UNIVERSITA' COMMERCIALE "LUIGI BOCCONI"

PhD SCHOOL

PhD program in Legal Studies

Cycle: 35th

Disciplinary Field (code): IUS/08 Constitutional law

The challenges of the General Data Protection Regulation to protect data subjects against the adverse effects of artificial intelligence

Advisor: Oreste POLLICINO

PhD Thesis by

Federico MARENGO

ID number: 3108406

Year 2023

Table of Contents

INTRODUCTION

Artificial intelligence (hereinafter AI) has brought many societal changes and is transforming the way to do business. AI systems are pushing the boundaries of machine capabilities, cutting down the time required to complete specific tasks, enabling the accomplishment of complex operations that exceed human capacity, and easing repetitive decision-making processes. In short, AI systems can provide a faster and cheaper method to solve everyday problems in various areas, for example, smart cities, fraud prevention, law enforcement, and autonomous driving. The opportunities offered by these technologies are countless. Yet, similar to what has happened with other innovations, AI solutions can have detrimental consequences for individuals and society. In particular, the processing of information using AI systems also entails risks to the rights and freedoms of individuals, such as the lack of respect for human autonomy, the production of material or moral harms, discrimination, and lack of transparency in the decision-making process.

Whereas these features reveal the potential of artificial intelligence to support most of the activities performed by individuals, it also triggers many crucial questions, in particular, regarding the adequacy and sufficiency of the EU legal framework on data protection to protect the rights and freedoms of individuals when their personal information is processed using AI systems.

This work attempts to discover the most critical challenges posed by the processing of personal data supported by AI systems, how the current European legal framework addresses these challenges, and it also proposes alternative pathways to better protect the rights of individuals without stifling innovation.

Since the entry into force of the General Data Protection Regulation (hereinafter GDPR) in 2018, there has been massive interest from researchers and industry stakeholders in data protection. The GDPR harmonised the regulatory landscape of data protection and introduced stringent requirements for the processing of personal data. Additionally, it is a technology-neutral regulation, which means that it was conceived to be flexible and adaptable to new technologies. It also included specific provisions concerning automated decision-making to provide more targeted protection against this particular way of processing.

When it comes to the interaction of AI and data protection, researchers dedicated much of their efforts to particular fields: the explainability of AI systems and the provisions concerning the right not to be subject to automated decisions established in the GDPR[1] and also concerning the fairness of the decisions taken using AI systems.[2]

However, while many works address both the challenges and solutions concerning the processing of personal data using AI systems, this work differs from others in two crucial aspects. First, most of the materials reviewed conceived AI systems in general without clearly explaining the concept of AI, the differences between the different AI systems or models, and how these differences impact the fundamental rights of individuals. This work additionally highlights the importance of not only evaluating the use cases of AI systems but also acknowledging that different AI systems pose various risks to individuals. Hence, a general understanding of the different methods, techniques, or approaches to AI systems is essential to further specify the risks posed

[1] See for instance, Antoni Roig, 'Safeguards for the Right Not to Be Subject to a Decision Based Solely on Automated Processing (Article 22 GDPR)' (2018) 8 European Journal of Law and Technology 1; Paul De Hert and Vagelis Papakonstantinou, 'The New General Data Protection Regulation: Still a Sound System for the Protection of Individuals?' (2016) 32 Computer Law & Security Review 179; Bryce Goodman and Seth Flaxman, 'European Union Regulations on Algorithmic Decision-Making and a "Right to Explanation"' [2017] *AI Magazine*; Andrew D Selbst and Julia Powles, 'Meaningful Information and the Right to Explanation' (2017) 7 International Data Privacy Law 233; Gianclaudio Malgieri and Giovanni Comandé, 'Why a Right to Legibility of Automated Decision-Making Exists in the General Data Protection Regulation' (2017) 7 International Data Privacy Law 243; 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; Maja Brkan, 'Do Algorithms Rule the World? Algorithmic Decision-Making and Data Protection in the Framework of the GDPR and Beyond' (2019) 27 International Journal of Law and Information Technology 91.

[2] See for instance, Michael Butterworth, 'The ICO and Artificial Intelligence: The Role of Fairness in the GDPR Framework' (2018) 34 Computer Law & Security Review 257; Danielle Keats Citron and Frank Pasquale, 'The Scored Society: Due Process for Automated Decisions' (2014) 89 Washington Law Review 1; Shea Brown, Jovana Davidovic and Ali Hasan, 'The Algorithm Audit: Scoring the Algorithms That Score Us' (2021) 8 Big Data and Society; Sandra Wachter, Brent Mittelstadt and Chris Russell, 'Bias Preservation in Machine Learning: The Legality of Fairness Metrics Under EU Non-Discrimination Law' (2021) 1 West Virginia Law Review 735; Philipp Hacker, 'Teaching Fairness to Artificial Intelligence: Existing and Novel Strategies Against Algorithmic Discrimination under EU Law' (2018) 55 Common Market Law Review 1143.

by these systems and to propose adequate and tailored measures according to the particular situation. Secondly, most reviewed materials provide highly detailed explanations about the legal concepts involved, present new interpretations, and offer recommendations about what should be done to improve individuals' general situation at the legislative level. However, there is in general a disconnection on how the legal concepts relate to operative aspects of data protection. For the adequate protection of individuals, the analysis cannot be only limited to legal texts and case law. It should also include non-binding documents such as guidelines, standards, and best practices that concretise and provide more detailed guidance on how the high-level principles stipulated in the laws and regulations governing the area should be translated into practical and operational requirements. This work attempts to bridge, on the one hand, the gap between the legislative and judicial interpretation and, on the other hand, the practical and operative aspects concerning the protection of personal data.

Against this backdrop, the general objective of this work is to evaluate the extent to which the processing of personal data using AI systems satisfies the requirements outlined in the GDPR. This work's central enquiry is: *how can individuals be better protected from the risks posed by artificial intelligence systems*?

To address this problem, the following questions will be explored:

What is artificial intelligence and how does it work? How can AI systems interfere with individuals' rights?

Is the GDPR suitable for regulating AI systems that process personal data?

How can be assured that AI solely uses data that is lawfully obtained, adequate, relevant, and limited to what is necessary for the purpose sought?

To what extent do decisions taken with AI systems satisfy the requirements of the GDPR? What is the nature of the right not to be subject to a decision based solely on automated means (art. 22 GDPR)? Under what circumstances is automated decision-making allowed and, in those cases, which are the safeguards data controllers must provide?

What are the legal issues stemming from algorithmic decision-making? Does Articles 13(2)(f), 14(2)(g), 15(1)(h), and 22(3) GDPR grant data subjects a right to explanation or only a right to information? How can organizations meaningfully inform data subjects about the logic involved, the significance, and the envisaged consequences of automated decisions without disclosing critical attributes of the

decisions or the processes by which the decisions were taken? Are the right to explanation and current explanation methods able to solve the legal issues that arise from automated decision-making of AI-powered decisions?

Is there any suitable mechanism to ensure that the risks posed by AI solutions are adequately identified and mitigated? How can data protection impact assessments (DPIA) help comply with the legal requirements? Which are other accountability mechanisms that can assist in protecting the rights of individuals?

Does the European data protection legal framework need to be updated to meet new technological developments? Would the AI Regulation (draft) contribute to protecting the fundamental rights of individuals? How can the limitations on the EU legislation be overcome?

This research involves doctrinal analysis, mainly arising from secondary sources of law. In particular, it primarily reviews international law, fundamental rights in the EU, privacy law, and law and technology journals. Then, it evaluates the abundant scholarly research about data protection and artificial intelligence, as well as relevant guidelines and standards issued by European institutions, national data protection authorities, and international standardisation organisations. The analysis of these materials will enable the identification of possible gaps in the literature and will serve as a possible analytical framework for addressing the specificities related to AI-powered processing of personal data.

This work is structured in 5 chapters as follows. Chapter I elaborates on the concept and importance of data, in particular personal data, for the development of AI systems and the conceptualization of artificial intelligence, considering its three most important techniques (machine learning, logic and knowledge-based approach, and statistical approaches). Chapter II explains how personal data is protected in Europe. Starting with an overview of the fundamental legal texts that govern the area, it then assesses more in detail the relevant provisions of the GDPR that has a bearing with regards to the protection of personal data where the processing is made using AI systems. In particular, it elaborates on the impact of AI systems on the data protection principles and the lawful basis to process personal data. Chapter III provides an in-depth appraisal of the protection of individual rights by the GDPR when personal data is processed using AI systems. Whereas a large part of this chapter is devoted to the rights related to automated decision-making (together with its conceptualization, the

exceptions and safeguards), the remaining rights listed by the GDPR are also evaluated. Particularly, rights derived from transparency obligations, the right to rectification, erasure, restriction, objection and portability are reviewed in the light of the challenges posed by the processing of personal data using AI systems. Furthermore, this chapter introduces the most important accountability mechanisms, which are also a crucial aspect regulated by the GDPR and essential to mitigating the risks posed by the processing of personal data using AI systems. Chapter IV provides more details about the limitations of the current regime and explains how the weaknesses previously identified could be overcome. It focuses on two of the most important risks to the rights and freedoms of individuals: algorithmic transparency and fairness and discrimination. Finally, Chapter V explores other governance mechanisms to further reduce the risks presented by the use of AI systems to process personal data. In particular, it explores alternatives such as the creation of registers for AI systems, the introduction of the role of the AI ethical officer, the benefits of relying on standards on AI systems to fill legislative gaps, certifications, and codes of conduct for AI system operators, the provision of more powers to data protection authorities and the reliance on privacy by design measures to reduce the risks of AI systems.

Chapter I

CONCEPTUALIZING ARTIFICIAL INTELLIGENCE

Introduction

Artificial intelligence will definitively reshape the world in which we live. Artificial Intelligence is the science of training machines to perform human tasks. It uses concepts from statistics, computer science and many other disciplines, to design algorithms that process data, make predictions, and help make decisions.[3] It establishes basic parameters regarding the data and trains the machine to learn by itself by identifying patterns using many layers of processing.

Though the term Artificial Intelligence was forged in 1956 by Minsky and McCarthy,[4] it was not until recently that it acquired its full significance. Discussions about the potentialities and fears around AI are not new, but technical features characterize and distinguish the current period. In particular, the surge in artificial intelligence is mainly due to the enormous increase in computational power and the access to huge amounts of data to train machine learning models. Recent technological developments have facilitated the transmission, processing and storage of huge amounts of information. The borderless nature of the Internet along with the vast volume of communications, create new regulatory issues for states, mainly regarding national security and data protection. These developments underpin the recent increase in machine learning capabilities and justify the wide public attention on the matter.

This chapter proceeds as follows. Firstly, it accounts for the importance of data and its free flow for the development of AI. It attempts to disentangle what is meant by 'data' and it pictures the importance of the free flow of information in the digital economy. Secondly, it assesses the intricacies of artificial intelligence. As there is no universally agreed definition of AI, it draws on the most common meaning of the term

[3] Michael I Jordan, 'Artificial Intelligence—The Revolution Hasn't Happened Yet' [2019] Harvard Data Science Review.

[4] Michael Haenlein and Andreas Kaplan, 'A Brief History of Artificial Intelligence: On the Past, Present, and Future of Artificial Intelligence' (2019) 61 California Management Review 5.

and the proposal made by the European Commission in the AI Regulation (hereinafter, AIA), as well as its foundations, capabilities and limitations. Thirdly, it reviews some algorithmic models that are covered under the umbrella term of AI. The purpose of this part is not to provide a complete understanding of the mathematical underpinning of the models. Instead, it leans on the assumption that there is a broad misunderstanding regarding the current methods comprising AI and it attempts to explain concisely the methods developers usually employ, leaving aside their mathematical foundations. For this purpose, the methods are classified according to their capacity to explain how they work and how they produce the results since it provides the groundwork to evaluate which methods could be more intelligible for users lacking specific technical background.

I.1.- The importance of data for the development of AI

I.1.1.- The concept of data

Throughout history, humans have kept the information they produced in diverse material means. Primitive societies painted walls to convey messages, and later written text was recorded on papyrus or paper. Technology allowed the digitalisation of information, which converted analogue into digital information. The process of digitalisation brought countless benefits to societies because digitalised information is easier to store, replicate and transmit. Data also changed the business environment in many ways. The most obvious outcomes of the digitalisation of the economy were the reduction of transaction and communication costs, the decrease in the time required to design, produce and deliver manufactured goods or provide services, and the creation of a whole new array of internet-enabled services.

Data can be defined as 'machine-readable encoded information'[5] or, more simply, as digitalised information. It is a recent discovery that data has an intrinsic capacity to generate wealth for the owner of the information and the society as a whole. In this

[5] Herbert Zech, 'Data as a Tradeable Commodity' in Alberto De Franceschi (ed), *European Contract Law and the Digital Single Market. The Implications of the Digital Revolution* (Intersentia 2016) 53.

sense, data was considered an asset in itself,[6] the lifeblood of the international trade,[7] the currency of the digital economy,[8] the world's most valuable resource,[9] the most valuable asset of tech companies,[10] the new oil,[11] -or, in reaction to the previous one and pointing out the detrimental implications of the cumulation of data, the new plutonium[12]-. For others, the application and use of data are the generators of value, mainly after merging and analysing it.[13] Therefore, there is an understanding that data is a resource that has inherent or potential value and, as such, it deserves appropriate protection, in particular, when the data is deemed personal data.

The acknowledgement of data as a valuable asset, or a tradable good,[14] demands the identification of its fundamental features. Several characteristics serve to contrast

---

[6] Paul M Schwartz, 'Property, Privacy, and Personal Data' (2004) 117 Harvard Law Review 2055, 2094.

[7] Organisation for Economic Co-operation and Development, 'Trade and Cross-Border Data Flows' (2019) 220 8.

[8] Diane A Macdonald and Christine M Streatfeild, 'Personal Data Privacy and the WTO' (2014) 36 Houston Jounal of International Law 625, 2; Susan Aaronson, 'Why Trade Agreements Are Not Setting Information Free: The Lost History and Reinvigorated Debate over Cross-Border Data Flows, Human Rights, and National Security' (2015) 14 World Trade Review 671, 695; Han-Wei Liu, 'Data Localization and Digital Trade Barriers: ASEAN in Megaregionalism', *ASEAN Law in the New Regional Economic Order* (Cambridge University Press 2019) 378.

[9] The Economist staff, 'The World's Most Valuable Resource is no Longer Oil, but Data' *The Economist* (6 May 2017) <https://www.economist.com/leaders/2017/05/06/the-worlds-most-valuable-resource-is-no-longer-oil-but-data> accessed 22 February 2020.

[10] John Naughton, 'Money's no object for Facebook, so hit it where it hurts', *The Guardian* (14 July 2019) <https://www.theguardian.com/commentisfree/2019/jul/14/facebook-google-fines-regulation> accessed 22 February 2020.

[11] Michael Haupt, '"Data is the New Oil" — A Ludicrous Proposition', *The Medium* (2 May 2016) <https://medium.com/project-2030/data-is-the-new-oil-a-ludicrous-proposition-1d91bba4f294> accessed 22 February 2020.

[12] Jim Balsillie, 'Data is not the new oil – it's the new plutonium', *Financial Post* (28 May 2019) <https://business.financialpost.com/technology/jim-balsillie-data-is-not-the-new-oil-its-the-new-plutonium> accessed 22 February 2020.

[13] Organisation for Economic Co-operation and Development, 'Digital Trade - Developing a Framework for Analysis' (2017) 205 205.

[14] Beate Roessler, 'Should Personal Data Be a Tradable Good? On the Moral Limits of Markets in Privacy' in Beate Roessler and Dorota Mokrosinska (eds), *Social Dimensions of Privacy: Interdisciplinary Perspectives* (CUP 2015) 142.

data with traditional assets. First, data is a *non-tangible asset*, as opposed to corporeal or material goods. At the same time, it allows physical storage, highlighting a difference with the traditional notion of services. Second, it is *easily transferable and replicable*. People can move information across borders, either in one direction or simultaneously to many places, or duplicate it very quickly without cost. Third, its *value increases with aggregation*. Many benefits emerging from data-enabled applications, such as big data algorithms or personal assistants, are based on their ability to make predictions.[15] The more information a company compiles from users' interactions, the more accurate the predictions that its applications or products can produce.[16] Fourth, it is a *non-rival and non-excludable good*. A person that uses a rival good (like every material good) prevents others from using it, hence data's non-rivalry nature allows simultaneous multi-party use.[17] This relates to the non-excludability of information since it is highly onerous for the generator of the information to avoid third-party unpermitted use and profit.[18] Fifth, the *generation of data is ubiquitous*. Albeit companies ask for users' consent to collect data, generally they gather information in a continuous and concealed way, and even without requiring permission for it.[19] Additionally, it is increasingly common to collect data directly from the human body.[20]

Before moving on, a crucial distinction must be made. Much of the data used to build AI tools is uncovered by the personal data protection regulations because that data is deemed as non-personal data. Non-personal data is out of the camp of GDPR, and it is out of the scope of this work. Hence, it is important to define what is personal data according to the GDPR since it will establish the boundaries of this work. In the next section, the concept of personal data is explained.

---

[15] Michael Mattioli, 'The Data-Pooling Problem' (2018) 32 Berkeley Technology Law Journal 1, 183.

[16] Maurice E Stucke and Ariel Ezrachi, 'How Digital Assistants Can Harm Our Economy, Privacy, and Democracy' (2018) 32 Berkeley Technology Law Journal 1239, 1251.

[17] European Commission, 'The Economics of Ownership, Access and Trade in Digital Data' (2017) 2017–01 12.

[18] Robert Heverly, 'The Information Semicommons' (2003) 18 Berkeley Technology Law Journal 1157.

[19] Lauren Willis, 'Why Not Privacy by Default?' (2014) 29 Berkeley Technology Law Journal 61, 64.

[20] Andrea M Matwyshyn, 'The Internet of Bodies The Internet of Bodies Repository Citation Repository Citation' (2019) 61 William & Mary Law Review 77, 86.

I.1.2.- Personal data

AI systems use data both in their development and deployment phase, but not every kind of data will trigger the application of the GDPR. Instead, the protection afforded by the GDPR will only be triggered where the processing operations involve personal data. The GDPR establishes that personal data is 'any information relating to an identified or identifiable natural person'.[21] From this definition it is possible to identify 4 constitutive elements: a) any information; b) relating to; c) an identified or identifiable; d) a natural person. Every constituent element of the given definition is evaluated below.

To begin with, the GDPR sets out that *'any information'* could be personal data. The scope of the concept is very broad, and it was considered that it encompasses all kinds of data insofar as they are related to the data subjects.[22] Regarding the nature of the information, it covers any kind of statement concerning an individual.[23] These statements can be objective information, like an ID number, or subjective information, like financial or working assessments. Then, the definition covers any kind of *format* in which the personal data could be available, for example, a sequence of letters or/and numbers, audiovisual information, etc. Finally, the content of the information that can be considered as personal data is also very wide, as it includes information touching upon any sort of activity performed by the data subject taking place in their private or public life or during the course of their professional activities. The GDPR provides an illustrative list of elements that can be considered personal data and it includes identifiers like a name, an ID number, location data or online identifiers (such as IP addresses, cookie identifiers or radio frequency identification tags (RFID)),[24] or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of the individual.[25] In particular, the CJEU found that different types of information could be considered personal data, like information

[21] Art. 4(1) GDPR.

[22] Case C-434/16 *Peter Nowak v Data Protection Commissioner*. [2017] EU:C:2017:582, para. 34

[23] Article 29 Data Protection Working Party, 'Opinion 4/2007 on the Concept of Personal Data' (2007) 6.

[24] Rec. 30 GDPR.

[25] Art. 4(1) GDPR.

pertaining to their working conditions and hobbies,[26] revenue and tax information,[27] passport information,[28] fingerprints,[29] images of individuals taken from CCTV cameras,[30] exam scripts and the explanations of exam evaluators,[31] handwriting traits or signature,[32] and traffic data collected from communications,[33] and dynamic IP addresses.[34]

Secondly, for the data to be personal data, the information must '*relate to the individual*', i.e. the data must be about the data subject. Yet, while the name or a picture of an individual undoubtedly relate to an individual, linking information to a person is not always a straightforward task. The CJEU set a position concerning the link of an individual to the data in *Nowak* since it stated that this element is satisfied 'where the information, by reason of its content, purpose or effect, is linked to a particular person'.[35] Hence, it could be argued that controllers should evaluate the content, the purpose or the effects (or results) of the data to conclude that the information relates to or links to the natural person. The *content* of the data relates to an individual when the data are about a particular person, irrespective of the purposes of the controller or the impact of the data on the individual.[36] The data subject can be related to the information by their name or online identifiers like advertising IDs or device fingerprints. The data are related to an individual because of its *purpose* where

---

[26] Case C-101/2001 *Bodil Lindqvist v Åklagarkammaren i Jönköping*. [2003] ECLI:EU:C:2003:596, para. 24.

[27] Joined cases C-465/00, C138-/01, and C-139/01 *Rechnungshof v Österreichischer Rundfunk*. [2003] ECLI:EU:C:2003:294, para. 64.

[28] Case C-524/06 *Heinz Huber v Bundesrepublik Deutschland*. [2008] ECLI:EU:C:2008:724, para. 31.

[29] Case C-291/12 *Michael Schwarz v Stadt Bochum*. [2013] ECLI:EU:C:2013:670, para. 27.

[30] Case C-212/13 *František Ryneš v Úřad pro ochranu osobních údajů*. [2014] ECLI:EU:C:2014:2428, para. 22.

[31] *Peter Nowak v Data Protection Commissioner* (n 22) para 36.

[32] ibid 37; Article 29 Data Protection Working Party, 'Opinion 3/2012 on Developments in Biometric Technologies' (2012) 27.

[33] Joined Cases C-293/12 and C-594/12 *Digital Rights Ireland Ltd v Minister for Communications*. [2014] ECLI:EU:C:2014:238, para. 26.

[34] Case C-582/14 *Patrick Breyer v Bundesrepublik Deutschland*. [2016] ECLI:EU:C:2016:779, para. 49

[35] *Peter Nowak v Data Protection Commissioner* (n 22) para 35.

[36] Article 29 Data Protection Working Party, 'Opinion 4/2007 on the Concept of Personal Data' (n 23) 10.

the information is employed, or is likely to be employed, to assess, treat in a specific manner or have an impact on the conduct of a person (e.g. call logs of a phone in a company assigned to a particular employee could give information about the people who was called).[37] Finally, the information can be deemed as being about an individual since its employment may affect the rights and freedoms of an individual. For the results of the data to be considered as being related to a data subject, it is necessary that the person can possibly be '*treated differently from other persons'* due to the processing operations carried out on that data.[38] For instance, setting up a system that monitors the geolocation of trucks to make the service more efficient (avoid congested routes or checking the speed) may have an adverse effect on the rights of drivers, since it allows surveillance of their performance at work.

Thirdly, the information qualifies as personal data if *an individual is identified or identifiable*. An individual is *identified* if he or she is differentiated from other persons in a group. The most common data element to identify a person is his or her name or unique personal ID. But information will be deemed as personal data where the person is directly or indirectly *identifiable* by the controller or another person. Direct identifiability is when to perform a person's identification no more than one identifier is necessary, e.g. personal ID or tax code. More challenging is the indirect identifiability of natural persons. Somebody is indirectly identifiable when an identifier could be shared by many people (e.g. the name and surname), so the identifier must be combined with another piece of information to distinguish the person (e.g. birth date, ZIP code). Through the searching of 'unique combinations'[39] among the different identifiers or information, which do not need to be held only by the controller,[40] the person can be differentiated from others. To achieve this task it should be evaluated all the means reasonably likely to be used, like singling out, and considered the time,

[37] ibid. It should be noted that the CJEU in Nowak changed its restrictive position concerning the link between data and the data subject. This is because in Case C-141/12 *YS v Minister voor Immigratie* [2014] ECLI:EU:C:2014:2081 considered that information provided by the applicant in its residence permit was personal data (e.g. name, birth date, nationality, gender, ethnicity, religion, language), but the legal assessment of the applicant's request is not in itself personal data.

[38] ibid 11.

[39] ibid 13.

[40] *Patrick Breyer v Bundesrepublik Deutschland* (n 34) para 48.

technology and money required for the identification.[41] Hence, for a person to be considered identifiable it needs more than the simple chance of being identified, and all the means reasonably available to be employed must be assessed. A particularly relevant factor for this evaluation is the purpose of the processing by the controller.[42] If the purpose of the processing of personal data consists in identifying natural persons (e.g. video surveillance), it is reasonable to conclude that identification is possible to achieve either by the controller or by a third party. However, this determination is not straightforward in all cases, so a case-by-case approach should be adopted.

Finally, the definition of personal data only concerns information related to human beings or *natural persons*, regardless of whether they are EU citizens or not. It excludes thus deceased persons or legal persons.[43]

As it can be seen from the previous analysis, the definition of personal data in the GDPR is very broad. The width of the definition of personal data, coupled with the ability of AI systems to find unexpected correlations in the datasets, means that the frontier that separates anonymous from personal data is moving toward the latter. It means that data that was previously considered non-identifiable or anonymous is, due to these novel technologies, easy to label as personal data.[44]

The interlink between personal data and AI systems is very close since there are many forms in which AI systems can work with personal data. In the development phase of an AI system, the system may include personal data in the training dataset. In the deployment phase, personal data can be used as input data to make a prediction about an individual and the outcome of this prediction will also be considered personal data. Finally, certain models by default include personal data in the model itself (like decision trees or support vector machines).

It is difficult to ignore the multiple scenarios under which AI systems may process personal data for their intended purposes. But for a better understanding of how AI

---

[41] Recital 26 GDPR.

[42] Article 29 Data Protection Working Party, 'Opinion 4/2007 on the Concept of Personal Data' (n 23) 16.

[43] Contrary to the Data Act draft that provides limited protection to the data of legal persons (see recital 30 Data Act draft).

[44] Lee A Bygrave and Luca Tosoni, 'Article 4(1). Personal Data' in Christopher Kuner and Lee A Bygrave (eds), *The EU General Data Protection Regulation (GDPR): A Commentary* (OUP 2020) 113.

systems process personal data, a deeper dive into the conceptualisation of artificial intelligence, AI models and learning paradigms (in the case of machine learning algorithms) is necessary. Without these conceptualisations, it may not be possible to interpret how AI systems use personal data, the risks associated with these processing activities, and how some of these shortcomings may be overcome.

The following section explores the concept of artificial intelligence, and it provides an overview of the most important models used by AI systems as well as the learning methods of machine learning algorithms.

I.2.- Artificial Intelligence

I.2.1.- Introduction

Artificial intelligence is considered a disruptive technology, which presents substantial benefits but at the same time, it poses major high risks under certain circumstances. AI systems are not solely taken in highly sophisticated fields. On the contrary, they are widely used and people nowadays inadvertently interact with them. Without exhausting the whole spectrum of areas where AI solutions are employed, suffice it here to mention that they are used in activities that have a negligible impact on data protection and on other rights and freedoms of individuals such as certain internet services (e.g. captchas,[45] chatbots[46]) or transportation (e.g. autonomous vehicles,[47] intelligent traffic lights,[48] optimization of routes and schedules of public transport services). But they are also broadly employed in fields that have a substantial

[45] Dennis Goedegebuure, 'You Are Helping Google AI Image Recognition' *Medium* (29 November 2016) <https://medium.com/@thenextcorner/you-are-helping-google-ai-image-recognition-b24d89372b7e> accessed 08/06/2021.

[46] Bernard Marr, 'How Artificial Intelligence Is Making Chatbots Better For Businesses' *Forbes* (18 May 2018), <https://www.forbes.com/sites/bernardmarr/2018/05/18/how-artificial-intelligence-is-making-chatbots-better-for-businesses/> accessed 08/06/2021.

[47] Bernard Marr, 'The Amazing Ways Tesla Is Using Artificial Intelligence And Big Data' *Forbes* (8 January 2018), https://www.forbes.com/sites/bernardmarr/2018/01/08/the-amazing-ways-tesla-is-using-artificial-intelligence-and-big-data/> accessed 08/06/2021.

[48] Francesca Baker, 'The technology that could end traffic jams' *BBC* (12 December 2018), <https://www.bbc.com/future/article/20181212-can-artificial-intelligence-end-traffic-jams> accessed 08/06/2021.

impact on data protection and other individual's rights like financial services (e.g. for mortgage forecasting based on customer profile analysis,[49] transaction monitoring to detect fraudulent activity based on consumption habits, automatic financial investments[50]), e-commerce and communications (e.g. product recommendations based on customer profile and analysis of their purchases and previous searches,[51] social network monitoring to targeting ads, virtual travel agents[52]), human resources (e.g. job application filtering and candidate selection)[53], public utilities (e.g. intelligent energy meters and prediction of customer consumption demand,[54] cost estimation of certain maintenance services), law enforcement and justice (e.g. automatic treatment of fines, decision support in administration of justice[55]), home appliances (e.g. smart assistants, smart mirrors, appliances, home security[56]), security (e.g. facial recognition, fingerprints, behavioural detection, border control, analysis of evidence of deception, intrusion detection, communication analysis), health and healthcare (e.g.

---

[49] Owen P. Hall, 'Artificial Intelligence Techniques Enhance Business Forecasts' (2002) 1 Graziadio Business Review 5 <https://gbr.pepperdine.edu/2010/08/artificial-intelligence-techniques-enhance-business-forecasts/> accessed 08/06/2021.

[50] Eleni Digalaki, 'The impact of artificial intelligence in the banking sector & how AI is being used in 2021' *Business Insider* (13 January 2021) <https://www.businessinsider.com/ai-in-banking-report?IR=T> accessed 08/06/2021.

[51] Google Cloud, 'Recommendations AI' <https://cloud.google.com/recommendations/> accessed 08/06/2021.

[52] Alexandr Bulanov, 'How Machine Learning and AI Can Improve Travel Services' *Towards Data Science* (3 October 2018) <https://towardsdatascience.com/how-machine-learning-and-ai-can-improve-travel-services-3fc8a88664c4> accessed 08/06/2021.

[53] Ben Dattner, Tomas Chamorro-Premuzic, Richard Buchband, and Lucinda Schettler, 'The Legal and Ethical Implications of Using AI in Hiring', *Harvard Business Review* (25 April 2019) <https://hbr.org/2019/04/the-legal-and-ethical-implications-of-using-ai-in-hiring> accessed 08/06/2021

[54] Franklin Wolfe, 'How Artificial Intelligence Will Revolutionize the Energy Industry', (*Harvard University Blog*, 28 August 2017) <http://sitn.hms.harvard.edu/flash/2017/artificial-intelligence-will-revolutionize-energy-industry/> accessed 08/06/2021.

[55] Elleora Thadanei Israni, 'When an algorithm helps send you to prision' (*New York Times,* 25 October 2017) <https://www.nytimes.com/2017/10/26/opinion/algorithm-compas-sentencing-bias.html> accessed 08/06/2021.

[56] Paul Sullivan, 'Can artificial intelligence keep your home secure?' (*New York Times,* 29 June 2018) <https://www.nytimes.com/2018/06/29/your-money/artificial-intelligence-home-security.html> accessed 08/06/2021.

allocation of beds and treatments in health services,[57] diagnosis based on image analysis,[58] prediction of patient readmission rates based on data analysis, health maps, mental health analysis, suicide prevention,[59] diagnosis by pathological sample analysis, natural language processing of medical records, genetic analysis, electrodiagnosis, development of vaccines and medications[60]), and education (e.g. content and training tailored to the needs of students, marking exams and essays, detection of plagiarism[61] or fraud in work, automatic tutoring). As seen from the previous collection, artificial intelligence solutions are widely employed in daily life, and while some use cases relate to non-intrusive activities, many of them touch upon important aspects of individuals' lives.

The fact that AI systems play an important and precious role in society should not leave unattended the risks these systems also create. Algorithms can cause harm to individuals[62] and also to societies (e.g. risking democracy itself)[63], since there is a fundamental lack of transparency in the collecting and processing of information[64] and

---

[57] Minoo Javanmardian and Aditya Lingampally, 'Can AI Address Health Care's Red-Tape Problem?' *Harvard Business Review* (5 November 2018) <https://hbr.org/2018/11/can-ai-address-health-cares-red-tape-problem> accessed 08/06/2021.

[58] Thomas Davenport and Ravi Kalakota, 'The Potential for Artificial Intelligence in Healthcare' (2019) 6 Future Healthcare Journal 94.

[59] Mason Marks, 'Suicide prediction technology is revolutionary. It badly needs oversight' *Washington Post* (20 December 2018) <https://www.washingtonpost.com/outlook/suicide-prediction-technology-is-revolutionary-it-badly-needs-oversight/2018/12/20/214d2532-fd6b-11e8-ad40-cdfd0e0dd65a_story.html> accessed 08/06/2021.

[60] Ethan Fast and Binbin Chen, 'Can artificial intelligence help us design vaccines?', *Tech Stream* (30 April 2020) https://www.brookings.edu/techstream/can-artificial-intelligence-help-us-design-vaccines/> accessed 08/06/2021.

[61] Mausumi Sahu, 'Plagiarism Detection Using Artificial Intelligence Technique In Multiple Files' (2016) 5 International Journal of Scientific and Technology Research 111.

[62] Oreste Pollicino and Gregorio De Giovanni, 'A Constitutional-Driven Change of Heart ISP Liability and Artificial Intelligence in the Digital Single Market' (2019) 18 The Global Community Yearbook of International Law and Jurisprudence 15.

[63] Karl Manheim and Lyric Kaplan, 'Artificial Intelligence: Risks to Privacy and Democracy' (2017) 21 Yale Journal of Law & Technology 106, 108.

[64] Pollicino and De Giovanni (n 62) 15.

the results of decisions taken with AI systems cannot always be predicted.[65] Additionally, cybercriminals can use AI systems to commit their actions.[66] Hence, all these aspects reinforce the need to become acquainted with the fundamentals of AI because such an understanding is essential for further assessing their data protection implications.

I.2.2.- Concept

There is no generally agreed-upon definition of AI. However, there have been many approaches and attempts to define artificial intelligence. There have been great efforts in academia,[67] governments,[68] international organizations,[69] NGOs[70] and companies[71] to draft a definition of this concept.

---

[65] Oreste Pollicino, Joe Cannataci and Valeria Falce, 'Introduction' in Oreste Pollicino, Joe Cannataci and Valeria Falce (eds), *Legal Challenges of Big Data* (Elgar 2020) 2.

[66] Europol, 'Malicious Uses and Abuses of Artificial Intelligence' (2020).

[67] See definitions from 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' (2019) 62 Business Horizons 15; David Poole and Alan Mackworth, *Artificial Intelligence: Foundations of Computational Agents* (2nd CUP 2017); Stuart Russel and Peter Norvig, *Artificial Intelligent. A Modern Approach* (3rd ed. Pearson 2010); John McCarthy, 'What is AI?', personal web page, <http://jmc.stanford.edu/artificial-intelligence/what-is-ai/index.html> accessed 10 June 2020; Hideyuki Nakashima, AI as Complex Information Processing (1999) 9 Minds and Machines 57; Nils Nilsson, *Artificial Intelligence: A New Synthesis* (Morgan Kaufman 1998); Roger Schank, 'What Is AI, Anyway?' (1987) 8 AI Magazine 59; Marvin Minsky (ed) *Semantic information processing* (MIT Press 1969);

[68] High Level Expert Group on Artificial Intelligence (HLEG), 'A definition of Artificial Intelligence: main capabilities and scientific disciplines' (2019) <https://ec.europa.eu/digital-single-market/en/news/definition-artificial-intelligence-main-capabilities-and-scientific-disciplines> accessed 08/06/2021; European Commission COM(2018) 795 on the Coordinated Plan on Artificial Intelligence

[69] Organisation for Economic Co-operation and Development, 'Recommendation of the Council on Artificial Intelligence, C/MIN(2019)3/FINAL' (2019).

[70] Virginia Dignum, Catelijne Muller and Andreas Theodorou, Final Analysis of the EU Whitepaper on AI, (2020 ALLAI) <https://allai.nl/wp-content/uploads/2020/06/ALLAI-Final-Analysis-of-the-EU-Whitepaper-on-AI-consultation.pdf> accessed 08/06/2021.

[71] McKensey, 'Artificial Intelligence. The Next Digital Frontier?' (2017) <https://www.mckinsey.com/~/media/McKinsey/Industries/Advanced%20Electronics/Our%20Insights/How%20artificial%20intelligence%20can%20deliver%20real%20value%20to%20companies/MGI-Artificial-Intelligence-Discussion-paper.ashx> accessed 06/06/2021.

To begin with, an *algorithm* is a sequence or string of orders that instructs a computer program on what must be accomplished.[72] An algorithm may have only a few or hundreds of lines of code, but the concept remains the same because these strings of computer code will give orders to achieve a certain objective.

Artificial Intelligence, on the other hand, encompasses a broader range of ideas. In the widest sense, AI can be defined as a part of computer science whose aim is to emulate intelligent behaviour. However, commentators and organisations have also provided their own definitions. Harry Surden defines artificial intelligence as the use of technology to automate assignments that regularly demand human intelligence.[73] The High-Level Expert Group on Artificial Intelligence (HLEG) provides the following definition of AI:

> *'Artificial intelligence (AI) systems are software (and possibly also hardware) systems designed by humans that, given a complex goal, act in the physical or digital dimension by perceiving their environment through data acquisition, interpreting the collected structured or unstructured data, reasoning on the knowledge, or processing the information, derived from this data and deciding the best action(s) to take to achieve the given goal. AI systems can either use symbolic rules or learn a numeric model, and they can also adapt their behaviour by analysing how the environment is affected by their previous actions.'*

The European Commission Joint Research Centre made a survey and they evaluated the definitions of AI in 55 documents. They concluded that the most important shared characteristics of the definitions surveyed acknowledged that AI systems: a) perceive the environment and evaluate the world complexity; b) gather, process and evaluate information; c) can take decisions, reason and learn from

---

[72] Céline Castets-Renard, 'Accountability of Algorithms in the GDPR and Beyond: A European Accountability of Algorithms in the GDPR and Beyond: A European Legal Framework on Automated Decision-Making Legal Framework on Automated Decision-Making' (2019) 30 Fordham Intellectual Property, Media and Entertainment Law 91, 97.

[73] Harry Surden, 'Artificial Intelligence and Law: An Overview' (2019) 35 Georgia State University Law Review 1305, 1307.

previous experiences, carry out different tasks (sometimes adapting themselves to the environment) with varying degrees of autonomy; d) can achieve specific objectives.[74]

However, a ground-breaking development was the 2021 Artificial Intelligence Regulation draft (hereinafter AIA) released by the European Commission.

Art. 3(1) AIA defines AI system as a:

> '*software that is developed with one or more of the techniques and approaches listed in Annex I [AIA] and can, for a given set of human-defined objectives, generate outputs such as content, predictions, recommendations, or decisions influencing the environments they interact with'*

Annex I AIA establishes the list of AI techniques and approaches referred to in Art. 3(1) AIA

> *(a) Machine learning approaches, including supervised, unsupervised and reinforcement learning, using a wide variety of methods including deep learning;*
> *(b) Logic- and knowledge-based approaches, including knowledge representation, inductive (logic) programming, knowledge bases, inference and deductive engines, (symbolic) reasoning and expert systems;*
> *(c) Statistical approaches, Bayesian estimation, search and optimization methods*.

There are a couple of preliminary remarks that can be made concerning the definition. From the outset, the aim of the AIA was to provide legal certainty, but at the same time to afford versatility to AI operators to create new applications and to promote innovation in this field. Additionally, it attempts to be technologically neutral to avoid the necessity to keep it updated in a field where innovation is particularly speedy. This seems to be linked to the breadth of the definition. The Commission preferred a broad definition of artificial intelligence, covering both machine learning

[74] European Commission - Joint Research Centre, 'AI Watch. Defining Artificial Intelligence. Towards an Operational Definition and Taxonomy of Artificial Intelligence' (2020) 8.

and locked algorithms.[75] It encompasses almost every system, technique or approach that in general would be considered as belonging to AI.

The definition is characterised by its functional features and by the techniques used. Yet, since the functional features required are vague and the techniques considered as AI solutions are very wide, the definition has expansive effects. Firstly, on *functional* characteristics, an AI system is defined by the capacity of the application to produce results that influence the environment according to predetermined objectives. As the draft explains, it is grounded on the main functional features of the software, chiefly the capacity to deliver certain outputs (suggestions, forecasts, etc) that have an impact on objects, persons, or even on other systems, with which the AI system interacts.[76] For the AIA draft, it does not matter whether the system is employed as a separate product or is a component of a product.[77] As the AIA draft is an overarching piece of legislation, and its scope is not limited to the AI system-to-person interaction. It also regulates AI system-to-machine interactions, because it is also applicable to AI solutions that are used as safety components of a product or is itself a product covered by certain listed EU regulations or directives.[78]

Secondly, apart from a certain functionality, the system must belong to a list of particular *techniques* or approaches, which are listed in Annex I AIA. The list covers a diverse range of techniques, including machine learning, logic and knowledge-based systems, and statistical techniques. While this is an attempt to limit the scope of the systems to be included under the term AI, these techniques are so broad that almost no AI approach is out of the scope of the regulation.

To obtain a better understanding of how AI systems work, in the following paragraphs some of the main categories of approaches and techniques will be

---

[75] Similar stance adopted Canada in its Directive on Automated Decision-Making (see Appendix A of the Directive). However, other countries took a more restrictive approach. See for instance, Brazilian Artificial Intelligence Bill, N. 21/2020 (draft), which does not apply systems which employs 'pre-defined programming parameters' without including 'the system's capability to learn and perceive'. Walter Gaspar, 'Non-Official Translation Of The Brazilian Artificial Intelligence Bill, N. 21/2020' *Cyberbricks* (25/10/2021) <https://cyberbrics.info/non-official-translation-of-the-brazilian-artificial-intelligence-bill-n-21-2020/> accessed 14/03/2022.

[76] Recital 6 AIA.

[77] Recital 6 AIA.

[78] Art. 6(1) AIA.

explained. This section provides a brief overview of the most important models used to deliver insights, inferences, profiles or decisions in artificial intelligence. The aim is to offer a basic understanding of the systems deployed to make decisions using AI systems, since knowing how they work, their main features, benefits and shortcomings will provide a basis for elaborating proposals on how to improve, in general terms, the major deficiencies identified in decision systems powered by AI techniques (for example, the lack of explainability and discriminatory outcomes).

I.2.2.1.- Machine learning

The first technique mentioned in Annex I AIA is machine learning. Nowadays, machine learning is one of the most important AI techniques since it supports most of automated individual decisions.[79] Machine learning is a subset of artificial intelligence, and it is constituted by a group of computer programmes that increase their performance at a particular task by experience.[80] It also entails finding correlations among different variables in a set of data, generally aiming at forecasting or predicting a result.[81] In other words, it generally entails making predictions and/or classifications and learning from experience. The system adapts its outputs to the new inputs provided, learning from previous experiences, to obtain more accurate results.

So machine learning techniques are adaptative systems whose performance to accomplish a certain task improves through the search for specific correlations of patterns and the inference of appropriate rules. Many processes and functions that are completely automated today rely on machine learning, such as language translation, self-driving cars and fraud detection.

Many of the well-known classifications relate to machine learning technics. In general, machine learning systems are classified according to how they learn from data.[82] In what follows a brief overview of the machine learning paradigms (supervised,

[79] Castets-Renard (n 72) 98.

[80] Tom Mitchell, *Machine Learning* (McGraw-Hill 1997) 2.

[81] David Lehr and Paul Ohm, 'Playing with the Data: What Legal Scholars Should Learn About Machine Learning' (2017) 51 UC Davis Law Review 653, 671.

[82] Office of the High Commissioner for Human Rights, 'Data Privacy Guidelines in Context of Artificial Intelligence' (2020) 3.

unsupervised and reinforcement learning) and their associated models (e.g. linear regression for supervised learning) is given.

i.- Supervised learning

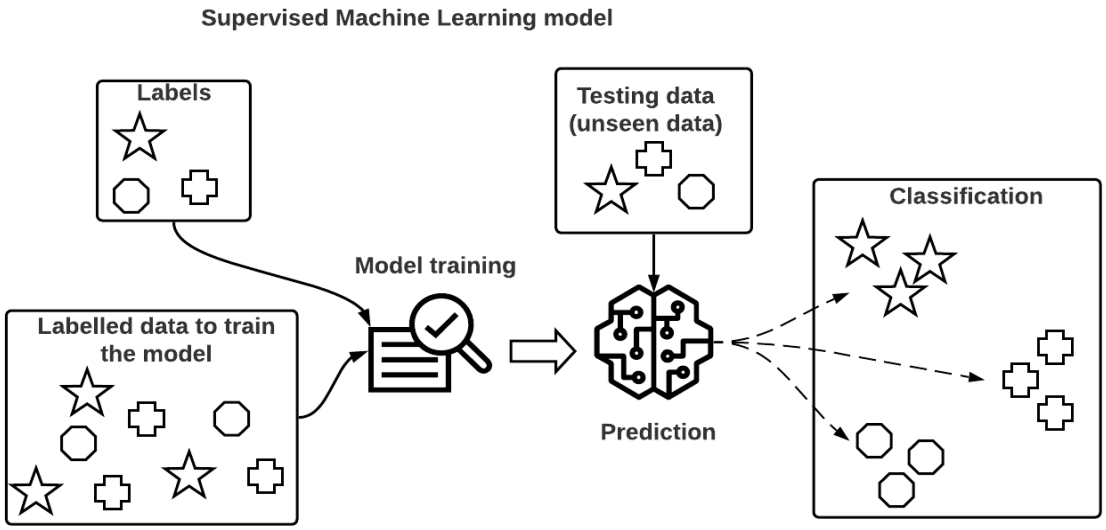
The first group concerns supervised learning systems. These systems are trained on human-labelled data. The data has been previously tagged (labelled) by humans and the system learns from the labelled examples. A 'teacher' provides training examples, each with a correct label. The system has input variables (x) and output variables (y) and the algorithm is used to make the prediction. The objective of the supervised learning model is to build a predictive model that connects features of the input data (v.gr. flat size, neighbourhood, zip code, amenities, etc) to an output value (v.gr. flat price).

Supervised learning can be then distinguished according to the type of task the systems are intended to perform or the kind of problem the systems are planned to solve. According to this distinction, problems to solve can be regression or classification problems. On the one hand, systems solving regression problems try to predict a continuous quantity.[83] The target type of regression is always a numerical value. Solving regression problems entails predicting a quantity, e.g. forecasting the price of a stock at a certain time. In these cases, the model produces a numerical prediction, not limited to a whole number (e.g. 3,15). Typical applications of regression applications are forecasting stock value, house prices, etc.

On the other hand, methods that solve classification problems try to predict a discrete[84] class label. The target type of classification problems is always categorical. Solving classification problems entails classifying data into one of two or more classes, e.g. categorizing emails as spam or not spam. When solving classification problems the system has to choose a class label for any value from a group (such as fraud/not fraud; spam/not spam; credit approved/rejected; image detected/not detected). Other typical applications are fraud detection, image classification, customer retention, and diagnostics.

[83] Continuous data can assume any possible value within a certain range.

[84] A discrete class means that it can take solely certain values, such as 1, 2, or 3, or yes/no values.



Graphic representation of a machine learning system using supervised learning for a classification problem.

Supervised learning uses a wide variety of methods to achieve its aims, but the more frequently used are linear regression (for regression problems) or logistic regression, decision trees, K-nearest neighbours, naive Bayes and support vector machines (for classification problems). Supervised learning also uses deep learning models or artificial neural networks. Specific details about these models are addressed below.

Supervised methods for regression problems: Linear regression

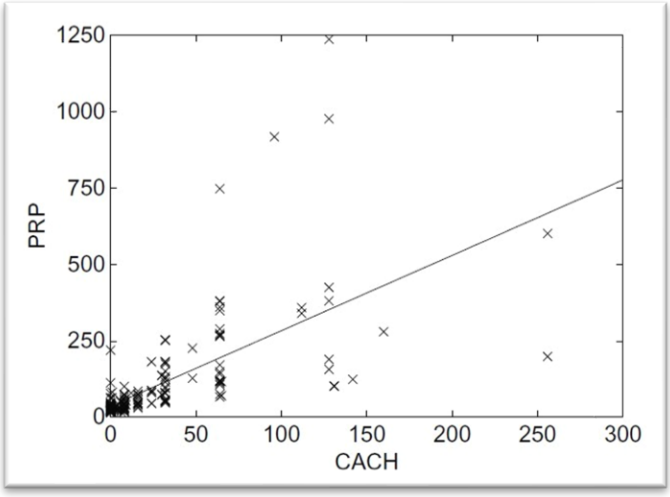
A linear regression algorithm forecasts the objective as a weighted sum of the input values.[85] It predicts a dependent variable[86] (in the 'Y axis' of the coordinate plane, e.g. stock price) based on an independent variable[87] (in the 'X axis' of the coordinate plane, e.g. time) values. It allows the prediction of continuous variables. It is a simple model

[85] Christoph Molnar, *Interpretable Machine Learning. A Guide for Making Black Box Models Explainable* (Leanpub 2020) 49.

[86] A dependent variable is the variable under evaluation. It is 'dependent' because it depends on the value of the independent variable. When the value of the independent variable changes it affects the dependent variable. It is conventionally recorded in the vertical axis of the bi-dimensional graphs and represented with an 'Y'.

[87] The independent variable is the variable that is under control to check the dependent variable. It is conventionally recorded in the horizontal axis of the bi-dimensional graphs and represented with an 'X'.

because it requires only two variables and the algorithm establishes the relationship between them. The relationship between the variables is linear (i.e. the output increase is always identical provided that the rise in the input feature is a fixed amount) and can be depicted as the best fitting straight line between the variables. It is a simple, easily interpretable, and efficient method to solve a wide array of problems. Linear systems produce truthful explanations, and linearity makes them more general and simpler.[88] While this algorithm is considered broadly interpretable, its interpretability could be impaired if the model has to calculate a large number of input features, thus becoming multi-dimensional.



Linear regression function for the CPU performance data (PRP = Published Relative Performance and CACH = Cache memory). Credits: Ian H Witten and others, *Data Mining. Practical Machine Learning Tools and Techniques* (4th edition, Elsevier 2017) 69

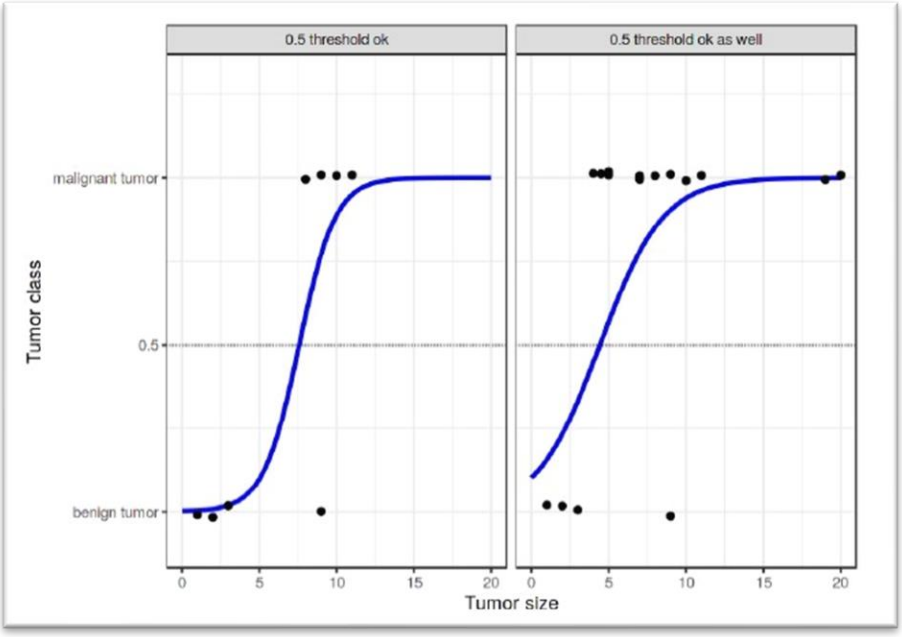
Among their common use cases can be mentioned forecasting monthly sales by assessing the relationship between online sales (dependent variable) and advertising costs (independent variable). It can also be employed in business domains (it is a reliable method to predict prices/costs of software, insurance, and real estate), forecasting sales and demand for products, and crime rates, among others.

