of information pursuant to Article 13-14 of the GDPR, including the information of Articles 13(2)(f) and 14(2)(g), should occur *before* a decision is made, that is, generally at the time of data collection. In other words, this is an *ex ante* explanation.[248] Therefore, any explanations provided pursuant to Articles 13(2)(f) and 14(2)(g) can, logically, only be about the *system functionality* as providing decisions on *specific decisions* is impossible prior to decision making.[249] 'Notably this cannot include any information about how a specific decision was made or reached, but rather addresses how the system itself functions, eg its decision tree or rules, or predictions about how inputs will be processed.' [250]

---

[245] Article 29 Data Protection Working Party, 'Guidelines on Automated individual decision-making and Profiling for the purposes of Regulation 2016/679' (WP 241, 3 October 2017), 25.
[246] See Chapter 3.5.
[247] Article 29 Data Protection Working Party, 'Guidelines on Automated individual decision-making and Profiling for the purposes of Regulation 2016/679' (WP 241, 3 October 2017), 25.
[248] GDPR, recitals 60-62.
[249] See Table 6.
[250] Sandra Wacther, Brent Mittelstadt and Luciano Floridi, '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, 76, 83.

4.2.2.2 Meaningful Information

The information provided pursuant to Articles 13(2)(f) and 14(2)(g) should be meaningful. 'The growth and complexity of machine-learning can make it challenging to understand how an automated decision-making process or profiling works.'[251] For this reason the GDPR requires data controllers to find

> simple ways to tell the data subject about the rationale behind, or the criteria relied on in reaching the decision. The GDPR requires the controller to provide meaningful information about the logic involved, not necessarily a complex explanation of the algorithms used or disclosure of the full algorithm. The information provided should, however, be sufficiently comprehensive for the data subject to understand the reasons for the decision.'[252]

Furthermore, Recital 58 of the GDPR even states that

> … this is of particular relevance in situations where the proliferation of actors and the technological complexity of practice make it difficult for the data subject to know and understand whether, by whom and for what purpose personal data relating to him or her are being collected ...[253]

Therefore, complexity of the methods or technologies used for the processing and for reaching the decision are no excuse for failing to provide the information required. On the contrary, the complex nature of AI imposes a greater obligation for transparency in relation to the decision-making process. Therefore, the data controller must find a way to provide *meaningful information*. However, one might ask, what is meaningful in this context? What is meaningful for one person might not be meaningful for the next person:

> The machine learning and legal communities have both taken relatively restricted views on what passes for an explanation. The machine learning community has been primarily concerned with debugging and conveying approximations of algorithms that programmers or researchers could use to understand which features are important[254] while law and ethics scholars have been more concerned with understanding the internal logic of decisions as a means to assess their lawfulness (e.g. prevent discriminatory outcomes), contest them, increase accountability generally, and clarify liability.[255]

---

[251] Article 29 Data Protection Working Party, 'Guidelines on Automated individual decision-making and Profiling for the purposes of Regulation 2016/679' (WP 241, 3 October 2017), 25.
[252] Article 29 Data Protection Working Party, 'Guidelines on Automated individual decision-making and Profiling for the purposes of Regulation 2016/679' (WP 241, 3 October 2017), 25.
[253] GDPR, recital 58.
[254] Various authors, 'Workshop on Explainable AI (XAI) Proceedings' (Melbourne Australia 20 August 2017) <http://www.intelligentrobots.org/files/IJCAI2017/IJCAI-17_XAI_WS_Proceedings.pdf> accessed 12 May 2019.
[255] Sandra Wacther, Brent Mittelstadt and Chris Russel '*Counterfactual Explanations without Opening the Black Box: Automated Decisions and the GDPR*' (2018) 31[2] Harvard Journal of Law & Technology 842, 6-8.

Therefore, increased transparency and the provision of *meaningful information* for programmers or researchers in the field of artificial intelligence might not be so meaningful for a law scholar or the data subject. While providing the source code might be the most transparent method it would most likely not be meaningful for the data subject. Also, due to the ever-changing nature of machine learning algorithms, providing the source code might be difficult or even impossible. As soon as the decision has been made, the underlying logic might have already changed. Further, full transparency might risk the disclosure of trade secrets or other intellectual property rights.[256] Full disclosure could also result in the data subject being able to utilize the information to *game the system*.[257] Consequently, full transparency is perhaps not the most optimal solution, as in most scenarios, full disclosure does not provide the most meaningful solution for the for the data subject. The data controller should always have the data subject in mind when providing information pursuant to Articles 13 - 14. He must ask himself *what is meaningful in the eyes of the data subject*?

Interestingly, the word *meaningful* is a polysemous word. It means both *intended to show the meaning* and *serious, important, useful*.[258] Therefore, some scholars have argued that this polysemy should be of guidance when interpreting the term *meaningful* within the meaning of Articles 13-14, '… information about algorithmic decision-making should be both 'relevant, significant, important' and 'intended to show the meaning'. In other words, explanation about

---

[256] Andrew D. Selbst and Julia Powels, 'Meaningful Information and the Right to Explanation' (2017) 7[4] International Data Privacy Law 233, 242; Lilian Edwards, and Michael Veale, 'Slave to the Algorithm? Why a 'Right to an Explanation' Is Probably Not the Remedy You Are Looking For' (2017) 16 Duke Law & Technology Review 19; Dr. Maja Brkan, 'Do Algorithms Rule the World? Algorithmic Decision-Making In The Framework of the GDPR and Beyond' ('Terminator or the Jetsons? The Economics and Policy Implications of Artificial Intelligence' conference, Washington February 2018) 21; Sandra Wacther, Brent Mittelstadt and Chris Russel 'Counterfactual Explanations without Opening the Black Box: Automated Decisions and the GDPR' (2018) 31[2] Harvard Journal of Law & Technology 842, 875; Sandra Wachter, Brent Mittelstadt and Luciano Floridi, 'Why a Right to Explanation of Automated Decision-Making Does Not Exist in the General Data Protection Regulation' (2017) 7 International Data Privacy Law 76, 99; GDPR, recital 63.

[257] Lilian Edwards, and Michael Veale, 'Slave to the Algorithm? Why a 'Right to an Explanation' Is Probably Not the Remedy You Are Looking For' (2017) 16 Duke Law & Technology Review 63; Tal Zarsky, 'The Trouble with Algorithmic Decisions: An Analytic Road Map to Examine Efficiency and Fairness in Automated and Opaque Decision Making' (2016) 41[1] Science, Technology & Human Values 118,125; Sandra Wacther, Brent Mittelstadt and Chris Russel 'Counterfactual Explanations without Opening the Black Box: Automated Decisions and the GDPR' (2018) 31[2] Harvard Journal of Law & Technology 842, 846; Joshua A. Kroll and others, 'Accountable Algorithms' (2017) 165 University of Pennsylvania Law Review 633, 658; Sandra Wachter, Brent Mittelstadt and Luciano Floridi, 'Why a Right to Explanation of Automated Decision-Making Does Not Exist in the General Data Protection Regulation' (2017) 7 International Data Privacy Law 76, 88.

[258] Cambridge Dictionary, 'Meaningful' <https://dictionary.cambridge.org/dictionary/english/meaningful> (accessed 24 June 2017).

automated decisions should be both *complete* and *comprehensible*.'[259] This interpretation ensures both transparency and comprehensibility.