Supervised methods for classification problems: Logistic regression

This method is used for classification problems and it calculates the probabilities that the event to be predicted falls in one over two possible outcomes. It is similar to

[88] Molnar (n 85) 63.

linear regression but applied to classification problems[89] since it converts the results of the linear regression method into estimations about labels. This model predicts a categorical dependent variable (Y) based on values of independent values (X). The dependent variable, in this case, will take categorical values such as YES or NO, O or 1, A or B. Hence the classificatory results will be binary. It calculates the probability that the event the designer tries to predict falls into any of the binary categories. The designer should also establish a threshold value, like 0.5 and if the probability of the event happening is higher than 0.5 the event is labelled as 1. On the contrary, if the probability of the event happening is lower than 0.5, the event is labelled as 0. This is a basically interpretable algorithmic model.



In these charts, the functions discover the decision boundary between malignant and benign tumours considering their sizes. Credits: Christoph Molnar, *Interpretable Machine Learning. A Guide for Making Black Box Models Explainable* (Leanpub 2020) p. 72

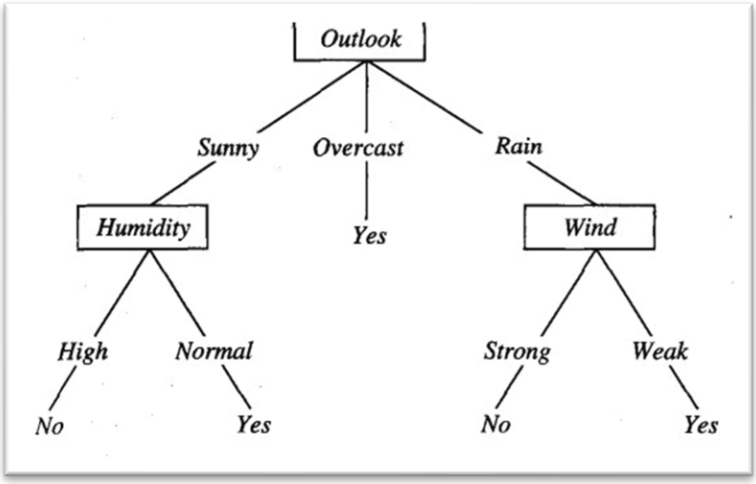
A use case could be stated as the prediction of whether an applicant to the Bocconi PhD program will be admitted (dependent variable) based on the applicant's English

[89] ibid 69.

test score (TOEFL). Fraud detection,[90] cybersecurity, and image processing are other useful applications of this algorithm.

Supervised methods for classification problems: Decision Trees

These models categorize instances by regrouping them from the root of the tree to a particular leaf node. This leaf node ultimately gives the categorization of the instance. Each node constitutes an evaluation of a feature of the instance, and each branch down the node constitutes a value for this particular feature.[91] The classification starts at the root node of the decision tree and then moves to the following nodes until it reaches the final nodes. The leaf node or final node represents the predicted outcome.[92]



A decision tree for the concept of playing tennis. Source: Tom Mitchell, *Machine Learning* (McGraw-Hill 1997) 62.

Decision trees work similarly to the decision-making process performed by human beings. Where humans have multiple features to use and make a decision, they make a mental model of the procedures needed to make the final decision. When using decision trees in machine learning, the algorithm decides which feature to use for the

[90] Fayaz Itoo, Meenakshi and Satwinder Singh, 'Comparison and Analysis of Logistic Regression, Naïve Bayes and KNN Machine Learning Algorithms for Credit Card Fraud Detection' (2020) 13 International Journal of Information Technology (Singapore) 1503.

[91] Mitchell (n 80) 53.

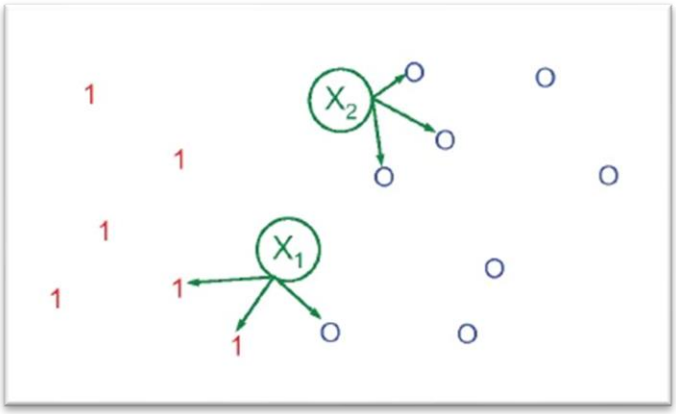
[92] Molnar (n 85) 102.

split of the layers and the threshold to use at each different layer based on the training data developers provide them.

Decision tree modes are very popular classification models since they are easily understandable. However, as seen in linear regression, its interpretability decreases if the model must calculate a large number of input features. Common applications for this method are selecting candidates in job applications[93] and approving or rejecting loans.

Supervised methods for classification problems: K-Nearest Neighbours

This model uses the closest neighbours of a data point to produce the prediction by associating data into clusters. K-nearest neighbours can be used to classify input data since it assigns the most frequent class of the closest neighbours of an instance.[94] The K in the nearest neighbours model denotes the amount of nearest neighbours that the classifier will use to make its prediction. In other words, the model checks the K points in the dataset that are closest to the instance (or input data).[95]



K-Nearest Neighbour (K-NN) representation. The 3 nearest neigbours (NN) of test point x1 have labels 1, 1 and 0, while the 3 NN of test point x2 have labels 0, 0, and 0. Source: Kevin Murphy, Machine Learning. A Probabilistic Perspective (MIT Press 2012) 16.

The interpretability is relatively straightforward because it predicts by looking at the nearest data in the training dataset. Among the use cases can be mentioned prediction

[93] A Liberman and T Rotarius, 'Pre-Employment Decision Trees: Job Applicant Self-Election.' (2000) 18 Health Care Manager 48.

[94] Molnar (n 85) 139.

[95] Kevin Murphy, *Machine Learning. A Probabilistic Perspective* (MIT Press 2012) 16.

of user behaviour in recommendation systems[96] or loan management systems, or to profile bank customers.[97]

As a downside, the use of this model requires keeping the original datasets. This is problematic not only because the high storage requirements and the time to deliver a prediction increase with the size of the training dataset. Additionally, keeping the dataset also has data protection implications. Since the dataset may contain personal data, additional privacy and security measures must be implemented.

Supervised methods for classification problems: Naïve Bayes

The Naïve Bayes classifier employs Bayes' rule of conditional probabilities,[98] to predict the probability that a feature fits within a given class. Naïve Bayes measures, for each input data, the likelihood that that instance belongs to a class according to the value of the input data.[99] It does not estimate the class probabilities for each feature conditionally to the value of the other related features. It is called naïve because the model assumes the independence of the features and it is employed in situations where the variables are not conditionally independent.[100] Naïve Bayes models are generally employed to predict credit scoring,[101] filter spam email,[102] make sentiment analyses in social networks,[103] or for recommendation systems.

---

[96] Gaowei Xu and others, 'A User Behavior Prediction Model Based on Parallel Neural Network and K-Nearest Neighbor Algorithms' (2017) 20 Cluster Computing 1703.

[97] Aida Krichene Abdelmoula, 'Bank Credit Risk Analysis with K-Nearest-Neighbor Classifier: Case of Tunisian Banks' (2015) 14 Journal of Accounting and Management Information Systems 79.

[98] Conditional probability calculates the probability that an event takes place knowing that another different even has taken place. See Yale University Department of Statistics and Data Science, past courses page http://www.stat.yale.edu/Courses/1997-98/101/condprob.htm, accessed on 30/01/2021.

[99] Molnar (n 85) 138.

[100] Stuart Russell and Peter Norvig, *Artificial Intelligence: A Modern Approach* (3rd edn, Pearson 2010) 499.

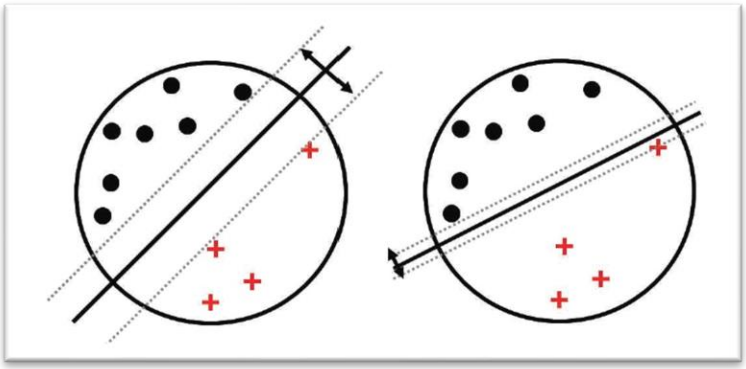
[101] AC Antonakis and ME Sfakianakis, 'Assessing Naïve Bayes as a Method for Screening Credit Applicants' (2009) 36 Journal of Applied Statistics 537.

[102] Aaron Massey and Travis D Breaux, 'Chapter 7: Interference' in Travis D Breaux (ed), *An Introduction to Privacy for Technology Professionals* (IAPP 2020) 316.

[103] Dhiraj Gurkhe, Niraj Pal and Rishit Bathia, 'Effective Sentiment Analysis of Social Media Datasets Using Naive Bayesian Classification' (2014) 99 International Journal of Computer Applications 1;

Supervised methods for classification problems: Support vector machines

Support Vector Machines (SVM) constitute a boundary (a line or a hyperplane) that best segregates two groups of features in a high-dimensional feature space.[104] The model first groups the training data and expands the borders of the decision boundaries between them.[105] Then it observes the edges of each cluster of data and draws a middle point between them as a threshold.



Support Vector Machines illustration. On the left, the hyperplane that constitutes the boundary has a wide margin, whereas on the right it has a small margin. Credits: Kevin Murphy, *Machine Learning. A Probabilistic Perspective* (MIT Press 2012) 500.

This method is an alternative to artificial neural networks and it is useful for data mining, for instance, to predict the decision of courts,[106] credit scoring,[107] question answering,[108] page rankings in search engines,[109] or spam filtering.[110]

Malhar Anjaria and Ram Mohana Reddy Guddeti, 'A Novel Sentiment Analysis of Social Networks Using Supervised Learning' (2014) 4 Social Network Analysis and Mining 1.

[104] Information Commissioner's Office, 'Explaining Decisions Made with AI' (2020) 118.

[105] ibid.

[106] Masha Medvedeva, Michel Vols and Martijn Wieling, 'Using Machine Learning to Predict Decisions of the European Court of Human Rights' (2020) 28 Artificial Intelligence and Law 237, 243.

[107] Xiujuan Xu, Chunguang Zhou and Zhe Wang, 'Credit Scoring Algorithm Based on Link Analysis Ranking with Support Vector Machine' (2009) 36 Expert Systems with Applications 2625.

[108] Show Jane Yen and others, 'A Support Vector Machine-Based Context-Ranking Model for Question Answering' (2013) 224 Information Sciences 77.

[109] Thorsten Joachims, 'Optimizing Search Engines Using Clickthrough Data', *ACM SIGKDD international conference on Knowledge discovery and data mining* (2002).

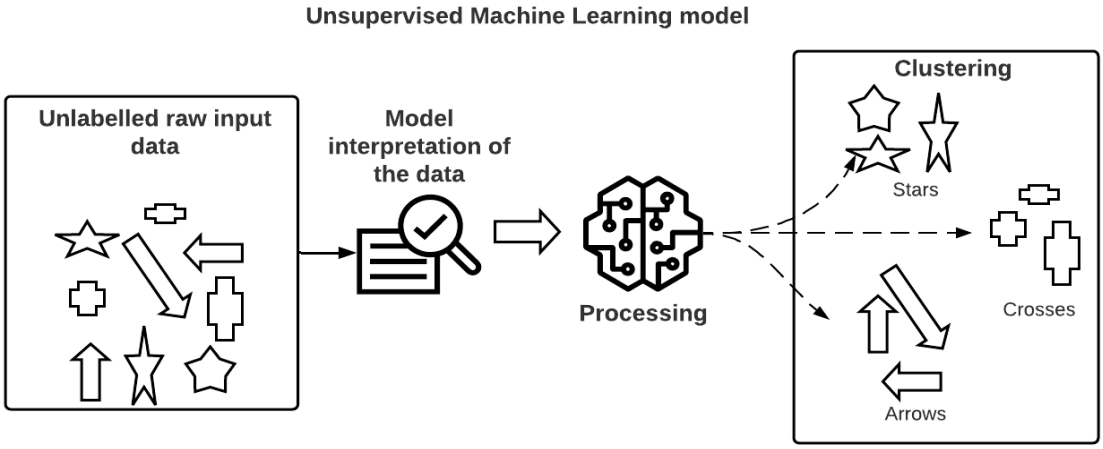
[110] Jenna Burrell, 'How the Machine "Thinks": Understanding Opacity in Machine Learning Algorithms' [2016] Big Data and Society 1, 7.

ii.- Unsupervised learning

Another kind of machine learning technique is constituted by unsupervised learning systems. In these algorithms the data provided is unlabelled. The system identifies patterns in the data without being provided with any explicit feedback.[111] Contrary to the supervised learning method, in unsupervised learning techniques there are no correct labels available for the training examples. The task of the algorithm is to detect or discover relevant patterns or any specific grouping or cluster behaviour within the observed data. Then, based on the relative distance between the observations data scientists can sort the outcomes into a few different groups or clusters, and these groupings allow them to conduct downstream analysis. For instance, clustering customers in accordance with the purchase preferences. *Clustering* is an application of unsupervised learning algorithms, whereby the data is gathered by taking into account how similar a single datum is from its neighbours and how different it is from anything else. The objects that share the most similarities are grouped together. For example, grouping individuals according to their purchasing history to predict their shopping behaviour uses unsupervised learning technics (in particular, the k-means clustering algorithm). Clustering algorithms are also used in recommendation systems, targeted marketing and customer segmentation. Clustering is the most common use of unsupervised learning but it is not the only one, since dimensionality reduction is made using unsupervised learning. *Dimensionality reduction* is used to reduce the number of input variables in the training data.[112] Some of the most common models used for unsupervised learning are k-means and Neural Networks.

[111] Russell and Norvig (n 100) 694.

[112] Murphy (n 95) 11.

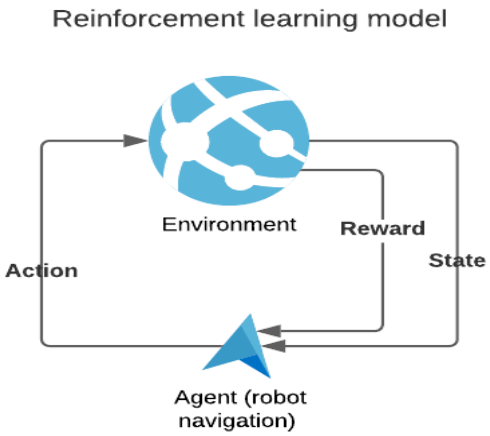


Graphic representation of a machine learning system using unsupervised learning for a clustering problem

iii.- Reinforcement learning

Finally, in reinforcement learning the agent learns how to behave according to the feedback received as a reward or punishment from the environment.[113] Where the agent performs desired actions it receives rewards (positive feedback) and undesired actions trigger penalties (negative feedback). To maximise the rewards received, the agent will try to perform the desired actions. Hence, the agent learns by experience and the type of feedback received. Robotic scientists generally employ this technique to teach robots the actions they should perform. Other use cases that employ reinforcement learning algorithms are robot navigation, autonomous driving, skill acquisition, learning tasks, game AI, and real-time decisions.

[113] Russell and Norvig (n 100) 695; Murphy (n 95) 2.



Graphic representation of a machine learning system using reinforcement learning.

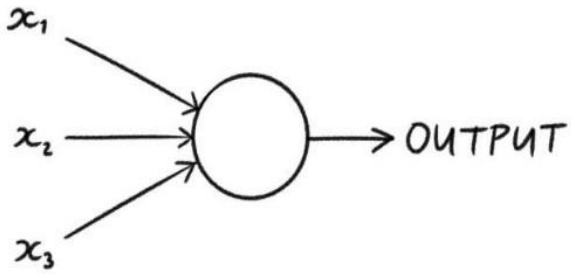
iv.- Artificial Neural Networks and Deep Learning

Finally, the definition of AI in the AIA draft includes deep learning methods. These techniques are used to create artificial neural networks, which are artificial emulations of the structure and function of the human brain. The concept of neural network was coined during the 1940s to 1960s when scientists tried to discover algorithmic representations of data processing in biological systems,[114] and they considered that the nodes resembled neurons, whereas the links between the nodes were similar to synapsis.

While artificial neural networks comprise a wide range of conceptualizations and models, in general, they are simply a group of units connected.[115] In its most basic formulation, an artificial neural network consists of nodes and connections among the nodes. The simplest neural network is the perceptron, which is a single-layer neural network that uses a list of input features (e.g. inputs x1, x2, x3)

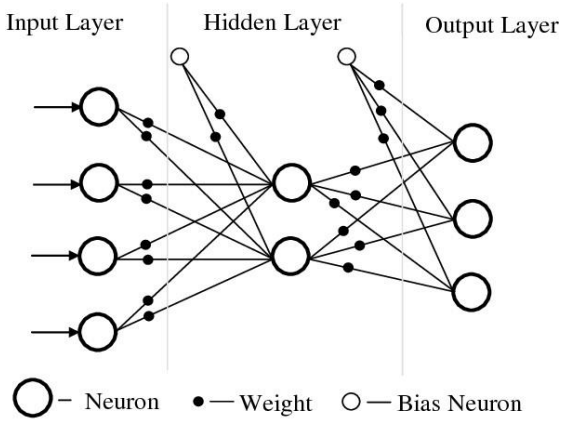
[114] Christopher Bishop, *Pattern Recognition and Machine Learning* (Springer 2011) 226.

[115] Russell and Norvig (n 100) 728.



Graphic representation of a perceptron. Credits Norwegian Data Protection Authority, 'Artificial Intelligence and Privacy' (2018) 14

The most salient feature of artificial neural networks is that they do not link directly input data to output data. Instead, they are constituted by one or more layers of processing. In general, they consist of an input layer (which has many input nodes) an output layer (which has many output nodes), along with different layers of nodes between the input and output layers (hidden layers) all connected by a web of connections between the layers which are weighted. Neural networks composed of more than three layers -input and output layers included- are considered deep learning algorithms.[116]



The basic structure of an artificial neural network. Note that this representation only has one hidden layer. Credits: Kulkarni, P., Londhe, S., & Deo, M.C., 'Artificial Neural Networks for Construction Management: A Review' (2017) 1 Journal of Soft Computing in Civil Engineering 71

[116] Eda Kavlakoglu, 'AI vs. Machine Learning vs. Deep Learning vs. Neural Networks: What's the Difference?' *IBM Cloud* (27/05/2020), <https://www.ibm.com/cloud/blog/ai-vs-machine-learning-vs-deep-learning-vs-neural-networks> accessed 14/03/2022.

Once the input data is fed into the model (e.g. a photo), the first layer processes the data and then passes its output to the next layer, and the latter performs the same actions, i.e., processing the output and passing it to the following layer. Hence, the classification of the data or the prediction of the output value is done by a feed-forward activation of input variables.[117] The task of the model is to fine-tune the weights given to each node so that the final output matches the initial example.

Artificial neural networks are currently applied in several fields and typical use cases are image recognition, handwritten recognition,[118] computer security,[119] detection and removal of inappropriate profiles in social networks,[120] chatbots,[121] law enforcement and financial services,[122] incident detection, and fraud detection, among others.

I.2.2.2.- Logic and Knowledge-based approaches

The definition of AI also covers logic-based and knowledge-based approaches, which include knowledge representation, inductive programming, knowledge bases, inference and deductive engines, and reasoning and expert systems.[123] In general, these systems are given certain rules that characterise the basic or fundamental logic and knowledge of any operation the developers attempt to model and automate.[124] This implies feeding the operational and decisional rules in advance so that the model can, where required, deliver the decision according to a pre-established set of parameters and instructions.

---

[117] Information Commissioner's Office, 'Explaining Decisions Made with AI' (n 104) 120.

[118] Burrell (n 110) 5.

[119] Halenar Igor and others, 'Application of Neural Networks in Computer Security' (2014) 69 Procedia Engineering 1209.

[120] Daniel Gorham, 'Keeping LinkedIn professional by detecting and removing inappropriate profiles' *Linkedin Engineering* (16/01/2020) <https://engineering.linkedin.com/blog/2020/keeping-linkedin-professional> accessed 14/03/2022.

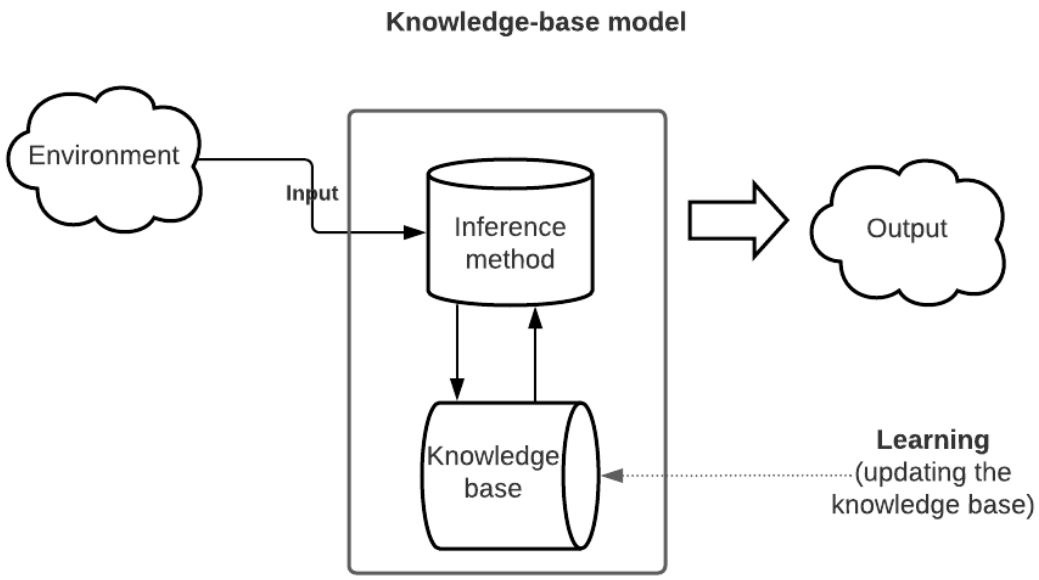
[121] IBM Cloud Education, 'Chatbots', *IBM Cloud* (09/05/2019) <https://www.ibm.com/cloud/learn/chatbots-explained> accessed 14/03/2022.

[122] IBM Cloud Education, 'Deep Learning', *IBM Cloud* (01/05/2020) <https://www.ibm.com/cloud/learn/deep-learning> accessed 14/03/2022.

[123] Annex I, point b, AIA.

[124] Surden (n 73) 1316.

Knowledge-based approaches involve the use of agents capable of keeping internal knowledge or information about the world and an inference engine. Knowledge-based agents receive input data from the environment, then process the information they have stored in the form of knowledge, make inferences using the inference engine and then use these inferences to choose the course of action (output).[125]



Graphic representation of an AI system based on a knowledge-based model.

Expert systems are a frequently used subcategory of logic and knowledge-based methods. In expert systems, developers together with experts in a particular area (e.g. tax, law, medicine), programme an algorithm with all the rules of the particular field in a computer-readable way, translating the expert knowledge in the area for the computer.[126] Examples of expert systems are Turbotax,[127] PXDES,[128] and MYCIN.[129]

---

[125] Russell and Norvig (n 100) 274.

[126] Surden (n 73) 1316–1317.

[127] Turbotax is an application for the preparation of American income tax returns. See https://turbotax.intuit.com/

[128] PXDES is an algorithm to forecast the degree and type of lung cancer.

[129] MYCIN is an expert system to diagnose and choose appropriate therapies for patients with bacterial infections. See William van Melle, 'MYCIN: A Knowledge-Based Consultation Program for Infectious Disease Diagnosis' (1978) 10 International Journal of Man-Machine Studies 313.

I.2.2.3.- Statistical approaches

Finally, the AIA draft includes among the different techniques that define an AI system the category of statistical approaches, Bayesian estimation, search and optimization methods.[130] In contrast to machine learning approaches, which adapt their behaviour according to previous experience, traditional statistical approaches only infer relationships between different variables.[131] Bayesian estimation is an instance of a statistical approach.[132] Search and optimization methods are used where solving a problem can be achieved by a series of predetermined operations in recognised, deterministic and observable settings.[133] Deep Blue, the computer that in 1997 defeated Gary Kasparov in chess, employed a searching algorithm to decide the movement of the pieces.[134]

The inclusion of this paragraph in the AIA draft was heavily criticised for over-stretching the definition of AI, and expanding the scope of application of the proposal. It is worth being noted that 'statistical approaches' and 'search and optimization methods' are very wide terms and they could cover many software elements or applications that currently are not generally under the remit of AI.[135] As it was claimed that this wide-encompassing formulation of the AI systems covered by the AIA draft would include not only very complex deep learning algorithms but also simple software solutions such as 'an Excel sheet' using a statistical formula to generate a result[136] or

---

[130] Annex I, point c, AIA.

[131] Hema Sekhar Reddy Rajula and others, 'Comparison of Conventional Statistical Methods with Machine Learning in Medicine: Diagnosis, Drug Development, and Treatment' (2020) 56 Medicina 455, 457.

[132] Bob Carpenter, 'EU proposing to regulate the use of Bayesian estimation', *Statistical Modeling, Causal Inference, and Social Science* (22/04/2021) <https://statmodeling.stat.columbia.edu/2021/04/22/eu-proposing-to-regulate-the-use-of-bayesian-estimation/> accessed 15/03/2022.

[133] Russell and Norvig (n 100) 120.

[134] Murray Campbell, Joseph Hoane and Feng-hsiung Hsu, 'Deep Blue' (2002) 134 Artificial Intelligence 57, 71.

[135] Michael Veale and Frederik Zuiderveen Borgesius, 'Demystifying the Draft EU Artificial Intelligence Act' (2021) 22 Computer Law Review International 97, 109.

[136] Nicolas Kayser-Bril, 'European Council and Commission in agreement to narrow the scope of the AI Act', *Algorithm Watch* (24/11/2021), <https://algorithmwatch.org/en/eu-narrow-scope-of-ai-act/> accessed 15/03/2022.

'task allocation systems' employed by organisations to perform back-office activities,[137] several institutions and commentators suggested removing the last paragraph.[138]

I.2.2.4.- Final observations

In the previous section it was mentioned that according to the definition contained in the AIA draft, the techniques and approaches that are deemed as AI systems are, broadly, machine learning approaches, knowledge-based approaches and statistical approaches. In addition, it was also pointed out that one of the most common classifications of machine learning models considers the way the algorithms learn, i.e, supervised by humans, not supervised by humans or reinforced through rewards or punishments for the desired actions. The table below (Table I) summarises the AI techniques and approaches employed by the AIA draft, as well as, where relevant, the learning paradigm, the problem to solve and the algorithm name as described in this work.

Table I Summary of the AI techniques and approaches described

Annex I AIA	AI techniques and approaches	Learning paradigm	Problem to solve	Algorithm name example
(a)	**2.2.1.- Machine Learning approaches**	**i.- Supervised learning**	Regression	Linear regression
				Artificial Neural Network
			Classification	Logistic regression
				Decision trees
				K-nearest neighbour

[137] Insurance Europe, 'Response to EC proposal for a Regulation on AI. Position paper' <https://www.insuranceeurope.eu/publications/2413/response-to-ec-proposal-for-a-regulation-on-ai/> accessed 15/03/2022.

[138] See for instance, Cailean Osborne, 'The European Commission's Artificial Intelligence Act highlights the need for an effective AI assurance ecosystem', *Centre for Data Ethics and Innovation* (11/05/2021) <https://cdei.blog.gov.uk/2021/05/11/the-european-commissions-artificial-intelligence-act-highlights-the-need-for-an-effective-ai-assurance-ecosystem/>; Raimond Dufour at al, 'AI or More? A Risk-based Approach to a Technology-based Society', *Oxford Business Law Blog* (16/09/2021) <https://www.law.ox.ac.uk/business-law-blog/blog/2021/09/ai-or-more-risk-based-approach-technology-based-society>; Christian Djeffal, 'The Regulation of Artificial Intelligence in the EU', *Heinrich Böll Stiftung Israel* (30/12/2021) <https://il.boell.org/en/2021/12/24/regulation-artificial-intelligence-eu>, all accessed 15/03/2022.

				Naive Bayes
				Support vector machines
				Artificial Neural Network
		ii.- Unsupervised learning	Clustering	K-means
			Dimensionality reduction	x
		iii.- Reinforcement learning	Learning by rewards	x
(b)	**2.2.2.- Logic-Knowledge-based approaches**	No learning	x	x
(c)	**2.2.3.- Statistical approaches**	No learning	x	x

However, this classification criterion, while useful in computer science, it is not entirely helpful for the purposes of this work. The core of this work concerns the protection of fundamental rights of individuals, hence the knowledge about how different machine learning techniques learn from the datasets to gain new insights would be hardly useful for that task. Then, for this work it is preferred to classify the systems according to the interpretability or opaqueness of the algorithms, i.e., on how complex to understand their functioning and the outcomes.

The concept of interpretability is subjective and there is no generally agreed definition or method to evaluate it.[139] To alleviate this problem, this work uses the classification made by the Information Commissioner's Office and the Alan Turing Institute in 'Explaining decisions made with AI'.[140] According to this guidance, algorithmic models can be classified into broadly interpretable and broadly non-interpretable or 'black-box' models.

Though most of the cases clearly fall into one group, sometimes the line between the interpretable and opaque models is blurred. Hence, it could be also understood as a degree of interpretability: at one end of the spectrum, algorithms are easily interpretable, whereas at the opposite end of the spectrum the models are highly

[139] Adrien Bibal and Benoît Frénay, 'Interpretability of Machine Learning Models and Representations: An Introduction', *European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning* (2016) 78.

[140] Information Commissioner's Office, 'Explaining Decisions Made with AI' (n 104) 67–68.

complex to understand, and, finally the middle ground between them is sometimes blurred and the models are relatively interpretable. Interpretability of AI systems can be understood as the degree to which an individual can comprehend the causes of a particular decision.[141] Yet, it is important to acknowledge that individuals should understand both the process through which the decision was taken and the decision itself. Understanding the reasons behind the decision is not trivial, since it may enable individuals to predict with a certain degree of certainty the outcomes of the algorithm and also challenge the decision made.

According to the classification previously mentioned, some of the broadly interpretable algorithms are linear regression, logistic regression, decision trees, k-nearest neighbours, and naïve Bayes.[142] Except for linear regression, the other models are classification algorithms (see Table I). Similarly, logic- and knowledge-based approaches[143] and traditional statistical approaches[144] are also generally interpretable AI approaches. These techniques and algorithms are considered interpretable because users can in a relatively easy manner discover and understand their work and outcomes. The interpretability of these models tends to decrease when more variables are included since with more variables the models become more complex.

However, some models are more complex, and their internal workings and foundations cannot be understood simply by assessing their parameters or features. Since they are opaque they are also usually labelled as 'black-box' algorithms. Some black-box algorithmic models are support vector machines and artificial neural networks.[145]

---

[141] Tim Miller, 'Explanation in Artificial Intelligence: Insights from the Social Sciences' (2019) 267 Artificial Intelligence 1, 8.

[142] In addition to those previously mentioned, the ICO-Alan Turing guide 'Explaining decisions made with AI' include as broadly interpretable models the following: Generalised linear model (GLM), Generalised additive model (GAM), Regularised regression (LASSO and Ridge), Rule/decision lists and sets, Supersparse linear integer model (SLIM), Case-based reasoning (CBR)/Prototype and criticism. However, the list was shortened since it is only an illustration of the most frequently used algorithms.

[143] Annex I, point b, AIA.

[144] Annex I, point c, AIA.

[145] In addition to those previously mentioned, the ICO-Alan Turing guide 'Explaining decisions made with AI' include as broadly 'black box' models the following: ensemble methods and random forest.

Table II classification of algorithms and AI techniques according to their interpretability

Broadly interpretable algorithms and techniques	Not explainable algorithms
• linear regression • logistic regression • decision trees • k-nearest neighbours • naïve Bayes • logic- and knowledge-based approaches • traditional statistical approaches	• support vector machines • artificial neural networks

As seen in the section, the term AI involves a wide range of approaches, methods and models and there is no single generally agreed-on definition that describes precisely which approaches, methods and models are included therein. The AIA draft is the first international normative approach to the topic and in its definition embraces machine learning techniques, logic and knowledge-based techniques and statistical techniques.

An understanding of the models is important to be aware that there are simpler ways to profit from the benefits that AI models have without impairing the ability of individuals to understand their inner workings and outcomes. In addition, controllers using AI systems to process personal data should be mindful that some models may contain personal data in the model itself, like decision trees,[146] support vector machines,[147] k-nearest neighbours, [148] and artificial neural networks. Furthermore, some models have their own particular GDPR-related problems. For instance, if rule-based expert systems contain mistakes in their conception, these errors will lead to erroneous inferences, hence the principle of accuracy is not respected.[149] Similarly, the accuracy of machine learning systems may be affected if the model suffers from concept drift[150] or if the machine learning is unable to model the desired processing.[151]

[146] Norwegian Data Protection Authority, 'Artificial Intelligence and Privacy' (2018) 10.

[147] Information Commissioner's Office, 'Guidance on AI and Data Protection' (2020) 55.

[148] ibid.

[149] Agencia Española de Protección de Datos, 'Adecuación Al RGPD de Tratamientos Que Incorporan Inteligencia Artificial. Una Introducción' (2020) 33.

[150] Concept drift is the modification in the links between input and output information over time.

[151] Agencia Española de Protección de Datos (n 149) 33.

Concerning artificial neural networks, while the vast amount of parameters gives these models better overall performance, this improvement can be hindered by the barriers to delivering easily understandable explanations.[152] Additionally, artificial neural networks can keep fragments of the dataset employed to train it for the further automatic production of text.[153] Hence, if a neural network contains personal data malicious actors could penetrate it and cause harm to data subjects. These issues are described below in the relevant sections in detail.

[152] Ivan Evtimov and others, 'Is Tricking a Robot Hacking?' (2019) 34 Berkeley Technology Law Journal 891, 899.

[153] Nicholas Carlini and others, 'The Secret Sharer: Evaluating and Testing Unintended Memorization in Neural Networks', *SEC'19: Proceedings of the 28th USENIX Conference on Security Symposium* (2019) 270.

Chapter II

THE PROTECTION OF PERSONAL DATA IN EUROPE. AN APPRAISAL OF THE PROVISIONS RELATED TO ARTIFICIAL INTELLIGENCE

Introduction

Artificial intelligence has the potential to harm different fundamental rights, like personal autonomy, freedom of expression and the prohibition of discrimination. However, this work mostly limits its scope of potential damages to privacy and data protection for two reasons. First, privacy and data protection can be two of the most particularly impaired rights by this technology and, second, assessing the whole array of fundamental rights and liberties potentially harmed by artificial intelligence would demand broader research leaving some particularities of data protection unattended. Nonetheless, the interconnections between the fundamental protections are strong and many references will be made to different rights that AI could potentially violate.

This chapter provides general background on the data protection legal framework in the European Union. A basic understanding of this fundamental protection is crucial to assess the application of AI systems in the EU and to provide some basis for the interpretation and application of the regulation on data protection to unforeseen applications of AI systems. In what follows, this chapter elaborates, first, on the regulation of data protection in the EU, the difference between data protection and privacy and the intertwining between the protection conferred by the EU system (e.g. Charter of Fundamental Rights of the EU, hereinafter, the Charter, and the Treaty of the European Union) and the Council of Europe (ECHR and Convention 108+). It is worth noticing that the main focus of this work is the assessment of the EU provisions, however, some references where appropriate are made to the ECHR and Convention 108 to highlight differences or to show alternative paths to solve the issues. Second, this chapter appraises the main data protection secondary legislation, i.e., the General Data Protection Regulation. Thirdly, this section dwells on the data protection principles and the basis for the lawful processing of personal data. Though addressed

in a relatively general fashion, this part is useful since it will shed light on opaque areas of data protection.

II.1.- Data protection as a fundamental right in Europe

While the acknowledgement of the protection of an individual's private life or the right to be let alone is not a contemporary creation,[154] the further distinction between the right to privacy and the protection of personal data is a relatively recent development.[155]

European countries consider personal data and privacy as fundamental rights. These rights are inherent to the human person and are essential protections that cannot be overridden, except under strictly limited circumstances. There are two different systems of protection of fundamental rights in Europe for the protection of personal data and privacy. One of them is the set of protections offered by the Council of Europe in the European Convention on Human Rights. The second is constituted by the safeguards granted by the European Union in the Charter or the TEU.

II.1.1.- Council of Europe legal framework

First, the European Convention of Human Rights protects the right to respect private life in Article 8(1): 'Everyone has the right to respect for his private and family life, his home and his correspondence'. It is worth noting that the ECHR does not expressly recognise the protection of personal data as a standalone right. The ECHR is centred on the negative dimension of the right to privacy,[156] i.e., the respect for private and family life. Nevertheless, a person's right to respect the processing of his or her personal data is part of the right to protect private and family life, which also includes his or her home and correspondence. In other words, processing personal data may

---

[154] Samuel D Warren and Louis D Brandeis, 'The Right to Privacy' (1890) 4 Harvard Law Review 193, 194.

[155] For a deeper understanding of development of both rights and their differences, see: Orla Lynskey, *The Foundations of EU Data Protection Law* (OUP 2016); Gloria Gonzalez Fuster, *The Emergence of Personal Data Protection as a Fundamental Right of the EU* (Springer 2014).

[156] Oreste Pollicino and Marco Bassini, '8. Protezione Dei Dati Di Carattere Personale' in Roberto D'Orazio and others (eds), *Codice della Privacy e Data Protection* (Giufrè Francis Lefebvre 2021) 38.

interfere with the data subject's right to respect for private life, as protected by Art. 8 ECHR.

Indeed, the ECtHR declared that the mere collection of data relating to the private life of an individual amounts to an interference with Art. 8 ECHR.[157] Additionally, the ECtHR has applied the ECHR in cases related to the collection of personal data (vg.r. location data,[158] health data,[159] interception of communications, phone tapping and secret surveillance,[160] surveillance of employees' computer use,[161] saliva samples,[162] voice samples,[163] video surveillance[164]), the storage and use of personal data (e.g. in the context of health,[165] in social insurance proceedings,[166] storage in secret registers,[167] telecommunication service providers' data[168]), disclosure of personal data,[169] access to personal data,[170] erasure or destruction of personal data.[171]

---

[157] *S. and Marper v the UK* Apps nos. 30562/04 and 30566/04 (ECtHR 4 December 2008).

[158] *Uzun v Germany* App No. 35623/05 (ECtHR 2 September 2010) and *Ben Faiza v France* App No 31446/12 (ECtHR 8 February (2018).

[159] *L.H. v Latvia* App no. 52019/07 (ECtHR 29 April 2014).

[160] *Malone v the UK* App no. 8691/79 (ECtHR 2 August 1984).

[161] *Bărbulescu v Romania* App no. 61496/08 (ECtHR 5 September 2017) and *Libert v France* App no. 588/13 (ECtHR 22 February 2018).

[162] *Dragan Petrović v Serbia* App no. 75229/10 (ECtHR 14 April 2020).

[163] *P.G. and J.H. v the UK* App no. 44787/98 (ECtHR 25 September 2001) and *Vetter v France* App no. 59842/00 (ECtHR 31 May 2005).

[164] *Peck v the UK* App no. 44647/98 (ECtHR 28 January 2003), *Antović and Mirković v Montenegro* App no. 70838/13 (ECtHR 28 November 2017) and *López Ribalda and Others v Spain* App no.1874/13 and 8567/13 (ECtHR 17 October 2019).

[165] *L.L. v France* App no. 7508/02 (ECtHR 10 October 2006) and *Mockutė v Lithuania* App no. 66490/09 (ECtHR 27 February 2018).

[166] *Vukota-Bojić v Switzerland* App no. 61838/10 (ECtHR 18 October 2016) and *Mehmedovic v Switzerland* App no. 17331/11 (ECtHR 11 December 2018).

[167] *Leander v Sweden* App no. 9248/81 (ECtHR 26 March 1987) and *Rotaru v Romania* App No. 28341/95 (ECtHR 4 May 2000).

[168] *Breyer v Germany* App no. 50001/12 (ECtHR 30 January 2020).

[169] *Z. v Finland* App no. 22009/93 (ECtHR 25 February 1997), *M.S. v Sweden* App no. 20837/92 (ECtHR 27 August 1997) and *Satamedia v Finland* App no. 931/13 (ECtHR 27 June 2017).

[170] *Gaskin v the UK* App no. 10454/83 (ECtHR 7 July 1989) and *Haralambie v Romania* App no. 21737/03 (ECtHR 27 October 2009).

[171] *Rotaru v Romania* App No. 28341/95 (ECtHR 4 May 2000).

Therefore, the ECtHR through the interpretation of Art. 8 has expanded the protection of personal data of individuals, even in cases where the individual's private life was unaffected.

Additionally, the Council of Europe adopted the Convention for the Protection of Individuals with regards to Automatic Processing of Personal Data (Convention 108) in 1981, which was modernized in 2018 (Convention 108+). The purpose of the Convention is to secure in the territory of every signatory state the respect for the individuals' rights and fundamental freedoms and in particular their right to privacy, with regard to automatic processing of personal data relating to them, which identifies as 'data protection' (Art. 1 Convention 108+). The importance of this convention was threefold: it protects in a legally binding instrument personal data, it links data protection to the safeguarding of rights and fundamental freedoms in general, and it also connects data protection to privacy and private life.[172] This convention also provided a fertile ground for the adoption and update of national data protection legal frameworks of EU countries.[173]

The ECtHR relied on Convention 108 for assessing the potential interference with the guarantees that protect personal data, as part of the right to respect for private life (Art. 8 ECHR).[174] This convention is expected to become the international standard on privacy in the digital age[175] and the amending protocol aims to be fully consistent with the GDPR, which will lead to the convergence of both regimes. This is particularly important since the Art. 8 ECHR is not fully applicable to private parties, it does not apply to every kind of personal data, and it covers a more restrictive range of activities and information rights,[176] than the European data protection legislation.

---

[172] Gonzalez Fuster (n 155) 89.

[173] ibid 92.

[174] *Z. v Finland* App no. 22009/93 (ECtHR 25 February 1997), para 95; *Malone v the UK* App no. 8691/79 (ECtHR 2 August 1984), concurring opinion of Judge Pettiti.

[175] At the time of writing, 55 States have ratified the Convention 108 including 9 States non-members of the Council of Europe <https://www.coe.int/en/web/conventions/full-list/-/conventions/treaty/108/signatures> accessed 03/03/2022.

[176] Orla Lynskey (n 155) 113–128.

II.1.2.- European Union legal framework: primary legislation

The European Union legal framework is composed of primary and secondary norms that have a bearing on the protection of personal information. The Treaty of Lisbon is an important achievement in the protection of personal data in the EU for two main reasons. First, the EU was given the competence to legislate on data protection on Art. 16 TFEU. This is a salient fact because before that treaty the legislation on data protection was enacted to harmonise the functioning of the internal market (Art. 114 TFEU).[177] Second, it granted binding status to the Charter of Fundamental Rights, which had already a provision safeguarding the right to data protection.

In the context of the EU, the most important primary norms related to data protection and privacy are found in the Charter of Fundamental Rights. This instrument has two different articles dealing with these topics. Art. 7 of the Charter provides for the respect for private and family life, home and communications, whereas Art. 8 of the Charter guarantees the protection of personal data. Whenever personal data are processed the right to data protection could be impaired, thus its scope of application is wider than the right to respect private life. This is an important difference between the protection granted by the Charter and the ECHR since the latter does not include a specific article concerning data protection.

Art. 52(3) of the Charter attempts to provide consistency between the two normative systems, i.e. the ECHR and the Charter. Hence, it stipulates that where the rights in the Charter correspond to rights protected by the ECHR, the meaning and scope of the former are the same as those of the latter.

The EU legal system, more importantly, establishes strong protection of personal data through secondary legislation. Though there were national laws that guaranteed the protection of personal data before the establishment of the EU,[178] under EU law data protection was first mandated in 1995 by the Directive 95/46/EC on the protection of individuals with regard to the processing of personal data and the free movement of such data (Data Protection Directive) and then the protection was reinforced in 2002 by the Directive 2002/58/EC concerning the processing of personal data and the

---

[177] Paul Craig and Gráinne de Búrca, *EU Law. Text, Cases and Materials* (7th edn, OUP 2020) 360.

[178] For instance in Germany, see J Lee Riccardi, 'The German Federal Data Protection Act of 1977: Protecting the Right to Privacy?' (1983) 6 Boston College International and Comparative Law Review 243.

protection of privacy in the electronic communications sector (e-Privacy Directive). However, due to the speedy technological developments and to adapt data protection to the digital era, in 2016 the EU enacted the General Data Protection Regulation, which entered fully into force in May 2018. This was a breakthrough in the data protection legislative framework.

The Charter does not use the notion of justified interference, as the ECHR does in Art. 8 regarding the right to privacy and family life. Instead, the Charter relies on the concept of lawful limitation of the rights, horizontal derogation applicable to every right. Hence, it states in Art. 52 that any limitation to fundamental rights, including personal data protection, can be lawful only if: a) it is in accordance with the law; b) it respects the essence of the right; c) it respects the principle of proportionality; e) it is necessary; and f) it pursues an objective of general interest recognised by the EU, or the need to protect the rights of others.

II.2.- The General Data Protection Regulation

The General Data Protection Regulation (GDPR) entered into force on May 25, 2018, and it was the most comprehensive and detailed data protection regime ever enacted. The idea behind the protection of personal data is to reduce the chances of causing damage to the rights and freedoms of individuals by misusing or mishandling their personal data. Whenever personal data[179] is processed the GDPR applies.[180] This legal framework regulates the collection, storing and processing of personal data, stipulating the obligations of the data controllers and processors, and it includes obligations concerning data mining, aggregation and international transfers.[181] It also grants strong rights to individuals (so-called data subjects), since it provides a minimum inalienable level of protection that they cannot trade away.[182]

However, the GDPR does not establish a prohibition on the processing of personal data. Instead, it allows controllers and processors to process personal data when they

[179] According to art. 4(1) GDPR, personal data is 'any information to an identified or identifiable natural person'. However, the material scope of application of the GDPR is limited by art. 2 GDPR.

[180] Unless the processing falls under the exceptions stated in Art. 2(2) GDPR.

[181] Oreste Pollicino and Fernanda G Nicola, 'The Balkanization of Data Privacy Regulation' (2020) 123 West Virginia Law Review 115, 62.

[182] Orla Lynskey (n 155) 40.

show that their processing is grounded on a legal basis[183] and the processing operations implement data protection safeguards.[184]

The GDPR applies to the processing of personal data totally or partially carried out by automated means.[185] This comprehensive demarcation of the material scope of application implies that using personal data to develop AI systems or in the deployment of AI systems (e.g. using AI solutions to assist human decision-making) are included in the realm of the GDPR.

There are some tensions between the core provisions stemming from the GDPR and artificial intelligence. Companies implementing AI systems can increase their efficiency and productivity, cut down costs, and deliver newer insights and more accurate decisions. At the same time, AI systems have the potential to produce biased or unfair outcomes, take decisions with little accountability, massively spread misinformation, or even threaten democratic regimes. Therefore, there is a need to consider more closely the risks posed by these technologies when they process personal data.

Before addressing the specific rights that the GDPR grant to data subjects it is important to describe the data protection principles, as they will guide the analysis of the relevant rights and provisions of the GDPR.

II.2.1.- Data protection principles

The principles concerning the processing of personal data are listed and explained in Art. 5 GDPR. This is one of the pillars of the data protection regime and the signpost for those who plan to process personal data. The principles of data protection are purpose limitation, lawful, fair and transparent processing, data minimisation, accuracy, storage limitation, data security, and accountability. In this section, the data protection principles will be explained, except for the principle of accuracy that will be evaluated together with the measures controllers must implement to demonstrate

[183] Art. 6 GDPR.

[184] Orla Lynskey (n 155) 30.

[185] Art. 2(1) GDPR. It also applies to manual processing of personal data if it forms of a filing system or it is intended to form part of a filing system.

compliance with the data protection provisions. Lawfulness and transparency of the processing will also be assessed further below.

II.2.1.1.- Purpose limitation

The principle of purpose limitation is listed in Art. 5(1)(b) GDPR, after the principle of lawfulness, fairness and transparency. Why should it then be explained in the first place? Because if data controllers do not have specific, explicit and legitimate purposes to process personal data, they cannot start collecting such data, to begin with.

From the outset, data controllers must specify the purposes for which they collect and process personal data. The principle of purpose limitation is intended to delimit the boundaries in which personal data gathered for a certain purpose can be processed and, eventually, used for another different purpose.[186] The determination of the purposes has a pivotal role since it affects the scope of application of the applicable legal regime and it dictates the person that will be deemed controller and processor.[187]

As a rule, controllers can only collect and process personal data if the purposes are specified, explicit and legitimate. First, the specification of the purpose means that any purpose must be demarcated clearly, thus delimiting the extent of the processing operations. Controllers must refrain from collecting personal data that are unnecessary, inadequate or irrelevant for those specific purposes. For instance, phone apps that offer location services to discover coffee bars in the vicinity, generating at the same time user profiles to target ads.[188] In this case, profiling users for advertising is a different purpose related to the original one, i.e., geolocation services. Second, a purpose is explicit when it is expressed clearly and there is no ambiguity regarding its meaning and scope. Nor must the purposes be hidden or opaque. The expression of the purpose must allow data subjects to understand how their data will be used and make informed choices. This is all the more important considering the complexity and

---

[186] Article 29 Data Protection Working Party, 'Opinion 03/2013 on Purpose Limitation' (2013) 4.

[187] Maximilian von Grafenstein, *The Principle of Purpose Limitation in Data Protection Laws* (Nomos 2018) 235.

[188] Article 29 Data Protection Working Party, 'Guidelines on Automated Individual Decision-Making and Profiling for the Purposes of Regulation 2016/679' (2018).

opaqueness of many processing operations carried out online. Third, the legitimacy of the purposes is related to the legal grounds of processing.[189] Yet, this requirement goes beyond the legal grounds of processing and encompasses all the applicable law in the broadest sense,[190] like the prohibition of discrimination.

Data controllers must also refrain from processing personal data if the purposes of processing can be achieved by different means.[191] Processing personal data for vague, unlimited or undefined purposes is not allowed. It is also forbidden the inclusion of imprecise formulas such as that the personal information 'could' or 'may' be used for a different purpose in the future.

However, under some circumstances, controllers can process personal data for different purposes than those initially collected. To begin with, further processing for archiving purposes in the public interest, scientific or historical research purposes, or statistical purposes may not be considered incompatible with the initial purposes.[192] The GDPR establishes a presumption of compatibility in these cases. While the GDPR does not define the term 'scientific research', this concept should be given a wide interpretation that includes technological development or fundamental research.[193] It is argued that the development of AI algorithms may occasionally be covered by the concept and, thus, providers of AI systems may enjoy this exception.[194] Then, the data subject's consent or EU or Member State law, as long as it is a necessary and proportionate measure to safeguard an important public interest,[195] may allow data controllers to process personal data for purposes different from those concerning the initial collection.[196] Finally, purposes are compatible if the data subject could reasonably consider the further processing as predictable, appropriate or non-

[189] Art. 6 GDPR.

[190] Article 29 Data Protection Working Party, 'Opinion 03/2013 on Purpose Limitation' (n 185) 20.

[191] Recital 39 GDPR.

[192] Art. 5(1)(b) GDPR.

[193] Recital 159 GDPR.

[194] Datatilsynet, 'Artificial Intelligence and Privacy' (2018) 17.

[195] Those important public interests are listed in art. 23(1) GDPR. See also *Heinz Huber v Bundesrepublik Deutschland* (n 28).

[196] Art. 6(4) GDPR.

objectionable.[197] The *compatibility assessment* should entail, among other things, an evaluation of any connection between the purposes, the context of data collection -in particular the relationship between the data subject and the controller-, the nature of the personal data -whether it is sensitive or not-, the consequences of the further processing, and the existence of appropriate safeguards.[198]

While it may seem relatively straightforward to comply with this requirement in ordinary processing activities, AI-powered systems pose new challenges and may hinder the unrestrained deployment of AI technologies.[199] This is because it may be difficult for controllers to specify at the beginning the possible uses of collected,[200] and oftentimes the processing purposes, instead of being specified (before collecting the data), explicit and legitimate, are unclear at the collecting stage, and then re-purposing is a very common practice among controllers.[201] Big data projects frequently collect vast amounts of data with some predefined purposes, but then the algorithm may be able to find unexpected correlations among the data. So, the initial purposes, for which consent or another legal base was used, differ from the posterior intended purposes.

Requiring developers of AI systems to specify all the purposes for which the data collected will be used before starting the processing operations may hinder innovation,[202] but this should not be the case. Companies can obtain value from personal data, yet the unreasonable re-utilisation of data is not allowed by the principle

---

[197] Council of Europe, 'Explanatory Report to the Protocol Amending the Convention for the Protection of Individuals with Regard to Automatic Processing of Personal Data' (2018) para 49.

[198] Art. 6(4) GDPR.

[199] Merel Elize Koning, 'The Purpose and Limitations of Purpose Limitation' (Radboud University - PhD Thesis 2020) 4.

[200] Asia Biega and Michèle Finck, 'Reviving Purpose Limitation and Data Minimisation in Personalisation, Profiling and Decision-Making Systems' (2021) 21–04 15; European Data Protection Board and European Data Protection Supervisor, 'Joint Opinion 5/2021 on the Proposal for a Regulation of the European Parliament and of the Council Laying down Harmonised Rules on Artificial Intelligence' 9.

[201] Manon Oostveen, *Protecting Individuals Against the Negative Impact of Big Data: Potential and Limitations of the Privacy and Data Protection Law* (Wolters Kluwer 2018) 157.

[202] Giovanni De Gregorio and Raffaele Torino, 'Privacy, Protezione Dei Dati Personali e Big Data' in Vincenzo Franceschelli and Emilio Tosi (eds), *Privacy Digitale. Riservatezza e protezione dei dati personali tra GDPR e nuovo Codice Privacy* (Guiffrè Francis Lefebvre 2019) 467.

of purpose limitation.[203] Posterior processing for a diverse purpose is not automatically forbidden, since the compatibility evaluation should be carried out on an individual basis.[204]

To evaluate whether the further processing would be compatible with the initial purposes, the Article 29 Data Protection Working Party (hereinafter WP29)[205] distinguished two situations. Firstly, where controllers aim at discovering correlations in the data. In this case, they must observe the principle of functional separation, clearly identifying and separating the different analytic operations. Secondly, where controllers aim at assessing or forecasting individuals' preferences and behaviours to make decisions that have an impact on them. The ulterior utilisation of personal data will not be deemed to be compatible in this second case if there is no 'free, specific, informed and unambiguous "opt-in" consent'.[206]

An alternative approach to appraising whether the intended further processing would be considered compatible is to evaluate the privacy impact of the new purpose for data subjects and their reasonable expectations concerning further utilisation. [207] In other words, if they could have a reasonable expectation that their information will be employed for this new purpose. Hence, the more unexpected for the individuals the new purpose is, the more likely it will be incompatible with the initial one.

II.2.1.2.- Lawful, fair and transparent data processing

The processing activities must comply with the law, be fair and transparent for data subjects (art. 5(1)(a) GDPR). In this section only fairness will be assessed, while 'lawfulness' and 'transparency' will be evaluated together with the legal basis for processing personal data (section II.2.2) and with the transparency obligations of controllers (sections III.2 and IV.1), respectively.

---

[203] Information Commissioner's Office, 'Big Data, Artificial Intelligence, Machine Learning and Data Protection Data Protection Act and General Data Protection Regulation' (2017) 37.

[204] Article 29 Data Protection Working Party, 'Opinion 03/2013 on Purpose Limitation' (n 185) 21.

[205] The Article 29 Data Protection Working Party was replaced by the European Data Protection Board (see Art. 94(2) GDPR).

[206] ibid 46.

[207] Information Commissioner's Office, 'Big Data, Artificial Intelligence, Machine Learning and Data Protection Data Protection Act and General Data Protection Regulation' (n 202) 38.

Fairness

Data subjects must be able to understand what controllers are doing with their personal data. Controllers must inform the data subject that they will process personal data related to him or her and how this processing will be carried out. The fairness requirement is linked to the need to process personal data ethically.[208] It requires an evaluation of both the expectations of individuals and the effects of the processing on natural persons.[209] Unfair processing can also be configured when the personal data were obtained or processed through unfair means, deception or concealment of the data subject.

Whereas the concept of fairness in the GDPR has different meanings,[210] fairness may entail two different aspects: informational and substantive fairness. To begin with, *informational fairness* is linked to the concept of transparency. For the processing of personal data to be fair, controllers must provide clear, concise, easily accessible and sufficient information to individuals about the processing operations on their personal data. Controllers should act in an honest manner and they cannot mislead individuals. They must communicate data subjects, for instance, the purposes of the processing, the storage period, the data subject's rights,[211] the categories of personal data and the sources of collection where the data is not provided by the data subject,[212] and the existence of automated decision-making, including profiling, along with meaningful information about the logic involved, and the significance and the envisaged consequences for them.[213] Fairness of the processing operations includes, for instance, the requirement to communicate to employees the automated processing operations aimed at scoring and ranking employees (riders) to assign priority time slots

[208] European Union Agency for Fundamental Rights and Council of Europe, *Handbook on European Data Protection Law* (2018) 119.

[209] Gianclaudio Malgieri, 'The Concept of Fairness in the GDPR: A Linguistic and Contextual Interpretation', *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency* (2020) 10.

[210] See Malgieri (n 207).

[211] Art. 13 and 14 GDPR.

[212] Art. 14(1)(d) and 14(2)(f) GDPR.

[213] Art. 13(2)(f) and 14(2)(g) GDPR.

for delivery orders.[214] The issues concerning informational fairness will be dealt together with the transparency section.

On the other hand, *substantive fairness* relates to the effects that the processing operations have on the data subjects. In this context, substantive fairness requires that providers and users of AI systems take all the measures necessary to avoid discrimination, biases or inaccuracies in the context of automated decision-making,[215] which includes both inferences and decisions taken by automated means.

The inferences and decisions about natural persons taken through automated means must be reliable, relevant and acceptable because individuals have a reasonable expectation that the outputs of an AI system will be statistically accurate and unbiased. For instance, in the context of gig workers, substantive fairness is compromised where the AI system treats equally those who do not participate in the booked session for trivial reasons and those who exercise a fundamental right (e.g., the right to strike, parental or sick leave).[216]

The importance of this principle is highlighted in the AIA draft which requires providers of AI systems to train, validate and test datasets to, among other purposes, examine possible biases.[217] Furthermore, it mandates that providers of AI systems

---

[214] Garante per la Protezione dei Dati Personali, *Ordinanza ingiunzione nei confronti di Foodinho SRL* (2021) 9675440 [3.3.1].

[215] Recital 71 GDPR.

[216] Tribunale Ordinario di Bologna (Sez. Lavoro), *FILCAMS CGIL Bologna e altri v Deliveroo Italia SRL* (2020) 2949/2019. The Bologna court declared Deliveroo's algorithm has discriminatory effects regarding the booking of sessions. Riders have two ways to receive trips. They can either book sessions via the self-booking service (SSB) in advance or via the free-login system in real-time. The self-booking service provides riders with a schedule of the incoming week's available slots to receive trip requests according to a ranking. The ranking was determined by the reliability index (number of times in which the rider had not participated in a session they already booked) and the peak participation index (number of times the rider is available for the most relevant times: 20:00 to 22:00 hrs from Friday to Sunday). The court considered that riders' fundamental rights were compromised since whenever a rider chooses to exercise certain fundamental rights it has a negative impact on his/her statistics, regardless of the justification. Hence the AI system conditioned future chances of work. The company's profiling system (based on reliability and peak participation) treated equally those who do not participate in the booked session for trivial reasons and those who are on strike (or are sick, are disabled, assist a disabled person or a sick child, etc.), discriminating against the latter.

[217] Art. 10(2)(f) AIA draft.

declare in the instructions of use the levels of accuracy and the relevant accuracy metrics of high-risk AI systems,[218] and these metrics should be informed to AI users.[219]

II.2.1.3.- Data minimisation

Data minimisation aims at constraining the unlimited collection of personal data, and it is the necessary corollary of purpose limitation.[220] The principle of data minimisation requires data controllers to collect and process only personal data that is adequate, relevant and limited to what is necessary for the purposes for which it is collected.[221] Controllers must refrain from collecting data that is not directly and strictly pertinent to the concrete purposes followed by the processing process.

Even if the purpose of processing is noble and genuine (e.g. fighting against serious crime), and necessary to achieve an objective of general interest, it does not automatically mean that the controller can collect more data than necessary. Indeed, whenever a measure covers, in a generalised manner all individuals, all means of electronic communications and all traffic data 'without any differentiation, limitation or exception' according to its objectives, it violates the principle of data minimisation.[222] More concretely, the principle of data minimisation is violated when an AI system, like a ride-hailing app, gathers and keeps all the information concerning the orders and it pre-authorises other operators to employ the data collected by the app.[223]

The principle of data minimisation may also clash with artificial intelligence models. The basic functioning of some AI models is grounded on their ability to learn from data. Instead of identifying the pertinent data points and then collecting them, as traditional prediction or classification systems do, in general, AI systems need to collect vast amounts of information, including personal data, from their inception.[224] This is because the more data is available to train the AI algorithms, the more statistically

---

[218] Art. 15(2) AIA draft.

[219] Recital 49 AIA draft.

[220] Biega and Finck (n 199) 25.

[221] Art. 5(1)(c) GDPR.

[222] *Digital Rights Ireland Ltd v Minister for Communications* (n 33) para 57.

[223] *Ordinanza ingiunzione nei confronti di Foodinho SRL* (n 212).

[224] Gregorio and Torino (n 201) 469.

accurate their predictions or classifications will be.[225] Thus, there is a need to collect as much data as possible to feed the system. Equally important, some models can draw unexpected inferences from data, adding more reasons to ingest all the data. Hence, reducing the amount of personal data collected collides with the very notion of AI-powered systems,[226] since massive data gathering is inherent in their business models.[227]

Not only does the principle of data minimisation concerns the quantity of data gathered, but also if the personal data processed is necessary for the purposes for which it was collected.[228] In other words, data controllers must process the least possible quantity of personal information to achieve their objectives.[229]

II.2.1.4.- Accuracy

Data controllers must ensure the data they store is accurate and up-to-date and, where needed, update the personal data. It applies to any processing of personal data falling under the GDPR and it covers factual and temporal accuracy.[230] To comply with this requirement controllers must take every reasonable step to guarantee that inaccurate personal data are erased or rectified immediately,[231] even using suitable mathematical or statistical procedures for the profiling and through the implementation of technical and organisational measures.[232]

The principle of accuracy is also related to fair processing. This obligation implies that controllers and processors may need to control the data regularly, otherwise, it may have detrimental effects on data subjects. That is to say, where the data is inaccurate, it can lead to unfair decisions, such as a low credit score if the data subject is mistakenly flagged as a debtor.

---

[225] Lehr and Ohm (n 81) 225.

[226] Oostveen (n 200) 157.

[227] European Union Agency for Fundamental Rights and Council of Europe (n 206) 356.

[228] Information Commissioner's Office, 'Big Data, Artificial Intelligence, Machine Learning and Data Protection Data Protection Act and General Data Protection Regulation' (n 202) 40.

[229] Biega and Finck (n 199) 27.

[230] Dara Hallinan and Frederik Zuiderveen Borgesius, 'Opinions Can Be Incorrect (in Our Opinion)! On Data Protection Law's Accuracy Principle' (2020) 10 International Data Privacy Law 1, 3–4.

[231] Art. 5(1)(d) and Recital 39 GDPR.

[232] Recital 71 GDPR.

In the context of AI solutions, it may be challenging for controllers to assure the accuracy of the information. It is important from the outset to differentiate between the principle of accuracy in data protection, which was previously outlined, and accuracy in AI or statistical accuracy which is the probability that the AI solution will deliver the right outcome.[233] Whereas some level of inaccuracy in the data used as input or the data produced as output is accepted in AI systems[234] (because they discover general tendencies or trends), such inaccuracies may harm individuals when they are employed to create profiles or deliver inferences about individuals.[235] Concerning the latter, it is worth mentioning that the AIA draft requires providers of AI systems to design and develop AI systems to achieve an adequate level of accuracy[236] and to declare the levels of accuracy and their relevant metrics,[237] in particular the overall expected level of accuracy in relation to the intended purpose of the AI system.[238]

There is a different kind of problems concerning accuracy. Firstly, issues concerning AI systems trained with inaccurate data. Oftentimes, the data employed to train AI systems come from diverse sources, i.e, third parties. As multiple sources provide information it is more likely to find inaccuracies.[239] Then, malicious actors may attack the algorithm and insert inaccurate data into the training dataset, resulting in erroneous outcomes. Secondly, problems related to the design of the algorithm. The AI solution may in itself be incorrectly designed or developed. Developers of rule-based expert systems may unintentionally design the algorithm with errors, thus producing flawed outcomes.[240] Thirdly, even if the system delivers accurate predictions at a certain moment, in real-life environments the groups of individuals to which the decisions are applied, or their characteristics, are dynamic and change over time. Consequently, the AI system may show an increasing inaccuracy rate as the underlying population or

---

[233] Information Commissioner's Office, 'Guidance on AI and Data Protection' (n 147) 35.

[234] Recital 71 acknowledges that these operations are prone to errors so it states the need to reduce them, not to eliminate them.

[235] Information Commissioner's Office, 'Big Data, Artificial Intelligence, Machine Learning and Data Protection Data Protection Act and General Data Protection Regulation' (n 202) 43.

[236] Art. 15(1) AIA draft.

[237] Art. 15(2) AIA.

[238] Annex IV, point 3 AIA draft.

[239] Gregorio and Torino (n 201) 470.

[240] Agencia Española de Protección de Datos (n 149) 33.

their behaviours change. This reduction in performance is called model or concept drift.[241] Finally, the uneven representativeness of the data has consequences for under-represented groups. AI systems tend to be more accurate for groups where data used to train the model is highly representative of them. Conversely, if training data is not representative of particular demographic groups (e.g. minorities, disabled people or women), the accuracy of the outcomes of AI systems (predictions or inferences) concerning underrepresented groups may greatly suffer.[242] For this reason, the AIA draft requires that providers of AI systems also inform the level of accuracy 'for specific persons or groups of persons on which the system is intended to be used'.[243]

II.2.1.5.- Storage limitation

Data controllers must not keep the personal data in a form that allows the identification of data subjects for a longer period than they need to fulfil the purposes for which it was originally collected.[244] Personal data may be stored for longer periods if it will be processed solely for archiving purposes in the public interest, scientific or historical research purposes, or statistical purposes, provided that appropriate safeguards are implemented.[245] Additionally, if the personal data has been anonymized storage limitations do not apply (because anonymized data is not personal data, thus falling outside the scope of application of the GDPR).

The retention of the personal data by the data controller must be proportionate to the objective of collection and restricted in time, especially in the police sector.[246] But proportionality also plays a role in the mass surveillance of citizens for security

---

[241] Information Commissioner's Office, 'Guidance on AI and Data Protection' (n 147) 38.

[242] Information Commissioner's Office, 'The Use of Live Facial Recognition Technology in Public Places' (2021) 21.

[243] Annex IV, point 3 AIA draft.

[244] Art. 5(1)(e) GDPR.

[245] Art. 5(1)(e) GDPR.

[246] *S. and Marper v the UK* Apps nos. 30562/04 and 30566/04 (ECtHR 4 December 2008) on the indefinite retention of biometric information of individuals who were arrested but then acquitted.

reasons.[247] Yet, where companies store many categories of personal data, gathered for diverse purposes, they must be careful and establish clear and the shortest possible retention periods to avoid administrative proceedings from supervisory authorities.[248]

In the realm of AI solutions, the principle of storage limitation may hinder the development of big data applications.[249] However, this limitation can be at least partially overcome if the personal information is used exclusively for statistical purposes.[250] *Statistical purposes* entail the collection and processing of personal data for statistical surveys or the production of statistical results[251] and analysis.[252] To be considered a statistical purpose, though, there are two further requirements. First, the outcome of the processing operations should deliver aggregate data, not personal data. This means that the personal data is collected and communicated in a summarised manner to conduct statistical analysis, for instance, to identify trends, and compare or gain insights. Second, the outcomes of the processing operation (e.g. trends, insights, comparisons) or the personal data used to carry out the statistical analysis should not be employed to support actions or decisions concerning any particular individual.[253]

This provision attempts to balance, on the one hand, the legitimate interests of those carrying out statistical analysis and, on the other, the fundamental rights of individuals to the protection of their personal data. However, the inclusion of this exception has limited practical implications for developers and users of AI systems

---

[247] *Digital Rights Ireland Ltd v Minister for Communications* (n 33) para 64. In this case the CJEU found that the storage of personal data for a period of 2 years was a breach of fundamental rights, if the retention policies do not make any distinction on the categories of data or individuals.

[248] *Ordinanza ingiunzione nei confronti di Foodinho SRL* (n 212). In this case the company, for instance, used to erase several categories of personal data 4 years after the expiry of the employment relationship; customer care call metadata after 4 years; and geolocation data after 10 months.

[249] Gianclaudio Malgieri, '5. Principi Applicabili Al Trattamento Di Dati Personali' in Oreste Pollicino and Roberto D'Orazio (eds), *Codice della Privacy e Data Protection* (Giufrè Francis Lefebvre 2021) 188.

[250] European Parliamentary Research Service, 'The Impact of the General Data Protection Regulation (GDPR) on Artificial Intelligence' (2020) 49.

[251] Recital 162 GDPR.

[252] Art 3(8) Regulation No 223/2009 on European Statistics.

[253] Recital 162 GDPR.

since they, almost in every situation, will need to implement adequate safeguards, i.e., technical and organisational measures to protect the rights of data subjects.

II.2.1.6.- Data security

Controllers and processors must process personal data ensuring appropriate security of the information. Data security is not only a core element of data protection[254] but also a precondition for lawful data processing.[255]

This obligation includes confidentiality, integrity and availability of personal data. *Confidentiality* relates to the provision of protection against unauthorised or unlawful disclosure, or access to, personal data processed,[256] which includes facilitating access to the equipment used for the processing.[257] Then, *integrity* concerns the provision of protection against accidental or unlawful modification or damage of personal data.[258] Finally, *availability*, while not mentioned in Art. 5(1)(f) GDPR, is an essential element of data security and it entails the protection against unintentional or unauthorised loss of access to, or destruction of, personal data.[259]

This is an obligation of means since both controllers and processors must use appropriate technical or organisational measures to ensure a level of security appropriate to the risk.[260] Absolute security against data breaches can never be guaranteed. Controllers and processors must consider, among others, the state of the art, the costs of implementation, the nature, scope, context and purposes of the processing, and the risks of varying likelihood and severity for the rights of data

---

[254] *Digital Rights Ireland Ltd v Minister for Communications* (n 33) para 40.

[255] European Union Agency for Cybersecurity, 'Artificial Intelligence Cybersecurity Challenges. Threat Landscape for Artificial Intelligence' (2020) 9.

[256] Art. 5(1)(f) GDPR.

[257] Recital 39 GDPR.

[258] Art. 5(1)(f) GDPR.

[259] Article 29 Data Protection Working Party, 'Guidelines on Personal Data Breach Notification under Regulation 2016/679' (2018) 7.

[260] Art. 5(1)(f) and 32(1) GDPR.

subjects.[261] Among the most important techniques to ensure a high level of security are encryption[262] and pseudonymisation[263] of personal data.

AI systems present new opportunities to attackers compared to traditional applications that process personal data. Traditional applications are easier to protect because developers have a clearer perspective of the different threats that can harm the system. However, the complexity of artificial intelligence, in particular machine learning solutions, provides more avenues for intrusions and security incidents. While the European Union Agency for Cybersecurity identified 96 threats that can affect machine learning algorithms,[264] malicious agents can exploit AI solutions' vulnerabilities in three main ways. First, malicious actors can attack the model itself by exploiting *algorithmic design flaws*. The mathematical underpinnings of the model may have some imperfections or vulnerabilities and these weaknesses may be profited by attackers, either by abusing them or extracting important parts of the algorithmic code.[265] Second, malicious actors can poison the data used to train or test the model. *Data poisoning* consists in inserting intentionally incorrect, altered or mislabelled data in the training or testing dataset, which especially affects machine learning techniques.[266] Third, bad actors can create malign *adversarial examples* to deceive the model and miscategorise or mislabel the algorithmic outcomes, altering the normal

---

[261] Art. 32(1) GDPR.

[262] Encryption is the cryptographic modification of original data into a form that hides the data's authentic meaning to avoid being discovered or employed. See NIST Special Publication 800-82 (Rev. 2). Guide to Industrial Control Systems (ICS) Security (2015) page 6-35, <https://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.800-82r2.pdf> accessed 18/03/2022.

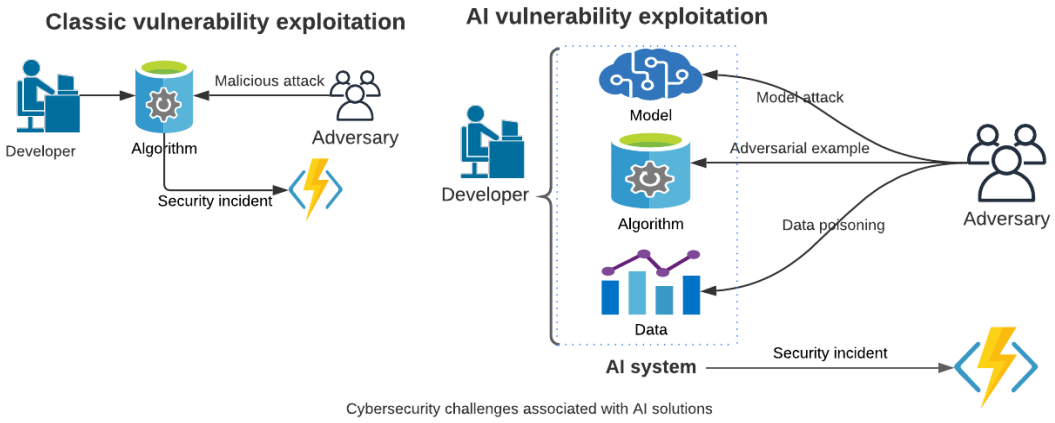
[263] Pseudonymisation is a processing technique whose objective is to de-identify the natural person. It involves the removal and change of certain data attribute (e.g. data subject's name) and the separation of the information that allows re-identification (see art. 4(5) GDPR).

[264] European Union Agency for Cybersecurity, 'Artificial Intelligence Cybersecurity Challenges. Threat Landscape for Artificial Intelligence' (n 253) 43–57.

[265] European Commission - Joint Research Centre, 'Artificial Intelligence - A European Perspective' (2018) 90.

[266] ibid; European Union Agency for Cybersecurity, 'Artificial Intelligence Cybersecurity Challenges. Threat Landscape for Artificial Intelligence' (n 253) 45. The AIA draft defines 'data poisoning' as interventions aimed at manipulating the training data, see art. 15(4) AIA draft.

functioning of the system.[267] Adversarial examples are constituted by altered or perturbed information that cannot be directly identified by humans, yet they greatly impair the performance of some AI algorithms, for instance, to identify or classify images.



Adapted from European Commission - Joint Research Centre, 'Artificial Intelligence - A European Perspective' (2018) 89

It is important to bear in mind that some models contain personal data by default within the model (e.g. support vector machines, k-nearest neighbours or decision trees) so whoever procures these models may access personal data contained in the dataset. However, this should not be regarded as an attack or a data breach. Instead, it should be considered lawful processing of personal data.[268]

Therefore, it is crucial that when deploying a machine learning algorithm adequate technical and organizational measures are put in place. It is important to highlight that the GDPR does not provide a precise or concrete list of actions that controllers or processors should implement to ensure security in the processing of personal data.[269] This principle is reaffirmed in the AIA draft where it requires that the measures aimed at guaranteeing the security of high-risk AI systems must be adequate to the

[267] European Commission - Joint Research Centre, 'Artificial Intelligence - A European Perspective' (n 263) 90; European Union Agency for Cybersecurity, 'Artificial Intelligence Cybersecurity Challenges. Threat Landscape for Artificial Intelligence' (n 253) 43. The AIA draft defines 'adversarial examples' as input data aimed at causing the algorithm to make a mistake, see art. 15(4) AIA draft.

[268] Information Commissioner's Office, 'Guidance on AI and Data Protection' (n 147) 55.

[269] Alessandro Mantelero and others, 'The Common EU Approach to Personal Data and Cybersecurity Regulation' (2021) 28 International Journal of Law and Information Technology 297, 301.

circumstances and risks.[270] Hence it is for the controllers and processors to determine, following a risk-based assessment, the particular measures to satisfy the requirements outlined in Art. 5(1)(f) and 32 GDPR.

Failing to provide adequate security to handle personal data constitutes a punishable action and is liable to fines. The case *Ticketmaster* exemplifies the importance of considering security aspects when processing personal data using AI systems.[271] Ticketmaster employed for its services an AI-powered chatbot and the latter suffered a cyberattack. The chatbot was designed by a third-party vendor (Inbenta Technologies) to interpret users' questions, automatically identifying relevant helpful articles or information. While the chatbot was hosted on the server of the third-party vendor, Ticketmaster (the data controller) offered its visitors assistance through the chatbot on various pages of its website, including the payment and checkout page. An attacker directed its attack at the third-party vendor's servers and inserted a malicious code into the chatbot. The malicious code gathered information provided by users, including financial information contained in payment cards such as names, payment card numbers, expiry dates and CVV numbers,[272] and forwarded it to the attacker (a technique known as 'scrapping'). Nearly 10 million customers were potentially affected by the incident and the company received a £1.25m fine.

On the other hand, another set of cases that failed to implement appropriate security measures concerns the ride-hailing apps Deliveroo and Glovo. In these cases, the Italian Data Protection Authority considered that providing by default access rights to operators who did not need access to the whole data of data subjects to perform their duties contravenes Art. 32 GDPR, since it does not ensure the confidentiality of the data.[273]

Hence, whenever companies implement AI solutions they must properly assess the risks of using these technologies and identify and implement adequate technical and organisational security measures to counter the increased risks posed by AI-enhanced technologies.

---

[270] Art. 15(4) AIA draft.

[271] Information Commissioner's Office, *Ticketmaster UK Limited* (2020) COM0759008.

[272] ibid 3.29.

[273] Garante per la Protezione dei Dati Personali, *Ordinanza ingiunzione nei confronti di Deliveroo Italy SRL* (2021) 9685994; *Ordinanza ingiunzione nei confronti di Foodinho SRL* (n 212).

II.2.2.- Lawful processing and lawful bases for processing

The processing of personal data is lawful only if it is carried out under one or more of the legitimate grounds provided for in the legislation.[274] Art. 6(1) GDPR indicates in a detailed fashion the objectives of general interest necessary to protect the rights of individuals in the field of data protection.[275] The legitimate grounds can be broadly divided into the existence of consensus or necessity. The first ground is the consent from the data subject. Processing personal data is lawful when the controller obtains the consent from the data subject. Then, when a situation of necessity is invoked, it could be grounded on the following cases: a) entering into or the performance of a contract; b) the compliance of a legal obligation; c) the protection of the vital interests of the data subject; d) the performance of a task carried out in the public interest or the exercise of the official authority; e) the legitimate interest of the controller or a third party.[276] The distinction has practical implications since except for the consent of the data subject, the controller must demonstrate the necessity of processing personal data to achieve its aims. While the six bases mentioned provide valid legal ground to process personal data, the two most important and currently invoked are consent and legitimate interests.[277]

For these purposes, it is helpful to distinguish between the two broad stages of AI: development and deployment.[278] This is because for each of these phases controllers may invoke different purposes. Assessing the purposes separately for the development and the deployment stages will be important, first, where the AI solution is designed for a general-purpose (facial recognition) but later implemented for

---

[274] Cécile de Terwangne, 'Article 5. Principles Relating to Processing of Personal Data' in Christopher Kuner, Lee A Bygrave and Christoper Docksey (eds), *The EU General Data Protection Regulation (GDPR): A Commentary* (OUP 2020) 314.

[275] Waltraut Kotschy, 'Article 6. Lawfulness of Processing' in Christopher Kuner, Lee A. Bygrave and Christopher Docksey (eds), *The EU General Data Protection Regulation (GDPR): A Commentary* (OUP 2020) 326.

[276] Art. 6 GDPR.

[277] Oostveen (n 200) 140.

[278] Information Commissioner's Office, 'Guidance on AI and Data Protection' (n 147) 29.

different objectives (e.g., unlocking the phone or law enforcement). Second, where the controller procures an AI solution from a vendor, the purpose for processing personal data of the latter (e.g., developing the algorithm) will be different from the former's (e.g. evaluating the creditworthiness of natural persons). Finally, if the system is intended to produce, in the deployment stage, automatic decisions that have a legal or similarly significant effect Art. 22(2) GDPR may apply, regardless of whether in the development of the algorithm personal data was processed.[279]

In the following sections, the basis for the legal processing of personal data will be assessed.

II.2.2.1.- Consent

The data subject must consent to the processing of his or her personal data for single or multiple purposes.[280] The GDPR establishes several conditions for considering valid consent. Consent is a freely given, specific, informed and unambiguous indication of the individual's wishes by which he or she agrees to the data processing.[281]

The consent must be freely given, which means that the person must be able to exercise a genuine choice and that he or she will not suffer deception, intimidation or serious detrimental consequences if he or she refuses or withdraw consent.[282] Consent is presumed to have been freely given where[283] a) it is unconditional or unbundled to the provision of services, except where the processing of personal data is necessary for the performance of the contract;[284] b) there is a balance of powers between the parties;[285] c) it is granular, meaning that the procedure for obtaining

[279] ibid 29–30.

[280] Art. 6(1)(a) GDPR.

[281] Art. 4(11) GDPR.

[282] Recital 42 GDPR.

[283] European Data Protection Board, 'Guidelines 05/2020 on Consent under Regulation 2016/679' (2020) 10–12.

[284] Art. 7(4) GDPR.

[285] Recital 43 GDPR.

consent permits the person consent separately for different data processing activities.[286]

Additionally, consent must be specific to the processing purpose. The specificity refers to the purpose of processing, irrespective of the companies through which the purposes are fulfilled,[287] or the Member State to which the data could be sent.[288] It must cover all processing operations and where the processing has multiple purposes, consent must be given for all of them.[289] Besides, consent must be informed, in the sense that data subjects must have sufficient information before granting consent to processing operations[290] and be fully aware of the consequences of consenting. Information must be provided in clear, concise and understandable language.[291] Finally, consent must be unambiguous, meaning that there must be no doubt whatsoever that the data subject signified agreement to the processing of his or her personal data. It is important to bear in mind that the data subject's silence or inactivity, like the omission to deselect a pre-checked checkbox to decline consent,[292] cannot constitute valid consent.[293] In other words, data subjects' consent is only valid where it is obtained through an opt-in mechanism.[294]

In the realm of AI systems, it may be difficult to obtain valid consent from data subjects due to the intrinsic features and operation of AI algorithms.[295] To begin with, consent can only constitute a lawful basis for processing personal data if the

---

[286] Recital 43 GDPR.

[287] Case C-543/09, *Deutsche Telekom AG v Bundesrepublik Deutschland*. [2011] ECLI:EU:C:2011:279, para 61.

[288] Case C-536/15, *Tele2 (Netherlands) BV and Others v Autoriteit Consument en Markt (ACM)*. [2017] ECLI:EU:C:2017:214, para 40-41.

[289] Recital 32 GDPR.

[290] Case C-40/17, *Fashion ID GmbH & CoKG v Verbraucherzentrale NRW eV*. [2019] ECLI:EU:C:2019:629, para 41.

[291] Recital 39 GDPR.

[292] Case C-673/17, *Bundesverband der Verbraucherzentralen und Verbraucherverbände v Planet49 GmbH*. [2019] ECLI:EU:C:2019:801, para 57.

[293] Recital 32 GDPR.

[294] Pollicino and Nicola (n 180) 38.

[295] Oostveen (n 200) 139.

individuals were clearly *informed* about the processing operations.[296] The Italian Court of Cassation ruled that consent is not valid if individuals subject to an automated decision-making system that may influence their rights are not adequately informed about the logic behind it. The Court of Cassation further considered that: a) data subjects must be informed about the essential elements of the processing for consent to be valid; b) adhering to a platform does not imply the acceptance of an automated system that makes a score of data subjects using their personal data if they are not aware of the 'executive scheme' (i.e. the logic involved) and the constitutive elements of the algorithm.[297]

As elaborated under the principle of transparency, obtaining informed consent from the data subjects is particularly challenging where their personal information is processed using AI systems. It may be problematic to ensure that the data subject's choice was genuine when it accepts the processing of his or her personal data for the complex processing activities involved in AI solutions[298] The consent in the online environment is generally obtained after having offered the data subjects the possibility to read long privacy policies, with no chance to negotiate their terms and conditions. Since individuals often do not read or, if they do, do not comprehend them, it is argued that online consent lacks the appearance of informational self-determination that once had.[299]

Then, the *specificity* of consent poses also hurdles when trying to obtain valid consent from individuals to process personal data by AI systems. Many artificial

---

[296] Human Rights Council, 'Artificial Intelligence and Privacy, and Children's Privacy. Report of the Special Rapporteur on the Right to Privacy, Joseph A. Cannataci. A/HRC/46/37' (2021) 5.

[297] Italian Court of Cassation, *Garante per la Protezione dei Dati Personali v Associazione Mevaluate Onlu* (2021) case 14381. The Italian Data Protection Authority had ordered the suspension of the implementation of a reputation rating system Mevaluate in 2016. A court in Rome partially ruled in favour of the company, but the Court of Cassation quashed the judgment. The system consisted of a web platform and an IT archive. It collected and processed personal data from documents uploaded voluntarily to the platform by the users or collected by the company from the web. Through an algorithm, the system would then assign a 'reputational rating', i.e., alphanumeric indicators capable of measuring the reliability of individuals in the economic and professional fields.

[298] Information Commissioner's Office, 'Guidance on AI and Data Protection' (n 147) 30.

[299] 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 and Technology Review 18, 66.

intelligence solutions cannot at the outset specify concretely every purpose for which the information collected will be used, and repurposing of data -i.e., employing personal data for a different purpose from which it was initially gathered- is frequent when developing AI solutions. The intrinsic nature of many artificial intelligence solutions is the discovery of unforeseen relationships in the datasets.[300] Then, if these new purposes were not detailed in the consent notice, it is argued that a renewed contact with the data subjects to obtain an updated consent would be required.[301]

II.2.2.2.- Necessity

Processing of personal data is lawful provided it is necessary to enter into or fulfil a contract with the data subject,[302] comply with a legal obligation,[303] protect vital interests of the data subject or another individual,[304] or perform a public interest task or an official authority task vested in the controller.[305]

The necessity for the performance of a *contract* should be interpreted restrictively.[306] In this case, the processing is considered necessary if the contract cannot be completed without the processing activities.[307] Hence, processing that is useful for the controller yet objectively unnecessary for fulfilling the contract is out of the scope of this provision.[308] It covers processing operations that are essential for managing the contract (e.g., billing and delivery). However, it does not cover further

---

[300] Michael Froomkin, 'Big Data: Destroyer of Informed Consent' (2019) 21 Yale Journal of Law and Technology 27, 32.

[301] Fred H Cate and Viktor Mayer-Schö, 'Notice and Consent in a World of Big Data' (2013) 3 International Data Privacy Law 67, 67.

[302] Art. 6(1)(b) GDPR.

[303] Art. 6(1)(c) GDPR.

[304] Art. 6(1)(d) GDPR.

[305] Art. 6(1)(e) GDPR.

[306] Article 29 Data Protection Working Party, 'Opinion 06/2014 on the Notion of Legitimate Interests of the Data Controller under Article 7 of Directive 95/46/EC' (2014) 16.

[307] Paul Voigt and Axel von dem Bussche, *EU General Data Protection Regulation (GDPR): A Practical Guide* (Springler 2017) 102.

[308] European Data Protection Board, 'Guidelines 2/2019 on the Processing of Personal Data under Article 6(1)(b) GDPR in the Context of the Provision of Online Services to Data Subjects' (2019) 8.

processing for service improvement,[309] profiling,[310] making predictions about the data subject,[311] or behavioural advertising since these activities are not objectively necessary for the performance of the contract.

It can be argued, though, that content personalisation can be covered by this provision if the processing is inherent to the service provided, which entails evaluating the nature of the service, the reasonable expectations of the users, and if it is not possible to personalise content in the absence of this processing.[312]

On the other hand, it is unlikely that *legal obligation, exercising a public task or protecting vital interests* provide a ground for developing an AI solution.[313] Whereas it is possible to conceive the deployment or use of an AI solution to, for instance, detect tax fraud (complying with a legal obligation or exercising a public task), developers of AI systems may exceptionally use these legal bases to process personal data in the developing stage of the AI system lifecycle.

II.2.2.3.- Legitimate interest pursued by the controller or by a third party

Processing also is legitimate when the controller or a third party pursues a legitimate interest (be it legal, economical, direct marketing, cybersecurity, etc), except where such interests are overridden by the interests or fundamental rights of the data subject which require protection of personal data.[314] It protects, for instance, purposes related to direct marketing, intra-group processing, and fraud prevention.[315]

Contrary to the five other grounds of processing, where the controller invokes legitimate interests as a legal basis to process personal data there is no presumption that the balance among the multiple interests concerned is fulfilled.[316] Hence, a balance must be carried out to assess whether the legitimate interests of the controller

---

[309] ibid 14.

[310] Article 29 Data Protection Working Party, 'Guidelines on Automated Individual Decision-Making and Profiling for the Purposes of Regulation 2016/679' (n 187) 13.

[311] European Parliamentary Research Service (n 248) 50.

[312] Information Commissioner's Office, 'Guidance on AI and Data Protection' (n 147) 31.

[313] ibid 32.

[314] Art. 6(1)(f) GDPR.

[315] Recital 47 GDPR.

[316] Article 29 Data Protection Working Party, 'Opinion 06/2014 on the Notion of Legitimate Interests of the Data Controller under Article 7 of Directive 95/46/EC' (n 304) 9.

or a third party prevail over the rights of the data subject.[317] This can be achieved by a three-pronged test generally referred to as 'legitimate interests assessment' (LIA),[318] whereby controllers must, first, recognise the legitimate interest pursued (purpose test), second, they must evaluate whether the processing is necessary for the purposes (necessity test), and, finally, they must assess the data subject's rights or interests (balancing test). The result of this test determines if this provision can be used as a legal basis for processing personal data.

The *evaluation of the controller's legitimate interests* must be made on a case-by-case basis. It requires a link between the processing operation and the interests pursued. However, the legitimate interests pursued by the controller must fulfil the following conditions: lawfulness, specificity and actuality. The legitimate interests must respect the applicable European or national legislation (lawfulness), they must be appropriately specific, precise and clear (specificity), and they must embody actual and real, and hypothetical, interests (actuality).[319]

Controllers can pursue a wide range of legitimate interests. Legitimate interests can be those that allow the exercise of a fundamental right, as enshrined in the Charter or in the ECHR, such as freedom of expression, freedom to conduct businesses, liberty, and the right to property, among others. Direct marketing purposes and ensuring network and information security are deemed legitimate interests of the controller.[320] Additionally, the legitimate interests of the controller can correspond to the public or societal interests or interests of the civil society (e.g., processing personal data for charitable reasons, combatting money laundering or financial fraud). As a general rule, the less compelling the legitimate interests of the controller are, the more likely the rights of the data subject will override them.[321]

---

[317] ibid.

[318] Information Commissioner's Office, 'How do we apply legitimate interests in practice?' <https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/legitimate-interests/how-do-we-apply-legitimate-interests-in-practice/> accessed 21/03/2022.

[319] Article 29 Data Protection Working Party, 'Opinion 06/2014 on the Notion of Legitimate Interests of the Data Controller under Article 7 of Directive 95/46/EC' (n 304) 25.

[320] Recital 47 and 49 GDPR.

[321] Article 29 Data Protection Working Party, 'Opinion 06/2014 on the Notion of Legitimate Interests of the Data Controller under Article 7 of Directive 95/46/EC' (n 304) 26.

Another crucial aspect is the *impact of the processing on the data subject*. In the evaluation of the impact, a controller must identify the sources that are likely to impact, the probability that the risk materialises and the severity of the results. The nature of the data is also another factor to consider. Where the processing includes sensitive personal data the balance assessment is more stringent because the possible damage to the data subject's rights could be more severe. Then, the reasonable expectations of data subjects and the uncertainty of the impact play a role too. The more uncertain the effects of the processing operations on the data subjects, the less likely the processing will be considered legitimate.[322]

The result of this balancing test must show that the legitimate interests of the controller are not overridden by the interests, rights and freedoms of the data subjects. It also means that the effects of the processing on the latter's interests and rights are minimised. In this sense, it is also important to note that the additional safeguards put in place by the controller to protect the rights and freedoms of the data subject may help tilt the scale in favour of the data controller.

Since legitimate interests can be invoked to base a wide range of purposes for processing personal data, it is a suitable legal basis for processing personal data for the development and deployment of AI systems. It is generally used by data controllers in this field because it is a flexible legal basis. Indeed, contrary to processing based on consent, processing based on the legitimate interests of the controller does not require active positive involvement of data subjects before the processing operations (i.e. consent). Certainly, controllers must demonstrate they have evaluated how the processing may affect the data subjects' rights and interests via the legitimate interests assessment. Equally important, controllers must inform data subjects, where possible before the processing takes place, and the latter can object to the processing based on legitimate interests at any time.[323]

It is important to note that in the design and development stage of the AI solutions, controllers may invoke purposes broadly (research, finding correlations in the data, development of a commercial product, etc.). Yet, in later stages of the AI system development and when more concrete purposes are recognised, controllers should

[322] ibid 37–40.

[323] See Arts. 12-14 and 19 GDPR, later commented in detail.

evaluate whether the initial legitimate interest assessment still reflects the situation at that moment or if, on the contrary, they should re-evaluate how the processing affects data subjects,[324] and inform them accordingly.

II.2.2.4.- Repurposing of personal data – the case of statistical research

According to the principle of principle limitation, personal data can only be collected for specified and predefined purposes. While repurposing of personal data is limited, it is not completely outlawed in the GDPR. Provided that the purposes are compatible, personal data can be used for a different purpose. So controllers have three options at hand when they plan to process personal data for a purpose different from which it was previously collected. Controllers may ask data subjects to renew their consent, evaluate the compatibility of the further use,[325] or rely on one of the presumed compatible purposes, such as the statistical purpose exception.[326]

In the latter scenario, it is worth distinguishing two kinds of situations. First, where developers of AI solutions aim to find statistical correlations concerning aggregated data without using this knowledge to take measures or decisions concerning specific individuals (functional separation).[327] In this case, repurposing of personal data may be legitimate, since additional processing for statistical purposes -among others- are compatible with the initial purposes,[328] as long as appropriate security and technical safeguards are implemented.[329]

Statistical purposes entail data processing for '*statistical surveys or for the production of statistical results*'.[330] While this definition does not provide a clear explanation of what concrete processing operations are covered by these purposes, it is arguable that acquiring and integrating individual-level data for aggregation and

[324] Information Commissioner's Office, 'Guidance on AI and Data Protection' (n 147) 33.

[325] Art. 6(4) GDPR.

[326] Art. 5(1)(b) and 89(1) GDPR.

[327] Article 29 Data Protection Working Party, 'Opinion 03/2013 on Purpose Limitation' (n 185) 30.

[328] Art. 5(1)(b) GDPR.

[329] Art. 89(1) GDPR.

[330] Recital 162 GDPR.

creating summary statistics are covered by this term.[331] No provision in the GDPR limits the scope of the statistical purposes. Hence, statistical purposes encompass both public and private statistics, even in pursuance of commercial profits[332] (for example, using data analytics on web pages or tools for market research). Therefore, if the outcome of the processing is aggregate data (for instance to gain insights at population level or identify trends in data), these data can be further employed for a different purpose.[333]

In this context, using aggregated personal data to build AI models is allowed under the GDPR. However, the need to employ adequate safeguards and the restriction on the use of statistical outputs to support decisions and measures affecting concrete individuals limit the scope of this exception. First, for the repurposing of personal data based on the exception of the statistical purposes to be accepted, controllers must put in place adequate safeguards. Data processed for statistical purposes is personal data under the GDPR,[334] so it is still required to guarantee the rights and freedoms of the data subjects. Therefore, controllers must implement technical and organisational measures to assure functional separation, including pseudonymisation if the statistical purposes can be achieved in that way, and -in every case- anonymise the data as soon as it can be done without impairing the purposes.[335]

---

[331] Travis Greene and others, 'Adjusting to the GDPR: The Impact on Data Scientists and Behavioral Researchers' (2019) 7 Big Data 140, 144.

[332] Nikolaus Forgó, Stefanie Hänold and Benjamin Schütze, 'The Principle of Purpose Limitation and Big Data' in Marcelo Corrales, Mark Fenwick and Nikolaus Forgó (eds), *New Technology, Big Data and the Law* (Springer 2017) 36; Viktor Mayer-Schönberger and Yann Padova, 'Regime Change? Enabling Big Data through Europe's New Data Protection Regulation' (2016) 17 Science and Technology Law Review 315, 326; Article 29 Data Protection Working Party, 'Opinion 03/2013 on Purpose Limitation' (n 185) 29.

[333] Recital 162 GDPR.

[334] Christian Wiese Svanberg, 'Article 89. Safeguards and Derogations Relating to Processing for Archiving Purposes in the Public Interest, Scientific or Historical Research Purposes or Statistical Purposes' in Christopher Kuner, Lee A Bygrave and Christopher Docksey (eds), *The EU General Data Protection Regulation (GDPR)* (OUP 2020) 1250.

[335] Art. 89(1) GDPR.

Secondly, the outcome or the personal data cannot be used as a support for 'measures or decisions' concerning concrete individuals.[336] Therefore, it is not allowed to repurpose personal data using the statistical purpose exception for gaining knowledge at the person-level,[337] reaching particular data subjects directly or indirectly[338] or for any kind of decision-making at the individual level. For example, an organisation may use the personal data of their clients for statistical purposes to predict the fidelization rates, but it could not use this information to forecast if a particular client may switch to another company[339] or to send targeted messages or advertising based on those insights.

Where controllers aim at processing personal data for evaluating individual preferences or behaviours of concrete persons and then use the knowledge gained to support 'measures or decisions' addressed to those particular data subjects, the repurposing of personal data is not covered by the exception of the statistical purposes. Hence, controllers must obtain new consent to process these data, since the purpose is not considered compatible with the initial one. This rules out of the scope of the statistical purposes exception most of the direct marketing techniques, tracking of individuals, and behavioural-based advertising.

[336] Recital 162 GDPR.

[337] Greene and others (n 329) 144.

[338] Tal Zarsky, 'Incompatible: The GDPR in the Age of Big Data' (2016) 47 Seton Hall Law Review 995, 1000.

[339] Rossana Ducato, '89. Garanzie e Deroghe Relative Al Trattamento a Fini Di Archiviazione Nel Pubblico Interesse, Di Ricerca Scientifica, o Storica o a Fini Statistici' in Oreste Pollicino and Roberto D'Orazio (eds), *Codice della Privacy e Data Protection* (Guiffrè Francis Lefebvre 2021) 962.

Chapter III

ENSURING INDIVIDUAL RIGHTS IN ARTIFICIAL INTELLIGENCE SYSTEMS

Introduction

The General Data Protection Regulation is the most important regulatory framework in Europe for the protection of fundamental rights of individuals when their personal data is being processed. This regulation establishes a comprehensive set of rules to guarantee that whenever personal information is processed, personal data is adequately protected. Where personal data is processed using automated means (and under certain circumstances under manual processing), the GDPR applies. Given that personal information may be processed at diverse stages in the lifecycle of AI systems, both providers and users of AI systems should follow the rules established therein, because they may be considered to be controllers or processors of personal data. For instance, during the development stage of AI systems, datasets employed to train, test and validate AI models may contain personal data. During the deployment of AI models, personal data may be used as input for predictions concerning individuals, and the outcome of a particular use of an AI system may also be personal data under the GDPR. Finally, some specific AI models are also composed of personal data, thus requiring personal data for their proper functioning.

While the GDPR is a technologically neutral legal framework,[340] which means that it applies regardless of the technologies used for the processing of personal data, the particular characteristics of the processing operations carried out with AI systems create some challenges that need to be carefully addressed.

This chapter elaborates on the rights of data subjects whose personal data is processed using AI systems, when confronted with automated decisions or profiling powered by AI systems and the challenges to ensuring individual rights in AI systems. The appraisal of the impacts that AI-assisted decision-making has on the enjoyment of those rights will be the crucial part of this section. The objective of this section is,

[340] Recital 15 GDPR.

thus, the identification of the merits and, especially, demerits of the current regulation. Only after properly evaluating the pitfalls of the current regulation the way will be paved for further research and working on proposals to overcome those limitations. This section also provides an overview of the GDPR general accountability mechanisms which have an impact on the issues related to the intersection between data protection and artificial intelligence. It will also address the particularities that these mechanisms require to tackle the issues highlighted when processing personal data using AI systems.

III.1.- Rights related to automated decision-making including profiling

III.1.1.- Introduction

The GDPR contains provisions that protect individuals against the possible harms caused by decisions taken or profiling performed through automated means.[341] As a matter of principle, EU data protection law forbids taking automated decisions that have legal or significantly similar effects on individuals. While it seems to be a technologically advanced provision, the right not to be subject to automated decisions is not new. In fact, the French Law on Computers, Files and Freedoms of 6 January 1978[342] already incorporated a similar right among its provisions. However, the sheer amount of data available, the increase in computational power and the improvement of the machine learning techniques have made the use of automation a common practice both for public and private organisations. Despite the benefits in terms of efficiency and cost reduction, automated decision-making can lead to uneven treatment and discrimination of data subjects or even to wrong decisions that can potentially violate their rights. Moreover, data subjects usually do not understand the algorithms that are used in the systems that deliver the decisions or profiling.

The rationale for the inclusion of the right not to be subject to automated decisions in data protection legislation does not exclusively relate to the protection of personal

[341] Related provisions: art. 4(4), 9, 12, 13, 14, 15, 21, 22, 35(1) and (3), recital 71.