On the other hand, the purpose of the right to information is to 'provide a simple and generic overview of the intended processing activities that aims to inform a general audience.'[260] This fact is best crystallized in Article 12(7): [261]

> The information to be provided to data subjects pursuant to Articles 13 and 14 may be *provided in combination with standardised icons* in order to give in an easily visible, intelligible and clearly legible manner a *meaningful overview* of the intended processing. Where the icons are presented electronically they shall be machine-readable. (emphasis added)

The fact that this *meaningful overview* of information could be provided via standardized icons is evident of that disclosure does not need to be detailed or highly technical. Full transparency is impossible using only standardized icons. A simple overview or simple icons are unsuitable for explaining the complex and highly technical rationale behind an automated decision. Consequently, and to sum up, Articles 13-14 only requires the data controller to provide a simple (yet comprehensive) and generic (yet complete) overview of the artificial intelligence *systems functionality* and the envisaged consequences of decisions made by this system.

4.2.2.3 The Logic Involved and Envisaged Consequence

The phrase *logic involved*, also needs to be further analyzed. Scholars have proposed that this phrase should be understood in a double way, specifically: *logic involved* as

a) the system functionality; or
b) the rationale and circumstances of a specific decision.[262]

However, considering again the complex nature of artificial intelligence, this interpretation can be tricky as:

> … the 'logic involved' in an algorithm is a mathematical concept which can be 'explained' in pure IT terms, ie the technical process that takes from input A to output B. However, in order to be 'meaningful' or – rephrasing the tone

---

[259] Andrew D. Selbst and Julia Powels, 'Meaningful Information and the Right to Explanation' (2017) 7[4] International Data Privacy Law 233, 234.
[259] 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, 257.
[260] Sandra Wacther, Brent Mittelstadt and Chris Russel 'Counterfactual Explanations without Opening the Black Box: Automated Decisions and the GDPR' (2018) 31[2] Harvard Journal of Law & Technology 842, 886.
[261] GDPR, recital 60.
[262] Sandra Wacther, Brent Mittelstadt and Luciano Floridi, '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 76, 78.

> at Article 7(2) – 'intelligible and easily accessible', 'using clear and plain language' we need to move beyond the mere mathematical functionality of an algorithm and ascertain for its contextual use, expected and actual impact, rationales, purposes, etc. … Often these pieces of information are widely unknown even to designers or to data controllers that use these algorithms.[263]

Therefore, the data controller must find another way than explaining the logic in purely IT terms or disclosing the algorithm, something more is required to meet the requirements of the GDPR.

In order to provide the information about the *logic involved* and *envisaged consequences*, the data controller must first understand the logic and consequences of the system himself. Therefore, he should perform audits of his systems. This duty is explicitly mentioned in Recital 71 of the GDPR which states that data controllers shall use appropriate mathematical or statistical procedures for the profiling, implement technical and organizational measures appropriate to ensure, that risk of errors are minimized. Regular audits are recommended to comply with the accountability principle.[264]

Therefore, the audit must address the architecture (technical) of the algorithm and its contextual implications (organizational).[265] Following the audit, data controllers are able to provide information about the intended or future processing, and how the automated decision might affect the data subject. In order to make the information meaningful and understandable, real, tangible examples of the possible effects should be given. In this respect the following excerpt from the Council of Europe is interesting:

> Data subjects should be entitled to know the reasoning underlying the processing of their data, including the consequences of such a reasoning, which led to any resulting conclusions, in particular in cases involving the use of algorithms for automated decision-making including profiling. *For instance in the case of credit scoring, they should be entitled to know the logic underpinning the processing of their data and resulting in a 'yes' or 'no' decision, and not simply information on the decision itself.* Without an understanding of these elements there could be no effective exercise of other essential safeguards such as the right to object and the right to complain to a competent authority.[266]

---

[263] 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, 258.
[264] GDPR, art 5(2).
[265] GDPR, art 5(2).
[266] Council of Europe, 'Draft Explanatory Report on the convention for the protection of individuals with regard to automatic processing of personal data [ETS No. 108], para 75. <https://rm.coe.int/CoERMPublicCommonSearchServices/DisplayDCTMContent?documentId=09000016806b6ec2> accessed 1 May 2019.

This broad interpretation, however, cannot be derived from the legal text itself, an *ex post* explanation cannot provide accurate information on the rationale of a specific decision as previously argued. However, such a broad interpretation offers better protection to the fundamental rights and freedoms of the data subject as its makes it easier for the data subject to exercise its other rights, such as the right to objection,[267] erasure,[268] or the right to complain to a competent authority.[269]

Lastly, the Article 29 Working Party in its guidelines on automated individual decision making have recommended that data controller provides the following examples of information in order to meet the requirements of Articles 13(2)(f) and 14(2)(g), none of them being an *ex post* explanation:

- 'the categories of data that have been or will be used in the profiling or decision-making process;
- why these categories are considered pertinent
- how any profile used in the automated decision-making process is built, including any statistics used in the analysis;
- why this profile is relevant to the automated decision-making process; and
- how it is used for a decision concerning the data subject.'[270]

This is evident of the Article 29 Working Party being of the same opinion as described hereabove. That is, to meet Article 13-14 obligations, data controllers must only provide a simple (yet comprehensive) and generic (yet complete) overview of the relevant factors of the *system's functionality* that are of importance for intended audience rather than a complex explanation in regard to a *specific decision*. In other words, to Articles 13(2)(f) and 14(2)(g) include a right to an *ex ante* explanation of the *system's functionality*.

4.2.3 Right to access (Article 15)

The third legal basis for a right to explanation can be found in Article 15(1)h (right to access) of the GDPR which contains the identical wording as in Articles 13(2)f and 14(2)g (right to information), that is it gives the data subject a right to obtain information about:

[267] GDPR, art 21.
[268] GDPR, art 17.
[269] GDPR, art 77.
[270] Article 29 Data Protection Working Party, 'Guidelines on Automated individual decision-making and Profiling for the purposes of Regulation 2016/679' (WP 241, 3 October 2017), 31.

> 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.[271]

Both the right to access and the right to information provide the data subject a right to this information, however, the only practical difference between the two rights is the point in time when the data subject is entitled to obtain the information. The data controller's obligation to disclose information pursuant to Article 13-14 is generally triggered at the time of collection of the data, that is, before the processing. On the other hand, the *right to access* can logically only be invoked by the data subject after the processing has initiated. Therefore, '[i]n contrast to the notification duties of data controllers in Articles 13–14, the right of access has to be invoked by the data subject.'[272]

This difference, between the two articles, is important when exploring the right to an explanation and its categorization of *ex ante* and *ex post* explanations. Given that the phrasing of Article 15(2)h is identical to its counterparts in to Articles 13(2)(f) and 14(2)(g), one could assume that the right of access similarly only grants access to an *ex ante* explanation of *system functionality*.[273] However, as the data subject can request this information at any time, including after a decision has been made, Article 15 does not have the same *timeline problem* as to Articles 13(2)(f) and 14(2)(g). Therefore, an *ex post* explanation is at least possible.[274]

Nonetheless, even though such an explanation might be possible in theory, when considering the language of the provision, it is unlikely that Article 15(1)h provides such an *ex post* right. 'The phrase 'envisaged consequences' is future oriented, suggesting that the data controller must inform the data subject of possible consequences of the automated decision-making before such processing occurs.'[275] By being future oriented, the data controller is only

---

[271] GDPR, art 15(1)(h).
[272] Sandra Wacther, Brent Mittelstadt and Luciano Floridi, '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, 76, 84.
[273] Sandra Wacther, Brent Mittelstadt and Luciano Floridi, '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, 76, 84.
[274] Sandra Wacther, Brent Mittelstadt and Luciano Floridi, '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, 76, 84.
[275] Sandra Wacther, Brent Mittelstadt and Luciano Floridi, '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, 76, 84.

required to predict and disclose the possible consequences of the processing. [276] Consequently, the right is limited to an *ex ante* explanation.