[342] Loi relative à l'informatique, aux fichiers et aux libertés du 6 janvier, art. 2 <https://www.legifrance.gouv.fr/jo_pdf.do?id=JORFTEXT000000886460&pageCourante=00227> accessed 13/06/2022.

information. Instead, the support of human dignity,[343] the fact that complete trust in an automated decision can be harmful to individuals, as well as giving individuals the chance to take a role in the decision process[344] are the main drivers behind this right. It is contrary to these fundamental protections to employ automated systems to make important decisions about persons simply because they represent a technically feasible and economically efficient tool for the user of the AI system.[345]

The *right not to be subject* to automated decision-making, including profiling, as provided for in Article 22 of the GDPR is currently a highly-debated topic in academia and policy-making. Article 22 GDPR bans taking decisions solely on automated processing, profiling included, whenever it produces legal effects concerning data subjects or affects them in a similarly significant way. This provision is applicable when no natural person has any decision-making power, regardless of whether in the decision-making process there has been some minor human involvement.[346] However, automated decision-making is allowed under certain circumstances. Using automated means to render decisions about individuals is allowed provided that it is: a) necessary for the entry into or performance of a contract; b) authorised by EU or Member State law applicable to the controller; c) based on the individual's explicit consent.[347] Yet, if data controllers decide to rely on this exception, they must place suitable safeguards to protect the rights and freedoms of data subjects.[348] The right not to be subject to a decision based solely on automated processing links to artificial intelligence, since a high number of decisions and measures are currently taken without human assistance and powered by algorithmic decision systems.

In the following sections, an attempt will be made to disentangle whether the bundle of rights enshrined in the GDPR (e.g. right not to be subject to a decision based solely on automated processing, the right to obtain human intervention, the right to express

---

[343] Orla Lynskey (n 155) 98–99.

[344] Lee A Bygrave, 'Minding the Machine: Article 15 of the EC Data Protection Directive and Automated Profiling' (Elsevier Advanced Technology, 1 January 2001) 17 18.

[345] Constitutional Court of the Slovak Republic, *Judgment no k PL ÚS 25 / 2019-117 - 492/2021 Coll* [125]. Translated using Google Translate.

[346] Voigt and von dem Bussche (n 305) 181.

[347] Art. 22(2) GDPR.

[348] Art. 22(3) GDPR.

his or her point of view, the right to challenge the decision, and eventually, the right to an explanation from Art 13, Art. 14, Art. 15, and, especially, Art. 22 GDPR) adequately solves legal problems that arise from automated decision-making.

III.1.2.- Content of the right

Article 22 GDPR establishes 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'.*

The first problem that should be solved relates to the *nature* of the provision in question. The question concerns whether the provision is a right that needs the data subject's involvement to be operative or, on the contrary, it attempts to establish a general prohibition addressed to data controllers and processors on the use of purely automated means to render decisions or profiling. The former, i.e. allowing data subjects to exercise their right or a right that individuals can exercise 'at their insistence'[349] implies that every time an individual is subject to a decision he or she should request the data controller to ascertain whether the decision or profiling is taken or generated exclusively without or with little human intervention.[350] This stance places an excessively heavy burden on data subjects, especially considering that generally individuals are unaware of the profiling and that oftentimes they do not read or do not understand the content of the privacy policies. It erodes all possible meaning to the provision and it does not seem to be the idea that European legislators had in mind when they drafted the regulation. The second position, i.e. establishing a general ban on controllers for decision-making or profiling based solely on automated processing, is more adequate to the current functioning of the data economy. It acknowledges the imbalance of power between the interested parties and places the burden on the party that has the technical and organizational tools and knowledge to implement solutions that respect fundamental rights. According to this position, the safeguards for data subjects are more effective, since from the outset they are protected against the

[349] Luca Tosoni, 'The Right to Object to Automated Individual Decisions: Resolving the Ambiguity of Article 22(1) of the General Data Protection Regulation' (2021) 11 International Data Privacy Law 145, 162.

[350] Wachter, Mittelstadt and Floridi (n 1) 94.

potential impairments that this type of processing may have on their rights and freedoms.[351] Additionally, a normative reason supports this interpretation. Recital 71 GDPR states that decision-making based on automatic processing 'should be allowed where' and then mentions the grounds on which it should be allowed. This means that, in principle, decision-making and profiling based on automated means is not permitted. On the contrary, it is only allowed in specific situations and following particular requirements.

Another question that requires an evaluation is the *scope of the prohibition*. This involves assessing whether the ban on deciding through automated means covers only individual decisions or profiles, or it also includes collective automated decision-making. The first option, limiting the bar to individual decisions or profiles, is supported by the majority of authors and is the position held by the WP29, since it considers that this interdiction solely applies to specific circumstances.[352] Not only is the article titled 'Automated *individual* decision-making, including profiling', which reflects the singular nature of the right, but also the right does not include any reference whatsoever to decisions taken against a collective of persons. However, some commentators questioned the limited scope of this prohibition, considering that collective automated decision-making should be covered as well.[353]

III.1.2.1.- Decisions based solely on automated processing or profiling

Decisions

The automated processing, including profiling, should be the basis on which a decision is made. A decision is a resolution arrived after examination or consideration.

---

[351] Article 29 Data Protection Working Party, 'Guidelines on Automated Individual Decision-Making and Profiling for the Purposes of Regulation 2016/679' (n 187) 20.

[352] ibid.

[353] Maja Brkan, 'The Concept of Essence of Fundamental Rights in the EU Legal Order: Peeling the Onion to Its Core' (2018) 14 European Constitutional Law Review 332, 11; Alessandro Mantelero, 'Personal Data for Decisional Purposes in the Age of Analytics: From an Individual to a Collective Dimension of Data Protection' (2016) 32 Computer Law & Security Review 238, 249.

Decisions should be given a broad meaning to encompass any resolution that has a binding effect on the data subjects and it may include measures.[354]

Determining when a decision is based solely on automated processing is a relatively straightforward operation if the processing operations that lead to the decision having legal or similarly significant effects on the data subject were undertaken by the same controller. However, there are situations in which this determination is not entirely clear. For instance, some credit information agencies provide financial institutions with scores that evaluate the creditworthiness of individuals, and then the financial institutions decide whether or not to grant the financial assistance requested by individuals (e.g. a loan). A literal interpretation of the legal provisions would suggest that there is no 'decision' in the sense of Art. 22 GDPR in this case. Neither do the credit information agency nor the financial institution make a decision based on automated processing. The credit information agency does not make a decision within the meaning of Art. 22 GDPR because its processing operations were limited to creating the profile of the individual and transferring the score (outcome or profile) to the credit institution without deciding on the loan. The financial institution does not decide solely based on automated processing because the score provided by the credit agency is only one element that the institution takes into account to decide whether to grant the loan or not. However, the Administrative Court of Wiesbaden in *Schufa Holdings AG* considered that it should not be the case.[355] In the preliminary reference to the CJEU, the Wiesbaden Court (referring court) held that the score created by the credit information agency (Schufa) represents a 'decision' in the sense of Art. 22(1) GDPR and not simply a preparatory profiling activity for the ultimate decision taken by the financial institution. To support its reasoning the court took into consideration factual aspects of the relationship between the alleged processor (credit information agency) and the controller (financial institution). The court pointed out that the credit information agency, together with the score, forwards a suggestion to the financial institution related to the suitability or unsuitability of the applicant to enter into a contract and that the applicant's score is a determinant factor for the success of the

[354] Recital 71 GDPR.

[355] Case C-634/21, *OQ v SCHUFA Holding AG and Land Hesse*. Request for a preliminary ruling from the Verwaltungsgericht Wiesbaden (Germany) lodged on 15 October 2021, para 21.

applicant's request. Concerning the latter, it is worth noting that whereas high scores do not guarantee approval of the financial assistance, lower scores are highly likely to result in a refusal.[356] If this interpretation is upheld by the CJEU it will definitively widen the meaning of 'decision' in the context of Art. 22(1) GDPR.

Based solely on automated processing

To determine whether a decision is based solely on automated processing, the level of human involvement should be assessed. It is generally agreed that a decision is not based only on automated processing if it is reached by a person that is empowered and competent to modify the decision and it is not a mere token gesture.[357] This means that a decision is not solely based on automated processing where the automated system is only employed to assist the human who makes the decision, for instance evaluating labour market opportunities for job applicants.[358] A controller may avoid this provision provided that an individual evaluates the substance of the pronouncement and he or she does not behave as a simple procedural step,[359] like rubberstamping.

Profiling

Profiling is any form of automated processing to evaluate, analyse or predict certain personal aspects relating to a natural person.[360] Though profiling entails the classification and grouping of individuals according to similar traits, the open texture of the provision ('any form') implies that clustering is not the only method controllers may use to profile and thus it could include any kind of automated processing leading to the creation of profiles. Within the definition of profiling, it should be included both

---

[356] ibid 22, 24–25.

[357] Article 29 Data Protection Working Party, 'Guidelines on Automated Individual Decision-Making and Profiling for the Purposes of Regulation 2016/679' (n 187).

[358] Austrian Federal Administrative Court (BVwG), *Public Employment Service Austria* (2020) W256 22353. The Austrian Public Employment Service measured the probability of job applicants of being employing in a defined timeframe, using candidates' personal data (age group, gender, educative level, disabilities, etc). Once the system measured the candidates' chances of being employed, it grouped candidates with similar scores (classification). To mitigate privacy risks, human reviewed the scores and followed dedicated institutional guidelines.

[359] Brkan (n 1) 11.

[360] Art. 4(4) GDPR.

the establishment of the parameters for the outcome of the assessment and the outcome itself.[361] In a nutshell, profiling is the creation and application of profiles.

The profiling process comprises three different stages.[362] First, data warehousing involves the transformation of individuals' features and registered behaviours into digital data, for its further collection and storage. Second, data mining is an analytical stage whereby the information collected is evaluated to obtain statistical correlations between the observed data from features, behaviours and the classifications (already created or new) that the analysts have. Third, in the inferential stage analysts deduct present or future behaviour from the observed characteristics of an individual. This means that profiling entails a certain evaluation of an individual.[363]

While the first two stages can be performed using anonymized or pseudonymized data, the last one necessarily requires the actual or potential identification of the individual.[364] For the GDPR the profiling must be carried out on personal data. However, there is no clarity on what happens with non-personal data (be it machine-generated or otherwise) that taken as a whole and grouped could lead to the creation of profiles. For example, data gathered from sensors that register energy consumption or different household habits, which in principle would fall outside of the definition of personal data, can be used to create a profile of an individual. Should the combination of non-personal data for the creation of profiles be made with standardised techniques and with current state-of-the-art technologies, the resulting information must be considered personal data as well.

Profiling aims at assessing personal aspects concerning a natural person and then applying it to make predictions about the behaviours of the individuals. Profiling always implies a kind of assessment or judgment of a person and the definition is qualified by

---

[361] *OQ v SCHUFA Holding AG and Land Hesse* (n 353) para 23.

[362] Council of Europe, 'The Protection of Individuals with Regard to Automatic Processing of Personal Data in the Context of Profiling. Recommendation CM/Rec(2010)13 and Explanatory Memorandum' (2011) 25.

[363] Article 29 Data Protection Working Party, 'Guidelines on Automated Individual Decision-Making and Profiling for the Purposes of Regulation 2016/679' (n 187) 7.

[364] Council of Europe, 'The Protection of Individuals with Regard to Automatic Processing of Personal Data in the Context of Profiling. Recommendation CM/Rec(2010)13 and Explanatory Memorandum' (n 360) 25.

the purpose. A mere classification or clustering of individuals according to personal characteristics or true features does not involve profiling for the GDPR.[365] However, if the controller or third parties plan to perform evaluations and then predict future actions, decisions or choices of individuals, for instance in targeted marketing for business or political campaigns, the GDPR protects the data subjects involved.

The list provided in art. 4(4) GDPR[366] is illustrative of which personal aspects could be assessed and predicted since the inclusion of the phrase 'in particular to' implies that this is a non-exhaustive list. For instance, the Amsterdam District Court considered, in a dispute concerning the extent of the information that the controller (ride-hailing app) must provide to data subjects (drivers), that an 'earning profile' was a form of automated processing of personal data that involved profiling since it was elaborated with variables like earnings, attendance, daily logging hours, driver's score or rating to predict future earnings.[367] The court also held that a 'fraud probability score' (i.e., the estimation of the likelihood a driver will commit fraud) was a form of automated processing of personal data that involved profiling since it forecasts the driver's behaviour and reliability (risk profile).[368] Profiling can be also used to infer special categories of personal data (e.g. propensity to vote for a certain candidate)[369] or to build scores estimating the probability that a person would repay a loan.[370]

III.1.2.2.- Producing legal or similarly significant effects

The decision in question must produce legal effects or affect the data subject in a similarly significant manner. A decision produces legal effects when it creates,

---

[365] ibid; Enza Pellecchia, 'Privacy, Decisioni Automatizzate e Algoritmi' in Vincenzo Franceschelli and Emilio Tosi (eds), *Privacy Digitale. Riservatezza e protezione dei dati personali tra GDPR e nuovo Codice PrivacyRiservatezza e protezione dei dati personali tra GDPR e nuovo Codice Privacy* (Guiffrè Francis Lefebvre 2019) 427.

[366] This provision mentions that profiling could be carried out to evaluate the individual's performance at work, economic situation, health, personal preferences, interests, reliability, behaviour, location or movements.

[367] Amsterdam District Court, *Ola Netherlands BV* (2021) C/13/68970.

[368] ibid.

[369] European Data Protection Board, 'Guidelines 8/2020 on the Targeting of Social Media Users' (2021) 32.

[370] *OQ v SCHUFA Holding AG and Land Hesse* (n 353).

modifies or extinguishes someone's legal status, rights or obligations. Where the decision produces merely trivial effects, it is out of the scope of this provision.

Although the decision taken does not create, modify or extinguish someone's rights and obligations (i.e., it does not produce legal effects), it may fall under the scope of Art. 22 GDPR provided that it generates a comparable or equivalent effect in its impact and scope. The threshold for the significant impact is rather high,[371] as is the case of credit determinations[372] or decisions taken within e-recruiting processes.[373]

The decision should potentially alter the circumstances, the conduct, and the searches of the person in a substantial manner, have a long-lasting effect on individuals, or, ultimately, exclude or discriminate natural persons without an objective reason.[374] For instance, an automated decision-making system that imposes discounts and fines on drivers registered on a ride-hailing app can affect their rights and alter their behaviour, thus producing effects significantly similar to legal effects.[375]

Since trivial effects are not covered by this provision,[376] this begs the question of whether targeted advertising should be covered as well. Some scholars, for instance, doubt whether race-targeted ads can have a significant impact on individuals.[377] However, while generally targeted advertising would not reach the threshold required, circumstances of mode (how the advertising is carried out or the spread of tracking activities) and person (benefiting from or exploiting someone's vulnerabilities) or social factors (for special groups of people) could make online behavioural advertising to fall into this prescription.[378] For instance, targeting economically vulnerable online

---

[371] Oostveen (n 200) 147.

[372] Recital 71 GDPR and *OQ v SCHUFA Holding AG and Land Hesse* (n 353).

[373] Recital 71 GDPR.

[374] Article 29 Data Protection Working Party, 'Guidelines on Automated Individual Decision-Making and Profiling for the Purposes of Regulation 2016/679' (n 187) 21.

[375] *Ola Netherlands B.V.* (n 365).

[376] Lee A Bygrave, 'Article 22. Automated Individual Decision-Making, Including Profiling' in Christopher Kuner and Lee A Bygrave (eds), *The EU General Data Protection Regulation (GDPR): A Commentary* (OUP 2020) 534.

[377] Edwards and Veale (n 297) 47–48.

[378] Article 29 Data Protection Working Party, 'Guidelines on Automated Individual Decision-Making and Profiling for the Purposes of Regulation 2016/679' (n 187).

gamblers individuals if it may lead to a significant detrimental economic impact on them,[379] should fall under the umbrella of Art. 22 GDPR.

But apart from specific cases of targeting advertising, there are a whole set of instances where the automated processing, due to its legal insignificance, will fall outside the realm of Art. 22 GDPR. The District Court of Amsterdam in *Uber B.V.*[380] considered that while some decisions taken by the company were automated and may have a direct effect on the driver's earnings, they did not reach the threshold required by Art. 22 (production of legal or significantly similar effects), such as the 'batched matching system'[381] and the upfront pricing system. Similarly, the temporary automatic blocking of the app following a fraud signal produced neither legal nor significantly similar effects on the drivers because the reactivation of their accounts is made after the driver makes contact with Uber.[382] In *Ola Netherlands B.V.* the court held that while the earning profile[383] can serve as a basis for receiving bonuses and may condition or determine the drivers' behaviour, it does not produce legal or significantly similar effects on them.[384] Similarly, the decision to allocate a passenger to an available driver is automatic but it does not produce legal or significantly similar effects on drivers.[385] Finally, while pre-granted loans, prices adjusted to the customer's profile, benefits and discounts imply profiling and may cause discriminatory effects, these were not considered as producing a legal or significantly similar effect.[386]

As seen from the previous collection of cases, the criterion to identify the systems capable of delivering decisions that produce similarly significant effects on data subjects is not entirely clear. In the majority of cases, the decision will be made on a

---

[379] European Data Protection Board, 'Guidelines 8/2020 on the Targeting of Social Media Users' (n 367) 24.

[380] District Court of Amsterdam, *Uber BV 1* (2021) C/13/68731.

[381] The batched matching system clusters the nearest drivers and passengers in a group and establishes the best matches between them. To perform this function the system employs geolocation, travel distance and direction, possible traffic congestions and personal preferences of Uber drivers.

[382] District Court of Amsterdam, *Uber BV 2* (2021) C/13/69200.

[383] Profile elaborated with variables like earnings, attendance, daily logging hours, driver's score or rating.

[384] *Ola Netherlands B.V.* (n 365).

[385] ibid.

[386] Agencia Española de Protección de Datos, *Caixa Bank SA* (2021) PS/00477/2.

case-by-case basis. However, it is important to understand the nuances in the interpretation of the concept 'similarly significant effects' since it could be a major source of failure of claims lodged before national data protection authorities. Two further examples can illustrate the complexity of this area.

In December 2021, the NGO None Of Your Business (NOYB)[387] filed two complaints against two tech companies based on Art. 22 GDPR. One complaint was filed against Amazon (Amazon Mechanical Turk) before the Luxembourg Data Protection Authority and it relates to the use of automated decision-making to accept or reject workers, without providing the required information to individuals and the safeguards required in Arts. 13, 14 and 22 GDPR.[388] The second complaint was lodged against Airbnb before the Data Protection Authority of Rheinland-Pfalz because it downgraded the platform user's rating as a host solely through an automated decision, without complying with the safeguard established in Art. 22 GDPR.[389]

While forecasting a decision is always a difficult task because it depends on a multiplicity of factors not always properly evaluated and the information publicly available is limited, the complaint against Amazon Mechanical Turk seems likely to succeed. Amazon Mechanical Turk is a crowdsourcing marketplace that enables business owners to outsource part of their internal processes, in particular simple repetitive or mechanical tasks, to a global workforce that can execute these activities.[390] Crucial for this assessment, Amazon Mechanical Turk can reject applications to participate in the program and the platform using fully automated means. Since the company uses automated decision-making to accept or reject workers, it can be covered by Art. 22 GDPR. Not only are 'e-recruiting practices without

---

[387] None Of Your Business is an NGO based in Vienna and founded by Maximilian Schrems to undertake strategic litigation and promote public awareness concerning digital rights, privacy and data protection in Europe.

[388] NOYB, 'Help! My recruiter is an algorithm!' (22/12/2022), <https://noyb.eu/en/complaint-filed-help-my-recruiter-algorithm> accessed 28/03/2022.

[389] NOYB, 'GDPR complaint: Airbnb hosts at the mercy of algorithms' (22/12/2022), <https://noyb.eu/en/complaint-filed-help-my-recruiter-algorithm> accessed 28/03/2022.

[390] See Amazon Mechanical Turk website for more information <https://www.mturk.com/> accessed 28/003/2022.

any human intervention' expressly mentioned in Recital 71 GDPR as an example of decisions having effects similarly significant to legal effects, but also the decisions taken by the Italian Data Protection Authority in Deliveroo[391] and Foodinho[392] support this stance. In these cases, the authority held that Deliveroo's and Glovo's booking systems, through which riders book the time slots predetermined by the company until saturation, fall within Art. 22 GDPR and they produced similarly significant effects because they allowed or denied access to job opportunities.

However, a more cautious approach should be taken concerning the complaint filed against Airbnb. Airbnb used automated means to delete a five-star review and, as a consequence, the platform downgraded the host's status from Superhost to 'normal host'. According to the claimant (NOYB):

> *'the decision to delete the 5-stars Review has had the effect of reducing the overall rating of the complainant as a host, which directly influences the Superhost status of the complainant and the contractual advantages that it provides (…). In other words, the complainant can lose her Superhost status and the substantial advantages that it confers on her'.*

To evaluate whether the decision to automatically delete a five-star review, which had the effect of downgrading the status of the host, produces similarly significant effects as required by Art. 22 GDPR it should be evaluated the potentially detrimental effects of those decisions. Airbnb Superhosts are more easily visible to potential guests, they earn a 'Superhost badge' which promotes even more visibility and trust, they obtain an additional 20% over the normal bonus if they refer new hosts, and they acquire a travel voucher equivalent to $100 after keeping the status of Superhost for at least sixteen consecutive months.[393]

This begs the following question: Had the automated decision to delete a five-star review, that downgraded an Airbnb host from Superhost to 'normal host', a similarly significant effect to a legal decision? It is doubtful. It is worth remembering that the

---

[391] *Ordinanza ingiunzione nei confronti di Deliveroo Italy SRL* (n 271) para 3.3.5.

[392] *Ordinanza ingiunzione nei confronti di Foodinho SRL* (n 212) para 3.3.6.

[393] See Airbnb's Superhost programme for more information <https://www.airbnb.com/d/superhost> accessed 28/03/2022.

decision in question must produce legal or similarly significant effects. A decision produces legal effects when it creates, modifies, or extinguishes somebody's legal rights, status or obligations. The decision may also fall under the scope of Art. 22 GDPR if it affects data subjects in a similarly significant manner. The WP29 considered that the effects of the processing must be sufficiently important to be 'worthy of attention'.[394] When applying this provision to particular cases, as mentioned before, courts and data protection authorities took a strict and narrow approach. For instance, targeting advertising (except where it targets vulnerable persons, like gamblers),[395] Uber's batched matching system and upfront pricing system and the temporary automatic blocking of the app to Uber drivers following a fraud signal,[396] the decision to allocate a passenger to an available driver or an earning profile,[397] and pre-granted loans, prices adjusted to the customer's profile, benefits and discounts[398] were all considered out of the scope of Art. 22 GDPR. Considering the high threshold set by courts and data protection authorities that interpreted this provision, it seems that the automatic decision to downgrade from Superhost to normal host, while surely detrimental for the host, cannot be considered as producing a similarly significant effect on the individual. Therefore, this decision may not be covered by Art. 22 GDPR.

III.1.3.- Exceptions from the prohibition

As a general rule, the GDPR forbids decisions based solely on automated processing, including profiling, when they produce legal or significantly similar effects on individuals. There are, however, exceptions to this rule and the legal framework authorises taking solely automated decisions of this kind when it is necessary for the performance of a contract, it is authorised by the EU or member state, or the data

---

[394] Article 29 Data Protection Working Party, 'Guidelines on Automated Individual Decision-Making and Profiling for the Purposes of Regulation 2016/679' (n 187) 21.

[395] ibid 22; European Data Protection Board, 'Guidelines 8/2020 on the Targeting of Social Media Users' (n 367) 24.

[396] *Uber B.V. 1* (n 378).

[397] *Ola Netherlands B.V.* (n 365).

[398] *Caixa Bank SA* (n 384).

subject explicitly consents to it. These exceptions provide flexibility to Member States and public and private data controllers.[399]

In the first place, despite the prohibition, individual decisions based solely on automated processing or profiling are allowed when they are necessary to enter into or for the performance of a *contract*. In this case, the necessity requirement should be understood as an 'enabling' requirement for the conclusion of the contract. However, if the same purpose can be achieved through another less privacy-intrusive method, this processing activity will likely not be considered necessary. Additionally, *EU or national law* to which the controller is subject may provide for exceptions and allow automated decisions, provided that it establishes suitable safeguards for the protection of the data subjects' rights. For instance, laws aimed at combatting fraud or tax evasion can include provisions allowing automated decisions.[400] On this matter, it is worth mentioning that the national law that should regulate the decision-making performed solely by automated means.[401] Finally, *explicit consent* of the data subject also permits taking automated decisions. Explicit consent imposes a higher standard than ordinary consent which is a 'statement or clear affirmative action'. The term 'explicit' qualifies how the consent is given, and it implies that the individual must provide an express statement of consent.[402] This can be achieved, for instance, by a written declaration, but also by compiling an online form, through an email or via two-stage verification (clicking on an 'Accept' button and then confirming the operation via email or SMS). In any case, it must represent a truly unconstrained and informed choice.

It is important to bear in mind that regardless of the legality of the processing operation of a controller under any of the legal basis for processing outlined in Art. 6 GDPR, whenever he or she subjects an individual to a decision based solely on automated processing or profiling, even using the same personal data, the former must rely on one of the three legal bases mentioned in Art. 22(2) GDPR. To be clear, this kind of processing is permitted only subject to the three exceptions previously mentioned. This means that controllers cannot rely on legitimate interests as a valid

---

[399] Castets-Renard (n 72) 118.

[400] Recital 71 GDPR.

[401] *OQ v SCHUFA Holding AG and Land Hesse* (n 353) para 38.

[402] European Data Protection Board, 'Guidelines 05/2020 on Consent under Regulation 2016/679' (n 281) 20.

legal basis to perform solely automated decision-making that causes legal or similarly significant effects on the data subjects.

III.1.4.- Automatic decision-making based on special categories of data

Automated decision-making cannot be based on special categories of personal data (i.e. personal data that can reveal race, and sexual orientation, among others), except where the data subject explicitly consents to it or the decision is necessary for reasons of substantial public interest based on EU or national law and includes suitable safeguards to protect individuals' rights.[403] This means that even if the processing could be necessary for entering into or for the performance of a contract, which is allowed under Art. 22(2) GDPR for ordinary (i.e., non-sensitive) data, where the controller processes special categories of personal data cannot rely on this legal basis.

While it may seem theoretically easy to apply this provision, in the context of AI the distinction between both categories is more nuanced. Indeed, AI systems blur the difference between ordinary personal data and special categories of data, because the latter can often be inferred from the former.[404] This feature of AI systems is problematic since ordinary data, i.e. data that in itself does not belong to a special category according to Art. 9(1) GDPR, like name-specific data relating to the spouse, cohabitee or partner,[405] zip codes, dietary preferences,[406] name or surname, etc, or even digital records and behaviour from social networks[407] can function as a proxy for the detection of sensitive data. For instance, an image of a person's face is non-sensitive personal data, but where these images are converted with computerised systems into numerical expressions to be related to others to establish their similarity,

---

[403] Art. 22(4) GDPR.

[404] Gregorio and Torino (n 201) 471.

[405] Case C-184/20 *OT v Vyriausioji tarnybinės etikos komisija*. [2022] ECLI:EU:C:2022:601, para. 119

[406] European Union Agency for Fundamental Rights, 'Preventing Unlawful Profiling Today and in the Future: A Guide' (2018) 117. Concerning dietary preferences of airline passengers as a proxy for religious beliefs.

[407] Michal Kosinski, David Stillwell and Thore Graepel, 'Private Traits and Attributes Are Predictable from Digital Records of Human Behavior' (2013) 110 Proceedings of the National Academy of Sciences of the United States of America 5802.

they can uniquely identify a person, thus becoming a special category of data (i.e., biometric data to uniquely identify an individual using facial recognition technologies). Additionally, where remote biometrical identification systems are placed in public spaces and used at public events, the processing may involve additional special categories of data, such as those that reveal political opinions or trade union membership.[408] As a consequence, controllers should be aware of the potential inferencing of sensitive data when processing ordinary personal data since stringent requirements apply to the processing of the former.

III.1.5.- Automated decisions and profiling of children

Children are especially vulnerable to being manipulated and their fundamental rights and freedoms can be easily impaired. They are less capable to understand the intricacies of AI, are less aware of the online risks and can be easily manipulated.

Except for children, there is no vulnerable group that was granted group-specific protection under the data protection framework. They are protected basically by particular consent and information requirements posed on controllers.[409] The processing of personal data related to a child and based on his or her consent can be lawful after the child is 16 years old (never less than 13 if a member state regulates child consent), and below that age, the consent must be given by the holder of parental responsibility.[410]

---

[408] *Garante per la Protezione dei Dati Personali, Parere sul sistema SARI Real Time* (2021) 9575877. The Italian Data Protection Authority gave an unfavorable opinion concerning the system SARI Real-time (25.03.21). SARI is a remote biometrical identification system (RBI) that allows through cameras installed in a predetermined and delimited geographical area to analyze in real-time the faces of the subjects filmed there and it compares them with a database ("watch-list"). Every time a face is matched with the faces from the database it generates an alert to the police. The Italian DPA considered that this system would also carry out automated processing on a large scale of people who are not under police search. And while these images would be immediately erased, the biometric data of every individual in the area would be automatically processed. As this implies a strong interference with private life, it must have an adequate legal base for its deployment. However, no such a legal base was found.

[409] Gianclaudio Malgieri and Jędrzej Niklas, 'Vulnerable Data Subjects' (2020) 37 Computer Law and Security Review 1, 12.

[410] Article 8(1) GDPR.

Automated decision-making concerning children presents some particularities due to the special risks posed to their rights. Recital 71 GDPR states that automated decision-making should not apply to a child. Yet, this restriction is not mirrored in Art. 22 GDPR or any other article of the GDPR. This is the reason why it has been argued that controllers should, as a best practice, refrain from subjecting children to automated decisions that have legal or similarly significant effects should be taken as a best practice[411] or, at least, this type of processing should not be the rule when it comes to processing children's data.[412] Hence, whilst it is not forbidden to use automated decision-making and profiling on children, these decisions could impact decisively on children's ability to choose and their behaviour, hence the threshold to consider the effects that similarly significantly affect them should be lower than compared to adults.

III.1.6.- Safeguards

When the automated decision-making is based on explicit consent, legal basis or contractual necessity controllers must implement suitable safeguards to protect the data subject's rights. Allowing the data subject to obtain human intervention before rendering the decision, to express his or her point of view and to contest the decision are mandatory safeguards.[413] Other compulsory safeguards relate to the provision of certain information to the data subjects, in particular informing them about the existence of automated decision-making, including meaningful information about the logic involved, together with the significance and the envisaged consequences of such processing,[414] and explaining the decision reached after such assessment.[415]

---

[411] Article 29 Data Protection Working Party, 'Guidelines on Automated Individual Decision-Making and Profiling for the Purposes of Regulation 2016/679' (n 187). The WP29 does not consider the restriction included in recital 71 GDPR as an absolute interdiction.

[412] Information Commissioner's Office, What if we want to profile children or make automated decisions about them? <https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/children-and-the-uk-gdpr/what-if-we-want-to-profile-children-or-make-automated-decisions-about-them> accessed 21/08/2021.

[413] Art. 22(3) GDPR.

[414] Art. 13(2)(f), 14(2)(g) and art. 15(1)(h) GDPR.

[415] Recital 71 GDPR.

However, this is only a minimum mandatory list of safeguards that controllers must implement whenever they deploy automated decision-making systems that fall within the remit of Art. 22 GDPR. They are not precluded to, and they are even encouraged to, implement other safeguards that they consider suitable to mitigate the impact of those decisions on the rights and freedoms of data subjects.

III.1.6.1.- Right to obtain human intervention

The right to obtain human intervention relates to the possibility to get human oversight over the automated decision. The decision is taken by the system, yet a person exerts to some extent significant influence on the decision-making process, which can include ignoring it altogether.[416]

There are different degrees of human involvement in the decision-making process. The intervention can be fulfilled by one of the following three mechanisms.[417] First, human-in-the-loop (HITL) means that the person in charge is involved in every decision the system delivers. This type represents the highest degree of interaction between humans and AI systems and would be out of the scope of Art. 22 GDPR since it applies to decisions taken solely by automated means.[418] Second, human-on-the-loop (HOTL) in which the person participates in the design of the AI system and controls the general operation of the system. Third, in human-in-command (HIC) a reviewer monitors the general functioning of the system and is empowered to determine the concrete cases and the modalities in which the AI solution should be used. However, the complexity of the AI systems may frustrate the idea of pursuing a complete review of the process[419] unless the person tasked to review the decision is familiar with data analysis to be able to identify pertinent associations in the data.[420]

---

[416] Kiel Brennan-Marquez, Karen Levy and Daniel Susser, 'Strange Loops: Apparent versus Actual Human Involvement in Automated Decision Making' (2019) 34 Berkeley Technology Law Journal 745, 749.

[417] European Commission's High-Level Expert Group on Artificial Intelligence, 'Ethics Guidelines for Trustworthy AI' (2019) 16.

[418] Information Commissioner's Office, 'Guidance on AI and Data Protection' (n 147) para 71.

[419] Castets-Renard (n 72) 121.

[420] Roig (n 1) 6.

There were many cases in which AI systems that automate decisions causing legal or significantly similar effects, while authorised under one of the exceptions of Art. 22(2) GDPR, failed to provide the required safeguards, in particular, the right to obtain human intervention on part of the controller. For instance, no human revision was found in an automated decision-making system that imposed discounts and fines on drivers,[421] or in automated systems that allocate rides to riders.[422] However, in another case, the requirement of a human reviewer was found. In Uber, the District Court of Amsterdam deemed as significant human intervention the fact that a decision (dismissal of an employee) was taken by two employees of a specialised team after an investigation carried out by another employee following fraud signals sent by the system. Additionally, in the event of a disagreement between the two reviewers, the decision of a third employee was envisaged.[423]

The AI Regulation draft establishes in detail the modalities of human intervention in the AI system's lifecycle. According to this proposal, human oversight must be embedded in AI systems from the design to reduce the risks to individual rights. Human oversight must be implemented either by the provider (i.e. AI system developer) or the user (i.e. the organisation that uses an AI system under its authority). The human implementer must be able to fully understand the limitations and monitor the operation of the AI system, avoid over-reliance on the system, correctly interpret the results, decide not to use the system or disregard the output, and interrupt the operation of the AI system. Finally, if the system refers to real-time remote biometric identification for law enforcement purposes, no decision can be taken based on the identification unless at least two persons verify it.[424] As seen from the proposal, the implementation of this safeguard is further detailed and standardised for all high-risk AI systems.

---

[421] *Ola Netherlands B.V.* (n 365).

[422] *Ordinanza ingiunzione nei confronti di Deliveroo Italy SRL* (n 271); *Ordinanza ingiunzione nei confronti di Foodinho SRL* (n 212).

[423] *Uber B.V. 2* (n 380).

[424] Art. 14 AIA draft.

III.1.6.2.- The right to contest the automatic decision

Additionally, data subjects have a right to express their point of view and to contest the decision.[425] Interestingly, the GDPR establishes a mechanism whereby it allows contesting a decision that if it had been taken completely by a human it would not have been challengeable.

This right has two implications. First, the controller must re-evaluate the automated decision. It does not mean that the controller must automatically discard the decision. Instead, the controller must evaluate all the relevant information and consider the arguments provided by the affected individual.[426] Secondly, the individual who seeks to challenge an automated decision must have information concerning the reasons that support the decision. This creates a link with the right to obtain an explanation. In *Uber* the court held that the company must indicate the particular fraudulent actions that have the consequence of deactivating Uber drivers' accounts, to evaluate the correctness and lawfulness of the processing of personal data.[427] The right to express their point of view and to challenge the decision was repeated also in *Foodinho* and *Deliveroo*.[428] It is clear that without this information data subjects cannot contest the automatic decisions they were subjected to.

III.1.6.3.- Right to obtain an explanation of the automated decision

The right to obtain an explanation of the automated decision is a debated topic in academia. To begin with, Recital 71 GDPR lists among the suitable safeguards against automated decisions the right to 'obtain an explanation of the decision reached after such assessment'. However, this entitlement is not mirrored in the text of the GDPR. Hence, this inconsistency created disagreements among commentators.

Many authors were not convinced that controllers must guarantee this right or the extent of this right. In fact, they argued that the GDPR does not provide for a right to

---

[425] Art. 22(3) GDPR.

[426] Article 29 Data Protection Working Party, 'Guidelines on Automated Individual Decision-Making and Profiling for the Purposes of Regulation 2016/679' (n 187) 27.

[427] *Uber B.V. 2* (n 380).

[428] *Ordinanza ingiunzione nei confronti di Foodinho SRL* (n 212); *Ordinanza ingiunzione nei confronti di Deliveroo Italy SRL* (n 271).

a precise explanation,[429] that this right has shaky foundations,[430] or that the right is limited to a prior explanation of the system functionality.[431] On the other hand, some commentators and public organisations supported granting a right to obtain an explanation. They based their positions on the fact that the GDPR grants the right to obtain meaningful information about the logic involved (Art. 13 and 14 GDPR),[432] that it is implicit in the right to contest the decision in Art. 22(3) GDPR,[433] that rejecting the right on grounds that it is contained only in the recitals is an over-formalistic stance,[434] and that recitals do shed light on the meaning and intention of binding legal provisions.[435]

Obtaining 'an explanation of the decision reached' entails providing the data subject relevant and sufficient information to know what the decision is about. Without some essential aspects of the automated decisions, other rights cannot be exercised, in particular the right to contest the decision. Hence, receiving information about the process, methodology, reasons and potential result or decision of the AI system is a 'necessary precondition' to contest an automated decision.[436] Additionally, while recitals are nonbinding, they provide critical interpretative guidance on the legislative text and are helpful to establish the nature of a legal provision.[437]

---

[429] Castets-Renard (n 72) 95.

[430] Edwards and Veale (n 297) 50.

[431] Wachter, Mittelstadt and Floridi (n 1) 79.

[432] Bryce Goodman and Seth Flaxman, 'European Union Regulations on Algorithmic Decision-Making and a "Right to Explanation"'; Selbst and Powles (n 1) 235; Margot E Kaminski, 'The Right to Explanation, Explained' (2019) 34 Berkeley Technology Law Journal 218, 210.

[433] Article 29 Data Protection Working Party, 'Guidelines on Automated Individual Decision-Making and Profiling for the Purposes of Regulation 2016/679' (n 187); Isak Mendoza and Lee A Bygrave, 'The Right Not to Be Subject to Automated Decisions Based on Profiling' (2017) 2017–20 University of Oslo Faculty of Law Legal Studies Research Paper Series 1.

[434] Brkan (n 1) 16.

[435] Information Commissioner's Office, 'Explaining Decisions Made with AI' (n 104) 13.

[436] Recommendation CM/Rec(2020)1 of the Committee of Ministers to member States on the human rights impacts of algorithmic systems 2020 para B.4.3.

[437] Tadas Klimas and Jurarate Vaiciukaite, 'The Law of Recitals in European Community Legislation ' (2008) 15 ILSA Journal or International and Comparative Law 61, 62–63. Also, Case C-355/95, *Textilwerke Deggendorf GmbH v Commission of the European Communities and Federal Republic of*

Certainly, giving this information to data subjects may be challenging[438] due to the different nature of the decisions, technical obstacles, intellectual property rights and state secrets. The different options to satisfy the transparency requirements will be addressed below.

III.1.6.4.- Additional safeguards

Whereas the right to obtain human intervention, to express their viewpoint and to challenge the automated decision are mandatory guarantees, the GDPR also requires controllers to protect data subjects' rights and freedoms by way of implementing other suitable safeguards. The inclusion of the phrase 'at least' before mentioning the safeguards in Art. 22(3) GDPR implies that these constitute an 'open list' that works as a baseline of protection, but controllers must, according to the particular circumstances, include more safeguards to protect individual rights.[439]

Among the measures that controllers may implement when carrying out automated decision-making are quality assurance checks, algorithmic assessments and auditing, vendor screening, establishing data minimisation measures and data retention periods, de-identifying personal data, certification mechanisms or codes of conduct, ethical review panels to evaluate potential harms,[440] human rights impact assessments[441] or ethical impact assessments.[442]

There have been only a few decisions interpreting this provision, in particular concerning the type of safeguards and the particular situations where they should be implemented, in addition to those expressly mentioned in art. 22(3) GDPR. For instance, the Italian Data Protection Authority fined two companies for not

---

Germany, [1997] ECLI:EU:C:1997:24, Case 215/88, *Casa Fleischhandels-GmbH v Bundesanstalt für landwirtschaftliche Marktordnung* [1989] ECLI:EU:C:1989:331.

[438] Wachter, Mittelstadt and Floridi (n 1); Sandra Wachter and Brendt Mittelstadt, 'A Right to Reasonable Inferences: Re-Thinking Data Protection Law in the Age of Big Data and AI' (2019) 2 Columbia Business Law Review 494; Castets-Renard (n 72); Brkan (n 1); Edwards and Veale (n 297).

[439] Kaminski (n 430) 198.

[440] Article 29 Data Protection Working Party, 'Guidelines on Automated Individual Decision-Making and Profiling for the Purposes of Regulation 2016/679' (n 187) 32.

[441] Recommendation CM/Rec(2020)1 of the Committee of Ministers to member States on the human rights impacts of algorithmic systems.

[442] UNESCO, 'Recommendation on the Ethics in Artificial Intelligence' (2021).

implementing technical and organisational measures to periodically verify the accuracy of the outcome of the algorithm, check the pertinency and accuracy of the input data in relation to the processing purposes, and reduce the risk of distortions and discriminatory effects of the platform, including the ranking system and the system to allocate the rides to riders.[443]

The alternative measures that can be implemented to reduce the harmful effects of automated decision systems and AI systems, in general, will be explored in further detail below.

III.2.- Rights derived from transparency obligations

III.2.1.- Information rights and right to access

Transparency is one of the main data protection principles.[444] Data processing is transparent if data controllers provide information about the collection and processing in a clear, adequate and timely manner. Transparency is an all-encompassing mandate that applies to every provision of the GDPR, in particular to the delivery of information to the data subject concerning the processing of his or her personal data,[445] his or her rights under the GDPR[446] and the facilitation of the exercise of these rights.[447] While the specific content of the information to be provided is regulated in Arts. 13-15 GDPR, Art. 12 establishes the minimum conditions of transparency of information. Transparency obligations establish the manner, the content and the timing of the information.

In general, privacy notices and other documents that provide information to data subjects about the processing of their personal data are extremely long and complex, containing abundant legal jargon. They seem to be drafted for lawyers to lawyers, and

[443] *Ordinanza ingiunzione nei confronti di Foodinho SRL* (n 212); *Ordinanza ingiunzione nei confronti di Deliveroo Italy SRL* (n 271).

[444] Art. 5(1)(a) GDPR.

[445] Arts. 13 and 14 GDPR.

[446] Arts. 15 to 22 and 34 GDPR.

[447] Art. 12 GDPR.

in many cases, their readability score is equivalent to academic publications.[448] This contravenes the goals of data protection law, which is to provide relevant information to data subjects so that they can understand the processing operations, their consequences and how they can exercise their rights. Therefore, the GDPR establishes that data controllers must communicate the processing purposes and the rights of data subjects in clear, easy and plain language,[449] according to their audience. Individuals should be able to understand what they are reading, the consequences of the processing and how they can exercise their rights.

Concerning the content of the privacy notices, data subjects should be aware of the identity of the controller, the purposes of the processing, the rights they enjoy and the risks posed by the processing activities, and whether the information is transmitted to other data processors, be the recipient a private or public institution.[450]

As a rule, this information must be provided before the data are obtained or at the collection point[451] or, if the data is gathered from other sources or third parties, shortly after having obtained the data and no later than one month.[452] Active requests of information from data subjects, through data subject access requests, can be exercised at any time and the information should be produced without undue delay.[453] Transparency constitutes the basis upon which the whole set of data subjects' rights is built. Without transparency the individuals' rights are illusory. Individuals are unable to enjoy their rights without a clear understanding of what data is processed and how their data is used. Hence, transparency is a precondition for the full exercise of their rights,[454] which is also applicable to the transparency of AI solutions.[455]

---

[448] Uri Benolie and Shmuel I Becher, 'The Duty to Read the Unreadable' (2019) 60 Boston College Law Review 2256, 2294.

[449] Recital 39 GDPR.

[450] In Case C-201/14, *Smaranda Bara and Others v Casa Naţională de Asigurări de Sănătate and Others*. [2015] ECLI:EU:C:2015:638, para 34.

[451] Art. 13(1) GDPR.

[452] Art. 14(3)(a) GDPR.

[453] *Haralambie v Romania*. App no. 21737/03 (ECtHR 27 October 2009).

[454] Gabriela Zanfir-Fortuna, 'Article 13 Information to Be Provided Where Personal Data Are Collected from Teh Data Subject' in Christopher Kuner, Lee A Bygrave and Christopher Docksey (eds), *The EU General Data Protection Regulation (GDPR)* (OUP 2020) 416.

[455] UNESCO (n 440) 37.

Compliance with transparency obligations is indeed a difficult endeavour. On the one hand, there is a duty to give detailed and sufficient information to completely understand the data processed and the purposes of processing and, other the other hand, the provision of information to comply with transparency obligations should be concise, transparent and understandable to laymen.[456] This conflict is exacerbated when controllers process personal data using complex technologies,[457] as is the case when the processing operations are performed with AI systems.

While transparency is not a one-off obligation and it affects every stage of the processing of personal data,[458] transparency obligations mainly emerge at two particular moments when the processing of personal data is carried out by AI solutions: when personal data is fed to the AI system and when the latter renders the result using personal data.[459] In other words, both the development of an AI system and its deployment carry their issues.

Where controllers plan to use personal data to develop AI systems they must communicate this purpose to the data subject (and find a proper legal basis to perform the processing). The problem here concerns the difficulty of establishing and providing individuals in advance with the legally required information. Even where controllers inform data subjects of their intention to employ personal data to develop their systems, machine learning models may find unexpected correlations in the data. Hence, if the specific purposes of processing are detailed beforehand, the information provided to the data subjects via the privacy notice should be updated accordingly at a later stage to reflect the changes in the purposes of the processing. At the deployment stage, there may be also issues concerning the transparency duties of controllers. The channel to adequately inform data subjects about the processing of personal data using certain AI systems may be difficult to implement (for instance, real-time remote biometric identification systems used in publicly accessible areas).[460]

---

[456] European Data Protection Board, 'Guidelines on Transparency under Regulation 2016/679' (2018) 18.

[457] Recital 58 GDPR.

[458] Agencia Española de Protección de Datos (n 149) 31.

[459] European Parliamentary Research Service (n 248) 53.

[460] European Data Protection Board and European Data Protection Supervisor (n 199) 11.

While the specific information controllers must provide is listed in Arts. 13-15 GDPR, sometimes the dividing line between mandatory information is not crystal-clear. Case law on AI-assisted platforms can prove helpful in determining the extent of this obligation. The provision of some type of information presents no difficulty: controllers must be transparent towards data subjects concerning their plans for processing location data, the categories of data gathered (especially, concerning chats, emails and/or phone calls with the call centre) and the assessment of riders by retailers and customers, retention periods must be detailed for every data category, and the DPO contact details.[461] These are all obligations stipulated by the General Data Protection Regulation.

Data subjects do not have an unrestricted right to obtain information from controllers and they may find limits to their requests. Their requests are confined to obtaining access to their personal data, meaning that the definition of personal data is of utmost importance. However, even the definition of what constitutes personal data could be a matter of disagreement. On the one hand, courts have found that it cannot be part of a data subject access request the petition to obtain the data subject's profile (driver)[462] if it is based on internal referrals and reports to Uber customer service employees, or labels created by Uber to evaluate the driver's conduct (e.g. 'inappropriate behaviour' or 'police tag'). This reasoning is in line with the restrictive approach to personal data held by the CJEU in *YS v Minister voor Immigratie,*[463] where the Court considered that the legal assessment of the defendant's profile, while may contain personal data, does not in itself constitute personal data. Yet, a more recent and expansive interpretation of the concept of personal data[464] requires that both data derived and inferred from other data should also be included in the data subjects' access requests.[465] Moreover, where the provision of the data may affect the privacy

[461] *Ordinanza ingiunzione nei confronti di Foodinho SRL* (n 212).

[462] *Uber B.V. 1* (n 378).

[463] Case C-141/12, *YS v Minister voor Immigratie*. [2014] ECLI:EU:C:2014:2081, para 39.

[464] *Peter Nowak v Data Protection Commissioner* (n 22).

[465] European Data Protection Board, 'Guidelines 01/2022 on Data Subject Rights - Right of Access' (2022) para 96. Data derived from other data is, for instance, a classification according to similar features of the individuals or state of residence derived from zip code. Data inferred from other data is, for instance, a credit score or a health evaluation inferred from health data.

rights of others, controllers should implement adequate measures to protect the rights of those third persons. For instance, Uber drivers (data subjects) had the right to access the reports based on feedback from passengers, the start and end location of the trip, rating history and the ratings given by individual passengers, but the company (controller) must de-identify the person who made the feedback or the trip.[466] Similarly, it was not deemed to be part of a data subject access request the provision of customer transactions, booking cancellation history and booking acceptance history, since this information collides with the privacy rights of the passengers (third parties not involved in the access request).[467]

III.2.2.- Information on automated decision making

Controllers may take decisions based solely on automated processing, including profiling, which produces legal effects concerning data subjects or affects them in a similarly significant manner. When controllers process personal information in this way, they have a qualified duty of transparency. In this case, apart from the information that they must ordinarily provide, they must inform the data subjects of the existence of the automatic decision-making and provide meaningful information about the logic involved and the significance and the envisaged consequences of such processing for the data subject.[468] While informing the existence of automatic decision-making is a straightforward obligation, it is not clear the extent of the information controllers must provide to fulfil the last two requirements: a) meaningful information about the logic involved; and b) the significance and the envisaged consequences of such processing. In the following sections, the information that controllers must provide is evaluated.

III.2.2.1. Meaningful information about the logic involved

Where controllers automate decisions that have legal or significantly similar effects on individuals, they must meaningfully inform data subjects about the logic involved in the functioning of the system.

---

[466] District Court of Amsterdam, *Uber B.V. 1* (n 378).

[467] District Court of Amsterdam, *Ola Netherlands B.V.* (n 365).

[468] Art. 13(2)(f), 14(2)(g) and 15(1)(h) GDPR.

Meaningful information should be read as information useful to satisfy the purposes of the norm. The information provided should be understandable to the intended individual.[469] It concerns the importance of the information provided to the data subject and it attempts to avoid the delivery of a huge amount of irrelevant information about the model that individuals lacking technical background would not understand. However, there is a conflict between sufficiency and conciseness. Giving sufficiently detailed information to data subjects is important since this is the only way that they can understand a decision that significantly affects them. Yet, overloading them with information undermines the whole rationale of the provision since it creates confusion and overwhelms individuals. Hence, the depth and the quality of the information to satisfy its purposes are important.

Meaningful information definitively concerns the content of the information. The information provided should be relevant and fit for the purpose. It should also be actionable, meaning that it must have practical value. But it also relates to how the information is delivered. For instance, information fully provided in plain text may be difficult to read and understand for a large part of the relevant stakeholders affected by the outcomes of the AI systems. However, the inclusion of interactive tools, graphics, images, and even a layered approach could achieve better results.

What information controllers must provide concerning the *logic involved* behind the decision is also open to debate. At the heart of this question is the difference between a general explanation of how the system works and an explanation regarding a particular decision. On one side of the spectrum, it has been argued that meaningful information about the logic involved should be interpreted as providing a general outline or synopsis of the system functionality in advance.[470] This means that individuals are not entitled to obtain an explanation or a justification of the reasons for the outcome of the decision. This seems to be the rationale behind the *Uber B.V*. case under the District Court of Amsterdam where the court rejected the request from Uber drivers to receive information about the 'upfront pricing system', i.e., information about how this system functions and the parameters it uses to determine the price. For the

[469] National Institute of Standards and Technology, 'Four Principles of Explainable Artificial Intelligence' (2021) 3.

[470] Wachter, Mittelstadt and Floridi (n 1) 82.

court, it was a way to obtain knowledge about the algorithm and it didn't relate to a data subject access request under Art. 15(1)(h) GDPR.[471]

However, this does not seem to be the rationale of the GDPR which aims at granting sufficient information to data subjects so that they would be able to challenge the decision whenever they consider it suitable. An alternative interpretation supports the idea of providing individuals with information concerning the reasons behind or the criteria on which the decision was based. The information that controllers are required to provide should be sufficient for an ordinary person to understand the rationale of the decision.[472] While it does not necessarily entail a complex explanation of the algorithms used or full disclosure of the AI system or algorithm,[473] it should be read as a requirement for the decision-maker to disclose the reasons or arguments behind the individual result.[474] For instance, in the case of gig workers, companies must clarify the specific calculation criteria adopted to establish the statistics of each worker.[475] This seemed to be the position of the Italian Court of Cassation in *Mevaluate* which considered that where controllers use automated means to score individuals, the latter should be aware of the executive scheme of the algorithm (which specifies the particular sequence of steps that must be performed) and its constitutive elements.[476] Yet, the debate on what constitutes the 'logic involved' remains unclear, since the Court of Cassation did not concretely delineate the constituent elements that a data subject is entitled to know to be deemed properly informed.

However, even under the strictest interpretation of the concept of the logic involved in the processing (as held in *Uber B.V.* by the District Court of Amsterdam), general information about the system should be provided. This means that the more

---

[471] *Uber B.V. 1* (n 378).

[472] Article 29 Data Protection Working Party, 'Guidelines on Automated Individual Decision-Making and Profiling for the Purposes of Regulation 2016/679' (n 187) 25.

[473] Zanfir-Fortuna (n 452) 430; Castets-Renard (n 72) 120.

[474] Finale Doshi-Velez and Mason Kortz, 'Accountability of AI Under the Law: The Role of Explanation' (2017) 2.

[475] *FILCAMS CGIL Bologna e altri v. Deliveroo Italia SRL* (n 214). The court held that the company did not disclose concrete aspects about the inner workings of the algorithm, because it only provided a generic reference of the parameters evaluated to take the automated decisions (concerning reliability and participation).

[476] *Garante per la Protezione dei Dati Personali v. Associazione Mevaluate Onlu* (n 295).

interpretable the model is, the easier will be to effectively provide relevant information about the logic involved in the processing. Hence, using broadly interpretable models (e.g. logistic regression, decision trees, k-nearest neighbours, and even logic- and knowledge-based techniques), instead of 'black-box' algorithms (like deep learning techniques), will contribute to improving the effectiveness and clarity of the information to convey to data subjects.

There should be noticed, though, that there is a difference between the breadth of information provided both before and after the decision was made. It is argued that Arts. 13 and 14 GDPR, which dictate the rules on the information to be provided before the processing is made, would allow individuals to have an overview of the automatic decision-making system in advance, while Art. 15 GDPR, which governs data subject access requests, requires controllers to reveal the information mentioned above,[477] i.e., the reasons behind the outcome or decision reached. The latter interpretation is more protective of the individual rights since without an understanding of the reasons or justifications of the decisions, challenging an automatic decision, which is a right expressly granted to individuals subject to the automatic decision in Art. 22(3) GDPR, would be illusory. Only after becoming acquainted with the arguments used to reach such a decision the data subject will be in a position to contest it, otherwise, it would force individuals to 'shoot in the dark'.

Where possible, the controller should be able to provide remote access to a secure system so that the data subject can have direct access to his or her personal data. That right should not adversely affect the rights or freedoms of others, including trade secrets or intellectual property and in particular the copyright protecting the software. However, the result of those considerations should not be a refusal to provide all information to the data subject. Where the controller processes a large quantity of information concerning the data subject, the former should be able to request that, before the information is delivered, the latter specifies the information or processing activities to which the request relates.

Finally, it is important to highlight that while intellectual property rights and trade secrets can constitute a hurdle for disclosure of some kind of information concerning

[477] Kaminski (n 430) 200.

the algorithm, individuals cannot be deprived of their fundamental rights (access to their personal information) for these reasons.[478]

III.2.2.2.- Significance and envisaged consequences of the processing

When it comes to communicating the significance and envisaged consequences of the processing, controllers must provide data subjects information about how the processing operations might influence the individuals' rights and freedoms.[479]

There is an open debate about what kind of consequences should be informed to the data subjects. On the one hand, it could be argued that only concrete significant consequences should be informed. For example, the (actual) refusal of an online credit application is an example mentioned in Recital 71 GDPR. On the other hand, the significance and envisaged consequences may refer to any potential or possible result. Only where an outcome is reasonably predictable as such for the data subjects can be qualified as significant.[480] This interpretation, apart from being more protective of the data subject rights', has also normative support. To begin with, both Arts. 13 and 14 GDPR refer to explaining the possible outcomes before the operations start. At this point, no concrete or realised consequence can be informed simply because there is no outcome of the AI system. Furthermore, the European Data Protection Board stated that this obligation is satisfied where the controller provides information about 'how the automated decision-making *might* affect the data subject'.[481] The use of the conditional 'might' suggests that the possibility is only potential or hypothetical. This is also corroborated by the example provided in the guidelines, concerning an insurance company that employs an automated decision-making system to establish

---

[478] *Judgment no. k. PL. ÚS 25 / 2019-117 - 492/2021 Coll.* (n 343) para 135; Recommendation CM/Rec(2020)1 of the Committee of Ministers to member States on the human rights impacts of algorithmic systems para B.5.2.

[479] Article 29 Data Protection Working Party, 'Guidelines on Automated Individual Decision-Making and Profiling for the Purposes of Regulation 2016/679' (n 187) 26.

[480] Reuben Binns and Michael Veale, 'Is That Your Final Decision? Multi-Stage Profiling, Selective Effects, and Article 22 of the GDPR' (2021) 11 International Data Privacy Law 319, 15.

[481] Article 29 Data Protection Working Party, 'Guidelines on Automated Individual Decision-Making and Profiling for the Purposes of Regulation 2016/679' (n 187) 26.

insurance premiums and describes that unsafe driving may result in a more expensive insurance premium.

III.3.- Other data subject rights

III.3.1.- Right to rectification

Data subjects can request their personal data to be rectified or completed if it is inaccurate or incomplete, respectively. This right is linked to the controller's obligation to keep data accurate and updated[482] and also to the right to effective legal protection established in Art. 47 CFR.[483]

The right to rectification can be invoked by individuals at both stages of the AI system lifecycle. Data subjects may claim their data to be rectified in the development stage. Where the aim of the processing is finding broad correlations in the datasets, some mistakes or errors in the personal data either in the model or used to train the model may not affect the overall statistical accuracy of the system, since models are built using large datasets. However, the fact that the statistical accuracy of the model remains unaffected by particular errors in the dataset does not allow controllers to deny data subject petitions concerning rectifications or updates of personal data contained in the training dataset or in the model itself.

More common will be cases where individuals contest the accuracy of the output of an AI system, i.e., during the deployment of the AI system. Admittedly, the results of the algorithm can be considered personal data, for instance, a profile or a categorization, because personal data includes also subjective information like opinions or evaluations insofar as it 'relates' to the individual.[484] Hence, the right to rectification covers profiling, inferences, categorizations, etc. However, it may be challenging to request a rectification of the outcome of an AI system, since these results are mere statistical predictions, not statements of fact. Statistical predictions

[482] Art. 5(1)(d) GDPR.

[483] Case C-362/14, *Maximillian Schrems v Data Protection Commissioner*. [2015] ECLI:EU:C:2015:650, para 95.

[484] *Peter Nowak v Data Protection Commissioner* (n 22) para 34.

allow a certain margin of error, which is not permissible in statements of fact, hence requests for rectification of the system's output are not likely to succeed.

III.3.2.- Right to erasure

Individuals have the right to request the controller the immediate erasure of their personal data.[485] Where a data subject exercises this right, controllers are under the obligation to not only erase the data directly processed by them, but they also must communicate with other known recipients of the personal data about the request, allowing the data subjects to wipe out their personal information from the offline and online environment.[486]

For this right to apply, the personal data must be unnecessary for the purposes it was collected,[487] unlawfully processed,[488] or collected from children.[489] In addition, it applies if the data subject withdraws his or her consent and there is no other legal basis to process the data.[490] Finally, this right can also be exercised by a data subject who objects to the processing and there are no overriding legitimate grounds for the processing.[491] The burden of proof to demonstrate the overriding legitimate grounds is on the controller.[492] This means balancing the rights of the data subject and the compelling legitimate grounds of the controller. Controllers may find it difficult to argue that a reduction of the accuracy constitutes a compelling interest. The removal of information about an individual from the training dataset does not generally have a detrimental effect on the capacity of the system to produce accurate predictions.[493]

---

[485] Art. 17(1) GDPR.

[486] Pollicino and Nicola (n 180) 41.

[487] Art. 17(1)(a) GDPR.

[488] Art. 17(1)(d) GDPR. This provision can be seen as a general clause that includes subparagraphs (a) and (b) of Art. 17(1). See Herke Kranenborg, 'Article 17. Right to Erasure ('right to Be Forgotten')' in Christopher Kuner, Lee A Bygrave and Christoper Docksey (eds), *The EU General Data Protection Regulation (GDPR)* (OUP 2020) 481.

[489] Art. 17(1)(f) GDPR.

[490] Art. 17(1)(b) GDPR.

[491] Art. 17(1)(c) GDPR.

[492] European Data Protection Board, 'Guidelines 5/2019 on the Criteria of the Right to Be Forgotten in the Search Engines Cases under the GDPR (Part 1)' (2020) 8.

[493] Information Commissioner's Office, 'Guidance on AI and Data Protection' (n 147) 67.

Datasets used to train AI models include a vast number of examples, so in general the statistical impact of removing the personal data from the individual who requested it is negligible.

There are many obstacles in general to the full implementation of this right,[494] many of which are related to the clash between privacy and transparency[495] and freedom of expression. In particular, concerning AI solutions, the first issue is related to the kinds of data that an individual could request deletion. In other words, whether the right to erasure should cover all types of personal data, i.e. provided, observed, derived and inferred. Closely related to this topic, the former WP29 considered that for the right to data portability only the first two types can be considered as "provided by the data subject" (actively provided and observed[496]), but not the last two.[497] While it has been argued that the right to erasure should follow this logic,[498] Art. 17 GDPR does not require that the data were provided by the data subject. Hence, there are reasons to consider that also data derived and inferred[499] from other data should be included in the erasure request. This interpretation is also in line with the *Google Spain* case where the CJEU requested the deletion of an inference from the search engine's algorithm, and the CJEU ruled in favour of the claimant.[500] This expansive interpretation has an impact on the workload controllers have to fulfil erasure requests.

As the erasure of personal data is a data subject's right, controllers must design their systems in a way they can honour deletion requests. Controllers cannot justify a denial of compliance with erasure requests due to problems related to the

---

[494] Oreste Pollicino and Virgilio D'Antonio, 'The Right to Be Forgotten in Italy' in Franz Werro (ed), *The Right To Be Forgotten. A Comparative Study of the Emergent Right's Evolution and Application in Europe, the Americas, and Asia* (Springer 2020) 170.

[495] Oreste Pollicino and Giovanni De Gregorio, 'Privacy or Transparency? A New Balancing of Interests for the "Right to Be Forgotten" of Personal Data Published in Public Registers' [2017] The Italian Law Journal 647, 648.

[496] See art. 7 ePrivacy Regulation draft on the possibility of deleting metadata and communication data.

[497] Article 29 Data Protection Working Party, 'Guidelines on the Right to "Data Portability"' (2017) 9.

[498] Edwards and Veale (n 297) 68–69.

[499] European Parliamentary Research Service (n 248) 57.

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

configuration of the software since, according to the principle of privacy by design,[501] they must adopt adequate technical and organisational measures in their processes to protect data subjects' rights. This does not mean that honouring data subjects' rights is a straightforward task when controllers develop or deploy AI systems. On the contrary, erasing information is a technically complex task in AI systems since the data is stored for processing at different locations of the system and it is also replicated across systems for backups.[502] Additionally, it is worth noticing that, in general, data from databases are not erased or destroyed, but solely identified as 'deleted' and hidden from the search indexes. Some time may be needed before the freed space is reused again with new data,[503] which effectively destroys the old data that was the matter of the initial erasure.

Yet another important problem concerning Art. 17 GDPR is that to entirely erase the personal data included in an all-encompassing request from current AI models it may be necessary to retrain the algorithm with the remaining data or amend the features of the system.[504] This is because even after the deletion of the specific personal data related to the data subject, data concerning the petitioner may remain in predictions produced by the AI models trained on the erased information as an 'algorithmic shadow'.[505] Retraining machine learning models in most cases is unfeasible due to the high computational and engineering cost[506] and time[507] required to do so, in particular for complex models like artificial neural networks.

---

[501] Art. 25(1) GDPR.

[502] Tiffany Li, Eduard Fosch Villaronga and Peter Kieseberg, 'Humans Forget, Machines Remember: Artificial Intelligence and the Right to Be Forgotten' (2018) 34 Computer Law & Security Review 308, 10.

[503] ibid 12.

[504] Eli MacKinnon and Dr. Jennifer King, 'Regulating AI Through Data Privacy' (Stanford University Human-Centered Artificial Intelligence, Jan 11th, 2022) <https://hai.stanford.edu/news/regulating-ai-through-data-privacy> accessed 14/04/2022.

[505] Tiffany Li, 'Algorithmic Destruction' (2022) Forthcomin SMU Law Review 1, 11.

[506] Zachary Izzo and others, 'Approximate Data Deletion from Machine Learning Models', *Proceedings of The 24th International Conference on Artificial Intelligence and Statistics* (2021) 1; Edwards and Veale (n 297) 70–71.

[507] Antonio Ginart and others, 'Making AI Forget You: Data Deletion in Machine Learning', *Proceedings of the 33rd International Conference on Neural Information Processing Systems* (2019) 3519.

III.3.3.- Right to restrict data processing

Another right granted to data subjects is the right to restrict processing, which entails marking stored personal data to limit their processing in the future.[508] Data subjects can request the restriction of the processing of their personal data for different reasons and for the time required to fulfil the specific objective. First, individuals may require the restriction where they contest the accuracy of the data and the restriction will last until the controller checks the correctness of the data processed.[509] Second, if the processing is unlawful and the data subjects, instead of requiring the deletion, demand the restriction of its use.[510] Third, where the data is no longer needed for the controller but the data subject plans to use the data to exercise legal rights. Forth, where data subjects object to processing on grounds concerning their particular situation, until the assessment of whether legitimate interests invoked by the controller to process the data is carried out.[511] Basically, this right allows the interruption of the processing activities while the assessment of those circumstances is being carried out. The controller must suspend the processing operations except for storage.[512]

This right is a substitute for the right to erasure, giving individuals a milder tool that keeps the personal data intact, and it is linked to the enjoyment of the right to rectification and objection[513] since the duration of the restriction depends on the outcome of those claims. Controllers may ensure this right by technical means like transitorily transferring the contested data to a different processing system only for storage, making the personal information not usable, or clearing the data made public from a website.[514] Since controllers must limit the processing operations carried out on the data, they can only store it, the issues they may face in the context of AI systems are similar to those related to the right to erasure.

---

[508] Art. 4(3) GDPR.

[509] Art. 18(1)(a) GDPR.

[510] Art. 18(1)(b) GDPR.

[511] Art. 18(1)(d) and 21(1) GDPR.

[512] Art. 18(2) GDPR.

[513] Gloria Gonzalez Fuster, 'Article 18. Right to Restriction of Processing' in Christopher Kuner and Lee A Bygrave (eds), *The EU General Data Protection Regulation (GDPR): A Commentary* (OUP 2020) 486.

[514] Rec. 67 GDPR.

III.3.4.- Right to object processing

Another way for data subjects to oppose the processing of their personal data is by the exercise of the right to object processing. Individuals can object to the processing of their personal data where processing is based on public interest or official authority vested in the controller or legitimate interest based on grounds related to their particular situation, including profiling based on those provisions.[515] The right also applies even where the processing is carried out for scientific research or statistical purposes in private commercial contexts.[516]

In these cases, controllers can oppose the data subject's objection if they demonstrate compelling legitimate grounds for the processing which prevail over the data subject's interests and rights. It should be borne in mind that, in general, individuals' rights and interests override the economic interests of the controllers. This requires an evaluation of the nature of the information and its sensitivity to data subjects' private life and, if the information was published, the assessment should also include the interest of the society in having that information.[517]

Additionally, individuals have an unconditional right to halt processing operations involving their personal data used for direct marketing, including profiling provided it is linked to direct marketing.[518] Therefore, controllers have no excuses to interrupt the processing where the individuals exercised this right and the processing is for direct marketing.

It gives data subjects the chance to oppose the processing of their personal data and it mirrors the right to withdraw consent where the processing is grounded on consent,[519] but in the context of Art. 6(1) lit. (e) and (g) GDPR and direct marketing. The fact that it requires an active behaviour from the data subject to contest the

---

[515] Art. 21(1) GDPR.

[516] Art. 21(6) GDPR a contrario sensu.

[517] *Google Spain SL and Google Inc. v Agencia Española de Protección de Datos (AEPD) and Mario Costeja González* (n 498) para 35.

[518] Art. 21(2) GDPR.

[519] Gabriela Zanfir-Fortuna, 'Article 21. Right to Object' in Christopher Kuner and Lee A Bygrave (eds), *The EU General Data Protection Regulation (GDPR). A commentary* (OUP 2020) 509.

processing, means that the exercise of this right reveals itself as an opt-out provision.[520]

III.3.5.- Right to data portability

The right to data portability allows individuals to obtain and reuse their personal data in different services. They can request from controllers their personal data and transmit the data to another controller under certain circumstances. This right is aimed at increasing individuals' autonomy, fostering competition and promoting innovation, without requiring individuals to terminate the relationship with the original controller.[521]

In the context of AI, it should be noted that the right to data portability has three important limitations. First, it proceeds only when the processing is based on consent or for the performance of a contract.[522] Whereas it is not mandatory to provide for this right when processing is based on any other of the four grounds, it has been suggested that as a good practice that controllers should allow individuals to obtain data in a portable format under a voluntary scheme.[523]

Secondly, the data must have been provided by the data subject to the controller.[524] This means that the information the controller must provide to comply with Art. 20 GDPR request is less comprehensive than information accessed under Art. 15 GDPR[525] or the personal data a data subject is entitled to request erasure under art. 17 GDPR.[526] By 'provided by the data subject' should be understood both information

---

[520] Michael Veale, Reuben Binns and Jef Ausloos, 'When Data Protection by Design and Data Subject Rights Clash' (2018) 8 International Data Privacy Law 105, 111.

[521] Josef Drexel, 'Legal Challenges of the Changing Role of Personal and Non-Personal Data in the Data Economy' in Alberto De Franceschi, Reiner Schulze and Oreste Pollicino (eds), *Digital Revolution - New Challenges for Law* (Nomos 2019) 38.

[522] Art. 20(2)(a) GDPR.

[523] Article 29 Data Protection Working Party, 'Guidelines on the Right to "Data Portability"' (n 495) 8; Article 29 Data Protection Working Party, 'Opinion 06/2014 on the Notion of Legitimate Interests of the Data Controller under Article 7 of Directive 95/46/EC' (n 304) 43.

[524] Art. 20(1) GDPR.

[525] European Data Protection Board, 'Guidelines 01/2022 on Data Subject Rights - Right of Access' (n 463) para 96.

[526] *Google Spain SL and Google Inc. v Agencia Española de Protección de Datos (AEPD) and Mario Costeja González* (n 498).

intentionally provided by the individual (e.g. when uploading photos) and the information the controller observed from the individual's behaviour or interaction with the service or device.[527] For instance, in the context of the use of virtual voice assistants, this information should include the data the user has transmitted using his or her voice (v.gr., the questions or orders made to the virtual assistant) and the information used to create the user account (such as email address, birth date, gender).[528] However, it should not be included within the data the individual is entitled to request in this right any information that is the consequence of a further assessment of the user's behaviour or interaction, i.e, data derived and inferred from the information given by the individual (v.gr. profiles or recommendations).[529]

Thirdly, the data must be interpretable, it should be provided in widely used public machine-readable formats like XML, JSON or CSV and it is up to the controller the decision about which format best satisfies compliance with this right. Yet, PDF files are unsuitable to fulfil this obligation since they do not allow the reuse of the information.[530]

In the context of AI systems, data controllers may process a huge amount of information from the data subject. Where the data subject exercises this right, he or she should be able to understand the data forwarded by the controller. Hence, the controller, as part of its accountability obligations, could allow data subjects to download part of the whole set of information.[531]

While the information in its raw form used to train an AI model should be considered as provided by the data subject -thus available to be ported-, where pre-processing activities substantially alter the original data provided by the data subject this kind of pre-processed information may not be included into the information that the controller must provide.[532] This information keeps the status of personal data and other rights can be exercised concerning them (e.g. right to access or erasure), but the data subject cannot switch it to a different service provider under the right to data portability.

---

[527] Article 29 Data Protection Working Party, 'Guidelines on the Right to "Data Portability"' (n 495) 10.

[528] European Data Protection Board, 'Guidelines 02/2021 on Virtual Voice Assistants' (2021) 37.

[529] Article 29 Data Protection Working Party, 'Guidelines on the Right to "Data Portability"' (n 495) 11.

[530] *Uber B.V. 1* (n 378).

[531] Article 29 Data Protection Working Party, 'Guidelines on the Right to "Data Portability"' (n 495) 18.

[532] Information Commissioner's Office, 'Guidance on AI and Data Protection' (n 147) 68.

Finally, it is worth noting that the outcome of an AI system is not subject to the right to data portability.[533] Whereas the inference, prediction or classification may be considered as personal data, since the information is derived or inferred from personal data provided by the individual, and not information knowingly provided or observed from him or her, the right to port this data is not applicable.

III.4.- General GDPR accountability mechanisms

The GDPR does not only grants subjective rights to individuals. The GDPR provides a long list of rights that individuals can exercise, but it also establishes a structure of control that protects the rights granted to data subjects by imposing accountability and oversight obligations to controllers. The objective behind this idea is that granting individual rights to data subjects is not enough to balance the power and information asymmetries vis-a-vis controllers. Even if data subjects do not exercise their rights, data controllers and processors have accountability obligations to comply with. Hence, the GDPR establishes also organisational obligations to controllers to make data protection rights more effective.

However, the GDPR does not concretely specify which safeguards should be applied in concrete cases, thus giving controllers discretionary powers to decide which particular accountability measure to apply to guarantee the data subjects' rights.[534] The GDPR regulates the data processing activities of controllers and processors by partially delegating regulatory functions also to non-public actors and by meta-regulating, a process that is known as 'collaborative governance'.[535] The gaps left by the regulation are filled by guidelines, standards, codes of conduct, best practices and other soft law instruments.

The principle of accountability for entities processing personal data was first introduced by the OECD 1980 Privacy guidelines.[536] Accountability can be seen as a

---

[533] ibid 69.

[534] Giovanni De Gregorio, *Digital Constitutionalism in Europe. Reframing Rights and Powers in the Algorithmic Society* (CUP 2022) 77.

[535] Margot E Kaminski and Gianclaudio Malgieri, 'Algorithmic Impact Assessments under the GDPR: Producing Multi-Layered Explanations' (2021) 11 International Data Privacy Law 125, 127.

[536] Art. 14 Annex to the Recommendation of the Council of 23rd September 1980: Guidelines Governing the Protection of Privacy and Transborder Data Flows of Personal data (OECD).

way of presenting how controllers are complying with their obligations and allowing others to verify this compliance.[537] It is worth remembering that data controllers are responsible for compliance and must be able to prove compliance with their data protection obligations, thus the principle of accountability reinforces the effective application of data protection provisions.

In this section, some of the most relevant accountability measures are introduced, as well as the particularities controllers have to comply with when they process personal data using AI systems.

III.4.1.- Records of processing activities

According to the GDPR, controllers and processors, and eventually, their representatives must keep records of their processing activities and make the records available to the data protection authorities upon request.[538] This is an accountability obligation whose purpose is to evidence compliance with the GDPR.[539]

There are some differences concerning the information that controllers and processors must register, but controllers must describe the central characteristics of the processing operations,[540] and in particular, the purposes of the processing must be clearly registered.[541] Recording the purposes of the processing could pose difficulties for controllers when developing AI systems. Sometimes the preliminary evaluation of the information is investigative and it may lack concretely defined purposes or needs, or the purposes of processing may change if the AI system finds unexpected correlations in the datasets.[542] This means that the registration of the processing operations when AI systems are employed may be challenging and controllers may need to update frequently the register of processing activities.

---

[537] Article 29 Data Protection Working Party, 'Opinion 3/2010 on the Principle of Accountability' (2010) 7.

[538] Art. 30 GDPR.

[539] Rec. 82 GDPR.

[540] Waltraut Kotschy, 'Article 30. Records of Processing Activities' in Christopher Kuner and Lee A Bygrave (eds), *The EU General Data Protection Regulation (GDPR). A commentary* (OUP 2020) 620.

[541] Art. 30(1)(b) GDPR.

[542] Information Commissioner's Office, 'Big Data, Artificial Intelligence, Machine Learning and Data Protection Data Protection Act and General Data Protection Regulation' (n 202) 51.

Mapping the processing activities is a fundamental task that any controller or processor processing personal data must fulfil properly. Not only do those who disregard this obligation are subject to disciplinary actions or sanctions from data protection authorities,[543] but the records of processing activities provide companies with valuable insights and constitute one of the most important tools for proper data management. However, building a register of processing activities can be burdensome for AI systems, since the processing operations carried out by AI systems may be extremely complex. For this reason, using specific software to perform this task (like data mapping software or privacy automation software) and asking for backup support from data science professionals may be very helpful.

III.4.2.- Data Protection Officer

Another accountability provision in the GDPR that certain controllers and processors must take into consideration is the duty to appoint a data protection officer (DPO).[544] A data protection officer is a data protection law and practice expert that helps the controller or processor in the supervision of their internal compliance with the GDPR[545] and in making sure that the processing of personal data does not violate the rights of the data subjects.[546] Besides, a data protection officer informs and advises controllers and processors about their data protection obligations, and must be consulted when carrying out the data protection impact assessment.

While controllers and processors are not under a general obligation to appoint a data protection officer, almost every public institution processing personal data must make the appointment.[547] Additionally, private controllers and processors whose core activities entail 'regular and systematic monitoring of personal data on large scale'[548]

---

[543] See art. 82(4)(a) GDPR. Also *Ordinanza ingiunzione nei confronti di Foodinho SRL* (n 212); *Ordinanza ingiunzione nei confronti di Deliveroo Italy SRL* (n 271). In these cases, the Italian Data Protection Authority found that the companies failed to include in their records of processing activities information about several categories of personal data, retention periods, and technical and organisational security measures.

[544] Art. 37(1) GDPR.

[545] Recital 97 GDPR.

[546] *CJEU (General Court), Oikonomopoulos* (2016) Case T-483.

[547] Except for courts acting in their judicial capacity, see Art. 37(1)(a) GDPR.

[548] Art. 37(1)(b) GDPR.

are obliged to appoint a data protection officer. National laws could also establish the mandatory designation of data protection officers. For instance, Spanish data protection law requires information society service providers to designate a DPO when they elaborate, on a large-scale, profiles of the users.[549]

It is important to bear in mind that 'core activities' do not relate to ancillary or auxiliary processing activities,[550] like payroll data processing. Instead, it concerns the primary processing activities, meaning the essential processing activities required to fulfil their objectives, or operations that are inextricably linked to the main activities.[551] Furthermore, for the obligation to be triggered the monitoring must be performed on large scale. Controllers and processors must factor in the number of individuals involved, the quantity of data, and the scope of the processing both in terms of time and geography to decide whether their activities entail large-scale processing.[552]

In this context, it is likely that controllers or processors developing or deploying AI systems or employing big data analytics to carry out online behaviour advertising, tracking individuals across the web or profiling individuals must appoint a data protection officer.[553] Since data protection officers must independently perform their duties, have expertise in data protection, be sufficiently resourced and report their activities to the highest management of the organisation where they work, they play a key role in GDPR compliance, particularly in the context of AI systems that process personal data. They can also assist in the demonstration of compliance with the legal framework.

The GDPR does not require specific qualifications or degrees, but it mandates that data protection officers have expert knowledge of the data protection legal framework, including its practical aspects, to be able to fulfil the functions legally assigned.[554] Moreover, their expertise should be valued in accordance with the data processing

---

[549] Art. 34(1)(d) Ley Orgánica 3/2018, de 5 de diciembre, de Protección de Datos Personales y garantía de los derechos digitales.

[550] Recital 97 GDPR.

[551] Article 29 Data Protection Working Party, 'Guidelines on Data Protection Officers' (2017) 7.

[552] ibid 8.

[553] Information Commissioner's Office, 'Big Data, Artificial Intelligence, Machine Learning and Data Protection Data Protection Act and General Data Protection Regulation' (n 202) 53.

[554] Art. 37(5) GDPR.

activities performed by the controller or processor,[555] in particular, it should be adequate to the quantity of data, to the complexity and nature of the processing operations.[556] This means that developers and deployers and AI systems should choose a person or organisation that, apart from being familiar with the data protection regulations, have a deep understanding of the features and issues related to AI systems.

III.4.3.- Data Protection by Design and by Default

Data controllers must guarantee the respect of the principles of data protection from the outset, which means that they must design, develop and deploy data processing operations taking care of the privacy and the protection of the personal information of individuals. The origin of this principle can be traced back to the 1990s when Ann Cavoukian, by then Ontario Privacy Commissioner, developed the idea of Privacy by Design.[557]

The GDPR establishes two related accountability obligations on data controllers: data protection by design and data protection by default. Data protection by design means that controllers must embed data protection principles into the design of their processing operations. Data protection by default, on the other hand, denotes the duty of the controller to preselect processing methods, values and alternatives that have the least data protection impact on individuals.

The thrust of this principle is to guarantee the adequate and effective processing of personal data by requiring controllers to implement technical and organisational measures and to accommodate the necessary safeguards into their processing activities to respect the rights of individuals, both at the moment they determine the means of processing and during the processing operations itself.[558] Arguably, the concept of technical and organisational measures is vague, which is detrimental to its enforcement, but at the same time, it is technology-neutral[559] and gives controllers

---

[555] Recital 97 GDPR.

[556] Article 29 Data Protection Working Party, 'Guidelines on Data Protection Officers' (n 549) 11.

[557] Ann Cavoukian, Scott Taylor and Martin E Abrams, 'Privacy by Design: Essential for Organizational Accountability and Strong Business Practices' (2010) 3 Identity in the Information Society 405, 407.

[558] Art. 25(1) GDPR.

[559] Oostveen (n 200) 158.

flexibility to comply with it. The scope of technical and organisational measures and necessary safeguards is very wide and in general, they should be regarded as 'any method or means that a controller may employ in the processing'.[560] And while the GDPR does not expressly require any particular technical and organisational measure, the measures must be fit to the intended purpose. Additionally, as data protection by design and by default is an accountability provision, controllers must demonstrate both the effective implementation of the measures and the suitability to their aims.

The processing of personal data using AI systems poses new challenges to the principle of privacy by design and by default,[561] because the logic and nature of the processing activities carried out by AI systems may contravene many of the most important data protection principles. Hence, data protection by design and by default plays an important role in ensuring the protection of data subjects.

The most well-known strategies to comply with data protection by design are pseudonymisation/anonymisation and encryption. Yet, there is a wide range of technical and organisational measures that allow controllers to comply with the principle of data protection by design and by default. In this sense, controllers should implement data protection by design strategies in the AI value chain. From the outset, employees must receive adequate training to guarantee that every person in the institution becomes acquainted with the necessity of and the risks linked to personal data processing activities.[562] Then, in the data collection stage AI systems operators can reduce the amount of information collected by clearly establishing the information that will be needed before starting the collection and making a selection (e.g. gathering fewer data points), they can aggregate data if personal data is not needed by using local anonymisation (anonymisation at the source) which erases all the personal data

[560] European Data Protection Board, 'Guidelines 4/2019 on Article 25 Data Protection by Design and by Default' (2020) 6.

[561] Ira S Rubinstein and Nathaniel Good, 'The Trouble with Article 25 (and How to Fix It): The Future of Data Protection by Design and Default' (2020) 10 International Data Privacy Law 37, 52.

[562] Norwegian Data Protection Authority (Datatylsinet), Software development with Data Protection by Design and by Default (2017) <https://www.datatilsynet.no/en/about-privacy/virksomhetenes-plikter/innebygd-personvern/data-protection-by-design-and-by-default/> retrieved on 01/06/2021.

before forwarding the information.[563] Additionally, controllers must notify individuals whose information is being used to perform data analytics and inferencing about these circumstances before the processing starts.[564]

Then, during the data analytics stage, the implementation of data aggregation or anonymisation techniques to avoid inferences or singling out could help ensure data protection by design in AI systems. AI systems also need to store data, and if controllers using AI systems process personal data, they should also implement data protection by design strategies. In this stage, processing data in a distributed manner, using separated or de-decentralised storage and processing facilities may help achieve this objective.[565] During the use of the AI solution, controllers must evaluate at regular periods whether the AI system is performing according to the intended purpose and, where necessary, make reasonable adjustments to guarantee fair processing and reduce biases.[566] During the whole processing of personal data, controllers must implement security measures to guarantee the confidentiality, integrity and availability of the information.[567] Implementing measures to limit the access of operators and workers to personal data according to their roles, and, in particular, external third parties, is important to comply with the principle of data protection by design and by default.[568]

III.4.4.- Data Protection Impact Assessment

A data protection impact assessment (hereinafter DPIA) is a procedure aimed at describing the processing operations, evaluating the necessity and proportionality of this processing, and aiding to mitigate the risks to the rights and freedoms of

---

[563] European Union Agency For Network And Information Security, 'Privacy by Design in Big Data. An Overview of Privacy Enhancing Technologies in the Era of Big Data Analytics' (2015) 24.

[564] European Data Protection Board, 'Guidelines 4/2019 on Article 25 Data Protection by Design and by Default' (n 558) 18.

[565] European Union Agency For Network And Information Security (n 561) 26.

[566] European Data Protection Board, 'Guidelines 4/2019 on Article 25 Data Protection by Design and by Default' (n 558) 18.

[567] Information Commissioner's Office, 'Big Data, Artificial Intelligence, Machine Learning and Data Protection Data Protection Act and General Data Protection Regulation' (n 202) 73.

[568] *Ordinanza ingiunzione nei confronti di Foodinho SRL* (n 212); *Ordinanza ingiunzione nei confronti di Deliveroo Italy SRL* (n 271).

individuals. This process is obligatory only for certain processing activities: those that are likely to result in a high risk to the rights of individuals.[569]

Carrying out a DPIA is not compulsory for every single processing operation, because they are mandatory only where the processing is likely to result in a high risk to the rights of individuals.[570] However, controllers developing or using AI systems that process personal data will be required to perform a DPIA in the majority of cases.[571] In particular, the GDPR establishes that a DPIA is required where the processing operations involve a systematic and extensive evaluation of personal aspects concerning individuals based on automated processing, including profiling, and on which decisions are based that produce legal or similarly significant effects on them.[572] Furthermore, where the processing operations involve innovative use of new technologies[573] or the processing operations combine different sets of data,[574] the risks for data subjects are higher. For example, the processing of personal data on a large scale, concerning a plurality of personal data (including geolocation, phone calls, chat and e-emails, and details relating to each phase of management of the orders), performed via a digital platform (app) which is based on an algorithmic system that links supply and demand was deemed to have an innovative character, thus requiring a data protection impact assessment.[575] The DPIA must be carried out before processing personal data, and it must be reviewed where the risks of the processing activities change.[576] In addition, the GDPR permits Member States to establish a list

[569] Art. 35(1) GDPR.

[570] Art. 35(1) GDPR.

[571] Information Commissioner's Office, 'Guidance on AI and Data Protection' (n 147) 4.

[572] Information Commissioner's Office, 'Big Data, Artificial Intelligence, Machine Learning and Data Protection Data Protection Act and General Data Protection Regulation' (n 202) 99. The ICO based its opinion on Art. 35(3)(a) GDPR.

[573] Art. 35(1) and Recs. 89 and 91 GDPR.

[574] European Data Protection Board, 'Guidelines on Data Protection Impact Assessment (DPIA) and Determining Whether Processing Is "Likely to Result in a High Risk" for the Purposes of the GDPR' (2018) 10.

[575] *Ordinanza ingiunzione nei confronti di Foodinho SRL* (n 212); *Ordinanza ingiunzione nei confronti di Deliveroo Italy SRL* (n 271).

[576] Art. 35(1) and 35(11) GDPR.

of activities that carrying out a DPIA is mandatory.[577] Using this authorisation, some data protection authorities compel data controllers to undertake a DPIA before processing personal data using some AI systems or solutions, for instance for scoring of individuals,[578] credit rating or solvency rating,[579] or innovative uses of data using AI applications.[580]

Carrying out a DPIA, apart from being mandatory when processing personal data using AI systems, is also advantageous for companies. DPIAs are generally seen as early warning systems[581] because potential difficulties can be discovered before the processing starts, thus solving these issues is easier and cheaper. Additionally, it fosters data protection awareness within the company, which in turn reduces the risks of breaching the legal framework. It can also help to build public trust in the AI systems since their privacy-related fears are addressed from its design.[582]

The GDPR does not establish a particular methodology to carry out the DPIA. There are many different methodologies to perform the DPIA: from guidelines and templates issued by the European Data Protection Board (EDPB)[583] and the European Data

---

[577] Art. 35(4) GDPR.

[578] See for instance Polish DPIA list, point 1 <https://edpb.europa.eu/sites/default/files/decisions/pl-dpia-list_monitor_polski.pdf> or Italian DPIA list, point 1 <https://www.garanteprivacy.it/documents/10160/0/ALLEGATO+1+Elenco+delle+tipologie+di+trattamenti+soggetti+al+meccanismo+di+coerenza+da+sottoporre+a+valutazione+di+impatto> both accessed 25/05/2022.

[579] See for instance Slovak DPIA list, points 6 and 7 <https://iapp.org/media/pdf/resource_center/slovakia_blacklist.pdf> accessed 25/05/2022.

[580] See for instance Greek DPIA list, point 3.1 <https://www.dpa.gr/sites/default/files/2020-12/article_35_dpia_list_en.pdf> accessed 25/05/2022.

[581] David Wright, 'The State of the Art in Privacy Impact Assessment' (2012) 28 Computer Law & Security Review 54, 55.

[582] David Wright, 'Making Privacy Impact Assessment More Effective' (2013) 29 The Information Society 307, 313.

[583] European Data Protection Board, 'Guidelines on Data Protection Impact Assessment (DPIA) and Determining Whether Processing Is "Likely to Result in a High Risk" for the Purposes of the GDPR' (n 572).

Protection Supervisor (EDPS)[584] or data protection authorities (such as Commission Nationale de l'Informatique et des Libertés (CNIL),[585] Agencia Española de Protección de Datos (AEPD),[586] Information Commissioner's Office (ICO),[587] Data Protection Commission (DPC)[588]), industry standards (e.g. ISO/IEC 29134:2017[589] or IAB Europe[590]). As seen from the list, controllers enjoy a wide margin of appreciation concerning the details of their practical implementation.

However, the GDPR mandates that certain information must always be included in the DPIA.[591] Firstly, it must include a systematic description of the processing operations and the purposes of the processing. Controllers must explain how they are going to process personal data and for which purposes. This must include relevant information concerning the processing, in particular about its nature (e.g. collection methods, sources of data, whether the data will be transferred to a third country), its scope (e.g. nature of data -whether it belongs to the special categories of data or not-, number of data subjects involved, geographical scope), its context (e.g. nature of the controller-individual relationship, data subjects expectations concerning the processing operations, whether processing includes data of children or other vulnerable groups) and its purposes (controllers' objectives, advantages of processing, effects on data subjects). In particular, where they rely on legitimate

---

[584] European Data Protection Supervisor, 'Accountability on the Ground Part I: Records, Registers and When to Do Data Protection Impact Assessments' (2019); European Data Protection Supervisor, 'Accountability on the Ground Part II: Data Protection Impact Assessments & Prior Consultation' (2019).

[585] Commission nationale de l'informatique et des libertés, 'Privacy Impact Assessment (PIA) 1: Methodology' (2018).

[586] Agencia Española de Protección de Datos, 'Guía Práctica Para Las Evaluaciones de Impacto En La Protección de Los Datos Sujetas Al RGPD' (2019).

[587] ICO, Data Protection Impact Assessments (DPIAs), available at <https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/data-protection-impact-assessments-dpias/> retrieved on 26/05/2021.

[588] Data Protection Commission, 'Guide to Data Protection Impact Assessments (DPIAs)' (2019).

[589] ISO/IEC 29134:2017 Information technology — Security techniques — Guidelines for privacy impact assessment, available at <https://www.iso.org/obp/ui/#iso:std:iso-iec:29134:ed-1:v1:en> retrieved on 26/05/2021.

[590] Interactive Advertising Bureau Europe, 'Guidance: GDPR Data Protection Impact Assessment (DPIA) for Digital Advertising under GDPR' (2020).

[591] Art. 35(7) DPIA.

interests to process personal data, they must concretely specify their legitimate purposes and balance them against the rights of data subjects.

However, describing systematically the processing operations and the purposes of processing may be challenging in AI systems. On some occasions, the discovery phase in the development of AI systems entails detecting unknown correlations between the data, which complicates the description of the data flows. Additionally, there are AI systems that at the early stages do not have clear purposes. In the latter, it is recommended that developers use only anonymised data and then, if useful correlations are found, narrow down and define the specific purposes for which the data will be processed.[592]

Secondly, they must evaluate the necessity and proportionality of the processing in light of the purposes. The evaluation needs to show that the use of the AI system is regarded as the fittest solution to achieve the objectives of the particular data processing operations. The controller should explain the reasons why a particular AI system was employed if they identified a less risky and privacy intrusive method to process personal data and the latter was discarded.[593] For instance, facial recognition technologies should not be used to monitor access and registration of students if the same purpose can be attained by a less intrusive method.[594] Additionally, the processing activities should be evaluated under the principles established in Art. 5 GDPR. That is to say, an appraisal of the processing operations in terms of its lawfulness, fairness, transparency, purpose limitation, data minimisation, accuracy, storage limitation, and security. For example, data collected for facial recognition

[592] Information Commissioner's Office, 'Big Data, Artificial Intelligence, Machine Learning and Data Protection Data Protection Act and General Data Protection Regulation' (n 202) 104.

[593] European Data Protection Suprevisor, 'Opinion 4/2020 on the European Commission's White Paper on Artificial Intelligence. A European Approach to Excellence and Trust' (2020) 15.

[594] Swedish Data Protection Authority, 'Supervision Pursuant to the General Data Protection Regulation (EU) 2016/679 – Facial Recognition Used to Monitor the Attendance of Students' (2019). See also CNIL, Expérimentation de la reconnaissance faciale dans deux lycées: la CNIL précise sa position (19/10/2019) <https://www.cnil.fr/fr/experimentation-de-la-reconnaissance-faciale-dans-deux-lycees-la-cnil-precise-sa-position?> accessed 25/05/2022.

purposes must be erased as soon as possible.[595] Furthermore, it must also include information about the lawful basis for processing, whether the processing attains the objectives and if there is another path to achieve the same results. It's been suggested that, under some circumstances, the evaluation should comprise also the entirety of the human rights limitation criteria outlined in Art. 52 CFR, i.e., legality, necessity and proportionality stricto sensu.[596]

Thirdly, controllers must appraise the risks to the rights of data subjects. In this context, risks should be considered as detrimental consequences that may emerge from the data processing operations. Controllers must evaluate both the likelihood of the risks, i.e. the probability that the situation takes place, and the severity of the risks, i.e. the significance of the consequences. A particular kind of risk that controllers should evaluate is the risk related to the fairness of the AI outcome which could be produced by errors in the performance of the AI solution. This idea is related not only to the principle of accuracy (reflected in Art. 5(1)(d) GDPR) concerning the lack of errors in the underlying data, but also to the concept of statistical accuracy which is linked to the output of the AI system, and the relationships between positive/negative results and false/true results.[597] Additionally, it should be borne in mind that the risks for data subjects are high where the processing activities concern vulnerable groups, the processing may reduce the job opportunities of gig workers,[598] or real-time remote biometric identification systems may reveal political opinions or trade union membership.[599]

---

[595] English and Wales High Court, *R (Bridges) v Chief Constable of South Wales Police and other*. [2019]. See also the decision from the Danish Data Protection Authority when granting permission to football club Brøndby IF to use facial recognition technologies. Jesper Lund, Danish DPA approves Automated Facial Recognition (European Digital Rights, 19/06/2019) <https://edri.org/our-work/danish-dpa-approves-automated-facial-recognition/> accessed 25/005/2022.

[596] Dariusz Kloza and others, 'Data Protection Impact Assessment in the European Union: Developing a Template for a Report from the Assessment Process' (2020) 29.

[597] This relationship is often reflected in a confusion matrix. According to this confusion matrix it can be measured: True Positives, True Negatives, False Positives, and False Negatives.

[598] *Ordinanza ingiunzione nei confronti di Foodinho SRL* (n 212); *Ordinanza ingiunzione nei confronti di Deliveroo Italy SRL* (n 271).

[599] *Garante per la Protezione dei Dati Personali, Parere sul sistema SARI Real Time* (n 406).

Then, controllers must recommend measures to address the risks and demonstrate compliance with the GDPR. Whereas many of these measures are discussed at length later when explaining the governance measures, suffice it to say here that some measures to mitigate the risks posed by AI systems are reducing the identifiability of personal data (through anonymisation, pseudonymisation or encryption), following codes of conduct and obtaining relevant certifications. Human oversight over the prediction or decision made by an AI can greatly reduce the risks posed by AI systems. As a caveat, the person responsible to oversee the measure should be capable of freely deciding whether to follow or not the recommendation or decision and even whether to use or not the AI system in a particular case. Additionally, a reasonable measure consists in registering any trade-off made during the AI model design or even during its deployment. This is, in particular, relevant when the trade-offs made affect some of the core data protection principles, for instance, opting for a more accurate model that may be difficult to explain (statistical accuracy vs explainability) or using a model that requires more data for better performance (statistical accuracy vs. data minimisation). The measures should be suggested for each data protection principle not satisfied.[600]

Finally, in some cases, the controller must consult individuals or their representatives concerning the planned processing operations.[601] The regulation does not impose a mandatory consultation. Instead, it requires this process 'where appropriate', meaning that where the controller considers that the input of affected or interested stakeholders could assist in the mitigation of the risks, a consultation should be carried out. However, while allowing controllers to decide when to involve affected stakeholders improves efficiency and celerity of the process, it leaves them a wide margin of interpretation and discretion concerning the situation which would trigger the consultation. And if even controllers decide to carry out the consultation, no guidelines are given with regards to the methodology and level of involvement of the stakeholders. This was an aspect heavily criticised by commentators[602] and a solution

[600] Dariusz Kloza and others (n 594) 30.

[601] Art. 35(9).

[602] Kaminski and Malgieri (n 533).

can be found through the implementation of impact assessments from other areas, such as human rights impact assessments. This alternative is discussed further below.

Chapter IV

OVERCOMING THE LIMITATIONS OF THE DATA PROTECTION REGULATION

Introduction

AI systems have revolutionised the world we live in. For good, they improved efficiency in the economy, allowed faster research processes, driverless cars, and many more. But they also modified our environment for the worst, where AI systems gathered incredible amounts of information about individuals, which allowed companies to create profiles with extreme precision, thus invading very intimate aspects of persons. Also, they profited from users' vulnerabilities, gained extremely valuable insights and profited from the individuals' data.

Many fundamental rights could be affected by automatic decisions and profiling assisted with AI. The respect for private life and protection of personal data, non-discrimination, rights of the child and elderly people, the right to good administration and the right to an effective remedy are among the most important fundamental rights that could be impaired through the use of AI systems.

Regardless of their merits and demerits, AI systems will be to an ever-increasing extent more relevant in our lives. Data protection and privacy regulations are not meant to stifle innovation. However, AI systems must be designed to protect fundamental rights. Hence, developers, deployers, data controllers and processors must ensure that AI systems are in line with fundamental values. This is not a question of banning the processing of personal data assisted by AI systems (except for some extremely harmful AI systems) but regulating them according to the intrinsic or potential risks of AI applications.

There are some initiatives to regulate AI systems, chiefly, the AI Regulation draft. It is not the scope of this work to make a thorough assessment of this draft piece of legislation but to evaluate how, with the current legislation and maybe with some new initiatives, AI systems will be safer for individuals, in particular from data protection and privacy perspective.

This section makes an in-depth assessment of some of the most acute problems concerning the processing of personal data using AI systems. In particular, the following problems are evaluated: first, the lack of transparency and the explainability issues associated with the processing of personal data using AI systems; second, the existence of biases in decision-making and the requirement of fair processing and non-discrimination. For every one of these issues, some proposals are made.

IV.1.- Addressing algorithmic transparency when processing personal data using AI systems

IV.1.1.- Outlining the problem of algorithmic transparency

Transparency is central to creating social trust in the use of AI systems. Individuals interacting with the AI systems should receive information about the methods of operation, the data used to make the predictions and the decisions themselves. Transparent and explainable AI solutions constitute 'essential preconditions' to guarantee fundamental rights,[603] including the right to privacy and data protection. This entails making data, features, models, algorithms, training methods and quality assurance processes available for external inspection. Transparent AI solutions allow individuals to know 'where, why and what data are collected',[604] in particular, if they collect and process personal data. In the absence of this information, a particular outcome produced by an AI system cannot be appropriately challenged.[605]

The lack of transparency of AI systems can be attributed to an AI system technically non-explainable,[606] the opaqueness of the sources of data to train the model, and organisational policies that refrain from disclosing information about the AI systems.[607]

---

[603] UNESCO (n 440) 37.

[604] International Standard Organisation, 'ISO/IEC TR 24028:2020 Information Technology — Artificial Intelligence — Overview of Trustworthiness in Artificial Intelligence' (2020) Subclause 9.2.

[605] European Commission's High-Level Expert Group on Artificial Intelligence (n 415) 13.

[606] See above the distinction between generally explainable models (e.g. linear and logistic regression, decision trees) and generally non-explainable models (e.g. support vector machines and artificial neural networks).

[607] International Standard Organisation, 'ISO/IEC TR 24028:2020 Information Technology — Artificial Intelligence — Overview of Trustworthiness in Artificial Intelligence' (n 602) s 8.6.

In the context of data protection, there is a stark problem. Whereas Arts. 12-14 GDPR establish the information and conditions under which controllers must inform data subjects of the processing operations, if a controller were to comply with these obligations, it will have little impact on data subjects. This is because, in general, data subjects very rarely read privacy notices.[608] Only about 1 to 10% of the individuals read them, and those who read them spend less than 2 minutes in this activity.[609] This means that long and detailed privacy policies are largely ignored. And there is no blame on data subjects since even if they would read the privacy notices, most of these documents are just simply too difficult to read, and often their readability score is equivalent to academic publications.[610] Hence, apart from requiring controllers to explain in a concise, transparent and intelligible format how they process data subjects' personal data, they should be encouraged to provide the information in a user-friendly manner.

By algorithmic transparency this section attempts to reframe the data controllers' obligations regarding the information they must provide to users. Articles 12-15 GDPR lay down in a detailed fashion the content, timing, and manner of the information that data controllers must facilitate to users. However, information rights under the GDPR are not effective enough for end-users of AI systems to understand how the systems work and the implications of some decisions taken through automated means.