Further, this is also supported by a comparison of the wording of Article 15(1)h and Recital 71:

Article 15(1)(h)	Recital 71
The data subject shall have … access to … the following information: (h) *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.	Which should include specific information to the data subject and the right to obtain human intervention, to express his or her point of view, *to obtain an explanation of the decision reached after such assessment and* to challenge the decision

TABLE 7

As stated above, the phrasing of Article 15(1)(h) is 'future oriented, and appears to refer to the existence and planned scope of decision-making itself.'[277] On the contrary, Recital 71 seems to refer to the circumstances of a specific decision, in other words an *ex post* explanation about a *specific decision.* If such a right were to be granted by Article 15(1)(h), why not use similar language as in Recital 71?

Lastly, Article 29 Working Party aligns with this interpretation. In its Guidelines on Automated individual decision making and Profiling, the working party states that data subjects are not entitled to any more information pursuant to Article 15 than they could have received by virtue of Articles 13-14.[278] This approach fatally damages 'the chances of generating a personalised ex post "right to an explanation" from art 15(h) without severe judicial disagreement with these guidelines.'[279] In conclusion, Article 15 only includes a right to an *ex ante* explanation of the *system's functionality*, in the same way as Articles 13(2)(f) and 14(2)(g).

[276] The German translation of the phrase (*angestrebten Auswirkungen*) further supports this. The phrase translates to *intended consequences.* The word *intended* being also future oriented and referring to something that has yet to take place.
[277] Sandra Wacther, Brent Mittelstadt and Luciano Floridi, '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, 76, 84.
[278] Article 29 Data Protection Working Party, 'Guidelines on Automated individual decision-making and Profiling for the purposes of Regulation 2016/679' (WP 241, 3 October 2017), 25.
[279] Michael Veale, Lilian Edwards, 'Clarity, surprises, and further questions in the Article 29 Working Party draft guidance on automated decision-making and profiling' (2018) 34 Computer Law & Security Review 398, 400.

In light of the foregoing, it is clear that a systematic analytic interpretation of a) Articles 13-14, b) Article 15 and c) Article 22 gives us an *ex ante* explanation of *system functionality* rather than an *ex post* explanation of a *specific decision*. However, some scholars have denied this restrictive interpretation and argued for a broader approach. They argue that the non-binding Recital 71 in conjunction with Articles 13-15 and 22 provide an *ex post* right to an explanation whereby data subjects can ask for an explanation of an algorithmic decision that was made about them.[280] In the following chapters I will explore the arguments of the opposing side.

4.3 Arguments for the Existence of an Ex Post Right to an Explanation

Although I align with the conclusion in the foregoing chapter, some scholars disagree[281] and argue for the existence of an *ex post* right, which opens the possibility of a right to an explanation of a *specific decision*. Their interpretation is supported by the following arguments.

Firstly, on the issue of timing. It is undisputed that to Articles 13(2)(f) and 14(2)(g) can only refer to the time before the decision is made (*ex ante*). However, in the case of Article 15, there is no deadline for accessing the data or information. 'Therefore ex post tailored knowledge about specific decisions made in relation to a particular data subject can be provided, i.e. "the logic or rationale, reasons, and individual circumstances of a specific automated decision." '[282] Followers of this interpretation have argued that this approach is promises an *ex post* right to an explanation and dismiss the importance of the future oriented phrasing of Article 15 as being *textual quibbles*.[283]

Secondly, some believe that 'Article 22, read in the light of Recital 71, in combination with Articles 13(2)(f), 14(2)(g) and 15(1)(h) GDPR, should be interpreted in a way that they give the data subject the right to an *ex post* explanation of the automated decision.'[284] This

---

[280] Bryce Goodman & Seth Flaxman, 'European Union regulations on algorithmic decision-making and a "right to explanation"' [2017] AI Magazine 50.
[281] Bryce Goodman & Seth Flaxman, 'European Union regulations on algorithmic decision-making and a "right to explanation"' [2017] AI Magazine 50; Isak Mendoza and Lee A. Bygrave, 'The Right not to be Subject to Automated Decisions based on Profiling' (2018) University of Oslo Faculty of Law Research Paper No. 2017-20; Michael Veale, Lilian Edwards, 'Clarity, surprises, and further questions in the Article 29 Working Party draft guidance on automated decision-making and profiling' (2018) 34 Computer Law & Security Review 398;
[282] Lilian Edwards, and Michael Veale, 'Slave to the Algorithm? Why a 'Right to an Explanation' Is Probably Not the Remedy You Are Looking For' (2017) 16 Duke Law & Technology Review 51.
[283] Lilian Edwards, and Michael Veale, 'Slave to the Algorithm? Why a 'Right to an Explanation' Is Probably Not the Remedy You Are Looking For' (2017) 16 Duke Law & Technology Review 52.
[284] Dr. Maja Brkan, 'Do Algorithms Rule the World? Algorithmic Decision-Making In The Framework of the GDPR and Beyond' ('Terminator or the Jetsons? The Economics and Policy Implications of Artificial Intelligence' conference, Washington February 2018) 15.

methodical grouping of different data protection provisions to provide the data subject with rights that are not specifically provided in the text of the data protection legislation itself, is not unusual in European jurisprudence. In *Google Spain*[285] the European Court of Justice relied on Article 12(b) on the right to access and Article 14(1)(a) on the right to object, of Directive 95/46[286] in order to construct the right to be forgotten, which had never been applied in earlier jurisprudence.[287] They argue that there should be no reason why the European Court of Justice would not apply a similar approach when interpreting the relevant provisions of the GDPR and read them together in order to construct a right to explanation.

Thirdly, they point out that Article 22(3) of the GDPR gives the data subject a right to contest an automated decision. In the absence of a right to explanation, a data subject's right to consent a decision would be entirely ineffective. [288]

> The term 'contest' connotes more than 'object to' or 'oppose'; in other words, a right of contest is not simply a matter of being able to say 'stop' but is akin to a right of appeal. If such a right is to be meaningful, it must set in train certain obligations for the decision maker, including (at the very least) an obligation to hear and consider the merits of the appeal.[289]

For this reason, a data subject should be able to at least get information about the input data and a reasonable explanation about why a certain decision was taken.[290] 'The right to contest is a right relating to substance of the decision and it would be an empty shell if the data subject was faced merely with a final decision without any explanation relating to it.'[291]

Fourthly, they highlight that, safeguards referred to in Article 22(3) are not the only possible safeguards. Article 22(3) states:

> the data controller shall implement suitable measures to safeguard the data subject's rights and freedoms and legitimate interests, *at least* the right to

---

[285] C-131/12 Google Spain SL and Google Inc. v Agencia Española de Protección de Datos (AEPD) and Mario Costeja González [2014] EU:C:2014:317.
[286] Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data [1995] OJ L 281/31.
[287] This right can now be found in Article 17 of the GDPR.
[288] Dr. Maja Brkan, 'Do Algorithms Rule the World? Algorithmic Decision-Making In The Framework of the GDPR and Beyond' ('Terminator or the Jetsons? The Economics and Policy Implications of Artificial Intelligence' conference, Washington February 2018) 15.
[289] Isak Mendoza and Lee A. Bygrave, 'The Right not to be Subject to Automated Decisions based on Profiling' (2018) University of Oslo Faculty of Law Research Paper No. 2017-20, 16
[290] Independent High-Level Expert Group on Artificial Intelligence - Set Up By The European Commission, 'Ethic Guidelines For Trustworthy AI' (European Commission 2018) 13.
[291] Dr. Maja Brkan, 'Do Algorithms Rule the World? Algorithmic Decision-Making In The Framework of the GDPR and Beyond' ('Terminator or the Jetsons? The Economics and Policy Implications of Artificial Intelligence' conference, Washington February 2018) 15.

> obtain human intervention on the part of the controller, to express his or her point of view and to contest the decision.