In particular, the GDPR fails to address the problems related to the different information to be provided to different interested parties. The audience is an important contextual factor when evaluating the kind and wealth of information to provide to individuals. Not only do different stakeholders need different kinds of information (for instance, regulators do not need the same information as individuals), but also within

---

[608] Jonathan A Obar and Anne Oeldorf-Hirsch, 'The Biggest Lie on the Internet: Ignoring the Privacy Policies and Terms of Service Policies of Social Networking Services' (2020) 23 Information, Communication & Society 128, 130.

[609] Georgina Kon, 'Does Anyone Read Privacy Notices? The Facts' (*Linklaters DigiLink*, 2020) <https://www.linklaters.com/en/insights/blogs/digilinks/does-anyone-read-privacy-notices-the-facts> accessed 27 December 2021; Pew Research Center, '4. Americans' Attitudes and Experiences with Privacy Policies and Laws' (*Americans and Privacy: Concerned, Confused, and Feeling lack of Control over their Personal Information*, 2019) <https://www.pewresearch.org/internet/2019/11/15/americans-attitudes-and-experiences-with-privacy-policies-and-laws/> accessed 27 December 2021.

[610] Benolie and Becher (n 446) 2294.

groups of individuals, their interests and level of prior knowledge vary (experts and non-experts), which requires controllers to provide information according to the intended group.

In the following sections, an attempt is made to address both the content of the transparency obligations and also the delivery methods adequate to satisfy the requirements of conciseness, accessibility, understandability and clarity of the information provided to individuals when interacting with certain AI systems.

IV.1.2.- Improving algorithmic transparency. Information to be provided before the AI-powered decision is made

There is no easy way to accomplish transparency in the data-driven world. For one, there is no generally agreed definition of what transparency means or which features a transparent AI system should possess. Secondly, while the conflict between clarity and conciseness of the information, on the one hand, and comprehensiveness or completeness of it, on the other, is not new, this mismatch is even more relevant when providing information related to AI systems. Thirdly, the interpretability or explainability of an outcome will be contingent to a great extent on the user to whom it is addressed.[611]

This section elaborates on the information to be provided before taking the decisions using AI systems. This kind of information relates to the datasets, the general functioning of the algorithms and the model itself. Not only does it highlight the content of the relevant information to be provided, but also it suggests innovative forms of delivering the information. Firstly, it provides an overview of the information that should be provided. Secondly, it argues that the relevant information must be provided in a written, graphic or animated fashion.

IV.1.2.1.- On the content of the information that should be provided to individuals

It has been discussed at length the information currently required by the GDPR, which is stated in Arts. 13 to 15 and 22 GDPR, along with the limitations and some

[611] National Institute of Standards and Technology, 'Psychological Foundations of Explainability and Interpretability in Artificial Intelligence' (2021) 2.

proposed interpretations of these legal provisions. The provisions of the AIA draft were also evaluated and their shortcomings were addressed.

While there is no clear answer on which information should be included both before and after the decision is made, the Information Commissioner's Office (ICO) guidelines 'Explaining Decisions Made with AI' are an excellent starting point to consider. The guidelines feature a framework to understand which information should be provided to individuals and it takes into account three aspects. First, whether the information is given before or after the decision is made. The former are explanations to understand the process through which the decision is made whereas the latter are directed to the decision itself. Second, the kind of explanation that should be provided to the individuals. This core aspect relates to the six most important types of explanations. Third, the contextual factors that should be considered to modulate the type of explanations to be given to individuals.

Concerning the nature of the explanation the ICO lists six different kinds of explanations[612] that may be provided: a) *rationale explanation*: which entails providing the reasoning or logic behind the decision; b) *responsibility explanation*: providing information about the persons involved in the design, development and use of the AI solution, and the persons that individuals can require assistance when needed; c) *data explanation*: this is information about the datasets employed for the development of the AI solution and the information used for the concrete decision that affects the individual; d) *fairness explanation*: entails showing individuals how it is guaranteed that the decisions taken by the system are fair and there is no inequitable treatment among groups or particular individuals. For this purpose, providers of AI systems may rely on different fairness metrics, such as statistical parity difference, equal opportunity difference, average odds difference or disparate impact; e) *safety and performance explanation:* providers and users of AI systems should give information about the safeguards put in place to ensure the accuracy, reliability, security and robustness of the AI system and its metrics. It may also show the level of statistical confidence resulting from the outcomes of the algorithm. The higher the statistical confidence, the more reliable the algorithmic outcomes would be.[613] It could be suggested that below

[612] Information Commissioner's Office, 'Explaining Decisions Made with AI' (n 104) 20.

[613] The industry standard for statistical confidence is 95%.

a certain threshold of statistical confidence algorithms should not be allowed or at least should be flagged as inadequate for their purposes.[614] It could also include information about whether the AI system outperforms a human being for an identical task;[615] f) *impact explanation:* controllers should inform end-users and society how the AI system may impact them and the safeguards put in place to mitigate these negative effects.

Since completeness of the information should be compatibilized with conciseness, clarity, intelligibility, and usefulness to the end-user,[616] the ICO suggests that the six explanation types should be modulated according to the *contextual factors* in which the decision is being given. Hence, AI operators should contextualise the information according to the domain, the impact, the data, the urgency, and the audience. First, organisations should consider the *domain* in which the AI system is deployed. The domain is the sector or area where the system is used. Different specifications or requirements may have a bearing on highly regulated sectors like healthcare, banking or insurance. For instance, in non-critical domains like spam filtering or ad targeting, a basic rationale and responsibility explanation would suffice. However, as the stakes are higher more explanations will be needed. For example, where the decisions may generate doubts regarding their fairness, a fairness explanation should be provided. Then, in AI systems where safety is a primary concern, like autonomous driving, AI operators should offer safety and performance explanations.[617] Second, the outcome of the AI solution will definitively have different *impacts* on data subjects and society. How the decision affects them should be factored in when evaluating which type of explanations to provide. While decisions having a low impact on individuals or society at large may not cause any issues, where the decisions may have a high impact on them, fairness, safety and performance, impact and responsibility explanations should

---

[614] Andrew Tutt, 'An FDA For Algorithms' (2017) 69 Administrative Law Review 84, 108. When it comes to assessing the performance of AI systems employed to screen job applicants the author suggests that, for instance, the algorithm should not have a dismissal rate any protected class (race, sex, ethnic group) of more than 20%, and the confidence of the results could be not below 95%.

[615] Paul Ohm, 'Chapter 12: Throttling Machine Learning' in Mireille Hildebrandt and Kieron O'Hara (eds), *Life and the Law in the Era of Data-Driven Agency* (Elgar 2020) 218. This is called Machine-to-human performance ratio (MHPR).

[616] Art. 12 GDPR.

[617] Information Commissioner's Office, 'Explaining Decisions Made with AI' (n 104) 34.

be prioritised.[618] Third, the *data* employed in the development of the AI system (in the training, validation and testing of the AI model) and in the use of the AI system to produce the particular decision, prediction or outcome is a contextual factor to evaluate. Where the AI system uses predominantly social data (identification, user interaction on the web, metadata, location, language, etc), the individuals should receive rationale, fairness and data explanations. On the other hand, if the AI systems use biophysical data (including, for instance, biometric, genetic and health data), explanations concerning the rationale, impact and safety and performance should be prioritised.[619] Fourth, how quickly the decision should be delivered is another factor to be weighted. If the decision should be provided *urgently*, the AI operator should prioritise impact and safety and performance explanations.[620] Fifth and lastly, the *audience* that will receive the information should be specially taken into consideration. This is because both the depth and the kind of explanation will also depend on the background knowledge of the persons that receive the information. As a rule of thumb, the explanations should be adapted to the requirement of the most vulnerable groups.[621] Where the decisions are addressed to persons without any particular background knowledge in the field, responsibility, rationale and safety and performance explanations should be prioritised. However, if the explanation is directed to expert persons, rationale and safety and performance explanations may be better suited.[622]

Table III factors to consider when delivering an explanation to individuals

Timing	Kind of explanations	Contextual factors
• before the decision is made • after the decision is made	• rationale explanation • responsibility explanation • data explanation • fairness explanation • safety and performance explanation	• domain where the decision is taken • impact of the decision • data processed • urgency to deliver the explanation • audience of the explanation

[618] ibid.

[619] ibid 35.

[620] ibid 36.

[621] For instance, Art. 12 GDPR expressly mentions the importance of accommodating both the kind of information and the way to deliver it to the needs of children where it is addressed to them.

[622] Information Commissioner's Office, 'Explaining Decisions Made with AI' (n 104) 37.

	• impact explanation	

The framework designed by the ICO constitutes a very useful tool to assist developers and deployers of AI systems to deliver the required information to data subjects, increasing accountability and building trust among end-users of the systems and society. Yet, this proposal is only a high-level framework to help those required to increase the transparency of their algorithms. More concrete applications and examples of the specific information that should be provided are evaluated below. These alternatives relate to the potential use of the AIA draft and standardisation.

IV.1.2.2.- AI Regulation to promote transparency of AI systems

In this section, the main provisions of the AIA draft related to transparency are evaluated, in addition to an appraisal of how these provisions may fit into the framework developed by the ICO to deliver explanations concerning AI systems.

Does the AI Regulation draft provide meaningful information to data subjects?

The AI Regulation draft establishes some transparency obligations for AI systems. It sets up three different levels of information. First, the AIA draft stipulates obligations for low-risk AI systems,[623] as these systems can create certain threats of impersonation or deception. Article 52 AIA draft requires that providers or users or low-risk AI systems (i.e., interactive systems, emotion recognition systems, biometric categorization systems and systems that produce deep fake content) communicate to individuals that they interacting with or exposed to an AI system or that the content was artificially created, unless it is clear from the circumstances and context or the systems are authorised by law for law enforcement purposes. Second, for high-risk AI systems (hereinafter HRIAS)[624] it requires that providers of AI systems design the AI

[623] Low-risk AI systems are AI systems that either: a) interact with individuals; b) are employed to perceive human emotions or perform biometric categorisation; or c) create or manipulate content (‘deep fakes’) (Art. 52 AIA draft).

[624] These obligations are also applicable to low-risk AI systems as defined in Art. 52 AIA, see Art. 52(4) AIA draft.

systems in a way they result easily understandable for AI users.[625] In particular, AI systems must contain information about the provider, the AI system capabilities and limitations (including their intended purpose, the level of accuracy, robustness and security, and the factors that may have an impact on these features), measures to ensure human oversight, and their expected lifetime. Third, providers of certain HRIAS must register the AI systems in a dedicated EU database before placing them on the market or putting them into service,[626] and the information contained therein will be open to the public.[627] In this database, providers must disclose information about themselves and their representative in the EU (where applicable), AI system trade name or identification, intended purpose, the status of the AI system, a copy of the certificate issued by the notified body (if applicable), information about EU Member States where the AI system is being deployed, the declaration of conformity as required in Art. 48 AIA draft, and instructions for use.[628]

While the AIA draft imposes heavy obligations to AI providers and AI users, it also falls short of providing full information to data subjects. The most important transparency obligations concern business-to-business relationships (i.e., AI providers to AI users), but not business-to-consumer relationships (i.e., AI providers or AI users towards end-users or consumers). The objective of the obligations established in Art. 13 AIA draft is to reduce the opaqueness of AI systems, helping AI users to interpret the results or predictions of the AI systems and to deploy them adequately and safely,[629] and it should not be regarded as the AIA equivalent of the provisions concerning transparency established in Arts. 12-15 GDPR.

Nevertheless, apart from minimum transparency obligations concerning certain low-risk AI systems, the AI Regulation draft lacks any comprehensive or detailed catalogue of obligations directed to AI system operators (chiefly, providers and users) to provide useful, concise and easily understandable information to individuals (end-users of AI

---

[625] Art. 3(4) AIA draft defines a 'user' of an AI system as any person that employs the AI system 'under its authority', except for those cases in which the system is employed 'in the course of a personal non-professional activity'.

[626] Art. 51 AIA draft.

[627] Art. 60(3) AIA draft.

[628] Art. 60(2) and Annex VIII AIA draft.

[629] See Recital 47 AIA draft.

systems). And even concerning these low-risk AI systems, the AIA draft exempts public authorities to disclose the use of AI systems for the detection, prevention, investigation or prosecution of crimes.[630] This carve-out was seen as too wide, and both the EDPB and the EDPS considered that some minimal safeguards should be put in place when it comes to using these systems to prevent and detect criminal offences since the presumption of innocence is at stake.[631]

Finally, it is worth noticing that while the AIA draft requires the registration of certain AI systems in a publicly accessible registry, the information open to the public is limited. This register will provide relevant information about, for instance, its intended purpose, a copy of the certificate issued by the notified body (if applicable), the Art. 48 AIA draft declaration of conformity, and instructions for use. However, it does not include information about, for instance, fairness metrics, sources of data, or other information that may be relevant for data subjects. The difference between the information that providers of AI systems must disclose to users of AI systems in the context of business-to-business relationships (listed in AIA Annex IV) and the information available to end-users of AI systems is noticeable. For instance, AI providers must include in the technical documentation for AI users:

> *"(2) (a) the methods and steps performed for the development of the AI system…*
> *(b) … the general logic of the AI system and of the algorithms; the key design choices including the rationale and assumptions made, also with regard to persons or groups of persons on which the system is intended to be used; the main classification choices; what the system is designed to optimise for and the relevance of the different parameters; the decisions about any possible trade-off made regarding the technical solutions adopted…*
> *(d) ... techniques and the training data sets used, including information about the provenance of those data sets, their scope and main characteristics; how the data was obtained and selected …*
> *(e) assessment of the human oversight measures …*

[630] Art. 52 AIA draft.

[631] European Data Protection Board and European Data Protection Supervisor (n 199) 19.

(g) … metrics used to measure accuracy, robustness, cybersecurity and … potentially discriminatory impacts; …

(3) Detailed information about … capabilities and limitations in performance, including the degrees of accuracy for specific persons or groups of persons on which the system is intended to be used and the overall expected level of accuracy in relation to its intended purpose; the foreseeable unintended outcomes and sources of risks to health and safety, fundamental rights and discrimination in view of the intended purpose of the AI system; the human oversight measures needed in accordance with Article 14, including the technical measures put in place to facilitate the interpretation of the outputs of AI systems by the users"

As illustrated by this excerpt from Annex IV AIA draft, the information is comprehensive and includes many aspects that may be of interest to data subjects. According to the current legislative framework, data controllers must inform data subjects about the existence of automated decision-making and provide meaningful information about the logic involved, and the significance and the envisaged consequences of the processing for them.[632] But this only applies if the AI system falls under the camp of Art. 22 GDPR, and even if it does, there is still uncertainty on what is the logic involved and the envisaged consequences for individuals. If the AI system is not covered by Art. 22 GDPR, the only information data subjects are entitled to receive is the fact that their personal information is being used to develop or use the AI system (i.e. train the model or make the prediction), the legal basis for processing, along with the information listed in Arts. 13-14 GDPR.

While the AIA includes some provisions on transparency, as mentioned before, it is mostly limited to the relationship between AI providers to AI users (business-to-business relationship) and it contains only a few provisions for individuals affected by the AI systems (e.g. the right to know they are interacting with low-risk AI systems in art. 52 AIA and the right to get access to the registry of AI systems in art. 60 AIA).

[632] Art. 13(2)(f), 14(2)(g) and 15(1)(h) GDPR.

How do the AIA draft information obligations align with ICO's categorization?

Previously it was considered that the ICO's guidelines 'Explaining Decisions Made with AI' constitute a solid starting point to evaluate how explanations to stakeholders should be delivered. Additionally, the provisions of the AIA draft were evaluated and it was concluded that the obligations established therein are insufficient to satisfy the transparency requirements for some stakeholders, chiefly end-users of the AI systems. However, it is possible to conduct a further assessment to see which particular areas should be improved.

The AIA draft requires the production of some information that may be aligned with the categories included in the ICO's guidelines and previously described. For instance, it requires to provide information concerning the following types of explanation: a) rationale explanation: the general logic of the AI system and of the algorithm, and other design choices including the rationale and assumption made, what the system is intended to optimise, the relevance of the different parameters, etc;[633] b) responsibility explanation: the AIA draft requires AI providers to give AI users information about the provider itself[634] and human oversight measures;[635] c) data explanation: AI providers must provide users of AI systems specifications of the training, validation and testing datasets, along with input data[636] and a description of the methodologies and techniques for training, information about the provenance of datasets, labelling procedures, data cleaning;[637] d) fairness explanation: providers of AI systems are expected to inform users of AI systems the metrics used to measure potentially discriminatory impacts[638] and the degrees of accuracy for different groups or persons;[639] e) safety and performance explanation: among the many provisions touching upon this kind of explanation, it is worth mentioning that providers of AI systems must inform users of AI systems about level of accuracy, robustness and

[633] Clause 2(b) Annex VI AIA draft.

[634] Art. 13(3)(a) AIA draft.

[635] Art. 13(3)(d) AIA draft.

[636] Art. 13(3)(b)(v) AIA draft.

[637] Clause 2(d) Annex IV AIA draft.

[638] Clause 2(g) Annex IV AIA draft.

[639] Clause 3 Annex IV AIA draft.

cybersecurity and any foreseeable circumstance that may affect them,[640] metrics used to measure accuracy, robustness and cybersecurity,[641] and the capabilities and limitations in performance;[642] f) impact explanation: AI providers must inform AI users about estimated risks that may emerge during the use of the high-risk AI systems,[643] along with the foreseeable unintended outcomes and sources of risks to individuals or society, such as to health and safety, fundamental rights and discrimination.[644]

Yet, there are some problems when trying to square the AIA draft to the ICO's explanation model for AI systems. First, the AIA draft, in general, does not consider the contextual factors that may require modulation of the explanations to provide according to the domain, the impact, the data, the urgency, and, crucially, the target audience of the explanation. It only requires that some particular AI systems should follow specific requirements as required in the applicable regulatory frameworks. Second, the duty to inform, as previously mentioned, only applies to high-risk AI systems. AI providers of AI systems not falling under this category are not compelled to produce the abovementioned information. Finally, the AIA draft requires AI providers to give AI users certain information. However, this information is not accessible to end-users or the general public.

IV.1.2.3.- The role of standards. Standardisation to solve the gaps in the legislation

While the role of standardisation to address the main problems related to AI systems will be discussed at length later (see section V.3), suffice it to introduce in this section a newly published standard that could fill the gaps left by the mandatory legislation. Institute of Electrical and Electronics Engineers (IEEE) 7001-2021 Standard for Transparency of Autonomous systems sets forth requirements to measure and evaluate the level of transparency of AI systems, taking into consideration both the intrinsic features of the AI solution and the kind of stakeholder

[640] Art. 13(3)(b)(ii) AIA draft.

[641] Clause 2(g) Annex IV AIA draft.

[642] Clause 3 Annex IV AIA draft.

[643] Art. 9(2)(b) and 9(4)(c) AIA draft.

[644] Clause 3 Annex IV AIA draft.

that demands information from the AI system.[645] It sets a progressive transparency level for AI systems, whereby AI systems containing no indication or explanation whatsoever are non-transparent (level 0) and those that fulfil every single requirement reach the highest possible level of transparency (level 5). At a higher transparency level, the standard compels developers not only to provide information to render the AI system more transparent but also to be explainable in itself, meaning that the information provided to the relevant stakeholders should be readily interpretable and accessible.[646] Interestingly, the standard establishes five different stakeholder groups, but the most important for the purposes of this work are those who are directly benefited from enhanced transparency, i.e., direct or end-users of the AI systems and the general public.[647] As the standard defines different requirements for every stakeholder group, transparency levels must be achieved for every single stakeholder group.

IEEE 7001 provides some examples of practical use cases to understand how it works. To evaluate concretely the information that should be provided in a certain situation, an assessment should be made concerning the criticality of the system and the information needed for the relevant stakeholders. Intuitive enough, direct or end-users of the AI systems need more information than the general public or bystanders as the former may have been more severely affected by the outcome of the AI system. The individuals belonging to the general public or bystanders do not directly enter into a relationship with the AI system. However, they should receive some information to make educated choices about whether or not they want to interact with the AI system as end-users of the system. An example illustrating the different transparency requirements for these two categories of stakeholders could help understand how it works. If it were considered an AI system for credit scoring, loan applicants would be considered 'users' (non-experts) and other clients of the bank or potential clients would be considered as 'general public'. Since their expectations and needs differ, the

[645] Institute of Electrical and Electronics Engineers, 'IEEE 7001-2021 Standard for Transparency of Autonomous Systems' (2022) 16.

[646] ibid.

[647] The other three categories of stakeholders are grouped under the label 'Expert stakeholders that work with the information provided as part of the transparency obligations' and includes certification or regulatory bodies, indecent investigators and expert advisors in administrative courts or litigation.

transparency requirements for them will differ as well. According to IEEE 7001 loan applicants should be provided with the information stipulated for level 3, whereas the general public should be provided with the information on level 1.

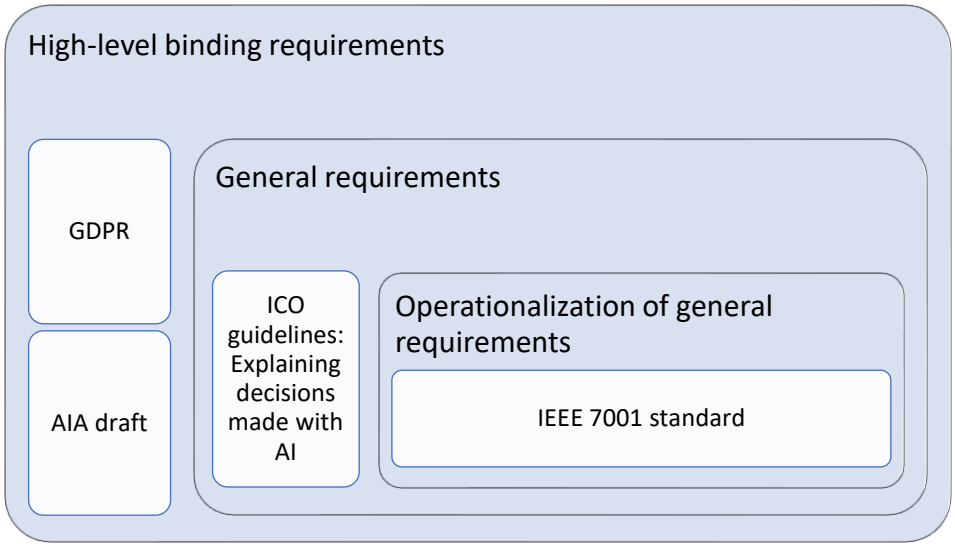
Credit scoring system (information required)				
Trans-prency levels	Stakeholders			
	Direct users (non-experts)		General public	
	E.g Loan applicant	Required?	E.g. Bank clients and potential applicants	Required?
Level 1	Illustrative scenarios with foreseen system behaviour, explanation of the system general principles of operation, data sources and potential biases affecting the functioning	yes	Clearly identify the system as an AI system (with messages or icons for instance)	yes
Level 2	Interactive training material for the user to rehearse its interactions with the system	yes	Warnings about any external data collected or otherwise recorded (geolocation data). Documentation or information explaining what form of sensor data are collected and how they are used, which should be publicly available.	no
Level 3	Brief and immediate explanation of the system's most recent activity. The system should answer to the question: *why did you do that?*	yes	High-level description of the system's intended purpose, contact details of the operator and responsible person	no
Level 4	Brief and immediate explanation of what the AI system does in a given situation, allowing to explore hypothetical 'what if' situations	no	Clear data-governance policy and operators shall accept and response data-governance related requests	no
Level 5	Continuous explanation of the behaviour that adapts the content and presentation of the explanation based on the user's information needs and context	no		

Source: adapted from IEEE 7001 standard for Transparency of Autonomous Systems, example B.4 page 38.

Relationship between legislation, guidelines and standards

Previously the transparency obligations contained in the GDPR and the AIA draft were evaluated. While these instruments have (or will eventually have, in the case of the AIA draft) binding force and establish requirements to safeguard public interests, regulations generally contain high-level goals along with some particular obligations for regulated actors to comply. However, it is unusual that regulations include specific provisions concerning how the goals established therein must be satisfied. To fill this gap both guidelines and standards can be extremely helpful. They specify the general

requirements established in the mandatory legislation. At the lowest level, standards like IEEE 7001 will concretise high-level requirements stipulated in the GDPR and, possibly, the AIA with specific examples and lay down the concrete information requirements that AI operators must satisfy to comply with their obligations.



Relationship between legislation, guidelines and standards to comply with transparency obligations

IV.1.2.4.- On the methods to deliver the information. Written, graphic or animated information about the inner working. Datasheets, Model cards, Factsheets

The content of the information is not the only important aspect to consider. As important as the content, the methods to convey the information should be redesigned, and innovative, proactive and adequate methods for providing information to end-users should be incorporated in the new AIA draft.[648] Where the processing of information relates to identified or identifiable persons, controllers should provide the information in an 'intelligible' and 'easily accessible form',[649] considering also relevant visualisation methods.[650]

[648] European Commission - Joint Research Centre, 'Artificial Intelligence - A European Perspective' (n 263) 19; European Data Protection Board and European Data Protection Supervisor (n 199) 19.

[649] Art. 7 GDPR regarding consent and Art. 12 GDPR regarding transparency.

[650] Recital 58 GDPR.

Providing relevant, clear and easy-to-understand information to individuals concerning AI systems is challenging because developers should find a balance between disclosing an amount of information that does not cause information fatigue to end-users, while at the same time translating complex engineering concepts into a common language that even a non-expert audience could understand. As a default position, information is provided in policies with varying degrees of length, which may not be completely suitable for individuals. As previously mentioned, in general users do not read privacy notices, and this lack of interest could be even more noticeable if the information they are trying to incorporate relates to complex algorithmic operations. Hence the need to find innovative methods to convey information. For instance, employing rating symbols, icons or marks in AI systems may be a better way to counteract the opaqueness of AI systems, in particular when they are used by certain vulnerable subjects, like children.[651] In 2020, Apple introduced new privacy details on the App Store, where app developers may voluntarily submit to the App Store the app's personal data handling, allowing users to check the app's data protection practices. Icons and short titles are used to increase the understandability of such practices.[652] In 2021, the Italian data protection authority opened a call to participate in the design of graphic proposals (icons) to render the information included in privacy notices simpler, clearer and easily understandable,[653] as allowed by Art. 12(7) GDPR. Then, the Swiss Digital Initiative is working on a Digital Trust Label which attempts to convey in simple, graphic and basic language the most important information to build trust in digital products.[654] These initiatives boost trust in consumers of digital products and empower them to make informed decisions, which is crucial when they are interacting with AI-powered systems.

There have been some initiatives to introduce innovative ways to convey huge amounts of complex information to individuals in the field of AI systems. In general,

---

[651] Clause 9.2 ISO 24028:2020.

[652] Apple, 'App privacy details on the App Store' (2020) <https://developer.apple.com/app-store/app-privacy-details/> accessed 13/05/2022.

[653] Garante per la Protezione dei Dati Personali, 'Informative Chiare. I vincitori del contest lanciato dal Garante Privacy.' (2021) <https://www.garanteprivacy.it/temi/informativechiare> accessed 13/05/2022

[654] Swiss Digital Initiative, 'The Digital Trust Label' (2021) <https://www.swiss-digital-initiative.org/digital-trust-label/> accessed 13/05/2022.

some initiatives are addressed to increase the transparency of the datasets with which the models are developed, whereas others concern with the documents to explain the models themselves.

A) For datasets used to build AI models

Datasheets for datasets[655] is one of the most well-known initiatives to increase transparency concerning the datasets used to develop an AI system. After the World Economic Forum's call to keep records of the provenance, generation and use of datasets in AI systems[656] a team of researchers led by Tinmit Gebru[657] proposed creating a datasheet for datasets. Taking as a model the electronic industry, where it is required a datasheet containing information about the product's operating features, evaluation results, and suggested use, among others, they proposed that a datasheet should be attached to every dataset employed to build machine learning models. The purpose of the datasheet is to make public and standardize information about datasets. The datasheet is primarily addressed to two different stakeholders. First, to the developers of the AI solution, since it compels them to prudently reflect on all the technical and ethical considerations of using a particular dataset. Second, to end-users, as the datasheet increases the transparency towards them. Datasheet for datasets includes questions concerning the motivation to create the dataset, the dataset composition, the collection of data to build the dataset, the pre-processing, the cleaning or labelling of the dataset, the foreseeable uses of the dataset, the distribution of the dataset, and the maintenance of the dataset.[658]

Another initiative is the Dataset Nutrition Label. The Data Nutrition project is a cross-industry collective whose objective is to develop a standardised label for evaluating

---

[655] Tinmit Gebru and others, 'Datasheets for Datasets' (2021) 64 Communications of the ACM 86.

[656] World Economic Forum Global Future Council on Human and Rights 2016-18, 'How to Prevent Discriminatory Outcomes in Machine Learning' (2018).

[657] Tinmit Gebru is an AI ethicist that was fired from Google after her refusal to withdraw a research paper explaining the hazards of large language AI models. See Cade Metz and Daisuke Wakabayashi, Google Researcher Says She Was Fired Over Paper Highlighting Bias in A.I., New Your Times, <https://www.nytimes.com/2020/12/03/technology/google-researcher-timnit-gebru.html> accessed on 10/02/2022.

[658] Gebru and others (n 653).

datasets. They proposed the Data Nutrition Label and, emulating the Nutrition Facts Label on alimentary products, this label summarises the main details of the datasets used to build the model. The idea is to provide transparency and, at the same time, reduce the risks posed by automated systems by giving concise and understandable information on the quality of the datasets.[659] Finally, a further proposal was made by Google in its Data Cards Playbook. Google's Data Cards are intended to allow designers and developers to keep a register of their datasets in an organised manner and facilitate the decision-making process on how to use the datasets.[660]

B) For AI models themselves

Providing information about the datasets is only one partial solution to address transparency concerns. It is also important to disclose, in a concise, understandable and easily accessible manner, information about the model or the AI system itself. Several initiatives have been developed for this purpose, such as Model Cards for Model Reporting and AI FactSheets.

Model Cards for Model Reporting[661] is one of the most well-known initiatives to disclose information about the AI model and it is a complementary method to the previously described about datasets. Prepared by a Google team, these model cards aim to constitute a referential standpoint for everyone, irrespective of their expertise in the field. In addition, they seek to standardise how developers communicate the most important characteristics of their models. Those features include providing basic details about the model (e.g. model type and responsible persons), intended use cases (including reasonable foreseeable misuses), factors that may influence in the performance of the model, metrics to assess the performance and the impact of the model (including decision thresholds), information about evaluation and (where possible) training data, ethical considerations, challenges and choices, and, finally, warnings and recommendations. The team foresees that where this one or a similar

---

[659] The Data Nutrition Project, 'The Dataset Nutrition Label' <https://datanutrition.org/labels/> retrieved on 19/01/2022.

[660] Google Research, 'Data Cards Playbook', <https://pair-code.github.io/datacardsplaybook/playbook> accessed 23/01/2022.

[661] Margaret Mitchell and others, 'Model Cards for Model Reporting', *FAT* '19: Proceedings of the Conference on Fairness, Accountability, and Transparency* (2019).

way to report the main features of the models turns into a regular practice, stakeholders may make comparisons among different AI systems with relevant and high-quality information.[662] This work was expanded and led to the proposal of Google's Model Cards.[663]

On the other hand, a team of researchers from IBM proposed using factsheets for AI models or services to increase AI transparency and governance (AI FactSheets 360).[664] Factsheets can be seen as a compilation of important information on the design, development and use of the AI system. The information contained in the AI factsheet will depend on the particular AI system, but as a general rule, clarifications concerning the purposes and envisaged application of the system, information about the model's basic performance (and how it was tested), safety, explainability, fairness, as well as about the lineage of the data and the model, should be included.[665] Interestingly, the AI FactSheets templates were developed following the supplier's declaration of conformity (SDoC), which is a common practice in certain highly regulated sectors to demonstrate that a product or service complies with a standard or technical specification. As in the supplier's declaration of conformity, the AI providers may self-report crucial information about the AI system. They propose that the factsheets should be voluntarily adopted, and they foresee that if this practice is widely adopted the AI Factsheets may turn into a default requirement of AI systems.[666] Another important point to mention concerning the AI Factsheets is that the AIA draft also requires AI providers to draw up a declaration of conformity for each high-risk AI system, assuming responsibility for compliance with the AIA requirements.[667] So by completing these voluntary AI factsheets, providers of AI systems are training and testing their processes for an eventual mandatory requirement imposed by the AIA if

---

[662] ibid 223.

[663] Google Cloud, 'Google Model Cards', <https://modelcards.withgoogle.com/about> accessed on 23/01/2022.

[664] IBM Research, 'Introduction to AI FactSheets' <https://aifs360.mybluemix.net/introduction> accessed on 08/04/2022.

[665] Matthew Arnold and others, 'FactSheets: Increasing Trust in AI Services through Supplier's Declarations of Conformity' (2019) 63 IBM Journal of Research and Development 1, 11.

[666] ibid 8.

[667] Art. 48 AIA draft.

enacted. Other examples of similar initiatives are the AI Ethics Label[668] and Facebook's System Cards,[669] which were used to explain the Instagram Feed Ranking.[670]

An evaluation

Using these model cards, factsheets and datasheets can be a method to improve information asymmetries among stakeholders. They summarise the most important information in short documents open to the relevant stakeholders, resembling the function of nutritional labels or energy efficiency labels. They are visually appealing and easy to understand, even for a non-expert user. Some potential drawbacks should also be considered. While there are many initiatives under development, there is no single, unified, way to disclose the information. This allows AI providers to present the information in a flexible manner, considering the nuances and particularities of their own systems, but it also creates uncertainty both among operators of the AI supply chain and end-users of AI systems. Wider development and adoption of transparency standards, like the IEEE 7001-2021 Standard for Transparency of Autonomous Systems,[671] could greatly contribute to the adoption of a harmonised set of information that should be provided to the relevant stakeholders. Additionally, most of them, if not all, rely on self-reporting, i.e., the information is collected, evaluated and disclosed by the AI provider. And while self-reporting is an agile procedure for this purpose, it requires trust in the AI provider that elaborates the factsheet, model card or datasheet, since no third party to assess the authenticity or accuracy of the information AI providers include in them. A partial solution to the issue of self-reporting is to rely on a certification mechanism, where an independent third party verifies the fulfilment of

---

[668] VDE & Bertelsmann Stiftung, 'From Principles to Practice. An Interdisciplinary Framework to Operationalise AI Ethics' (2020).

[669] Meta AI, 'System Cards, a new resource for understanding how AI systems work' (23 Feb 2021) <https://ai.facebook.com/blog/system-cards-a-new-resource-for-understanding-how-ai-systems-work/> accessed 08/04/2022.

[670] Meta AI, 'Instagram Feed Ranking System Card' (23 Feb 2022) <https://ai.facebook.com/tools/system-cards/instagram-feed-ranking/> accessed 08/04/2022.

[671] Institute of Electrical and Electronics Engineers (n 643).

the requirements. An evaluation of certification mechanisms as a method to improve the accountability of AI system operators is provided below.

IV.1.3.- Information to be provided after the automated decision or profiling

Whereas many of the transparency principles evaluated concerning the information to provide to stakeholders before the AI-assisted decisions are made, individuals would be interested in a different set of information once the decision is rendered. Following the classification previously given,[672] concrete information that AI providers release may concern primarily with rationale, data and fairness explanations.

The rationale explanation should include information that falls within the meaning of 'meaningful information about the logic involved' from Arts. 13-15 GDPR. That is, in the rationale for a particular decision, controllers should explain what features were considered to make the decision and their relative importance. Controllers should be able to explain technical aspects concerning the model's internal working in common non-technical language.[673] The explanation of the data employed to make the decision or prediction is another important aspect to inform. In particular, the input data and the sources should be disclosed. Moreover, providing counterexamples or hypothetical counterfactual scenarios[674] to show how the result would have differed had the data subject provided different data or information will help to understand the result. Finally, in the fairness explanation controllers should communicate the relevant fairness metrics, along with information about the performance of alike people.

Providing this information not only complies with the transparency obligations but also is crucial to allow another right: the right to contest an automated decision that produces legal or similarly significant effects on the data subjects (Art. 22(3) GDPR).

---

[672] Taken from ICO's document 'Explaining Decisions Made with AI'.

[673] Information Commissioner's Office, 'Explaining Decisions Made with AI' (n 104) 24.

[674] Institute of Electrical and Electronics Engineers (n 643) s 5.1.1.

IV.2.- Addressing fairness and non-discrimination when processing personal data using AI systems

IV.2.1.- Outlining the problem of algorithmic bias and fairness

Non-discrimination is a fundamental right in the EU and Article 21 of the Charter forbids discrimination based on any ground. However, oftentimes algorithmic biases affect the performance of AI systems and it may result in unfair or discriminatory outcomes. This section will provide an overview of the problems related to algorithmic biases that deliver unfair results. Then, a basic overview of the legislative framework that protects against unfair discrimination. Finally, some alternatives will be offered as a way to mitigate the negative impact of biases in automated decision-making and promote algorithmic fairness and non-discrimination.

Biases exist not only in automated decision-making systems. Human decision-making is also influenced by biases. However, the problem with AI systems is that they automatise and potentiate biases already shared by humans. Additionally, it is important to note that biases are not only reflected in systems that continue to learn after being put into the market. Knowledge and logic-based systems can also permeate biases held by experts who developed the systems.[675]

There are different sources of biases in AI models, but they are generally generated in the early stages of AI system development. In particular, biases in AI can occur either in the data gathering or in the data preparation. Concerning *data collection*, two issues may arise, namely, when data lack statistical representativity or when it mirrors existing social preconceptions. The *lack of representativity* of the datasets with which AI systems are trained is one of the most frequent problems of AI systems. In this case, the AI system will deliver notably less accurate predictions to individuals that belong to underrepresented groups. This is because the performance of some AI systems is dependent on the training data, in particular for machine learning algorithms.

Facial recognition systems constitute a compelling example of the different accuracy scores across different ethnic groups. To begin with, most of the images that

[675] National Institute of Standards and Technology, 'NIST Special Publication 1270. A Proposal for Identifying and Managing Bias in Artificial Intelligence.' (2022) 25.

AI systems use for training come from specific locations around the world. ImageNet is the most common dataset for pre-training AI models. The vast majority of the images from this dataset come from the 'west': US 45.4%, UK 7.6%, Italy 6.2%, Canada 3%, rest of the world 37.8%.[676] The lack of diversity in the input with which AI systems are trained has a clear impact on the system's outputs. For instance, a study of facial-analysis software showed an error rate of 0.8% for light-skinned men versus 34.7% for dark-skinned women.[677] In automated dermatologists, deep learning neural networks were used to identify skin cancer from photographs. From a dataset that included 129.450 images, 60% of them scraped from Google Images, less than 5% of them were of dark-skinned people.[678] This imbalance also had significant consequences on the accuracy of the predictions.

The lack of representativity is by no means confined to image recognition systems, and it also happens in other fields, like the under-representation of different ethnicities in biobanks. The UK Biobank is one of the largest banks of genetic information in the world. However, it poorly represents people from minority groups. Within this biobank, 94.6% of participants are of white ethnicity. Compared with the general population, participants of the biobank are more likely to be older, be female, and live in less socioeconomically deprived areas. Moreover, compared with the general population, participants are less likely to be obese, smoke, drink alcohol daily, and self-report health conditions.[679] If AI systems are trained using this databank, they will be more accurate for people that belong to the majority group (which in general are white and do not suffer from economic hardship), than for people with different skin tones.

Apart from the lack of representativity in datasets, the *reflection of socially rooted biases or preconceptions* is another frequent problem that appears in the data

---

[676] James Zou and Londa Schiebinger, 'AI Can Be Sexist and Racist — It's Time to Make It Fair' (2018) 559 Nature 324, 324.

[677] Joy Buolamwini and Timnit Gebru, 'Gender Shades Intersectional Accuracy Disparities in Commercial Gender Classification', *Conference on Fairness, Accountability and Transparency* (2018); Zou and Schiebinger (n 674).

[678] Andre Esteva and others, 'Dermatologist-Level Classification of Skin Cancer with Deep Learning Neural Networks' (2017) 542 Nature 115.

[679] Anna Fry and others, 'Comparison of Sociodemographic and Health-Related Characteristics of UK Biobank Participants with General Population' (2017) 186 American Journal of Epidemiology 1026, 1027.

collection stage. If there are biases in the data entered into the model, the model will reproduce these biases. For instance, Amazon started using an applicant tracking system (an algorithm that handles the initial stages of the recruiting process automatically, vetting and ranking candidates), but the company noticed that the algorithm was gender-biased. The reason behind the disparity in the ranking between male and female candidates was found in the hiring history of the company, where men outnumber women in technical roles. Hence, the algorithm learned to 'discriminate' and downgrade CVs which contained words like 'women's'.[680] Another example of the amplification of socially grounded biases by AI systems is the use of COMPAS risk assessment tool to calculate recidivism.[681] Since these algorithms are not able to make holistic evaluations and consider every circumstance that may affect human decisions and behaviours, decisions to arrest or imprison a human should not be left completely to AI systems.[682]

Finally, developers can introduce biases when they prepare the data for the AI system. In the *preparation stage*, developers choose the features that they want the AI system to evaluate. This process is called feature engineering. Selecting the features or attributes the AI systems will take into account critically affects the precision or accuracy of the system's outcomes. In automated tracking systems, the applicant's gender, experience, and education may constitute different features or attributes the developers select to scan and rank candidates with the information provided in their CVs.

IV.2.2.- Addressing algorithmic biases and fairness

[680] Jeffrey Dastin, 'Amazon scraps secret AI recruiting tool that showed bias against women', Reuters, 11 October 2018, <https://www.reuters.com/article/us-amazon-com-jobs-automation-insight-idUSKCN1MK08G> last consulted 01/09/2021.

[681] Supreme Court of Wisconsin, *State v Loomis*. [2016] No. 2015AP157–CR. COMPAS risk assessment is a software solution that uses an algorithm to predict the risk of recidivism of a person taking into account the results of an interview with the individual and data collected from the person's criminal record.

[682] Michael L Rich, 'Machine Learning, Automated Suspicion Algorithms, and the Fourth Amendment' (2016) 164 University of Pennsylvania Law Review 871, 893–901.

IV.2.2.1.- Countering discrimination in the law

Before evaluating the methods to achieve fair processing of personal data when using AI systems, it is necessary to address another difficulty: the definition of fairness. Fairness is not defined in the legal system and there are dozens of different definitions of fairness in computer science.[683] This is the reason why one way to address this point is to rely on the concept of non-discrimination.

The principle of non-discrimination is deeply rooted in the EU tradition and it is expressly established in many EU fundamental provisions. To provide more tools to authorities for dealing with discriminatory data processing, one of the avenues could be given by the law. The legislation may be enacted at different levels, such as European, national, regional or municipal levels.

To begin with, the Charter incorporates provisions banning discrimination in Art. 20 which establishes the principle of equality before the law, and Art. 21, which sets forth the principle of non-discrimination. Then, the Treaty of the European Union contains provisions on non-discrimination in Art. 2 (values of the EU) Art. 3(3) (combatting social exclusion and discrimination in the internal market) and Art. 9 (democratic equality). The Treaty on the Functioning of the European Union also prohibits discrimination in Art. 10 (combatting discrimination when drafting and implementing EU policies and activities). Furthermore, the European Convention of Human Rights also forbids discrimination in Art. 14 by stating that the enjoyment of the rights established in the ECHR must be guaranteed without discrimination on any ground, and the ECHR Protocol nº 12 widens the scope of the non-discrimination principle to cover the enjoyment of any right, including those guaranteed under national laws.[684] Apart from these general provisions contained in the EU primary law and the legal framework of the Council of Europe, EU secondary legislation also guarantees the equal treatment of individuals. For instance, general provisions on equality and non-discrimination can be found in the Employment Equality Directive (2000/78/EC), Racial

---

[683] Arvind Narayanan, '21 Fairness Definitions and Their Politics', *Conference on Fairness, Accountability, and Transparency* (2018).

[684] However, not every EU Member State has ratified the Protocol No. 12 to the Convention for the Protection of Human Rights and Fundamental Freedoms (ETS No. 177). See the Chart of signatures and ratifications of Treaty 177, <https://www.coe.int/en/web/conventions/full-list?module=signatures-by-treaty&treatynum=177> accessed 06/01/2022.

Equality Directive (2000/43/EC), Gender Goods and Services Directive (2004/113/EC), Gender Equality Directive (recast) (2006/54/EC). Furthermore, non-discrimination based on nationality and immigration status is guaranteed in the Directive on the right to family reunification (2003/86/EC) and the Directive on long-term legally resident third-country nationals (2003/109/EC).

But crucially for this work is the GDPR, which also contains specific provisions addressed to the particular intersection between data protection and discrimination. Discrimination of data subjects is included among the elements that constitute a high risk to carry out a data protection impact assessment (DPIA). While there is no clear-cut distinction in the GDPR between high and low-risk processing, it is considered that processing operations that, as a consequence, may lead to discrimination of data subjects, must be considered as high-risk processing.[685] Additionally, discrimination is one of the harmful consequences that a data breach may have on data subjects.[686]

However, the most important provision is Art. 5(1)(a) GDPR where the principle of fair processing is established. People are treated fairly when they are considered equals, with no favouritism or discrimination. In previous sections it was evaluated the principle of fairness and its two components, informative fairness and substantive fairness. The potential discrimination of data subjects relates to the latter, that is, substantive fairness.

For the purposes to guarantee substantive fairness in the processing of personal data, the GDPR also compels data controllers to employ:

> '*appropriate mathematical or statistical procedures for the profiling, implement technical and organisational measures appropriate to ensure (…) that factors which result in inaccuracies in personal data are corrected and the risk of errors is minimised, secure personal data in a manner that takes account of the potential risks involved for the interests and rights of the data subject and that prevents, inter alia, discriminatory effects on natural persons on the basis of racial or ethnic origin,*

[685] Recital 75 GDPR.

[686] Recital 85 GDPR.

political opinion, religion or beliefs, trade union membership, genetic or health status or sexual orientation, or that result in measures having such an effect.'[687]

Hence, the GDPR itself not only recognizes the problems linked to discrimination or substantial fairness, but it also requires controllers to employ adequate scientific mechanisms (e.g. mathematical or statistical) to perform profiling activities, implement adequate measures (technical, organisational, contractual, etc) to reduce errors, inaccuracies and, ultimately, discrimination.

However useful these provisions are, they are not always easy to apply. Firstly, some of the antidiscrimination provisions established in the primary EU legislation are addressed to the States (when they legislate, design policy programs, or implement legislation) and not to individuals, like data controllers.[688] This means that data subjects may not directly rely on them to exercise their rights before data protection authorities. Second, non-discrimination provisions contained in the Charter or the ECHR may not be applicable in every case, since they generally require some degree of interpretation. This interpretation is primarily made by the judiciary, but the case law is scarce and does not encompass many cases data subjects daily face. Third, Art. 5(1)(a) GDPR on fairness is also open to interpretation. There is neither a clear criterion on what fairness means nor which the most appropriate fairness metrics are.[689] Fourth, the data subject's right to not be subject to a decision based solely on automated processing, including profiling, which produces legal or similarly significant effects concerning them can be considered also as a tool that fosters fair processing of personal data. However, while Art. 22 GDPR provides some assistance for data subjects' rights, the protection it provides is limited. As previously evaluated, there is room for interpretation and debate as to which, how and when the safeguards apply in many real case scenarios (for instance, what kind of information should be provided, how can data subjects contest automated decisions and which type of human intervention is required). Additionally, Art. 22 GDPR applies only to a limited set of AI systems that process personal data, since it only covers decisions taken solely using

[687] Recital 71 GDPR.

[688] Art. 51 Charter of Fundamental Rights.

[689] Wachter, Mittelstadt and Russell (n 2) 764.

automated processing and in so far as the decision produces legal or significantly similar effects. Finally, whereas Recital 71 GDPR provides a clear indication concerning some of the measures or steps controllers should take to mitigate discrimination, recitals are not binding even though they can function as an interpretative guide for authorities. Therefore, the current legislation offers less than optimal protection to counteract the discriminatory effects of AI systems. While general anti-discrimination principles are in place, there is little guidance on how to implement those principles in particular cases.

One of the most important recent legislative initiatives (both for its coverage and substance) is the AIA draft and it has been previously mentioned how the AIA draft can assist in this task. The AIA is particularly taxing when it comes to data governance issues. It compels providers of AI systems to train, validate and test datasets to identify possible biases.[690] It also requires relevancy, representativeness, accuracy, and completeness of datasets, which must also have adequate statistical properties.[691] Crucially, datasets must consider specific 'geographical, behavioural or functional' elements in which the AI solutions are planned to be deployed.[692] Furthermore, it mandates that providers of AI systems declare in the instructions of use the levels of accuracy and the relevant accuracy metrics of high-risk AI systems,[693] and these metrics should be informed to AI users.[694]

But despite the efforts, there are some problems. To begin with, as previously mentioned, there is no clear definition of fairness since it depends on social and cultural factors. Then, the regulation does not mention which fairness metrics developers of AI systems must use (e.g. precision or sensitivity), giving them wide discretion to choose them. It should be taken into account that the selection of a particular set of fairness metrics is not neutral, but a decision concerning the kind of biases that will be accepted.[695] In other words, establishing fairness metrics will give

[690] Art. 10(2)(f) AIA draft.

[691] Art. 10(3) AIA draft.

[692] Art. 10(4) AIA draft.

[693] Art. 15(2) AIA draft.

[694] Recital 49 AIA draft.

[695] Wachter, Mittelstadt and Russell (n 2) 48.

different results, because different fairness metrics matter to different stakeholders.[696] Hence, the metrics employed should be tailored to the problem the AI system is called to solve.

IV.2.2.2. Impact assessments and audits to mitigate risks not covered by DPIAs

Evaluating the risks posed by AI systems is crucial for the responsible use of these technologies. However, there are multiple ways to evaluate the wide diversity of risks posed by AI systems. AI solutions can be subject to different kinds of impact assessments such as stakeholder impact assessments,[697] environmental impact assessments, responsible innovation assessments,[698] ethical impact assessments, human rights impact assessments, or algorithmic impact assessments. However, these assessments are generally performed on a voluntary basis by AI operators.

The GDPR requires controllers only to carry out data protection impact assessments (DPIA) for their processing activities involving personal data.[699] DPIA must include at least a complete description of the processing and the purposes, an evaluation of the necessity and proportionality of the processing, and the risks to individuals, along with measures to mitigate the risks identified.[700]

However, two major setbacks should be considered regarding DPIAs. First, DPIAs are mostly limited to identifying and evaluating only some risks posed by the processing operations, namely, threats to the rights to data protection and respect the

---

[696] Narayanan (n 681).

[697] David Leslie, 'Understanding Artificial Intelligence Ethics and Safety. A Guide for the Responsible Design and Implementation of AI Systems in the Public Sector' (2019) 26; International Standard Organisation, 'ISO/IEC TR 24027:2021 Information Technology — Artificial Intelligence (AI) — Bias in AI Systems and AI Aided Decision Making' (2021) s 8.2.5.

[698] See for instance standard CEN/CWA 17796:2021 Responsibility-by-design - Guidelines to develop long-term strategies (roadmaps) to innovate responsibly, available at <https://standards.cencenelec.eu/dyn/www/f?p=205:110:0::::FSP_PROJECT:74562&cs=13B49C595A36D9E4066EDD7D09FA02FFF> accessed on 11/02/2022; Emad Yaghmaei and Ibo van de Poel, *Assessment of Responsible Innovation. Methods and Practices* (Rutledge 2020).

[699] Art. 35 GDPR. As previously mentioned, it is highly likely that when controllers use AI systems to process personal data controllers they must carry out a DPIA.