Article 22(3) provides data subjects with *at least* those safeguards. The provision does not limit the possibility of adding more safeguards, including the right to an explanation from Recital 71 being added though judicial interpretation. This would not be contrary to the clear wording of the provision. 'It seems that the wording was explicitly left open not to entirely preclude a possibility of judicial interpretation giving the data subject the right to explanation.'[292]

To summarize, the followers of an *ex post* right to an explanation argue that even though such a right cannot be found directly in the operative text of the GDPR, a less restrictive interpretation is required to ensure the proper protection of the right of the data subject. Further, such a right is not in contradiction to the language of the GDPR and therefore it should be interpreted in this way. Consequently, they argue that an *ex post* right to an explanation of both the *system functionality* and a *specific decision* exists within the GDPR and that his right is essential for data subjects to contest automated decisions.

5 Possible Solutions

In line with the conclusion of the foregoing Chapter, data subjects are at least entitled to meaningful information about the logic involved in automated decision making. In the realm of computer science, researchers have been seeking evolving techniques to make explainable artificial intelligence (XAI) since the 1970's.[293] This field of computer science is constantly growing 'as AI becomes more ubiquitous, complex and consequential'[294] and 'the need for people to understand how decisions are made and to judge their correctness becomes increasingly crucial due to concerns of ethics and trust.'[295] In this Chapter I will introduce two methods of explanations which aim to give meaningful information to the data subject in line with the GDPR's requirements. First, I will explore *subject-centric explanations*. Next, I will discuss *counterfactual explanations.*

---

[292] Dr. Maja Brkan, 'Do Algorithms Rule the World? Algorithmic Decision-Making In The Framework of the GDPR and Beyond' ('Terminator or the Jetsons? The Economics and Policy Implications of Artificial Intelligence' conference, Washington February 2018) 17.

[293] Lawrence M. Fagan, Edward H. Shortliffe and Bruce G. Buchanan, 'Computer-Based Medical Decision Making: From MYCIN to VM' (2010) 242

[294] IJCAI 'IJCAI 2019 Workshop on explainable Artificial Intelligence (XAI)' (*IJCAI,* 2019) <https://sites.google.com/view/xai2019/home> accessed 12 May 2019.

[295] IJCAI 'IJCAI 2019 Workshop on explainable Artificial Intelligence (XAI)' (*IJCAI,* 2019) <https://sites.google.com/view/xai2019/home> accessed 12 May 2019.

5.1 Subject-Centric Explanations

First, the idea of subject-centric explanations (SCEs). SECs focus on particular regions of a model around a query and show a lot of promise for interactive exploration, 'as do explanation systems based on learning a model from outside rather than taking it apart.' In other words, SCEs explain only the relevant parts of the model, one at a time, for a more meaningful explanation.[296] In theory, it is possible to give this kind of explanation both before[297] and after a decision is made. Therefore, it can be used to provide both an *ex ante* and *ex post* explanation. In order to understand this method of explanations, let us first consider a concept from computer science, *the curse of dimensionality*: [298]

> Data can be thought of geometrically: with two numeric variables, you can display all data on a two-dimensional scatter plot. With three variables, a three-dimensional one. Conceptually, you can scale this up to however many variables you have in your data. As you increase the dimensions (i.e., the number of variables) the number of ways that all potential values of them can be combined grows exponentially. It is this dynamic which makes the data especially complex to comprehend.[299]

It is easy for the human eye to comprehend two-dimensional scatter plots (X and Y axis), and even three-dimensional (X, Y and Z axis). Any more, and data controllers start having problems displaying the data and humans comprehending it through visualization.

In line with the foregoing, decision making models can be explained via scatter plots. The following Figure 4 is a simple example of a model with two variables for deciding whether a loan application should be accepted. For the sake of argument, I will keep the sample algorithm simple. The underlying logic will be as follows:

> If the applicant's income is equal or higher than its expenses, then the algorithm shall accept the application.
>
> However, if the applicant's income is lower than its expenses, then the algorithm shall reject the application.

The Y-axis representing the income of the applicant and X-axis representing its expenses and the blue points representing applicants (hypothetical or historical). The green triangle represents the boundary where applicants are accepted or rejected. When inside the boundary, the applicant will be accepted for the loan. When outside the boundary, the applicant will be

---

[296] Edwards, Lilian and Veale, Michael, 'Slave to the Algorithm? Why a 'Right to an Explanation' Is Probably Not the Remedy You Are Looking For' (2017) 16 Duke Law & Technology Review 56.
[297] If the explanation is to be provided beforehand, the explanation would use historic or hypothetical input values.
[298] Edwards, Lilian and Veale, Michael, 'Slave to the Algorithm? Why a 'Right to an Explanation' Is Probably Not the Remedy You Are Looking For' (2017) 16 Duke Law & Technology Review 56.
[299] Edwards, Lilian and Veale, Michael, 'Slave to the Algorithm? Why a 'Right to an Explanation' Is Probably Not the Remedy You Are Looking For' (2017) 16 Duke Law & Technology Review 56 – 57.

rejected. Getting this kind of visual explanation can really provide meaningful information to the data subject.

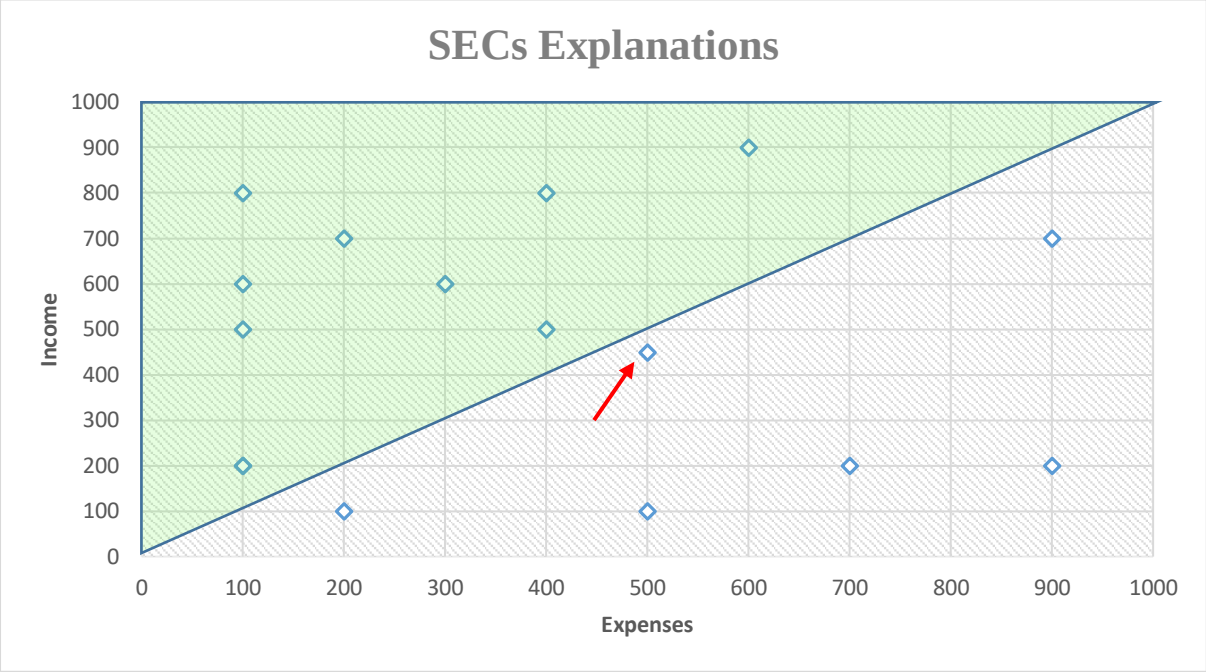

FIGURE 4

Now let us consider the applicant represented by the blue point highlighted by the red arrow. By receiving this SECs explanation, the applicant knows exactly where he stands. He knows that he is required to do in order to get accepted. In this case he has two options, he could either increase their income or reduce their expenses. By receiving this simple scatter plot, the data subject gets information about the factors that influenced the decision as well as an insight into the importance or weight of those factors. With this knowledge, the data subject will easily be able to spot any inaccurate information in relation to him which enables him to object or contest the decision. This is indeed meaningful information.

SCE's are far from being a perfect solution. They work well when the system only uses a 'few input variables that are combined in relatively straightforward ways, such as increasing or decreasing relationships.'[300] While it may be easy to break down and visualize such simple models it gets more complex with increasing variables.

> LinkedIn, for example, claims to have over 100,000 variables held on every user that feed into ML modelling. Many of these will not be clear variables like "age," but more abstract ways you interact with the webpage, such as how long you take to click, the time you spend reading, or even text you type

[300] Edwards, Lilian and Veale, Michael, 'Slave to the Algorithm? Why a 'Right to an Explanation' Is Probably Not the Remedy You Are Looking For' (2017) 16 Duke Law & Technology Review 59.

> in a text box but later delete without posting. These variables may well hold predictive signals about individual characteristics or behaviours, but we lack compelling ways to clearly display these explanations for meaningful human interpretation.[301]

How can data controllers possibly provide meaningful explanations for decisions based on vast amount of data like this? 'Do we even possess the mental vocabulary of categories and concepts to grasp the important aspects in the data?[302]

5.2 Counterfactuals

In this chapter I will introduce counterfactual explanations as a way to provide meaningful information of the logic of automated decisions. Using counterfactuals for this purpose was originally proposed in the paper Counterfactuals Explanations Without Opening the Black Box: Automated Decisions and the GDPR.[303]

The goal of counterfactual explanations is to enable the data controller to provide meaningful information to the data subject by giving them a helpful insight into the internal decision-making process of the algorithm rather than disclosing the algorithm itself:

> The machine learning and legal communities have both taken relatively restricted views on what passes for an explanation. The machine learning community has been primarily concerned with debugging and conveying approximations of algorithms that programmers or researchers could use to understand which features are important while law and ethics scholars have been more concerned with understanding the internal logic of decisions as a means to assess their lawfulness (e.g. prevent discriminatory outcomes), contest them, increase accountability generally, and clarify liability.[304]

Therefore, counterfactuals, as explanations, attempt to provide something new that lies outside of the previous taxonomies of explanations in the machine learning and legal literature. But what are counterfactual explanations?

---

[301] Edwards, Lilian and Veale, Michael, 'Slave to the Algorithm? Why a 'Right to an Explanation' Is Probably Not the Remedy You Are Looking For' (2017) 16 Duke Law & Technology Review 59-60.
[302] Edwards, Lilian and Veale, Michael, 'Slave to the Algorithm? Why a 'Right to an Explanation' Is Probably Not the Remedy You Are Looking For' (2017) 16 Duke Law & Technology Review 59-60
[303] Sandra Wacther, Brent Mittelstadt and Chris Russel 'Counterfactual Explanations without Opening the Black Box: Automated Decisions and the GDPR' (2018) 31[2] Harvard Journal of Law & Technology 842.
[304] Sandra Wacther, Brent Mittelstadt and Chris Russel 'Counterfactual Explanations without Opening the Black Box: Automated Decisions and the GDPR' (2018) 31[2] Harvard Journal of Law & Technology 842, 848-849.

In simple terms, counterfactual explanations can be provided in a form of statement followed by a counterfactual, that is, a statement describing how the data subject would have to act or behave differently to reach the desired outcome.[305] Here is an example:

> You were denied a loan because your annual income was € 40,000. If your income had been € 45,000, you would have been offered a loan.[306]

Multiple counterfactuals are possible, as a decision is usually made from multiple variables that affect the outcome. Therefore, there may be several ways for the data subject to achieve any of the desirable outcomes.

> The concept of the 'closest possible world,' or the smallest change to the world that can be made to obtain a desirable outcome, is key throughout the discussion of counterfactuals. In many situations, providing several explanations covering a range of diverse counterfactuals corresponding to relevant or informative "close possible worlds" rather than "the closest possible world" may be more helpful. Knowing the smallest possible change to a variable or set of variables to arrive at a different outcome may not always be the most helpful type of counterfactual. Rather, relevance will depend also upon other case-specific factors, such as the mutability of a variable or real world probability of a change. [307]

Therefore, going back to the example of a loan application, the data controller could have given a different explanation simultaneously:

> You were denied a loan because your annual expenses were € 45,000. If your expenses had been € 40,000, you would have been offered a loan.

One of the greatest advantages of counterfactuals is that they 'bypass the substantial challenge of explaining the internal workings of complex machine learning systems.'[308] Those kinds of highly complex and technical explanations usually generate little to no value for the data subject and should be avoided. 'In contrast, counterfactuals provide information to the data subject that is both easily digestible and practically useful for understanding the reasons for a decision, challenging them, and altering future behaviour for a better result.'[309] In addition,

---

[305] Sandra Wacther, Brent Mittelstadt and Chris Russel 'Counterfactual Explanations without Opening the Black Box: Automated Decisions and the GDPR' (2018) 31[2] Harvard Journal of Law & Technology 842, 847-849.
[306] Sounds familiar? Essentially, counterfactual explanations are a lot similar to the SECs solution introduced by Edwards & Veale.
[307] Sandra Wacther, Brent Mittelstadt and Chris Russel 'Counterfactual Explanations without Opening the Black Box: Automated Decisions and the GDPR' (2018) 31[2] Harvard Journal of Law & Technology 842, 847-849.
[308] Sandra Wacther, Brent Mittelstadt and Chris Russel 'Counterfactual Explanations without Opening the Black Box: Automated Decisions and the GDPR' (2018) 31[2] Harvard Journal of Law & Technology 842, 864.
[309] Sandra Wacther, Brent Mittelstadt and Chris Russel 'Counterfactual Explanations without Opening the Black Box: Automated Decisions and the GDPR' (2018) 31[2] Harvard Journal of Law & Technology 842, 864.

from a technical standpoint, counterfactuals are easy to implement.[310] Therefore, they are 'an easy first step that balances transparency, explainability, and accountability with other interests such as minimising the regulatory burden on business interest or preserving the privacy of others, while potentially increasing public acceptance of automatic decisions.' [311]

Obviously, both the counterfactual explanation and the SECs are a minimal form of explanation and might not be appropriate in all scenarios.[312] However, both methods are an interesting and promising alternative in the way of explaining automated decisions. But do these kinds of explanations provide the meaningful information required by the GDPR? In the following Chapter I will explore if that is the case.

5.3 Do the Proposed Solutions Meet the Requirements of the GDPR?

5.3.1 Meaningful Information About the Logic Involved

As previously discussed, in the cases of automated decisions the GDPR requires that the data controller discloses to the data subject meaningful information about the logic involved in the process, in a way that is both complete and comprehensible to the data subject.

Neither the counterfactual explanations nor the SECs provide any technical information about the system functionality of the algorithm. However, they provide valuable information to the data subject in relation to the underlying variables used in the decision-making process. By understanding the correlation and the relationship between the variables and how the smallest change in the value of the variables will affect the outcome, the data subject actually gains useful insight into the system's functionality. One could even argue that this information provides more complete and comprehensible information to the data subject than a generic technological overview of the system functionality, or even full disclosure of the algorithm itself, which might only be comprehensible to high level computer scientists or engineers.