[700] Art. 35(7) GDPR.

private and family life. Granted, the GDPR does not restrict the evaluation of the risk posed by data processing operations to those previously enumerated. On the contrary, it only mentions that the assessment should include the 'risks to the rights and freedoms of data subjects' without any further clarification. In addition, Art. 1(2) GDPR expressly recognises that it safeguards 'fundamental rights and freedoms of natural persons and in particular their right to the protection of personal data'.[701] This should lead to the conclusion that the assessment of the risks to the rights and freedoms of data subjects should be broader and include other rights and freedoms going beyond those related to data protection and the protection of private and family life. In other words, while the DPIA chiefly relates to safeguarding personal data and privacy, it could additionally cover other fundamental rights.[702]

However, this interpretation contrasts with the recommendations and guidelines from EU or national supervisory authorities and with the practice of privacy professionals.[703] For instance, the European Data Protection Board and the Commission Nationale de l'Informatique et des Libertés mention the following risks: non-legitimate access, unwanted modification, and data loss.[704] Similarly, the Irish Data Protection Commissioner in the relevant guidelines under the title 'data protection and related risks' provides an illustrative list of the risks that data controllers should evaluate.[705] This list includes relevant risks to individuals such as data breach, personal data used in a way not expected by individuals, excessively intrusive uses of data, and excessively long retention periods, among others. As seen, all of them relate

---

[701] See also Heleen L Janssen, 'An Approach for a Fundamental Rights Impact Assessment to Automated Decision-Making' (2020) 10 International Data Privacy Law 76, 85. The author finds support in Recital 2 and 75 GDPR.

[702] European Data Protection Suprevisor (n 591) 15.

[703] See for instance, Public DPIA Teams OneDrive SharePoint and Azure AD carried out by the Government of the Netherlands <https://www.rijksoverheid.nl/documenten/publicaties/2022/02/21/public-dpia-teams-onedrive-sharepoint-and-azure-ad> accessed 22/02/2022.

[704] European Data Protection Board, 'Guidelines on Data Protection Impact Assessment (DPIA) and Determining Whether Processing Is "Likely to Result in a High Risk" for the Purposes of the GDPR' (n 572) 22; Commission Nationale de l'Informatique et des Libertés, 'Privacy Impact Assessment (PIA). Knowledge Bases' (2018) 3.

[705] Data Protection Commission (n 586) 18–19.

primarily to data protection and privacy risks. This means that while the scope of the evaluation is not restricted to a specific set of rights, in practice, professionals and data protection authorities pay little attention to rights and freedoms not related to the protection of personal data and private and family life.

Second, the GDPR does not require data controllers or AI providers to involve data subjects or other potentially affected stakeholders in the production of the DPIA. Whereas the GDPR mandates consulting with data subjects or their representatives about future processing activities,[706] this consultation has two limitations. First, it only requires this consultation 'where appropriate', meaning giving data controllers wide discretion to merit the appropriateness of the consultation or to value the circumstances under which such feedback would be useful. Second, the discretion to seek feedback from data subjects is further widened, since the GDPR allows controllers to evaluate the necessity to engage in such consultation against the need to protect business or public interests, or the security of the processing. It is worth noting that, in general, to uncover a full range of risks to individuals it is not enough to enumerate the persons or groups of people potentially affected without receiving their feedback.[707] Finally, the GDPR does not mandate disclosing the results of the DPIA to society.[708] Public disclosure of the results of the DPIA, even summaries of them, would contribute to the transparency of the processing operations.

These are crucial downsides for the clear identification of the risks posed by the processing operations, the proposal of mitigation measures, and to keep data controllers fully accountable. In this context, there is a need to incorporate processes that ensure full transparency and consider a wide range of fundamental rights that AI systems are capable to impair, which are not limited to data protection and the protection of private and family life. Below some of the most important tools to mitigate the harmful impacts of AI systems that process personal data are evaluated, namely, Human Rights Impact Assessments, Ethical Design of AI Systems and Algorithmic Audits.

---

[706] Art. 35(9) GDPR.

[707] International Standard Organisation, 'ISO/IEC TR 24027:2021 Information Technology — Artificial Intelligence (AI) — Bias in AI Systems and AI Aided Decision Making' (n 695) s 8.2.5.

[708] Kaminski and Malgieri (n 533) 133.

A) Human rights impact assessments

Concept. While DPIAs are very important tools to identify, evaluate and mitigate the risks posed by the processing operations carried out using AI systems, this accountability tool has its limitations. Human rights impact assessments[709] can provide an alternative to overcome the limitations of DPIAs.

Human rights impact assessments are tools whose objective is to find, interpret, evaluate, and mitigate the possible or current negative effects of any kind of measure, process or business on individuals or groups of people, and to guarantee that the measures, processes or businesses under evaluation are compliant with international human rights obligations.[710] These assessments generate crucial information to consider a wide range of viewpoints and help to reach better decisions concerning activities originated in the public or private sector that could negatively affect the enjoyment of fundamental rights.[711] This is because they take into account human rights obligations as the benchmark and should actively collect insights and feedback from affected individuals and groups of rights-holders. The stakeholder participation is not only for the diagnostic phase, but participants are also encouraged to propose methods or ways to mitigate the impacts of the process or activity under evaluation. Moreover, the results of the human rights impact assessments are disclosed to the public, which ensures broader accountability of the whole process.

The outcome of human rights impact assessments can play a crucial role in AI governance since it is suggested that AI systems should not be deployed if assessors identify significant risks to fundamental rights that cannot be reduced.[712]

Requirement. The current regulatory framework does not require the performance of human rights impact assessments for the design, development and deployment of

---

[709] Office of the United Nations High Commissioner for Human Rights, 'Guiding Principles on Business and Human Rights. Implementing the United Nations "Protect, Respect and Remedy" Framework' (2011); Organisation for Economic Co-operation, 'Guidelines for Multinational Enterprises' (2011).

[710] Human Rights Council, 'Guiding Principles on Human Rights Impact Assessments of Economic Reforms. A/HRC/40/57' (2019) para 6.

[711] Nora Götzmann, 'Introduction to the Handbook on Human Rights Impact Assessment: Principles, Methods and Approaches' in Nora Götzmann (ed), *Handbook on Human Rights Impact Assessment. Research Handbooks on Impact Assessment series* (Elgar 2019) 4.

[712] Recommendation CM/Rec(2020)1 of the Committee of Ministers to member States on the human rights impacts of algorithmic systems para B.5.4.

AI systems. However, there is a growing consensus among academics,[713] civil society organisations,[714] and independent expert groups set up by the European Commission[715] that they should be part of the AI processes.

Additionally, the Council of Europe has taken a clear stance on the necessity of human rights impact assessments for the design, development and deployment of AI systems. To begin with, the Council of Europe Commissioner for Human Rights recommended member states to require the performance of Human Rights Impact Assessments when public authorities build or use AI systems.[716] In the same vein, the Committee of Ministers of the Council of Europe recommended member states to evaluate the convenience of building legal frameworks that safeguard the fundamental rights of individuals against the deleterious effects of targeting using automated

---

[713] Alessandro Mantelero and Maria Samantha Esposito, 'An Evidence-Based Methodology for Human Rights Impact Assessment (HRIA) in the Development of AI Data-Intensive Systems' (2021) 41 Computer Law & Security Review 1; Céline Castets-Renard, '6 - Human Rights and Algorithmic Impact Assessment for Predictive Policing' in Oreste Pollicino and Hans-W Micklitz (eds), *Constitutional Challenges in the Algorithmic Society* (CUP 2019).

[714] Danish Institute for Human Rights, Guidance on Human Rights Impact Assessment of Digital Activities: Introduction (2020) <https://www.humanrights.dk/publications/human-rights-impact-assessment-digital-activities> accessed 21/02/2022; European Digital Rights (EDRi), An EU Artificial Intelligence Act for Fundamental Rights. A Civil Society Statement (30/11/2021) <https://edri.org/wp-content/uploads/2021/12/Political-statement-on-AI-Act.pdf> accessed 21/02/2022; European Centre for Not-for-Profit Law, Mandating Human Rights Impacts Assessments in the AI Act <https://ecnl.org/sites/default/files/2021-11/HRIA%20paper%20ECNL%20and%20Data%20Society.pdf> accessed 21/02/2022; Statewatch, EU: Artificial Intelligence Act must put human rights first (30/11/2021) <https://www.statewatch.org/news/2021/november/eu-artificial-intelligence-act-must-put-human-rights-first/> accessed 21/02/2022; Access Now, Here's how to fix the EU's Artificial Intelligence Act (07/09/2021) <https://www.accessnow.org/how-to-fix-eu-artificial-intelligence-act/> accessed 21/02/2022; Center for Democracy and Technology, EU Tech Policy Brief: July 2021 Recap (06/08/2021) <https://cdt.org/insights/eu-tech-policy-brief-july-2021-recap/> accessed 21/02/2022.

[715] High-Level Expert Group on AI, 'The Assessment List for Trustworthy Artificial Intellignece (ALTAI)' (2020) 5.

[716] Council of Europe Commissioner for Human Rights, 'Unboxing Artificial Intelligence: 10 Steps to Protect Human Rights' (2019) 7.

systems '*beyond current notions of personal data protection and privacy*'.[717] In particular, the Committee of Ministers of the Council of Europe recommended that Human Rights Impact Assessments should be conducted before: a) engaging in computational experimentation or research that may have a substantial impact on fundamental rights;[718] b) developing and procuring any AI system with the capacity to produce potentially significant human rights impact or carrying high risks to fundamental rights;[719] and c) developing and procuring of high risks AI systems, and at regular intervals during their lifecycle.[720] Additionally, impact assessments should be carried out to evaluate the specific risks of profiling using AI systems.[721]

As seen from the previous list, a human rights impact assessment should not be performed for every single AI system. It seems unreasonable to require the performance of human rights impact assessments to minimal or low-risk AI systems. Just as the DPIA is required for particularly intrusive processing operations, a similar criterion should be agreed upon to require human rights impact assessments for AI systems. Among the relevant criteria for requiring human rights impact assessments to AI systems could be included the purpose or the intended use of the AI system, the potential to impact individuals' fundamental rights, and the number of individuals potentially affected by the system.[722]

It should be noted though that the recommendations from the Committee of Ministers of the Council of Europe are only suggestions, guidelines and best practices that lack binding effect on the Member States. Nevertheless, the recommendations offer a reference policy framework and roadmap to the Member States, and they could guide legislative development and judicial interpretation of the international obligations

---

[717] Council of Europe, 'Declaration by the Committee of Ministers on the Manipulative Capabilities of Algorithmic Processes - Decl(13/02/2019)1' (2019) para 9.

[718] Recommendation CM/Rec(2020)1 of the Committee of Ministers to member States on the human rights impacts of algorithmic systems para B.3.1.

[719] ibid B.5.1.

[720] ibid B.5.2.

[721] Recommendation CM Rec(2021)8 protection of individuals with regard to automatic processing of personal data in the context of profiling 2021 para 7.9.

[722] Council of Europe - Ad Hoc Committee on Artificial Intelligence (CAHAI), 'Human Rights, Democracy and Rule of Law Impact Assessment of AI Systems' (2021) 22.

of Member States. Member States may eventually be invited to communicate the steps taken concerning the recommendations issued by the Committee of Ministers.[723]

However, the persuasive effects of these recommendations are noticeable in the decision 492/2021 Coll from the Constitutional Court of the Slovak Republic.[724] In this case, the Constitutional Court relied heavily on Recommendation CM Rec(2020)1 issued by the Committee of Ministers of the Council of Europe to request public authorities to conduct human rights impact assessments when implementing AI systems that could potentially affect a broad range of fundamental rights.[725] Interestingly, the protected fundamental rights were not limited to those directly addressed by the GDPR. The court considered that the automatic evaluation of individuals can not only interfere with their right to informational self-determination. It held that even in those cases where no personal data is being processed, thus falling out of the scope of the GDPR and Art. 19(3) and 22(1) Slovak Constitution -which relate to the right to data protection and privacy, respectively-, the right to a fair trial, to freedom of expression and assembly, and the prohibition of discrimination and unequal treatment could be impaired by automated decision systems.[726]

Therefore, even though human rights impact assessments are not mandatory under the current legislative framework, it seems likely that the performance of an evaluation of the fundamental rights implications to deploy AI systems will be a requirement for trustworthy AI in the Council of Europe. Likewise, more courts could take a similar approach to the Constitutional Court of Slovakia, taking the Council of Europe's recommendations as a benchmark for the minimum requirements developers of AI systems should implement. In this way, the guidance from the Council of Europe spills over the European legal framework. This could take place in particular when evaluating the design and use of AI systems by public authorities. However, this kind of assessments could also relate to activities carried out by private actors. For instance, in March 2022 two Dutch MPs called to require human rights impact assessments before deploying algorithms for evaluations or decisions about

[723] Article 15(b) Statute of the Council of Europe.

[724] *Judgment no. k. PL. ÚS 25 / 2019-117 - 492/2021 Coll.* (n 343).

[725] ibid 134.

[726] ibid 131.

people.[727] And while this is only a motion, it shows that parliamentarians around Europe have started discussions about this topic.

Methodology. The concrete aspects concerning the methodology to perform a human rights impact assessment should also be agreed upon. Whereas there are several well-established methodologies to carry out impact assessments (privacy impact assessments, data protection impact assessments, environmental impact assessments, etc), the different objectives and scope of this kind of evaluation require a tailored methodology.

While there are many toolkits for performing human rights impact assessments of AI systems, which have been developed both by public and private institutions,[728] the most comprehensive methodology so far developed for this purpose was prepared by the Alan Turing Institute. This is a proposal to analytically evaluate the human rights, democracy and rule of law implications of AI systems[729] and it was recently submitted to the Council of Europe's Ad Hoc Committee on Artificial Intelligence (CAHAI) for consideration. The Human Rights, Democracy, and the Rule of Law Assurance Framework for AI Systems has four stages.[730] Firstly, the Preliminary Context-Based Risk Analysis has the objective to give an overview of the risks that the AI system may cause to human rights, democracy and rule of law, and helps define the level of stakeholder involvement. Secondly, the Stakeholder Engagement Process completes the definition of stakeholders invited to provide feedback and then establishes the

---

[727] House of Representatives of the Netherlands, 'Motie van de leden Bouchallikh en Dekker-Abdulaziz over verplichte impactassessments voorafgaand aan het inzetten van algoritmen voor evaluaties van of beslissingen over mensen' (29th March 2022) <https://www.tweedekamer.nl/kamerstukken/moties/detail?id=2022Z06024&did=2022D12329> accessed 13/04/2022. Translated with Google Translate.

[728] See for instance the methodology developed by the Dutch Government (Rijksoverheid) Impact Assessment for Human Rights in the Use of Algorithms (*Impact Assessment Mensenrechten en Algoritmes*, IAMA) which can be consulted in the official webpage <https://www.government.nl/documents/reports/2021/07/31/impact-assessment-fundamental-rights-and-algorithms> accessed 13/05/2022.

[729] David Leslie and others, 'Human Rights, Democracy, and the Rule of Law Assurance Framework for AI Systems: A Proposal Prepared for the Council of Europe's Ad Hoc Committee on Artificial Intelligence' (2022).

[730] ibid 11–12.

mode and depth of the stakeholder engagement. It analyses the stakeholders, invites team members to reflect on their personal characteristics and social status and how these features impact their decisions, defines the objective of the stakeholder engagement and the engagement methods (e.g. in-person interviews or focus groups) and, then, receives the stakeholder feedback. Thirdly, the Human Rights, Democracy, and the Rule of Law Impact Assessment. This stage aims at precisely assessing the probable and current negative effects that the design, development and deployment of the AI system may have on fundamental rights, which were provisionally considered in the first step, and draws a plan to mitigate those impacts. Crucially, it also requires evaluating and addressing all the negative effects that the AI solution may produce across the value chain.[731] Finally, the Human Rights, Democracy, and the Rule of Law Assurance Case serves to ensure relevant stakeholders that their concerns about the potential negative impacts of the AI system have been addressed and documented. In particular, it determines the risk management strategy and concludes the impact mitigation plan and it describes detailed steps to make those objectives operative.

B) Addressing ethical concerns in the design and use of AI systems

Abiding by mandatory regulations is necessary but not always enough to design, develop and deploy safe and trustworthy AI systems. Ethical principles and norms may play a role when the law is either not enacted or insufficiently applied. To establish what is good or ethical it is necessary to understand the ethical theories that are employed to make decisions, and which are the social and individual underlying values of the persons in charge of the design, development and deployment of AI systems.[732] The ethical implications of the AI systems should be discussed in every stage of the AI system lifecycle, but in particular during the initial stages.[733] This is because the initial steps set the groundwork upon which the following stages are built.

At the same time, there have been numerous initiatives worldwide to address the concerns related to the development and deployment of AI systems. While there is a

[731] ibid 245.

[732] Virginia Dignum, *Responsible Artificial Intelligence. How to Develop and Use AI in a Responsible Way* (Springer 2019) 35. The main ethical theories are consequentialism, deontology and virtue ethics.

[733] Mark Coeckelbergh, *AI Ethics* (MIT Press 2020) 165; Luciano Floridi and Andrew Strait, 'Ethical Foresight Analysis: What It Is and Why It Is Needed?' (2020) 30 Minds and Machines 77.

growing consensus regarding the main principles that should guide the design and use of AI systems,[734] in general, these guidelines are mostly theoretical and only to a limited extent they operationalise the ethical principles or fundamental values developed therein.[735] In other words, they do not fully explain how these values or principles should be implemented in practice or which concrete steps professionals should take when developing and deploying AI systems.

Therefore, there is a need to explore methodologies to assist organisations to address ethical concerns in the design and use of AI systems. The UNESCO recently suggested that member states should require AI operators to carry out ethical impact assessments before using AI systems.[736] The UNESCO neither defines ethical impact assessments nor provides a methodology to guide professionals on how they could be carried out. On the contrary, it only mentions some important aspects that should be included,[737] and it considers that data protection impact assessments are an integral part of ethical impact assessments.[738] Hence, more guidance about ethical impact assessment should be sought in standards or guidelines.

In general, ethical impact assessments are methodologies that, in close consultation with relevant stakeholders, aim at identifying and evaluating the negative ethical impacts of human activities and elaborating remedial measures to reduce those ethical risks.[739] Its objective is to look beyond the clear and immediate potential risks of technologies[740] and foresee a wide range of consequences not only for the

---

[734] See for instance Jessica Fjeld and others, 'Principled Artificial Intelligence: Mapping Consensus in Ethical and Rights-Based Approaches to Principles for AI' (2020) 64.

[735] Coeckelbergh (n 731) 165.

[736] UNESCO (n 440) para 50.

[737] ibid 50–53. For instance, identification of impacts on human rights, testing high-risk AI solutions before commercializing them, implementation of suitable measures during the whole AI lifecycle, adherence to human rights obligations, and multi-stakeholder participation, including a horizontal gender viewpoint (see para. 87).

[738] ibid 72.

[739] European Committee for Standarization, 'CWA 17145-2:2017 (E) - Ethics Assessment for Research and Innovation - Part 2: Ethical Impact Assessment Framework' (2017) s 2.5.

[740] Rasmus Øjvind Nielsen, Agata M Gurzawsk and Philip Brey, 'Satori Project. Principles and Approaches in Ethics Assessment. Ethical Impact Assessment and Conventional Impact Assessment. Annex 1.A' (2015) 5.

individuals affected by the use of the AI systems but also for society at large. They evaluate broader socio-economic effects than data protection impact assessments.

Specifically for AI systems, a definition can be found in the standard *CAN/CIOSC 101:2019 - Ethical design and use of automated decision systems*. According to this standard, an ethical impact assessment is:

> '*a framework to help organisations better understand and reduce the potential ethical risks associated with automated decision systems and to provide the appropriate governance, oversight and reporting/audit requirements that best match the type of application being designed'*.[741]

CAN/CIOSC 101:2019 standard requires implementing a risk management framework, which should be integrated into the company's compliance program, designate professionals to monitor the process, implement a measurement system, and carry out an ethical impact assessment in consultation with adequately skilled professionals.[742] In addition, it requires embedding ethical considerations into the design of AI systems (i.e., evaluating the ethical impacts on relevant stakeholders) during the whole AI system lifecycle, as well as considering possible stages of human intervention (such as human-in-the-loop or human-on-the-loop), properly describe and curate the data used to train the AI system, and monitor potential biases in the data.[743] Finally, it establishes adequate mechanisms for the use of the AI system (e.g. training of staff and setting reliability metrics) and for the surveillance and maintenance of the AI systems (e.g. setting a monitoring procedure and re-evaluating relevant metrics at appropriate intervals).[744]

Additionally, numerous initiatives were recently developed to encourage ethical thinking from the earliest stages of AI development. For instance, the IEEE 7000-2021 Standard Model Process for Addressing Ethical Concerns During System Design is another tool that operationalises and provides practical guidance for organisations that

---

[741] Standards Council of Canada, 'CAN/CIOSC 101:2019 - Ethical Design and Use of Automated Decision Systems' (2019) s 3.

[742] ibid 4.1.1 to 4.1.15.

[743] ibid 4.2.1 to 4.2.10.

[744] ibid 4.3 and 4.4.

plan to include processes to instil ethical principles into the design and development of AI systems.[745] Moreover, the Dutch Government recently published the Ethically Responsible Innovation Toolbox which assists developers and administrators on what is needed for ethically responsible innovation, it considers important public values and fundamental rights and provides concrete recommendations for each ethical principle.[746]

Hence, both ethical impact assessments and similar frameworks that assist in the ethical design of AI systems are then important tools to foresee the consequences of the development and use of AI systems, as well as to reduce risks, enhance stakeholder engagement, and design strategies to mitigate the impact for the individuals that may be unfavourably affected by the AI system.

C) Algorithmic audits

Another mechanism to ensure adequate functioning and regulatory compliance is the algorithmic audit. Audits are a 'systematic, independent and documented process for obtaining objective evidence and evaluating it objectively to determine the extent to which the audit criteria are fulfilled'.[747] Auditing requires independency from the auditors. In general audit frameworks require that auditors are certified by independent institutions, employ a third-party set of rules to perform the audit, and abide by rules that ensure independence in the performance of their tasks (like those stated in Sarbanes-Oxley Act).

While the GDPR does not explicitly require controllers to audit AI systems for compliance, it is possible to infer the obligation to perform this activity from the articles and principles of the GDPR. The GDPR requires controllers to adopt adequate technical and organisational measures to guarantee and be able to prove that the

---

[745] Institute of Electrical and Electronics Engineers, 'IEEE 7000-2021 Standard Model Process for Addressing Ethical Concerns During System Design' (2021).

[746] Dutch Government, Toolbox Ethically Responsible Innovation, <https://www.digitaleoverheid.nl/overzicht-van-alle-onderwerpen/nieuwe-technologieen-data-en-ethiek/publieke-waarden/toolbox-voor-ethisch-verantwoorde-innovatie/> accessed on 13/04/2022. Translation using Google Translate.

[747] ISO 19011:2018 - Guidelines for auditing management systems, Clause 3.1

processing operations are carried out in line with the GDPR,[748] and audits constitute a way to demonstrate compliance. Additionally, the processor must contribute to the performance of 'audits' carried out by the controller,[749] controllers and processors must develop a process to routinely test and evaluate the effectiveness of the technical and organisational measures to guarantee the security of the processing (i.e., carry out periodic audits),[750] and one of the functions of the data protection officer is the supervision of 'audits'.[751] Likewise, the principle of accountability[752] requires the adoption of a system of continuous improvement including the performance of the pertinent periodic reviews or audits. Finally, binding corporate rules[753] must specify the procedures to verify compliance with them, including 'data protection audits'.[754]

Data protection audits in AI systems allow controllers to effectively prove compliance with their statutory obligations, but also to have more control over the processing of personal data, detect vulnerabilities and non-conformities in the management of information systems promptly, and their correction and monitoring. Finally, it is worth noting that audits constitute an opportunity for routinely improvement by developing action plans.

However, auditing AI systems is more complex than a mere GDPR audit. An algorithm audit consists in gathering information about the characteristics and behaviour of the AI system to use the collected information to evaluate the negative effects, if any, on the rights of individuals or society as a whole.[755] In other words, it focuses on uncovering algorithmic actual or potential problematic behaviour.[756]

---

[748] Art. 24 GDPR.

[749] Art. 28(3)(h) GDPR.

[750] Art. 32(1)(d) GDPR.

[751] Art. 39(1)(b) GDPR.

[752] Art. 5(2) GDPR.

[753] Binding Corporate Rules (BCR) are one of the mechanisms that can be used to transfer personal data to non-EU countries in the absence of an EU Commission adequacy (see Arts. 4(20), 46(2)(b) and 47 GDPR).

[754] Art. 47(2)(j) GDPR.

[755] Brown, Davidovic and Hasan (n 2) 2.

[756] Jack Bandy, 'Problematic Machine Behavior: A Systematic Literature Review of Algorithm Audits', *Proceedings of the ACM on Human-Computer Interaction Volume 5 Issue CSCW 1* (2021) 4.

Algorithmic audits can be carried out in different depths and, accordingly, they require different degrees of access to the system's technical elements. While basic audits entail reviewing the documentation provided by the auditee, more complex audits include data and code inspection and reproduction of the system or parts of the system to evaluate the model's behaviour and performance.[757]

A particular kind of algorithmic audit is related to the search for biases in algorithmic decision-making. Contrary to wide-range AI auditing, bias audits focus on a specific aspect of the AI system, i.e., the evaluation of input and output data to determine if the AI system produces unfairly biased predictions, decisions or other outcomes, for instance in recruiting.[758] However, there are also compliance audits and ethical audits, among others. In contrast to impact assessments, audits are performed after the AI system is in operation or deployment,[759] which allows for drawing more precise and concrete conclusions about the real effects of the AI systems on individuals and society.

Algorithmic audits constitute a solution to mitigate the risks posed by AI systems and several problematic behaviours of AI systems were discovered via algorithmic audits. Algorithmic audits have shown how some AI systems, like those present in search engines or recommender systems, could manipulate or alter the underlying facts, leading for instance to disinformation or 'echo chambers'.[760] Additionally, algorithmic audits uncovered AI systems that unproperly employed user-generated content or inferred personal data from non-sensitive data[761] and algorithms whose erroneous predictions, outcomes or classifications lead to misjudgment of individuals,

---

[757] Supreme Audit Institutions of Finland, Germany, the Netherlands, Norway and the UK 'Auditing Machine Learning Algorithms. A White Paper for Public Auditors' (2020) 16–17.

[758] Richard Landers and Tara Behrend, 'Auditing the AI Auditors: A Framework for Evaluating Fairness and Bias in High Stakes AI Predictive Models' [2022] American Pshycologist 1; Emre Kazim and others, 'Systematizing Audit in Algorithmic Recruitment' (2021) 9 Journal of Intelligence 46.

[759] Ada Lovelace Institute, 'Examining the Black Box. Tools for Assessing Algorithmic Systems' (2020) 5.

[760] Lucas D Introna and Helen Nissenbaum, 'Shaping the Web: Why the Politics of Search Engines Matters' (2000) 16 The Information Society 169.

[761] Nicholas Vincent and others, 'Measuring the Importance of User-Generated Content to Search Engines', *Proceedings of the International AAAI Conference on Web and Social Media* (2019).

in particular in matters pertaining to criminal justice[762] and marketing. Crucially, many algorithmic audits revealed discriminatory behaviours of algorithms. Wide-known examples of these audits include uncovering discriminatory behaviour in advertising (gender-based discrimination in the display of STEM job vacancies),[763] in the performance of facial analysis algorithms (in particular concerning dark-skinned women),[764] in online booking websites (leading to price discrimination or differentiation),[765] and search engines (like exaggerating gender stereotypes in image search results).[766]

However, performing an algorithmic audit does not guarantee that the auditors have complete access to the inner workings of the AI systems and even if they are granted full access, it may not be possible for them to understand how the model works and why it delivers particular outcomes.[767] A strategy to overcome this issue is to consider the AI system's internal processes to decide as a black box, but this type of audit provides little information concerning the reasons the discriminatory behaviour was detected.[768]

E) Conclusions

In this section, some of the most important accountability mechanisms have been evaluated. These mechanisms are voluntary non-binding accountability mechanisms

---

[762] Lauren Kirchner, Surya Mattu, Jeff Larson, and Julia Angwin, 'Machine Bias' (ProPublica 2016) <https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing> accessed 16/05/2022.

[763] Anja Lambrecht and Catherine Tucker, 'Algorithmic Bias? An Empirical Study of Apparent Gender-Based Discrimination in the Display of STEM Career Ads' (2019) 65 Management Science 2966.

[764] Joy Buolamwini and Gebru (n 675).

[765] Thomas Hupperich and others, 'An Empirical Study on Online Price Differentiation', *Proceedings of the Eighth ACM Conference on Data and Application Security and Privacy* (2018).

[766] Matthew Kay, Cynthia Matuszek and Sean A Munson, 'Unequal Representation and Gender Stereotypes in Image Search Results for Occupations', *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems* (2015).

[767] Balazs Bodo and others, 'Tackling the Algorithmic Control Crisis –the Technical, Legal, and Ethical Challenges of Research into Algorithmic Agents' (2019) 19 Yale Journal of Law & Technology 133, 144.

[768] Joshua Kroll and others, 'Accountable Algorithms' (2017) 165 University of Pennsylvania Law Review 633, 651.

that organisations may implement to address ethical and fundamental rights concerns. They cannot be enforced in case of non-application or non-compliance of them. They are soft governance 'post-compliance' mechanisms since they acknowledge what should or should not be done only after compliance with legally binding regulations has been ensured.[769]

Currently, where AI developers intend to build ethical AI systems or attempt to mitigate the impacts on fundamental rights of individuals and society they follow voluntary (non-binding) guidelines or standards. As seen before, there were lots of academic initiatives to develop ethical guidelines and to provide support for the responsible use of AI systems. However, some obstacles still make it difficult in practice. Not only does the lack of binding effects disincentivise their wider adoption, but also the guidelines often address different aspects of the AI systems and there is no harmonization of the requirements that developers should abide by or follow. This intricate scenario creates perplexity among AI operators and constitutes a barrier to the wider implementation of good practices in the AI ecosystem. In addition, while there are dozens of guidelines mapping the ethical and human rights implications, only a handful of attempts were made to translate those high-level principles into operative, ready-to-use, tools, guidelines or standards for those evaluating AI systems.

However, there are some indications that this situation may change. Even though these instruments are not mandatory some of the largest tech companies have already commissioned human rights impact assessments to evaluate their processes. For instance, Intel assessed the probable risks linked to emerging technologies, like autonomous driving and drones,[770] Google commissioned human rights impact assessments for some of its facial recognition technologies,[771] Yahoo for their search

---

[769] Luciano Floridi, 'Soft Ethics and the Governance of the Digital' (2018) 31 Philosophy & Technology 1, 4; Jakob Mökande and Maria Axente, 'Ethics-Based Auditing of Automated Decision-Making Systems: Intervention Points and Policy Implications' [2021] AI & Society 1, 3.

[770] Article One, Human Rights Impact Assessment on Intel's products (2018) <https://www.articleoneadvisors.com/intel-hria> accessed 1704/2022.

[771] BSR, Google Celebrity Recognition API Human Rights Assessment (2019) <https://www.bsr.org/reports/BSR-Google-CR-API-HRIA-Executive-Summary.pdf> accessed 17/04/2022.

engine technologies,[772] Meta (ex-Facebook) for the use of the platform in Indonesia,[773] Cambodia[774] and Sri Lanka,[775] and Microsoft AI for technologies in general.[776] These initiatives may create incentives to other companies to follow suit. Additionally, according to some commentators, there will be a convergence and integration among the different impact assessment methodologies (e.g. data protection impact assessments and ethical impact assessments)[777], and there may be a merge between current mandatory impact assessment and proposed auditing technics.[778]

Furthermore, legislation addressing the problematic aspects of AI systems may soon be enacted. Lawmakers in different jurisdictions have already introduced specific laws addressing particular problems in sensitive sectors like insurance,[779]

---

[772] Yahoo, Yahoo Business & Human Rights Program (2016) <https://www.ohchr.org/Documents/Issues/Expression/Telecommunications/Yahoo.pdf> accessed 17/04/2022.

[773] Facebook, Assessing the Human Rights Impact on Facebook's Platform in Indonesia (2018) <https://about.fb.com/wp-content/uploads/2020/05/Indonesia-HRIA-Executive-Summary-v82.pdf> accessed 17/04/2022.

[774] Facebook, Human Rights Impact Assessment. Facebook in Cambodia (2019) <https://about.fb.com/wp-content/uploads/2020/05/BSR-Facebook-Cambodia-HRIA_Executive-Summary2.pdf> accessed 17/04/2022.

[775] Facebook, Assessing the Human Rights Impact on Facebook's Platform in Sri Lanka (2018) <https://about.fb.com/wp-content/uploads/2020/05/Sri-Lanka-HRIA-Executive-Summary-v82.pdf> accessed 17/04/2022.

[776] Article One, Human Rights Impact Assessment on Microsoft's AI Products (2018) <https://www.articleoneadvisors.com/case-studies-microsoft> accessed 07/04/2022

[777] David Wright and Michael Friedewald, 'Integrating Privacy and Ethical Impact Assessments' (2013) 40 Science and Public Policy 755, 757.

[778] Emre Kazim and others, 'AI Auditing and Impact Assessment: According to the UK Information Commissioner's Office' (2021) 1 AI and Ethics 1, 1.

[779] See for instance, Colorado Law SB21-169 which limits the ability of insurer's to rely on consumer data from external sources and to use predictive models that employ external consumer data < https://www.leg.colorado.gov/sites/default/files/2021a_169_signed.pdf> accessed 16/05/2022.

recruitment,[780] automatic scoring for contractual purposes,[781] or facial recognition in open spaces.[782] Even some countries have enacted[783] or proposed[784] legislative acts which impose certain organisations to perform an algorithmic impact assessment before deployment of AI systems. These regulations could be seen as stepping stones for the wider adoption of similar laws that include more sectors and more jurisdictions.

Another important initiative is the AIA draft that, while it does not imposes carrying out a human rights impact assessment, requires the establishment, implementation, and documentation of a risk management system for high-risk AI systems.[785] However, this risk assessment will be carried out only by the provider of the AI system. The user of the AI system, which will generally act as a controller under the GDPR, will not have any obligation to perform a risk assessment under the AIA, but it may be required to undertake a DPIA according to the GDPR. Since the risks are different for

---

[780] See for instance, New York Law 2021/144 which requires a bias audit before using automated employment decision tools < https://legistar.council.nyc.gov/LegislationDetail.aspx?ID=4344524&GUID=B051915D-A9AC-451E-81F8-6596032FA3F9> accessed 16/05/2022. Also Illinois Law IL HB0053 which requires to those who use AI to determine whether job applicants move to the interview phase report demographic information to the Department of Commerce and Economic Opportunity to assess potential racial biases < https://legiscan.com/IL/text/HB0053/2021> accessed 16/06/2022.

[781] German Federal Law on Data Protection, paragraph 31. See also *OQ v SCHUFA Holding AG and Land Hesse* (n 353).

[782] Italian Law 205/21 suspends until 31//12/2023 the use CCTV systems equipped with facial recognition technologies in spaces open to the public, except when used for the prevention of crimes.

[783] Canadian Treasury Board Directive on Automated Decision-Making (2019) <https://www.tbs-sct.canada.ca/pol/doc-eng.aspx?id=32592> accessed 17/05/2022.

[784] Algorithmic Accountabiltiy Act of 2022, proposal made by Democrat legislators in the USA <https://www.wyden.senate.gov/download/algorithmic-accountability-act-of-2022-bill-text> accessed 17/05/2022; Canada Digital Charter Implementation Act 2022, which incluces a 'Artificial Intelligence and Data Act' <https://www.parl.ca/DocumentViewer/en/44-1/bill/C-27/first-reading> accessed 20/06/2022; American Data Privacy and Protection Act, which proposes conducting algorithmic impact assessments in Section 207 <https://www.congress.gov/117/bills/hr8152/BILLS-117hr8152ih.pdf> accessed 21/08/2022.

[785] Art. 9 AIA draft.

these two AI operators, it was suggested that the AIA draft should incorporate a clear obligation for AI users to perform some sort of algorithmic impact assessment.[786]

In the foregoing, it has been explained not only the most important academic and policy developments to address the issues concerning the use of AI systems to process personal data, but also some of the legislative proposals to promote the responsible use of AI systems and provide reassurance to individuals and society. Nonetheless, more work needs to be done. New legislative initiatives should establish binding requirements regulating the requirements that operators of AI systems should follow and the specific procedures to demonstrate compliance with them, such as impact assessments, assurance, certifications, or audits. The following chapter explains how other governance and accountability mechanisms can help to identify, assess and mitigate the risks associated with the use of AI systems when processing personal data.

[786] Martin Ebers and others, 'The European Commission's Proposal for an Artificial Intelligence Act - Critical Assessment by Members of the Robotics and AI Law Society (RAILS)' (2021) 4 J 589, 597.

Chapter V

GOVERNANCE MECHANISMS TO FURTHER MITIGATE THE RISKS POSED BY AI SYSTEMS

Introduction

More extensive and clearer transparency and fairness obligations constitute a necessary but not sufficient condition to mitigate the risks posed by the processing of personal data using AI systems. Enhanced accountability obligations will pave the way for better development and use of AI systems[787] and reduce the impacts on fundamental rights, in particular those concerning the protection of personal information and the private life of individuals.

Apart from the strategies suggested above to tackle specific risks posed by AI systems, there is a wide range of governance mechanisms to consider to further reduce them. In this chapter, some of the most important governance and accountability mechanisms to enhance the level of protection of fundamental rights are evaluated. First, public registers of AI systems are proposed, as a way to enhance the transparency of these solutions towards individuals and society. Second, it is proposed that a specialized person or organization takes the role of AI ethical officer to operationalize AI-related corporate values and guarantee a trustworthy development and use of AI across the organization. Third, the process of standardisation and certification of AI systems is discussed, as a way to strengthen the protection of individuals and establish a common set of rules to protect personal data. Fourth, it evaluates the codes of conduct as a method to set suitable compliance and ethical rules for institutions working in a specific sector. Fifth, the role of supervisory authorities is assessed, how their powers could be interpreted and whether there are mechanisms that they can use to monitor the compliance of data protection obligations more effectively. Sixth, some particular measures to

[787] Rebecca Slaughter, 'Algorithms and Economic Justice: A Taxonomy of Harms and a Path Forward for the Federal Trade Commission' (2021) Special Pu Yale Journal of Law & Technology 1, 51.

operationalize the principle of privacy by design, in particular those concerning the reduction of the identifiability of personal data.

V.1.- Register of AI systems or AI providers

Transparency is a core aspect of the development and use of AI systems. A way to enhance the transparency of AI systems is through the creation of a register of AI systems or AI providers. Such a register is a public repository of AI systems or AI providers, whose main features can be scrutinised by users of AI systems or the public at large. Transparency and accountability are not the only principles that can be improved through the registration of AI systems. Transparency is fundamental to building people's trust and these initiatives can also support social awareness about the employment of AI systems, enhancing citizens' literacy in AI and digital matters and triggering public debate and social participation. This is particularly the case when AI systems are employed by governmental agencies for granting or denying benefits for citizens.

There are currently some registers of AI systems used in the public sector. The cities of Amsterdam,[788] Helsinki,[789] Nantes,[790] Antibes,[791] and New York[792] have

---

[788] Gemeente Amsterdam, City of Amsterdam Algorithmic Register Beta, <https://algoritmeregister.amsterdam.nl/en/ai-register/> accessed on 27/01/2022. This register included 3 AI systems: an AI system to for automated parking control, another to report issues in public, and finally another for illegal holiday housing rental.

[789] City of Helsinki, City of Helsinki AI Register, <https://ai.hel.fi/en/ai-register/> accessed on 27/01/2022. This register includes 5 AI systems: three chatbots (for a health centers, maternity clinics and parking), and two recommendation systems for the public libraries.

[790] City of Nantes, Nantes Metropole Open Data, <https://data.nantesmetropole.fr/pages/algorithmes_nantes_metropole/> accessed on 27/01/2022. Two AI systems concerning social tariffication (for public transportation and water and sewages services) are included in this register.

[791] City of Antibes, Access to Administrative Documentation, <https://www.antibes-juanlespins.com/administration/acces-aux-documents-administratifs> accessed on 27/01/2022. The inventory of the algorithms employed by the municipality <https://en.calameo.com/read/002074504242548f87596> accessed on 27/01/2022.

[792] New York City, Algorithms Management and Policy Officer, <https://www1.nyc.gov/assets/ampo/downloads/pdf/AMPO-CY-2020-Agency-Compliance-Reporting.pdf> accessed on 27/01/2022.

already implemented a register of some AI systems. The registers of the cities of Amsterdam and Helsinki provide a wide range of information about listed algorithms. They share information about the datasets employed to train the model, the data processing activities (i.e. information about the logic and reasoning of the system), information on whether there is a risk of discrimination, human oversight measures, and risk management measures.

In France, the Code des Relations entre le Public et l'Administration,[793] requires public administrations to fulfil certain transparency obligations if they use algorithmic processing (including automated processing and decision support tools), and the systems are employed to make individual decisions concerning natural or legal persons.[794] The ex-ante transparency obligations consist of: a) providing a general notice: they must make available online the rules defining the main processing operations used in the performance of their missions when they are used to base individual decisions (Article L.312-1-3); b) including an explicit mention: they must publish online and in other documents (notices, notifications) the following information: the administration responsible for the decision, the purpose of the processing, a reminder of the right to obtain communication of the rules defining this processing and the main characteristics of its implementation, the procedures for exercising this right. (Article L.311-3-1). While the personal scope of application is limited (it only applies to public administrations and private parties fulfilling missions in the public interest), this mandatory information could be gathered and centralised in a public register.

Not only did public bodies create registers of AI systems, but also some initiatives have emerged in the private sector concerning the registration of AI providers and AI systems.[795] Lloyd's Register opened the first AI register to keep a record of the company's certified AI systems and providers of AI systems for the maritime sector.

---

[793] Code des Relations entre le Public et l'Administration (CRPA) <https://www.legifrance.gouv.fr/codes/texte_lc/LEGITEXT000031366350/2022-01-27/> accessed 27/01/2022.

[794] Art. L.300-2 and L311-3-1 Code des relations entre le public et l'administration, https://www.legifrance.gouv.fr/codes/article_lc/LEGIARTI000033205535/

[795] Llyod's Register, Lloyd's Register launches industry-first Artificial Intelligence Register, 22nd Nov 2021, in < https://www.lr.org/en/latest-news/lloyds-register-launches-industry-first-artificial-intelligence-register/> accessed on 10/02/2022.

Finally, the AIA establishes the creation of an EU database for high-risk AI systems[796] listed in AIA Annex III. This database must contain certain information and the information will be publicly available. AI providers must include in the register[797] the contact details of the AI provider and, if applicable, the EU representative, AI system name and identification, its intended purpose, EU countries where the system is operating, and instructions for use. But the most comprehensive pieces of information that AI providers must provide are a copy of the certificate issued by the notified body and a copy of the EU declaration of conformity. According to AIA Annex V, the EU declaration of conformity must include, among other information, the AI system name and type, a statement that the AI system complies with the AIA, and a statement of the relevant standards or specifications followed.

While the provisions contained in Art. 60 AIA draft are welcomed they are not sufficient to address the lack of transparency in AI systems. First, the information contained in the proposed register does not allow individuals to scrutinise, at least minimally, certain features of the system. The inclusion of the EU declaration of conformity does not solve the opacity problem. Suffice it to compare the information included in the registers built by Amsterdam and Helsinki with the required information of the central register created by the AIA draft. In line with the approaches taken by these cities, some organisations are calling for the mandatory registration of all AI systems used in the public sector or by public authorities.[798] To increase algorithmic transparency, not only a broader range of AI systems should be registered, but also more information should be included as mandatory and freely accessible to the public.

V.2.- AI Ethical Officer to overcome limitations from DPOs

The GDPR requires controllers carrying out certain processing operations to appoint a data protection officer as an accountability measure. While the appointment of this professional is welcomed, it may be necessary to designate a professional with particular expertise in the field to assist data AI providers and AI users. This role can be fulfilled by the AI Ethical Officer.

---

[796] Art. 60 AIA draft.

[797] Annex VIII AIA draft.

[798] Algorithm Watch, "Our response to the European Commission's consultation on AI" (2020) https://algorithmwatch.org/en/response-european-commission-ai-consultation accessed 09/04/2022.

An AI Ethical Officer can boost compliance, promote ethical awareness and oversee AI ethics within organisations developing or using AI systems. The AI Ethical Officer is responsible for offering sound advice on ethical AI practice while guarding the organization against bias and ensuring accountability.[799] The professional or group of professionals entrusted with this role should combine technical know-how and strong awareness of the ethical and human rights issues surrounding the development and deployment of AI systems. An AI Ethical Officer may help build the ethical framework or guidelines that will rule the way the organisation employs the AI systems. They can assist in the definition of AI ethics goals and advise the organisation on how to achieve them.

The AI Ethical Officer can have an advisory function. In fact, it may communicate and advise entities developing and using AI systems on the most convenient standards to adopt according to their AI systems and the intended uses, as well as on their obligations and duties under the applicable regulatory and ethical frameworks and standards. Additionally, they may provide advice and support on impact assessments required or recommended for a trustworthy use of AI systems, such as ethical impact assessment, Algorithmic impact assessment, Human rights impact assessment, data protection impact assessment, stakeholder impact assessment, and responsible innovation impact assessment.[800] They can also provide assistance when carrying out AI-related audits.

Monitoring compliance is another potential function that can be attributed to AI Ethical Officers. They may supervise compliance with existing legal frameworks that regulate the development and deployment of AI systems, as well as advise on the application of guidelines, codes of conduct and ethical frameworks, and on the

[799] UNESCO (n 440) para 58.

[800] Adele Tharani and others, 'The COMPASS Self-Check Tool. Enhancing Organizational Learning for Responsible Innovation through Self-Assessment' in Emad Yaghmaei and Ibo van de Poel (eds), *Assessment of Responsible Innovation. Methods and Practices* (Routledge 2020); Andrea Porcari and Elena Mocchio, 'Managing Social Impacts and Ethical Issues of Research and Innovation: The CEN/WS 105 Guidelines to Innovate Responsibly' in Emad Yaghmaei and Ibo van de Poel (eds), *Assessment of Responsible Innovation. Methods and Practices* (Rutledge 2020).

preparation for the entry into force of certain legal frameworks that regulate the design and use of AI systems.[801]

The AI Ethical Officer can manage the organization's communications with stakeholders. They may handle communications with supervisory authorities and cooperate at the latter's request, and can assist when the organisation needs to communicate with individuals interested or affected by the decisions taken by the AI system. Finally, the AI Ethical Officer can help develop an AI ethics culture in the organisation, raising awareness of the staff and preparing training activities for those involved in any of the stages of the lifecycle of AI systems.

Similar to the position entrusted to DPOs, the AI Ethical Officer is a single point of responsibility for overseeing compliance with and establishing a culture of responsible use of AI systems, who can aid in building trust in how the entity develops and uses the AI systems.[802] To date, many large companies have appointed persons to fill positions with similar tasks, albeit there is no uniformity in the title or functions of such a role.[803] Most of them belong to the C-suite[804] range and it may be burdensome for smaller companies or start-ups to hire another internal full-time C-suite position. Hence, the option could be an external professional or group of professionals to satisfy this need, just as in the case of a data protection officer. Outsourcing the AI Ethical

---

[801] Such as the AI Regulation draft.

[802] Mark Minevich and Francesca Rossi, Why you should hire a chief AI ethics officer, World Economic Forum Agenda <https://www.weforum.org/agenda/2021/09/artificial-intelligence-ethics-new-jobs/> accessed on 03/01/2022.

[803] Different organisations have called this position differently. For example, Chief AI Ethics Officer (US Army's AI Task Force), Global AI Ethicist (DataRobot), Chief Ethics Officer (Hypergiant), AI Ethics Global Leader (IBM), Chief Ethical and Humane Use Officer (Salesforce), Chief AI Ethics Officer (BCG), Chief AI Ethics Advisor (Paravison), Chief Responsible AI (US Department of Defense, Joint AI Center), Head of Responsible AI & Data (H&M Group), Chief Responsible AI Officer (Microsoft), Responsible AI Leader (PriceWaterCoopers), Responsible AI/Machine Learning, Acting Head of ML Strategy (BBC), Lead for Responsible AI (Accenture), AI/Tech Ethics Lead (Deloitte). See <https://www.forbes.com/sites/markminevich/2021/08/09/15-ai-ethics-leaders-showing-the-world-the-way-of-the-future/> accessed on 03/01/2022. Other common denominations include: Data Ethics Consultant, ML Ethicist, Data Ethics Lead, AI Policy Coordinator, AI Ethics Manager, AI Ethicist, and AI Governance Manager. See <https://analyticsindiamag.com/tech-firms-are-racing-to-hire-ai-ethicists/> accessed on 02/03/2022.

[804] 'C-suite' is referred to the executive-level managers in an organisation.

Officer may help reduce costs and is flexible enough to allow small companies or start-ups to obtain qualified advice without increasing the payroll.

It is worth noticing that the AIA draft does not require the appointment of an AI Ethical Officer. However, further revisions of the AIA draft may include the mandatory designation or appointment of an AI Ethical Officer for providers and users of high-risk AI systems, and a voluntary scheme for those developing or using low or minimum-risk AI systems, as a method for continuous supervision of AI systems.[805]

V.3.- Standardisation of AI systems. Industry standards as a method to fill legislative gaps

Since the current legal framework governing AI does not seem sufficient to appropriately tackle the risks and other deleterious effects of AI systems,[806] standards may seem an alternative to achieve this aim. Standards are documents prepared by experts which contain rules, recommendations or requirements for products or processes to achieve the highest degree of order in a particular field.[807] They are intended to explain the most convenient methods to carry out a particular process or create a product.

Whereas the drafting of standards for AI systems is still in its infancy, in the last couple of years the number of initiatives has skyrocketed. To date, there are many standards published or under development by international or national certification organisations that can match the core requirements for trustworthy AI systems as detailed in the AI Regulation draft. There are standards covering the AIA requirements

[805] European Commission for the Efficiency of Justice, 'Possible Introduction of a Mechanism for Certifying Artificial Intelligence Tools and Services in the Sphere of Justice and the Judiciary: Feasibility Study' (2020) 7.

[806] Martin Ebers, 'Standardizing AI - The Case of the European Commission's Proposal for an Artificial Intelligence Act' in Larry Di Matteo, Cristina Poncibò and Michel Cannarsa (eds), *The Cambridge Handbook of Artificial Intelligence: Global Perspectives on Law and Ethics* (CUP) 5.

[807] The definition and the differences between International Standard (IS), Technical Specification (TS), Technical Report (TR) can be consulted here: https://www.iso.org/deliverables-all.html accessed 08/01/2022.

for data and data governance,[808] technical documentation,[809] record-keeping,[810] transparency and provision of information to users of AI systems,[811] human oversight,[812] accuracy, robustness, and cybersecurity,[813] risk management system,[814] and quality management system.[815] There are also standards related to the ethical or

[808] Art. 10 AIA draft establishes that high-risk AI systems must be developed based on training, validation, and testing datasets that meet a set of quality criteria. The relevant standards that could be used for data governance and to evaluate the data quality are: ISO/IEC TR 24027:2021, ISO/IEC TR 24029-1:2021, ISO/IEC 38507:2022, ETSI SAI 005 and ETSI SAI 002 (published) and ISO/IEC TS 4213, ISO/IEC 5259-2, ISO/IEC 5259-3, ISO/IEC 5259-4, ISO/IEC 5338, ISO/IEC 5469, ISO/IEC 23894.2, ISO/IEC 24668, ISO/IEC 42001 (under development).