5.3.2 Envisaged Consequences

The GDPR requires that the data controller discloses understandable, real, tangible examples of the possible effects of the decision. Additionally, followers of an *ex post* right to explanation

---

[310] Sandra Wacther, Brent Mittelstadt and Chris Russel 'Counterfactual Explanations without Opening the Black Box: Automated Decisions and the GDPR' (2018) 31[2] Harvard Journal of Law & Technology 842, 885.
[311] Sandra Wacther, Brent Mittelstadt and Chris Russel 'Counterfactual Explanations without Opening the Black Box: Automated Decisions and the GDPR' (2018) 31[2] Harvard Journal of Law & Technology 842, 889.
[312] Sandra Wacther, Brent Mittelstadt and Chris Russel 'Counterfactual Explanations without Opening the Black Box: Automated Decisions and the GDPR' (2018) 31[2] Harvard Journal of Law & Technology 842, 888.

argue that the GDPR also requires disclosure of information about the rationale and circumstances of a *specific decision*. This is exactly what both counterfactuals and SECs are designed to achieve. They provide real examples of how the smallest change in a single variable could affects the decision.

Theoretically it is possible to give these explanations both 'before or after a "decision" as discussed in the sense of data protection, if access to the model is provided.'[313] Firstly, the data controller can use hypothetical examples[314] before the *real* processing occurs to provide the data subject with explanation *ex ante*. This provides the data subject with highly valuable information about the *system functionality*. Additionally, if the hypothetical variables are similar or identical to the data subject's variables, they also provide information about a *specific decision.* Secondly, the data controller can use the real data, after the processing happens to provide an *ex post* explanation which contains both information about the *specific decision* and the *system functionality*. This important fact makes it so that an *ex ante* explanation of a *specific decision* is possible with these explanations. Therefore, they are suitable to provide the required information about the envisaged consequences of the processing.

5.3.3 Easily Understandable

Providing a single counterfactual explanation or SCE in a concise transparent intelligible and easily accessible form, using clear and plain language is a fairly easy task. However, this task quickly becomes more complex when an algorithm depends on multiple variables which are all connected.[315] Even though the basic idea behind these kind of explanations is to give the data subject the knowledge about the smallest required changes to reach the desirable outcome, it is possible to imagine a scenario where the smallest required change depends on the change of multiple variables. In those scenarios, the data controller would have to provide multiple instances of explanations. This might result in confusion and an overwhelming instance of explanations, difficult for the data subject to comprehend, an information overflow. Therefore, I argue that the counterfactual explanations and SCEs provide an excellent solution in the case of simple algorithmic decision. However, they might not be the option in more complex decisions relying on many variables.

---

[313] Lilian Edwards, and Michael Veale, 'Slave to the Algorithm? Why a 'Right to an Explanation' Is Probably Not the Remedy You Are Looking For' (2017) 16 Duke Law & Technology Review 56.
[314] Alternatively, he could use the training data.
[315] See the example of LinkedIn on page 59.

5.4 Ethical Requirements to Explainability

All of the discussion has been focused on the right to explanation within the GDPR. However, the subject of the transparency and explainability of artificial intelligence is not only a subject of data protection, nor is it only a topic for legal scholars. 'Given the scale of the challenge associated with AI, the full participation of all actors including businesses, academics, consumer organisations, trade unions, policy makers and representatives of civil society is essential.'[316] For this reason, the European Commission has tasked the AI HLEG to implement the European Union's strategy on AI.[317] The AI HLEG has highlighted the importance of accountability, transparency and explainability of automated decisions.

On 8 April 2019 the AI-HLEG published its Ethic Guidelines for Trustworthy AI.[318] In the guidelines, explicability of artificial intelligence is said to be crucial for building and maintaining users' trust in artificial intelligence systems. They propose that the processes need to be transparent, the capabilities of AI systems disclosed, 'and decisions – to the extent possible – explainable to those directly and indirectly affected. Without such information, a decision cannot be duly contested.' [319] They admit that 'an explanation as to why a model has generated a particular output or decision (and what combination of input factors contributed to that) is not always possible.' But highlight that in these so called "black box" cases, other explicability measures should be implemented to ensure protection of the fundamental rights of the individual. Further on the topic of explainability, the High-Level Expert Group states that 'whenever an AI system has a significant impact on people's lives, it should be possible to demand a suitable explanation of the AI system's decision-making process.'[320] Artificial intelligence users should therefore 'establish mechanisms to inform (end-)users on the reasons and criteria behind the AI system's outcomes'[321] and 'ensure explanation as to why the system

[316] European Commission 'Call for a High-Level Expert Group on Artificial Intelligence' (*European Commission,* 9 March 2018) <https://ec.europa.eu/digital-single-market/en/news/call-high-level-expert-group-artificial-intelligence> accessed 11 May 2019.
[317] European Commission 'Call for a High-Level Expert Group on Artificial Intelligence' (*European Commission,* 9 March 2018) <https://ec.europa.eu/digital-single-market/en/news/call-high-level-expert-group-artificial-intelligence> accessed 11 May 2019.
[318] Independent High-Level Expert Group on Artificial Intelligence - Set Up By The European Commission, 'Ethic Guidelines For Trustworthy AI' (European Commission 2018).
[319] Independent High-Level Expert Group on Artificial Intelligence - Set Up By The European Commission, 'Ethic Guidelines For Trustworthy AI' (European Commission 2018) 13.
[320] Independent High-Level Expert Group on Artificial Intelligence - Set Up By The European Commission, 'Ethic Guidelines For Trustworthy AI' (European Commission 2018) 18.
[321] Independent High-Level Expert Group on Artificial Intelligence - Set Up By The European Commission, 'Ethic Guidelines For Trustworthy AI' (European Commission 2018) 29.

took a certain choice resulting in a certain outcome that all users can understand?'[322] These disclosure requirements, provided by the guidelines seem similar (but not identical) to the requirements of the GDPR.

It is out of the scope of this thesis to compare the requirements suggested by the AI HLEG to the disclosure requirements of the GDPR any further. However, the interest of this highly political body and the importance the AI HLEG has given to the explainability of artificial intelligence is not without value. This is evident of how important it is for data controllers to be able to explain their automated decisions in order to build and maintain ethical[323] and trustworthy[324] artificial intelligence systems, regardless if providing such an explanation is required by the GDPR.

6 Conclusion

In this thesis I have explored the functions of artificial intelligence systems in order to highlight some of the inherent data protection issues and risks often found in such systems, mainly biases, inaccuracies and errors relating to the opaqueness of the systems. Scholars have pointed out that the GDPR includes a safeguard against these risks, in the form of a right to an explanation of automated decisions.

I have analyzed the scope and extent of this right by performing a systematic interpretation of Articles 13 – 14, Article 15 and Article 22 of the GDPR. By performing this exercise, I have noted that this obligation is only triggered whenever the decision is based solely on automated processing and produces legal effects concerning the data subject or similarly significantly affects the data subject. I have also noted that the explanation must be provided in a meaningful and understandable way for the data subject. The explanation must include a simple (yet comprehensive) and generic (yet complete) overview of the relevant factors of the *system's functionality* that are of importance for intended audience. In conclusion,

---

[322] Independent High-Level Expert Group on Artificial Intelligence - Set Up By The European Commission, 'Ethic Guidelines For Trustworthy AI' (European Commission 2018) 29.

[323] Independent High-Level Expert Group on Artificial Intelligence - Set Up By The European Commission, 'Ethic Guidelines For Trustworthy AI' (European Commission 2018) 11-13, the guidelines list 'four ethical principles, rooted in fundamental rights, which must be respected in order to ensure that AI systems are developed, deployed and used in a trustworthy manner. They are specified as ethical imperatives, such that AI practitioners should always strive to adhere to them.' The principle of *explicability* being one of those four ethical principles.

[324] Independent High-Level Expert Group on Artificial Intelligence - Set Up By The European Commission, 'Ethic Guidelines For Trustworthy AI' (European Commission 2018) 14, the guidelines list seven requirements in order to implement and realize trustworthy AI, *transparency* (including explainability) being one of those seven requirements.

data controllers shall provide an *ex ante* explanation of the *system functionality* whenever the processing falls within the scope of Article 22(1) of the GDPR.

I have also assessed how data controllers can comply with this obligation by introducing two methods of explaining automated decision making in a meaningful way, *SECs* and *counterfactual explanations*. Both methods are promising and could provide a possible solution for data controllers in order to comply with the requirements of the GDPR. However, they are not perfect for every situation. There is no *one-size-fits-all* solution when it comes to providing meaningful information. What is meaningful in the eyes of one person might not be meaningful in the eyes of another. Therefore, data controllers have to take consideration to the data subject and the nature of the processing when they choose *how* they provide explanations.

Lastly, due to the rapid technological advances of artificial intelligence systems in the last sixty years, such systems are getting more powerful and more popular than ever. For this reason, researchers and political institutions are calling for increased transparency and explainability of artificial intelligence. Transparency and explainability are becoming cornerstones of trustworthiness and accountability of artificial intelligence. Therefore, in order to maintain ethical, accountable and trustworthy artificial intelligence systems, data controllers might want to consider explaining their automated decisions, regardless if providing such an explanation is required by the GDPR.

The greater the power, the more need there is for transparency,
because if the power is abused, the result can be so enormous.

- Julian Assange

Bibliography

Angwin J and others, ‘Machine Bias: There’s software used across the country to predict future criminals. And it’s biased against blacks’ (ProPublica, 23 May 2016) <https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing> accessed 11 May 2019

Article 29 Data Protection Working Party, ‘Guidelines on Automated individual decision-making and Profiling for the purposes of Regulation 2016/679’ (WP 241, 3 October 2017)

Article 29 Data Protection Working Party, ‘Guidelines on Personal data breach notification under regulation 2016/679’ (WP 250rev.01, 6 February 2018)

Article 29 Data Protection Working Party, ‘Guidelines on transparency under Regulation 2016/679’ (WP 260, 11 April 2018)

Article 29 Data Protection Working Party, ‘Opinion 03/2013 on purpose limitation’ (WP 203, 2 April 2013)

Article 29 Data Protection Working Party, ‘Opinion 3/2010 on the principle of accountability’ (WP 3/2010)

Baratta R, ‘Complexity of EU Law in the Domestic Implementing Process’ (2014) 2 The Theory and Practice of Legislation, 293

Bartlett R.P and others, ‘Consumer Lending Discrimination in the FinTech Era’ (2017) UC Berkley Public Law Research Paper

Binns R, ‘Algorithmic Accountability and Public Reason’ (2017) 31[4] Philosophy & Technology 1

Bovens M, Schillemans T, and Goodin R.E, *The Oxford Handbook of Public Accountability* (Oxford University Press 2014)

Brkan M, ‘Do Algorithms Rule the World? Algorithmic Decision-Making In The Framework of the GDPR and Beyond’ (‘Terminator or the Jetsons? The Economics and Policy Implications of Artificial Intelligence’ conference, Washington February 2018)

Cambridge Dictionary, ‘Meaning of algorithm in English’ <https://dictionary.cambridge.org/us/dictionary/english/algorithm> accessed 11 May 2019

Cambridge Dictionary, 'Meaningful' <https://dictionary.cambridge.org/dictionary/english/meaningful> (accessed 24 June 2017).

Chabert J.L. and others, *A history of Algorithms: From the Pebble to the Microchip* (Springer 1999)

Council of Europe and others, *Handbook on European data protection law: 2018 edition,* (Publications Office of the European Union 2019)

Council of Europe, 'Draft Explanatory Report on the convention for the protection of individuals with regard to automatic processing of personal data [ETS No. 108], para 75 <https://rm.coe.int/CoERMPublicCommonSearchServices/DisplayDCTMContent?documentId=09000016806b6ec2> accessed 1 May 2019

Danziger S, Levav J and Avnaim-Pesso L, 'Extraneous factors in judicial decisions' (2011) 108 PNAS 6889

Domingos P, *The Master Algorithm: How the Quest for the Ultimate Learning Machine Will Remake Our Worlds* (Basic Books 2015)

Edwards L, and Veale M, 'Slave to the Algorithm? Why a 'Right to an Explanation' Is Probably Not the Remedy You Are Looking For' (2017) 16 Duke Law & Technology Review

EUROPA, 'Guide to the Approximation of EU Environmental Legislation ANNEX I' (Environment, 2015) <http://ec.europa.eu/environment/archives/guide/annex1.htm> accessed 19 April 2019

European Commission 'Call for a High-Level Expert Group on Artificial Intelligence' (European Commission, 9 March 2018) <https://ec.europa.eu/digital-single-market/en/news/call-high-level-expert-group-artificial-intelligence> accessed 11 May 2019

European Commission, 'Regulation of the European Parliament and the Council on the Protection of Individuals with Regard to the Processing of Personal Data and on the Free Movement of Such Data (General Data Protection Regulation)' (European Commission 2012) 2012/0011 (COD) <http://ec.europa.eu/justice/data-protection/document/review2012/com_2012_11_en.pdf> accessed 19 April 2019

European Parliament Committee on Civil Liberties, Justice and Home Affairs, 'Report on the Proposal for a Regulation of the European Parliament and of the Council on the Protection of

Individuals with Regard to the Processing of Personal Data and on the Free Movement of Such Data (General Data Protection Regulation) - A7-0402/2013' (European Parliament, 2013) A7–0402/2013 <http://www.europarl.europa.eu/sides/getDoc.do?pubRef=-//EP//TEXT+REPORT+A7-2013-0402+0+DOC+XML+V0//EN> accessed 19 April 2019

Fagan M, Shortliffe E.H and Buchanan B.G, 'Computer-Based Medical Decision Making: From MYCIN to VM' (2010) 242

Goodman B and Flaxman S, 'European Union regulations on algorithmic decision-making and a "right to explanation"' [2017] AI Magazine 50

Griffin A, 'Facebook's Artificial Intelligence Robots Shut Down After They Start Talking To Each Other In Their Own Language' (Independent, 31 July 2017) <https://www.independent.co.uk/life-style/gadgets-and-tech/news/facebook-artificial-intelligence-ai-chatbot-new-language-research-openai-google-a7869706.html> accessed 11 May 2019

Have I been pwned, 'have i been pwned?' <https://haveibeenpwned.com/> accessed 1 May 2019

Holmes O.W. Jr., *The Common Law* (first published 1881, Paulo J. S. Pereira & Diego M. Beltran 2011)

IJCAI 'IJCAI 2019 Workshop on explainable Artificial Intelligence (XAI)' (IJCAI, 2019) <https://sites.google.com/view/xai2019/home> accessed 12 May 2019

Independent High-Level Expert Group on Artificial Intelligence - Set Up By The European Commission, 'A Definition of AI: Main Capabilities And Disciplines' (European Commission 2018)

Independent High-Level Expert Group on Artificial Intelligence - Set Up By The European Commission, 'Ethic Guidelines For Trustworthy AI' (European Commission 2018)

Information Commissioner's Office 'Big data, artificial intelligence, machine learning and data protection' (Information Commissioner's Office, 2017) 7 <https://ico.org.uk/media/for-organisations/documents/2013559/big-data-ai-ml-and-data-protection.pdf> accessed 11 May 2019

Information Commissioner's Office, 'Guide to the General Data Protection Regulation (GDPR): The principles' (Information Commissioner's Office, December 2018) <https://ico.org.uk/for-organisations/guide-to-the-general-data-protection-regulation-gdpr/principles/> accessed 11 May 2019

Information Commissioner's Office, 'Guide to the General Data Protection Regulation (GDPR): Principle (a): Lawfulness, fairness and transparency' (Information Commissioner's Office, December 2018) <https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/principles/lawfulness-fairness-and-transparency/> accessed 11 May 2019

Information Commissioner's Office, 'Guide to the General Data Protection Regulation (GDPR): Principle (b): Purpose limitation' (Information Commissioner's Office, December 2018) <https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/principles/purpose-limitation/> accessed 11 May 2019

Information Commissioner's Office, 'Guide to the General Data Protection Regulation (GDPR): Principle (b): Purpose limitation' (Information Commissioner's Office, December 2018) <https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/principles/purpose-limitation/> accessed 11 May 2019

Information Commissioner's Office, 'Guide to the General Data Protection Regulation (GDPR): Principle (c): Data minimisation' (Information Commissioner's Office, December 2018) <https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/principles/data-minimisation/> accessed 11 May 2019

Information Commissioner's Office, 'Guide to the General Data Protection Regulation (GDPR): Principle (e): Storage limitation' (Information Commissioner's Office, December 2018) <https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/principles/storage-limitation/> accessed 11 May 2019

Information Commissioner's Office, 'Guide to the General Data Protection Regulation (GDPR): Principle (f): Integrity and confidentiality (security)' (Information Commissioner's Office, December 2018) <https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/principles/integrity-and-confidentiality-security/> accessed 11 May 2019

Information Commissioner's Office, 'Guide to the General Data Protection Regulation (GDPR): The right to be informed – When should we provide privacy information?' (Information Commissioner's Office, December 2018) <https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/the-right-to-be-informed/when-should-we-provide-privacy-information/> accessed 11 May 2019

Information Commissioner's Office, 'Guide to the General Data Protection Regulation (GDPR): The right to be informed – Are there any exceptions?' (Information Commissioner's Office, December 2018) <https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/the-right-to-be-informed/are-there-any-exceptions/> accessed 11 May 2019

Infosys, 'Amplifying Human Potential – Toward purposeful Artificial Intelligence: A Perspective for CIOs' (Infosys, 2017) <https://www.infosys.com/aimaturity/Documents/amplifying-human-potential-CIO-report.pdf> accessed 11 May 2019

Infosys, 'Amplifying Human Potential – Toward purposeful Artificial Intelligence' (Infosys, 2017) <https://www.infosys.com/aimaturity/Documents/amplifying-human-potential-CEO-report.pdf> accessed 11 May 2019

Jolls C, Sunstein C.R. and Thaler R, 'A Behavioral Approach to Law and Economics' (1988) 50 Stanford Law Review 1471

Kaplan A and Haenlein M, 'Siri, Siri, in my hand: Who's the fairest in the land? On the interpretations, illustrations, and implications of artificial intelligence' 62 (2019) Business Horizons 15

Knuth D, The Art of Computer Programming: fundamental algorithms (3rd edn. 1997)

Kozinski A, 'What I Ate for Breakfast and Other Mysteries of Judicial Decision Making' (1993) 26 Loyola LA L Rev 993

Kroll J.A and others, 'Accountable Algorithms' (2017) 165 University of Pennsylvania Law Review 633

Krutz R.L. and Vines R.D, *Cloud Security: A Comprehensive Guide to Secure Cloud Computing* (Wiley 2010)

Li D and Du Y, *Artificial Intelligence With Uncertainty* (2nd edn, CRC Press, Taylor & Francis Group 2017)

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

McCarthy J. and others 'A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence' (31 August 1955) <http://jmc.stanford.edu/articles/dartmouth/dartmouth.pdf> accessed 1 May 2019

McNamee J, 'Is Privacy Still Relevant in a World of Bastard data?' (Edri, 9 March 2016) <https://edri.org/enditorial-is-privacy-still-relevant-in-a-world-of-bastard-data> accessed 11 May 2019

Meadow W and Sunstein C.R, 'Statistics, Not Experts' (2001) 51 Duke Law Journal 629

Mendoza I and Bygrave L.A., 'The Right not to be Subject to Automated Decisions based on Profiling' (2018) University of Oslo Faculty of Law Research Paper No. 2017-20

Merriam-Webster Dictionary, 'Definition of algorithm' <https://www.merriam-webster.com/dictionary/algorithm> accessed 11 May 2019

Moore G.E., 'Cramming more components onto integrated circuits' (1965) 38 Electronics

Nissenbaum H, 'How computer systems embody values' (2001) 34[3] Computer 119

Panetta K, '5 Trends Emerge in the Gartner Hype Cycle for Emerging Technologies, 2018' (Smarter with Gartner, 16 August 2018) <https://www.gartner.com/smarterwithgartner/5-trends-emerge-in-gartner-hype-cycle-for-emerging-technologies-2018/> accessed 8 December 2018

Polli F, 'The Dark Side Of Artificial Intelligence' (Forbes, 5 December 2017) <https://www.forbes.com/sites/fridapolli/2017/12/05/the-dark-side-of-artificial-intelligence/#3fe68e4a1261> accessed 11 May 2019

Russel S and Norvig P, *Artificial Intelligence: A Modern Approach* (3rd edn, Prentice Hall 2010)

Selbst A.D. and Powels J, 'Meaningful Information and the Right to Explanation' (2017) 7[4] International Data Privacy Law, 233

Stanford University, ‘Machine Learning’ (Coursera, undated) <https://www.coursera.org/learn/machine-learning/home/info> accessed 11 May 2019

Takahasi D, ‘Forty years of Moore’s Law’ (The Seattle Times, 18 April 2005) <https://www.seattletimes.com/business/forty-years-of-moores-law/> accessed 11 May 2019

Techopedia, ‘Garbage In, Garbage Out (GIGO)’ <https://www.techopedia.com/definition/3801/garbage-in-garbage-out-gigo> accessed 11 May 2019

Tricentis, ‘AI Approaches Compared: Rule-Based Testing vs. Learning’ (Tricentis, undated) <https://www.tricentis.com/artificial-intelligence-software-testing/ai-approaches-rule-based-testing-vs-learning/> accessed 11 May 2019

Turing A.M, ‘Computing Machinery and Intelligence’ (1950) 49 Mind 433

Various authors, ‘Workshop on Explainable AI (XAI) Proceedings’ (Melbourne Australia 20 August 2017) <http://www.intelligentrobots.org/files/IJCAI2017/IJCAI-17_XAI_WS_Proceedings.pdf> accessed 12 May 2019

Veale M, Edwards L, ‘Clarity, surprises, and further questions in the Article 29 Working Party draft guidance on automated decision-making and profiling’ (2018) 34 Computer Law & Security Review 398

Vincent J, ‘Google “fixed” its racist algorithm by removing gorillas from its image-labeling tech: Nearly three years after the company was called out, it hasn’t gone beyond a quick workaround’ (The Verge, 12 January 2018) <https://www.theverge.com/2018/1/12/16882408/google-racist-gorillas-photo-recognition-algorithm-ai> accessed 11 May 2019

Wachter S, Mittelstadt B and Floridi L, ‘Why a Right to Explanation of Automated Decision-Making Does Not Exist in the General Data Protection Regulation’ (2017) 7 International Data Privacy Law 76

Wacther S, Mittelstadt B and Russel C, ‘Counterfactual Explanations without Opening the Black Box: Automated Decisions and the GDPR’ (2018) 31[2] Harvard Journal of Law & Technology 842

Zarsky T, ‘Incompatible: The GDPR in the Age of Big Data’ (2017) 47 Seton Hall Law Review 995

Zarsky T, ‘The Trouble with Algorithmic Decisions: An Analytic Road Map to Examine Efficiency and Fairness in Automated and Opaque Decision Making’ (2016) 41[1] Science, Technology & Human Values 118