[809] Art. 11 AIA draft requires providers of AI systems to draw up technical documentation before the AI system is placed on the market. The relevant standards that could be used for this purpose are: ISO/IEC TR 24027:2021 (published) and ISO/IEC 23894.2, ISO/IEC 42001 (under development).

[810] Art. 12 AIA draft establishes that high-risk AI systems must enable the automatic recording of events or logs. The relevant standard for record keeping is ISO/IEC 23894.2 (under development).

[811] Art. 13 AIA draft establishes that high-risk AI systems must guarantee that their operation is transparent to enable users of AI systems to interpret the system's output and use it appropriately. Relevant standards for this purpose are ISO/IEC TR 24027:2021, ISO/IEC TR 24028:2020, ISO/IEC 38507:2022, IEEE 7001-2021 Standard for Transparency of Autonomous Systems, UK Central Digital and Data Office 2021 Algorithmic Transparency Standard <https://www.gov.uk/government/collections/algorithmic-transparency-standard> accessed 07/01/2022 (published), and, ISO/IEC 23894.2, ISO/IEC 42001 (under development).

[812] Art. 14 AIA draft establishes that high-risk AI systems must supervised by humans while they are in use. Relevant standards for human oversight are: ISO/IEC 38507:2022 (published) and ISO/IEC 23894.2, ISO/IEC 42001 (under development).

[813] Art. 15 AIA draft requires high-risk AI systems to achieve adequate levels of accuracy, robustness and cybersecurity. Relevant standards for to achieve accuracy, robustness and cybersecurity are: ISO/IEC TR 24029-1:2021, ETSI SAI 002, ETSI SAI 003, ETSI SAI 005, ETSI SAI 006, DIN SPEC 92001-2 (published) and ISO/IEC TS 4213, ISO/IEC 5338, ISO/IEC 5469, ISO/IEC 23894.2, ISO/IEC 24668, ISO/IEC 42001 (under development).

[814] Art. 9 AIA draft requires providers of AI systems to implement, document and maintain a risk management system for high-risk AI systems. Relevant standards to achieve this objective are: ISO/IEC 38507:2022, IEEE 7010-2020 - Recommended Practice for Assessing the Impact of Autonomous and Intelligent Systems on Human Well-Being (published) and ISO/IEC 5338, ISO/IEC 5469, ISO/IEC 23894.2, ISO/IEC 42001 (under development).

[815] Art. 17 AIA draft requires providers of high-risk AI systems to implement a quality management system to comply with the AIA. Relevant standards for this purpose are: ISO/IEC TR 24029-1:2021 and

responsible development and use of AI systems.[816] This list illustrates that there are several relevant standards to develop and deploy AI systems. Whereas many of them are still under development, in particular many ISO/IEC standards, others were already published and are ready to be implemented.

The benefits of relying on standards to fill the gaps left by the AI regulatory framework are clear. First, the process of developing, drafting and publishing standards are faster than binding legislation. Standards can quickly transfer technology from research to industry.[817] They specify important conditions for the adequate development and use of AI systems, in particular concerning data quality, robustness, transparency, fairness, and cybersecurity. All these topics would require extensive debate in legislatures, but standardisation bodies can quickly accomplish these objectives. So these documents can address the most significant problems identified in the development and use of AI. Second, standards implement in greater detail legal provisions and clearly delineate some of the obligations of providers and users of AI systems. This is because it is very challenging for laws and statutes to describe with the required level of detail the obligations and requirements to comply with certain provisions or concerning how to achieve appropriate levels of functioning of the AI systems. Third, standards are developed by experts in the fields, which ensure a deeper knowledge of the subject matter and a closer contact with the current problems AI systems may create. Fourth, many of the most well-known standard-setting institutions are independent organisations and they are supported by long-standing expertise and reputation in the field. Among these organisations are the

DIN SPEC 92001-1:2019-04 (published), ISO/IEC 5259-3, ISO/IEC 5259-4, ISO/IEC 5338, ISO/IEC 23894.2, ISO/IEC 38507, ISO/IEC 42001 (under development).

[816] National Standard of Canada, CAN/CIOSC 101:2019, Ethical design and use of automated decision systems, IEEE 7007-2021 - Ontological Standard for Ethically Driven Robotics and Automation Systems and IEEE 7000-2021 - Standard Model Process for Addressing Ethical Concerns during System Design (published).

[817] German Institute of Standarisation, 'German Standardization Roadmap on Artificial Intelligence' (2020) 4.

International Standards Organisation, the European Standardization Organizations,[818] and the German Institute of Standardisation,[819] among others.

Finally, the AIA draft relies heavily on the use of standards to demonstrate compliance with the requirements established therein. Standardisation will play an important role in providing technical solutions to providers of AI systems to guarantee compliance with the AIA.[820] This is chiefly because there is a presumption of conformity[821] with the AIA if the high-risk AI systems satisfy harmonised standards.[822] Hence, one of the pillars of the AIA to address high-risk AI systems is the successful publication of harmonized standards developed by the three EU standardisation organisations[823] (i.e. European Committee for Standardization (CEN), European Committee for Electrotechnical Standardization (CENELEC) and European Telecommunications Standards Institute (ETSI)). Additionally, the AIA draft acknowledges that standardisation may be employed to foster fundamental rights[824] and it calls for the development of technical standards addressed to high-risk AI systems which should be compatible with the Charter of Fundamental Rights.[825]

However, there are still some problems concerning the reliance on standards to regulate the details of the development and use of AI systems. To begin with, while standardisation organisations are very active in developing new standards and some standards could be currently implemented, several years are still needed to achieve full standardisation of the requirements established in the AIA. In particular, the development of standards was uneven across the different sub-requirements of the

---

[818] The European Standardization Organizations include the European Committee for Standardization (CEN), the European Electrotechnical Committee for Standardization (CENELEC), and the European Telecommunications Standards Institute (ETSI).

[819] Deutsches Institut für Normung e. V. (DIN).

[820] Rec. 61 AIA draft.

[821] Art. 41 AIA draft.

[822] Harmonised standards are EU standards implemented based on a request made by the Commission for the application of Union harmonisation legislation (Art. 3(27) AIA and Art. 2(1)(c) of Regulation (EU) No 1025/2012).

[823] Ebers (n 803) 14.

[824] Mark McFadden and others, 'Harmonising Artificial Intelligence: The Role of Standards in the EU AI Regulation' (2021) 19.

[825] AIA draft Recital 13.

AIA and more work should be done in some areas.[826] Additionally, AI systems are evaluated for a clearly defined set of uses, and the standards mirror the requirements for these specific uses. However, many times AI systems are employed in manners not totally foreseen or expected by the developers, thus reducing the effectiveness of the standards. Furthermore, new AI systems are introduced routinely and AI systems or their features are modified at a very fast pace. The astonishing speed with which such changes operate may obsolete standards in a very short period, thus requiring continuous amendments or updates of standards. This is particularly problematic when it comes to developing harmonised technical standards to be cited in the Official Journal of the EU (which provides a presumption of conformity with EU legislation) because after CJEU's decision on *James Elliot*[827] longer drafting and publication timeframes are expected. Finally, it is debatable the extent to which standardisation bodies can carry out regulatory functions. Standardisation organisations inherently lack democratic accountability. Whereas the rulemaking function is always held by States within the limits of their jurisdiction, the implicit delegation the AIA draft grants to standardisation bodies to fill the gap in the legislation can bring problems related to accountability and democratic deficit of these institutions.

V.4.- Certification of AI systems

The GDPR establishes that certification is another optional mechanism to assist controllers and processors to demonstrate compliance with the regulation.[828] But demonstrating compliance, while a crucially important objective of these provisions, is not the only aim of certification mechanisms. Certification in the GDPR also serves to enhance transparency and compliance as well as to allow individuals to quickly evaluate the data protection level of products and services placed on the market.[829]

---

[826] European Commission - Joint Research Centre, 'AI Watch: AI Standardisation Landscape State of Play and Link to the EC Proposal for an AI Regulatory Framework' (2021) 54. The areas in which more standardisation efforts should be committed are: Data and data governance, Technical documentation, and Risk management system.

[827] Case C-613/14, *James Elliott Construction Limited v Irish Asphalt Limited*. [2016] ECLI:EU:C:2016:821.

[828] Art. 42(1) and 42(3) GDPR.

[829] Recital 100 GDPR.

A certification is an attestation or declaration made by independent third parties that refers to products, processes and services.[830] The products, processes and services to certify in this context are related to the processing operations of controllers and processors. The independent third parties conducting the certification issue a certificate, which is a statement of conformity with the relevant requirements. Additionally, the GDPR regulates the use of seals and marks. These are graphical representations (e.g. a logo) and their inclusion in a certified product, process or service signifies that the latter has successfully undergone a certification procedure and that they comply with the requirements of the specific certification method.[831] Certifications under GDPR can be issued for a maximum, renewable, period of 3 years and should be withdrawn where the certified organisation no longer comply with the requirements.[832]

It is important to highlight that certification mechanisms do not demonstrate by themselves that the processing operations are carried out in line with the GDPR since they do not lessen the responsibility of certified organisations to comply with the GDPR.[833] Instead, certifications are simply additional factors on which controllers and processors may rely to prove the required compliance.[834] Yet, there are many benefits and incentives for companies to obtain a certification. To begin with, it is a tool that can help to improve the organisations' image toward customers. Organisations can leverage this competitive advantage over other companies since it enhances customers' trust and creates a positive image of certified organisations.[835] Additionally, certifications can reduce compliance risks. While certifications do not reduce the responsibility of controllers or processors for compliance, they considerably reduce the chances of being subject to enforcement actions by data protection authorities or,

---

[830] International Standard Organisation, 'ISO/IEC 17000:2020(En) Conformity Assessment — Vocabulary and General Principles' (2020) s 7.6.

[831] European Data Protection Board, 'Guidelines 1/2018 on Certification and Identifying Certification Criteria in Accordance with Articles 42 and 43 of the Regulation' (2018) 7.

[832] Art. 42(7) GDPR.

[833] Art. 42(4) GDPR.

[834] European Data Protection Board, 'Guidelines 1/2018 on Certification and Identifying Certification Criteria in Accordance with Articles 42 and 43 of the Regulation' (n 829) 7.

[835] Voigt and von dem Bussche (n 305) 77.

if found non-compliant, the amount of the fines imposed. This is because certifications constitute a clear commitment to comply with regulations and evidence of the technical and organisational measures put in place to achieve the maximum level of compliance.[836] Finally, it eases the relationships with vendors, since certified organisations offer a high assurance of compliance, which reduces the need to engage in time-consuming and costly privacy audits before entering into a commercial relationship between them.

Whereas there is no express mention of processing personal data using AI systems in these provisions, the certification mechanism can help to increase the levels of compliance with regulations and foster a culture of responsible use of AI. Adherence to approved certification mechanisms[837] is an element to demonstrate compliance with the mandatory requirements for controllers,[838] the obligations under the principle of data protection by design,[839] the guarantee requirements before engaging data processors,[840] the security measures needed to process personal data,[841] and to transfer data to countries where no adequacy decision has been issued.[842] Finally, adherence to certifications is an element that supervisory authorities should consider when evaluating the imposition of administrative fines for non-compliance with the regulation.[843]

Many EU institutions have supported initiatives to develop certification mechanisms addressed to AI systems, that range from the creation of a general certification scheme for trustworthy AI systems[844] to the requirement for AI providers selling applications to

---

[836] Giovanni Maria Riccio and Federica Pezza, 'Certifications Mechanism and Liability Rules under the GDPR. When the Harmonisation Becomes Unification' in Alberto De Franceschi, Reiner Schulze and Oreste Pollicino (eds), *Digital Revolution - New Challenges for Law* (Beck 2019) 150.

[837] These also applies to the adherence to code of conduct under Art. 40 GDPR.

[838] Art. 24(3) GDPR.

[839] Art. 25(3) GDPR.

[840] Art. 28(5) GDPR.

[841] Art. 32(3) GDPR.

[842] Art. 46(2)(f) GDPR.

[843] Art. 82(2)(j) GDPR.

[844] European Economic and Social Committee, The EESC proposes introducing EU certification for "trusted AI" products (14/11/2019) <https://www.eesc.europa.eu/en/news-media/news/eesc-proposes-introducing-eu-certification-trusted-ai-products> accessed 08/06/2022.

public authorities to obtain a 'data hygiene certificate'.[845] However, no common certification mechanisms have been agreed upon so far.

In the context of AI systems, providers or users of AI systems may be able to certify the AI solution as a product (software) either by indicating the required specifications for the design or the processes of its design (e.g. the certification scheme issued by the French Laboratoire National de Métrologie et d'Essais (LNE) which covers the design, development, evaluation and maintenance of AI systems in operational conditions)[846] or by requiring certain performance or accuracy conditions so that the outputs of the AI system can be supervised and assessed. Additionally, AI systems may be certified in the context of their deployment, considering training datasets, input data, outputs and the contextual or sectoral regulations or good practices applicable to the system.

Certification schemes are perfectly suitable to improve accountability on algorithms. While output-based certification would be the best solution to improve the algorithmic trustworthiness in terms of fairness and discrimination, as previously highlighted, finding common grounds under which to evaluate the fairness of the algorithmic results is not always a straightforward task. Hence, process-based certifications (such as the certification scheme developed by the French LNE) can represent a strategic solution to the difficulties in reaching an agreement on the definition of fairness and the relevant metrics to measure discriminatory outputs, which are exacerbated by the wide disparities concerning the different sectors of applications of AI systems.

V.5.- Codes of Conduct for AI operators

Another governance tool that can be used to better protect the rights of individuals concerning the use of AI systems is the code of conduct. Codes of conduct are optional or non-mandatory accountability mechanisms that establish the most adequate compliance and ethical rules for organisations operating in a certain domain or sector.[847]

---

[845] European Parliamentary Research Service, 'Artificial Intelligence: From Ethics to Policy' (2020) 26.

[846] Laboratoire National de Métrologie et d'Essais, 'Certification Standard of Processes for AI. Design, Development, Evaluation and Maintenance in Operational Conditions' (2021).

[847] European Data Protection Board, 'Guidelines 1/2019 on Codes of Conduct and Monitoring Bodies under Regulation 2016/679' (2019) 7.

While the AIA draft touches upon the drafting of codes of conduct for AI systems, it leaves this area vastly unregulated. The AIA draft establishes that the Commission and the Member States should support the drafting of codes of conduct for the voluntary application of the requirements for high-risk AI systems[848] to limited-risk and low-risk AI systems, adapting the requirements where appropriate.[849] Every four years the Commission will evaluate the effectiveness of codes of conduct drafted for these purposes.[850]. Additionally, codes of conduct may encourage the voluntary application to any AI system of requirements concerning environmental standards, accessibility for handicapped people, stakeholder involvement in all stages of the AI lifecycle and diversity in teams engaged in the design of AI systems.[851] It is worth noticing that, contrary to the GDPR,[852] the AIA draft allows drawing codes of conduct to individual providers of AI systems.[853]

But even though the AIA draft is relatively silent on this matter, the GDPR can provide a suitable legal basis for the elaboration of codes of conduct where the AI systems process personal data, either as a principal or ancillary activity. These tools can support the application of the General Data Protection Regulation, considering the nuances of particular data processing operations or industry domains, including the needs of SMEs. Additionally, codes of conduct give wide autonomy to controllers to agree on best practices for their sectors and they constitute pragmatic solutions to issues detected in their domains.[854] This is particularly relevant in the field of AI, where the processing operations involving personal data carried out with AI systems show special characteristics that should be evaluated by organisations with know-how in the field.

---

[848] The requirements for HRIAS are listed in art. 8 to 15 AIA and include risk management system, data governance, human oversight, transparency, accuracy, robustness, etc.

[849] Art. 69(1) AIA draft.

[850] Art. 84(4) AIA draft.

[851] Art. 69(2) AIA draft.

[852] Art. 40(2) GDPR only allows associations and other bodies representing categories of controllers or processors to prepare codes of conduct, but not individual controllers or processors.

[853] Art. 69(3) AIA draft.

[854] European Data Protection Board, 'Guidelines 1/2019 on Codes of Conduct and Monitoring Bodies under Regulation 2016/679' (n 845) 9.

A final advantage for organisations that abide by approved codes of conduct relates to the concept of accountability and how they prove compliance with the Regulation,[855] particularly since adherence to these codes eases the requirements in terms of security[856] and the evaluation of the impact of the processing activities[857] and, finally, this fact should be considered by national data protection authorities when assessing the imposition and amount of an administrative fine.[858]

Codes of conduct are not intended to repeat the obligations already stated in the GDPR. Instead, they must contain clauses concerning how they concretely fulfil specific needs of the industry domain or processing operation and regarding the domain-specific application of the GDPR.[859] These clauses may concern risk detection, evaluation (probability and severity) and the measures to avoid the risk.[860]

Finally, codes of conduct must include adequate safeguards to reduce the risks posed by the data processing operations,[861] which should be more taxing or more specific than those already included in the GDPR, as well as effective procedures to monitor and enforce the clauses established therein by accredited monitoring bodies.[862] For instance, the EU Cloud Code of Conduct, which in 2020 received a favourable opinion from the EDPB and was approved by the Belgian Data Protection Authority in May 2021, establishes thirteen security objectives which were drafted based on internationally recognised standards in information security such as ISO 27001.[863] In other words, it specifies which concrete security requirements are needed to demonstrate compliance with the regulatory framework.

---

[855] Art. 24(3) and 28(5) GDPR.

[856] Art. 32(3) GDPR.

[857] Art. 35(8) GDPR.

[858] Art. 83(2)(j) GDPR. See also European Data Protection Board, 'Guidelines on the Application and Setting of Administrative Fines for the Purposes of the Regulation 2016/679' (2017) 15.

[859] European Data Protection Board, 'Guidelines 1/2019 on Codes of Conduct and Monitoring Bodies under Regulation 2016/679' (n 845) 14–15.

[860] Rec. 77 GDPR.

[861] Art. 40(5) GDPR.

[862] Art. 40(4) and 41(1) GDPR.

[863] Scope Europe, 'EU Cloud Code of Conduct' (2021) 18. Nowadays, a large portion of the data processing operations using AI systems are conducted relying on cloud service providers, like Google Cloud, Amazon Sagemaker (Amazon Web Services), Microsoft Azure.

So far, most of the so-called 'codes of conduct' for AI systems are mostly a catalogue of good practices, but there are some prominent examples, such as the Code of Conduct for AI systems used by the NHS,[864] Technology Code of Conduct by the World Bank,[865] and Code of Conduct on Artificial Intelligence in Military Systems,[866] that should be taken into consideration for further development of more specific codes of conduct to demonstrate compliance with a future regulation on AI systems.

V.6.- Empowerment of Supervisory Authorities

V.6.1.- Supervisory authorities evaluating AI systems

The monitoring of the data protection provisions is a crucial aspect of the data protection regime. National public authorities should have corrective powers to supervise the effective application of the law and act accordingly. If law-breaking behaviours were not followed by a sanction, laws would not have teeth and hence they could be violated without any consequence.[867] The GDPR provides specific powers to national supervisory authorities which go beyond monitoring the GDPR and privacy regulations at the national level, and also include investigative powers (e.g. requiring information, carrying out data protection audits),[868] advisory powers (e.g. issuing opinions, advising controllers on DPIAs)[869] and a whole suite of corrective powers.[870] The corrective powers are the most important powers that supervisory authorities have to force controllers and processors to comply with the data protection regulations. Data protection authorities can issue warnings or reprimands to controllers or processors

---

[864] UK Department of Health & Social Care, New code of conduct for artificial intelligence (AI) systems used by the NHS <https://www.gov.uk/government/news/new-code-of-conduct-for-artificial-intelligence-ai-systems-used-by-the-nhs> accessed on 18/04/2022; Liesbeth Venema, 'Code of Conduct for Using AI in Healthcare' (2019) 1 Nature Machine Intelligence 265.

[865] World Bank Group, 'IFC Technology Code of Conduct — Progression Matrix — Public Draft' (2020).

[866] Center for Humanitarian Dialogue, 'Code of Conduct on Artificial Intelligence in Military Systems' (2021).

[867] Bernard Schwartz, *Administrative Law* (3rd edn, Little Brown & Co 1991) 91.

[868] Art. 58(1) GDPR.

[869] Art. 58(3) GDPR.

[870] Art. 58(2) GDPR.

as well as order them to comply with the data subject's requests, to bring processing operations into compliance, or to suspend international transfers of data. Additionally, they can impose a ban on processing and fine data controllers or processors.

The supervisory powers of data protection authorities are not limited by the means through which data processing operations are carried out. National data protection authorities currently monitor data processing activities performed using AI systems, and they are well-positioned to address the challenges posed by AI technologies on fundamental rights, in particular to data protection and privacy.[871] So far, national supervisory authorities have taken an active role in monitoring the application of data protection provisions where the processing is carried out using AI systems. Where controllers or processors have violated the applicable legal regime, supervisory authorities have chiefly ordered the suspension and the ban of processing activities as well as the imposition of administrative fines. For instance, a company called Clearview AI scrapped social media networks to collect images of human faces from social networks for the development and deployment of a facial recognition algorithm. After investigations, the company was fined and was ordered to suspend the data processing activities by the data protection supervisory authorities of France,[872] Italy,[873] the United Kingdom,[874] and Greece.[875] It was also ordered to comply with the erasure request by the Hamburg data protection authority,[876] as well as to discontinue the unlawful processing operations and delete the data by provincial data protection authorities of Canada.[877]

---

[871] European Data Protection Board and European Data Protection Supervisor (n 199) 13.

[872] Commission Nationale de l'Informatique et des Libertés, *Clearview AI*. [2021].

[873] Garante per la Protezione dei Dati Personali, *Ordinanza ingiunzione nei confronti di Clearview AI*. [2022].

[874] Information Commissioner's Office, *Clearview AI Inc*. [2022] This was a joint investigation carried out with the Office of the Australian Information Commissioner (OAIC)

[875] European Data Protection Board, Hellenic DPA fines Clearview AI 20 million euros (20 July 2022) <https://edpb.europa.eu/news/national-news/2022/hellenic-dpa-fines-clearview-ai-20-million-euros_en> accessed 21/08/2022.

[876] Hamburgische Beauftragte für Datenschutz und Informationsfreiheit, *Clearview AI Inc*. [2020]

[877] Joint investigation of Clearview AI Inc. by the Office of the Privacy Commissioner of Canada, the Commission d'accès à l'information du Québec, the Information and Privacy Commissioner for British

There is, however, a remedy they have not already employed that could be effective to tackle not only non-compliance with the regulations but also discouraging the taking of a purely economic analysis of the consequences of the behaviours: algorithmic disgorgement.

VI.6.2.- Algorithmic disgorgement. Destroying AI systems that used ill-gotten or tainted data for training

One of the most powerful tools that supervisory authorities have is algorithmic disgorgement. Disgorgement implies that the organisation profiting from unlawful actions relinquishes any advantage or profit realised as a consequence of the unlawful behaviour. Algorithmic disgorgement means that companies must delete the algorithm where the algorithm was developed using data illegally collected or processed.

The rationale behind this tool is to disincentivise the commissioning of wrongful acts by blocking unfair enrichment. Several long-established legal doctrines inspire the idea of algorithmic disgorgement, like unjust enrichment and traditional disgorgement (from contract law) and the fruit of the poisonous tree (from criminal law).[878] But algorithmic disgorgement does not only imply forfeiting the fruits of the deception. Since the model itself may leak personal information (like support vector machines, decision trees, or K-nearest neighbours algorithm) deleting the data unlawfully collected that was used to train the algorithm may not be sufficient to avoid future data breaches. Hence, deleting the algorithm, apart from being an exemplar punishment measure, in some cases constitute the only way to fully eradicate the illicit behaviour.

The Federal Trade Commission (FTC) is the USA's federal agency in charge of supervising and enforcing US customer and privacy federal regulations. Recently, the

---

Columbia, and the Information Privacy Commissioner of Alberta (2nd Feb 2021) <https://www.priv.gc.ca/en/opc-actions-and-decisions/investigations/investigations-into-businesses/2021/pipeda-2021-001/> accessed 27/05/2022. In the United States of America, while Clearview AI Inc. was not fined by any supervisor authority, the company settled a dispute with the State of Illinois and the American Civil Liberties Union (ACLU) and, as part of the settlement, the company refrains from making its faceprint database available to most businesses and other private actors in the USA. See American Civil Liberties Union, ACLU v Clearview AI Inc (updated 11th May 2022) <https://www.aclu.org/cases/aclu-v-clearview-ai> accessed 27/05/2022.

[878] Li (n 503) 21.

FTC has had the opportunity to enforce federal laws against companies that processed personal information using AI systems. While the FTC had a conservative position in early cases and allowed both Google[879] and Facebook[880] to keep the AI systems and related technologies that were developed using data unlawfully collected, this stance seems to have recently changed.

This new trend started with the case of Facebook & Cambridge Analytica. According to the FTC, Cambridge Analytica developed an app that allowed users to reply to questions about personality and gathered data such as 'likes' of public Facebook pages by the app's users and by their connections on Facebook. This method enabled Cambridge Analytica to harvest information from 250.000 app users and at least 30 million of those users' Facebook friends in the USA. The data was later employed to train an algorithm that created personality scores of those individuals. With these scores, and after linking them with USA voter records, Cambridge Analytica profiled voters and targeted political advertising and messages. As part of the settlement, which also included a record-breaking fine of $5bn to Facebook for allowing access to users' data, the FTC ordered Cambridge Analytica to delete all personal data illegally collected from users and their friends and to destroy 'any information or work product, including any algorithms or equations' whose origin relates to that information.[881]

The second time this tool was implemented concerns the Everalbum case. The company developed an app ('Ever') where users could upload photos and videos and it also allowed them to tag friends by their names and cluster users' images by the faces of those who appear in the photos. While Everalbum stated that facial recognition was allowed on an opt-in basis, according to the FTC's investigation facial recognition had been activated by default for all users until mid-2019. Additionally, the company merged millions of Ever app users' faces with images that the company collected from open datasets to generate training datasets for their facial recognition algorithm. While the company did not share Ever app users' personal data with external companies, it provided facial recognition solutions to third parties (using the algorithms developed with the personal information provided by Ever's users). Finally,

---

[879] FTC, *Google LLC and YouTube LLC*. Federal Trade Commission File No. 1723083 (September 4, 2019).

[880] FTC, *Facebook Inc*. Federal Trade Commission File No. 1823109 (July 24, 2019).

[881] FTC, *Cambridge Analytica LLC*. Federal Trade Commission File No. 1823107 (July 24, 2019).

the company did not honour promises made concerning retention periods and erasure of personal data once users cancelled their accounts. Considering the FTC findings, the company and the authority signed a settlement according to which the company accepted to erase not only the personal information of Ever's users that switch off their accounts, but also to destroy all facial templates ('face embeddings') generated from users' photos for automated recognition and any 'affected work product', i.e. models or algorithms created totally or partially using biometric data gathered from Ever app users.[882]

Finally, in early 2022 the FTC again made use of the algorithmic disgorgement when it ordered WW International and Kurbo Inc to destroy the mathematical models trained on misbegotten personal data. WW International and Kurbo placed into the market a weight loss app addressed to children (as young as 8 years old) and collected personal data without the consent of their parents or the holder of their parental responsibility. From 2014 to 2019 the companies offered a weight-management and tracking app ('Kurbo') to be used by children, teenagers, and their families. Kurbo app collected personal data about their food intake, physical activity, and weight, along with other data points like names, email addresses, and birth dates. Until 2020 Kurbo app was used by nearly 280.000 users, from whom nearly 20.000 were children under 13 years old. According to the FTC, the company failed to provide the required information concerning the data processing activities, failed to collect the required parental consent for children under the age of 13, and failed to delete personal information as required under Children's Online Privacy Protection Rule (COPPA). For these reasons, the companies and the FTC settled the dispute and it was agreed that the organisations will pay a $1,5 million penalty, erase the data unlawfully collected and processed, and destroy any affected work product (mathematical models or algorithms) that were developed with the data illegally collected from children in violation of COPPA.[883]

As shown in these cases, algorithmic disgorgement is an extreme penalty that was applied in cases where the law-breaking conducts were exceptionally grave. It

[882] FTC, *Everalbum Inc*. Federal Trade Commission File No. 1923172 (January 8, 2021).

[883] FTC, *Kurbo Inc & WW International Inc*. Federal Trade Commission File No. 1923228 (March 4, 2022).

functions as a deterrent for companies engaging in the processing of personal data through AI systems, since, first, the development of AI systems requires months or years of intensive work and it may cost millions of Euros, even exceeding the amount of the administrative fines imposed. Second, the consequences of these enforcement mechanisms are not always limited to a single algorithm. If the dataset that has been generated using ill-gotten data is employed for the training of several algorithms, a whole set of models could be potentially at risk. Similarly, if a model was trained with tainted data, and this model is employed to develop another model, the deletion of the whole chain of algorithms may be ordered. Consequently, organisations should balance the risks of illegally collecting data to develop AI systems against the potentially severe consequences. It creates an additional compliance burden on developers of AI systems, in particular, if the datasets they use to train AI systems are created by third parties. In these cases, they must implement due diligence measures, as well as carry out an integral vendor assessment to evaluate how the dataset was created.

There are, however, some grey areas concerning algorithmic disgorgement as ordered by the FTC. First, the FTC did not address any specific methodology to carry out the deletion or destruction of the algorithm.[884] Secondly, it is not clear whether individuals affected by the unlawful processing of personal information are entitled to request the imposition of this penalty on controllers or processors. According to commentators, The power to request the algorithmic disgorgement remains entirely on the FTC. Some authors even propose to introduce algorithmic disgorgement as a free-standing right for data subjects.[885]

VI.6.3.- Could algorithmic disgorgement be applied in the EU?

At this point, it is fair to inquire whether this enforcement mechanism could be applied in Europe. In other words, could EU data protection authorities order companies to delete or destroy algorithmic or mathematical models built using ill-gotten personal data? Does this remedy fall under the powers of supervisory

[884] Kate Kaye, The FTC's 'profoundly vague' plan to force companies to destroy algorithms could get very messy (Protocol, 17/03/2022) <https://www.protocol.com/enterprise/ftc-algorithm-data-model-ai> accessed 02/05/2022.

[885] Li (n 503) 23.

authorities? It is worth noticing that the FTC based its powers to order the algorithmic destruction on Section 5 of the FTC Act which bans 'unfair or deceptive acts or practices in or affecting commerce',[886] and the FTC has the power to order a remedy 'reasonably tailored' to the law-breaking conduct.[887] The FTC does not have express powers to destroy the algorithm, but it was understood that this remedy was suitable according to the particular circumstances of the cases.

The powers that EU data protection authorities have to fulfil their mission are established in Art. 58 GDPR. While this article provides an express list of powers for data protection supervisory authorities, questions concerning the extent of the powers granted to them were several times discussed by the CJEU.[888] A close and literal evaluation of the wording of the article reveals that EU data protection authorities would not have express power to destroy algorithms developed and trained using tainted personal data. Firstly, supervisory authorities have the power to temporarily or definitively order controllers to limit or ban the processing operations being carried out.[889] Supervisory authorities can order a ban on the processing, which would force the controller to halt the use of the algorithm, but not its destruction. Secondly, supervisory authorities may request the deletion of personal data to controllers,[890] which may potentially lead to the algorithm deletion. However, the applicability of this path is limited. Not only is this remedy applicable if the data subject exercised his or her rights to erasure pursuant to Art. 17 GDPR, but also the erasure of the algorithm following a request would be effective only for those algorithms that contain personal data within the model itself like support vector machines, decision trees or K-nearest neighbours. Yet, this is an indirect effect concerning some particular models and not a horizontal power granted to supervisory authorities applicable to any algorithm.

---

[886] 15 U.S. Code § 45 - Unfair methods of competition unlawful; prevention by the Commission

[887] Slaughter (n 785) 39–40.

[888] Oreste Pollicino and Marco Bassini, 'Bridge Is Down, Data Truck Can't Get Through...A Critical View of the Schrems Judgment in the Context of European Constitutionalism' in G Ziccardi Capaldo (ed), *The Global Community Yearbook of International Law and Jurisprudence 2016* (Oxford University Press 2017) 250.

[889] Art. 58(2)(f) GDPR.

[890] Art. 58(2)(g) GDPR.

However, a provision may allow data protection authorities to order the algorithmic destruction. Supervisory authorities may request the controller or processor 'to bring processing operations into compliance' with the legal framework 'in a specified manner and within a specified period'.[891] This is a generic power for authorities to compel controllers to adequate their processing operations, modify their unlawful course of operations and proceed in line with the legal framework. It also allows supervisory authorities to establish how the controller should comply with the order, including the timeframe. Taking into account the width of this power it seems plausible that supervisory authorities may find inspiration in this power to compel companies to destroy the algorithms developed using tainted data. Supervisory authorities may consider that to restore the legality of the processing operations, the erasure of the algorithm developed using misbegotten data is the most suitable remedy. In the transition from the 'world of atoms to the world of bits'[892] and to model a stronger framework for the protection of personal data and privacy, data protection authorities play a crucial role. Allowing the use of algorithmic disgorgement to supervisory authorities will enhance their toolkit to oversee the correct application of the data protection framework in Europe.

However, relying upon the erasure of the algorithm should come after a thorough assessment of all the circumstances. In addition to the factors listed in Art. 83(2)(a) to (k) GDPR to impose administrative fines, the supervisory authority seeking to request the erasure of the algorithm should also consider the egregiousness of the behaviour of the controller or processor and if other equally compelling suitable measures are available.

V.7.- Privacy by Design measures: reducing the identifiability of data

Processing personal data entails privacy and security risks for data subjects, so carrying out personal data processing operations should be done with close supervision of the respective legal frameworks. The most common alternatives to mitigate the risks posed by data processing operations are encryption or de-

[891] Art.58(2)(d) GDPR.

[892] Oreste Pollicino, 'The Transatlantic Dimension of the Judicial Protection of Fundamental Rights Online' (2021) 1 The Italian Review of International and Comparative Law 277, 295.

identification of personal data. Another alternative method is to generate new data from real datasets, a process that is called synthesising data. These techniques are evaluated below.

7.1.- Anonymisation and pseudonymisation

In previous sections, the principle of data protection by design and by default was addressed. It was explained that, according to the principle of data protection by design and by default, controllers must ensure compliance with the data protection legal framework from the inception, which means conducting legally adequate data processing activities when designing, developing and deploying AI systems. However, developing adequate technical and organisational measures to translate these high-level principles into concrete specifications and controls is a challenge for controllers and processors.[893] Two of the most important techniques with which this principle can be ensured are anonymisation/pseudonimisation and encryption. In this section, it is evaluated the former whereas the latter is explained in the following section.

Anonymous information is data unrelated to an identified or identifiable individual, as well as personal data that was transformed in a manner that the individual cannot be identified anymore.[894] No element should be left in the data which could, through reasonable effort, help re-identify the individual concerned. The identification of the individual must be prevented irreversibly,[895] which means that the transformation of the data is one-way only. The regulation does not require absolute anonymity. Anonymous data is so where it prevents identification using 'all means likely reasonably to be used' to re-identify. This means that a contextual element should be considered, which includes cost, time, know-how and computational power, evaluating also the likelihood and severity of the consequences of re-identification. Additionally,

[893] European Union Agency for Cybersecurity, 'Data Protection Engineering. From Theory to Practice' (2022) 5.

[894] Recital 26 GDPR and Article 29 Data Protection Working Party, 'Opinion 05/2014 on Anonymisation Techniques' (2014) 5.

[895] ibid 6; International Standard Organisation, 'ISO/IEC 29100:2011 Information Technology — Security Techniques — Privacy Framework' (2011) s 2.2.

the means likely reasonably can be used either by the controller itself or by a third party.[896]

The main benefit of processing anonymised data is that the GDPR does not apply to this kind of information.[897] However, where technological improvements render it possible to transform anonymous data into personal data again, the resulting data will be considered personal data for the purposes of the GDPR and, thus, the latter will apply to it.[898] This acknowledges the fact even anonymous data can be linked to individuals and a risk-based approach that evaluates the probability and impact of the re-identification should be adopted.[899]

Whereas anonymisation is an effective technique to remove the compliance burden when carrying out data processing operations powered by AI systems, this methodology is not always suitable, because strong de-identification of the personal data may hinder its utility for further processing[900] and oftentimes personal data is needed either as an input or resulting as an output of the processing. Hence, an alternative to increasing the utility of the data is to pseudonymise personal data.

Pseudonymised data aims to mask the identities of identified or identifiable natural persons and its generation entails the substitution of personal identifiers (such as the name, surname, or zip code) with a different attribute (such as an alphanumeric code).[901] Through pseudonymisation, the personal data may only be assigned to a precise individual if supplementary information is provided.[902] Therefore, two conditions are required to consider personal data as pseudonymised. First, personal

---

[896] Article 29 Data Protection Working Party, 'Opinion 05/2014 on Anonymisation Techniques' (n 891) 9.

[897] Recital 26 GDPR.

[898] See Recital 9 Regulation (EU) 2018/1807 of the EU Parliament of the Council of 14 November 2018 on a framework for the free flow of non-personal data in the European Union.

[899] Thiago Guimarães Moraes and others, 'Open Data on the COVID-19 Pandemic: Anonymisation as a Technical Solution for Transparency, Privacy, and Data Protection' (2021) 11 International Data Privacy Law 32, 42.

[900] European Union Agency for Cybersecurity, 'Data Protection Engineering. From Theory to Practice' (n 890) 10.

[901] Luca Tosoni, 'Article 4(5). Pseudonymisation' in Christopher Kuner, Lee A Bygrave and Christopher Docksey (eds), *The EU General Data Protection Regulation (GDPR). A commentary* (OUP 2020) 133.

[902] Art. 4(5) GDPR.

data should be de-identified. Data controllers must remove the connections between the data and the data subject. This process is generally performed by removing one feature in the data (e.g. data subject's name) and changing it with another. Second, there should be a separation of the data needed to re-identify the individual. This supplementary information must be maintained separately and the controller must put in place technical and organisational measures to guarantee that the personal data is not assigned to an individual.[903]

Pseudonymisation is explicitly mentioned in the GDPR several times. It is a safeguard measure that can be considered when evaluating the compatibility of further processing of personal data[904] and processing personal data for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes.[905] Additionally, it is a measure that contributes to the fulfilment of the privacy by design principle, since pseudonymization helps to minimise the personal identifiers in the data processed.[906] Finally, it helps ensure a higher level of security in the processing of personal data.[907]

As seen, both anonymisation and pseudonymization are measures that can mitigate the risks posed by the processing of personal data using AI and they may assist controllers to comply with data protection regulations.[908] While anonymisation is the preferred option to protect the rights of individuals, pseudonymisation can provide a fair balance between the protection of the data subjects and the usability of the information for analytics when using AI systems. While pseudonymous data is still personal data, the identification of individuals in the original dataset requires more computational power, resources and time. Data pseudonymisation thus constitutes a measure to comply with the principle of data minimisation since pseudonymised data

[903] Art. 4(5) GDPR.

[904] Art. 4(6)(e) GDPR.

[905] Art. 89(1) GDPR.

[906] Art. 25(1) GDPR.

[907] Art. 32 GDPR.

[908] Michèle Finck and Frank Pallas, 'They Who Must Not Be Identified — Distinguishing Personal from Non-Personal Data under the GDPR' (2020) 10 International Data Privacy Law 11, 35.

expose less information about data subjects, both in terms of the quantity and the nature of the data.[909]

7.2.- Encryption

Encryption is another method to keep the confidentiality of personal data. Through encryption unencrypted data (plaintext) is transformed into encrypted data (ciphertext).[910] The main goal of encryption is to safeguard stored information (data at rest) and transmitted information (data in transit). The encrypted information or ciphertext should not provide any information about the original information or plaintext.

For the process of encryption of personal data, a key is needed both to encrypt the original data (i.e. to transform plaintext into ciphertext and referred to as 'encryption key') and to decrypt the encrypted data (i.e. to restore the original plaintext from the ciphertext and referred to as 'decryption key'). Without the encryption or decryption key, the process of re-identification is extremely difficult, since it requires systematically trying different encryption/decryption keys until the right one is found (brute force attack). If identical keys are employed for the encryption and decryption process, the process is called 'symmetric encryption'. On the other hand, where different keys are employed for the encryption/decryption process, it is an 'asymmetric encryption' system.

As mentioned before concerning pseudonymisation, encryption is an example of a measure that can be taken to mitigate some risks of the data processing activities as exemplified by the GDPR. It is a safeguard measure that can be considered when evaluating the compatibility of further processing of personal data,[911] it helps ensure a higher level of security in the processing of personal data[912] and it could exempt controllers from their obligation to notify data subjects in case of a data breach.[913] Whereas not explicitly mentioned in the text of the GDPR, just like pseudonymisation,

---

[909] Datatilsynet (n 193) 18.

[910] Clause 3.12 ISO/IEC 18033-1:2021(en) Information security — Encryption algorithms — Part 1: General

[911] Art. 4(6)(e) GDPR.

[912] Art. 32 GDPR.

[913] Art. 34(3)(a) GDPR.

encryption can be considered as a measure to implement the principle of privacy by design.[914]

Encryption has been used to secure communications and some encryption algorithms and techniques constitute industry standards. For instance, HyperText Transfer Protocol Secure (HTTPS)[915] is a standard protocol to secure communications on the Internet, and it is currently the standard encryption technique in web browser communications. Since it uses an encryption mechanism for the data in transit, it provides a higher level of confidentiality than its predecessor (i.e. the HTTP). So important is keeping the confidentiality of the communications that supervisory authorities have started to enforce GDPR provisions where security standards are not satisfied, and this protocol was also subject to evaluation. For example, an organisation was fined because of the use of the HTTP protocol instead of the HTTPS protocol on its website. According to the supervisory authority, the organisation failed to provide adequate security measures to its users since the HTTP enables third parties to intercept the information transferred from the user's device to the web server.[916] Another example of the use of encryption is the end-to-end encryption (EE2E) protocol which is the standard for communications between users of messaging applications. According to a recent survey from the Hong Kong Office of the Privacy Commissioner for Personal Data, except for WeChat, all other major instant messaging applications (including Facebook, Facebook Messenger, Instagram, Linkedin, Twitter, Skype and WhatsApp) implemented end-to-end encryption for the transmission of private messages between users.[917]

The most common and widely used encryption techniques only support the transformation of plaintext into ciphertext for storage (data at rest) and transmission

---

[914] Art. 25(1) GDPR.

[915] In the web browsers it is referred as https:// and it comes before the domain address. For example: https://www.unibocconi.it/

[916] Agencia Española de Protección de Datos, *Procedimiento Nº: PS/00185/2020*.

[917] Hong Kong Office of the Privacy Commissioner for Personal Data, PCPD Releases Report on "Comparison of Privacy Settings of Social Media" (12/04/2022) <https://www.pcpd.org.hk/english/news_events/media_statements/press_20220412.html> accessed 12/05/2022.

(data in transit). Hence, to perform any other processing operations[918] (such as consultation, analysis, alteration or to train an AI model) the ciphertext should be transformed again into plaintext by a decryption algorithm, and the processing operations are carried out on the original unprotected plaintext (data in the clear). This poses a risk since the security that encryption provides disappears once the data is transformed again into plaintext for further processing. To overcome this limitation, a special kind of cryptographic algorithm has been under development: homomorphic encryption.

Homomorphic encryption allows undertaking data processing operations on encrypted information (data in use), without the need to convert it into plaintext before processing.[919] This means that with this technique it is possible to perform useful processing activities on cyphertext without the need to decrypt the encrypted data and without holding the decryption key. It may potentially solve the privacy problems related to the processing of personal data via cloud service providers since the processing can be undertaken without revealing the original data. A cloud client should carry out the encryption and, keeping locally the encryption keys, transmits the encrypted data to the cloud provider for further processing. The cloud provider performs the computations on encrypted information and then sends back the processed encrypted information to the cloud client. Finally, the cloud client decrypts the data sent by the cloud provider.[920] This is also very important when the processing operations are carried out with AI systems, in particular, if the processes are performed in the cloud since models can make inferences, predictions and any other evaluations on homomorphically encrypted data and the model owner (cloud service provider) would not be able to see the original data.

However, this technique is currently experimental and cannot be applied at scale, since it is computationally expensive and it creates substantial operative costs.

---

[918] It is worth remembering that for the purposes of the GDPR, both storage and transmission of personal data constitute 'processing' of personal data (see Art. 4(2) GDPR).

[919] European Union Agency for Cybersecurity, 'Data Protection Engineering. From Theory to Practice' (n 890) 14.

[920] Kristin Lauter, 'Private AI: Machine Learning on Encrypted Data' in Tomás Chacón, Rosa Rebollo and Inmaculada Higueras Donat (eds), *Recent Advances in Industrial and Applied Mathematics. SEMA SIMAI Springer Series* (Springer 2022) 110.

Additionally, if fully homomorphic encryption were carried out with conventional devices, operations that normally take milliseconds would be completed in weeks.[921] Moreover, while this technique seems a promising measure to enhance privacy protections when processing personal data using AI systems, there are reasonable concerns about the possibility that in the not too distant future quantum computers will break encryption protection very easily.[922] However, quantum computing may also open door to a new frontier of data security: post-quantum cryptography. Post-quantum cryptography is an active area of research not only by public institutions[923] but also by private actors[924] and it can provide in the future a whole new set of privacy and security protections.

7.3.- Synthetic data

Many times anonymisation/pseudonymisation or encryption is not feasible, too expensive to implement or they do not provide adequate protection. In those cases, synthetic data can be an alternative. Synthetic data is data generated from real data which attempts to emulate the statistical proprieties of real datasets. Provided that the statistical properties of the original data are appropriately replicated, synthetic data

---

[921] Defence Advance Research Projects Agency, DARPA Selects Researchers to Accelerate Use of Fully Homomorphic Encryption (08/03/2021) <https://www.darpa.mil/news-events/2021-03-08> accessed 12/05/2022.

[922] Frederik Armknecht and others, 'General Impossibility of Group Homomorphic Encryption in the Quantum World' in H Krawczyk (ed), *Public-Key Cryptography – PKC 2014. PKC 2014. Lecture Notes in Computer Science* (Springer 2014).

[923] See for instance, National Institute of Standards and Technology, 'Status Report on the Second Round of the NIST Post-Quantum Cryptography Standardization Process' (2020); European Union Agency for Cybersecurity, 'Post-Quantum Cryptography: Current State and Quantum Mitigation' (2021); German Federal Office for Information Security, 'Quantum-Safe Cryptography – Fundamentals, Current Developments and Recommendations' (2021); Agence nationale de la sécurité des systèmes d'information, 'Avis Scientifique et Technique de l'ANSSI Sur La Migration Vers La Crypographie Post-Quantique' (2022).

[924] See for instance, IBM, 'What Is Quantum-Safe Cryptography, and Why Do We Need It?' (10 March 2022) <https://www.ibm.com/cloud/blog/what-is-quantum-safe-cryptography-and-why-do-we-need-it> Microsoft, 'Cryptography in the era of quantum computers' <https://www.microsoft.com/en-us/research/project/post-quantum-cryptography/>; Huawei, 'Post Quantum Cryptography' <https://www.huawei.com/it/trust-center/post-quantum-cryptography> all accessed on 30/05/2022.

can be used as a representation of real data in certain cases. Synthetic data appropriately replicates the original data if the persons or systems that evaluate the resulting dataset draw similar conclusions as they would have reached after the assessment of the original dataset.[925]

But while synthetic data should resemble and keep the statistical properties of the original dataset, it cannot be identical to it because the identifiability of data subjects needs to be reduced. Identifiability can be thought of as the probability of linking the true identity to a record in a dataset. At one extreme of the identifiability spectrum is perfect identifiability, which means that the overlap between real and synthetic data is equal to one. At the opposite end of the spectrum, it is impossible to correctly attribute an identity to a record, meaning that the probability to identify a record is zero. Any synthetically produced dataset will have a probability of re-identification along this spectrum.[926] Consideration should also be made to the trade-off between data privacy and data utility. This is because to reduce the identifiability of the data (which increases privacy protection), there should necessarily be a correlative reduction of the utility of the data. Contrarily, the maximum utility would be represented by the original dataset (without any transformation or control), but this will mean relinquishing any additional gain in terms of privacy.

Since synthetic data is not really personal data, privacy risks are mitigated. If the process that leads to the generation of synthetic data is done adequately, there is no one-to-one link or mapping between the synthetic data records and those pertaining to the natural persons included in the original dataset. Therefore, it can be considered de-personalised or not personal data.

Retaining the statistical properties of a dataset while removing risk factors associated with the identifiability of natural persons brings many benefits. First, synthetic data constitutes a suitable alternative to accessing large datasets when the consent of the data subjects is difficult or impossible to obtain. Accessing huge amounts of data is critical to developing AI systems. Large datasets are needed to train, test and validate AI models, as well as test AI software applications or

---

[925] European Data Protection Supervisor, 'Synthetic data' <https://edps.europa.eu/press-publications/publications/techsonar/synthetic-data_en> accessed 09/05/2022.

[926] Khaled El Emam, Lucy Mosquera and Richard Hoptroff, *Practical Synthetic Data Generation. Balancing Privacy and the Broad Availability of Data* (O'Reilly 2020) 24.

applications that incorporate AI models. In general, data is collected for a purpose, so if developers plan to use it for other purposes, they must obtain consent from individuals or have another valid legal basis that allows the repurposing to take place. Data synthesis can give the analysts data that mimics real datasets to work with. This is particularly important when the information is considered sensitive data, as in healthcare.[927] Secondly, synthetic data allows better analytics in cases where there is a lack of real datasets (for example, when developing a new solution for which there is no suitable dataset available) or where data exists but it is insufficient (for instance to train robots to undertake complex tasks in the production line or warehouses). Finally, synthetic data can alleviate the regulatory compliance burden and cut down the costs of implementing controls to comply with regulations. Using personal data to develop AI systems requires compliance with data protection and privacy regulations, which considering the amount of personal information required to train AI models may entail an economic challenge for controllers. However, if the synthetic data does not allow the identification of the data subjects belonging to the original dataset, data protection regulations do not apply. The benefits of reducing the compliance burden should not be underestimated. In 2021 the Norwegian Data Protection Authority (Datatilsynet) fined an organisation for a data breach that involved the online disclosure of personal information belonging to 3.2m individuals (around 450.000 were children) following an error that occurred when solutions were tested in connection with moving the database from a physical server to a cloud environment. The exposed personal information included names, dates of birth, addresses, telephone numbers, and email addresses. The Norwegian supervisory authority suggested that to avoid the data breach, and the subsequent sanction, the testing could have been carried out using synthetic data instead of original records.[928] In other words, had the organisation

[927] Allan Tucker and others, 'Generating High-Fidelity Synthetic Patient Data for Assessing Machine Learning Healthcare Software' (2020) 3 npj Digital Medicine 1; Zahra Azizi and others, 'Can Synthetic Data Be a Proxy for Real Clinical Trial Data? A Validation Study' (2021) 11 BMJ Open 1.

[928] Janne Stang Dahl, Vedtak om overtredelsesgebyr til Norges idrettsforbund for mangelfull testing (Datatilsynet official website 11/05/2021) <https://www.datatilsynet.no/aktuelt/aktuelle-nyheter-2021/vedtak-om-overtredelsesgebyr-til-norges-idrettsforbund-for-mangelfull-testing/> accessed 09/05/2022 (translated using Google Translate).

concerned employed synthetic data to test the cloud solution, it may have avoided the economic and reputational consequences of the sanction.

Yet, synthetic data does not mitigate every single risk concerning the processing of personal data for the development and deployment of AI systems. While it reduces to a great extent the identifiability of natural persons within the dataset, the risk of reidentification persists. The more a synthetic dataset emulates the original dataset, the more utility it will have for analysts but the more personal data it might reveal. This increases the compliance burden and the need to implement more controls to protect personal data.[929] Additionally, it is worth noticing that there may be some cases where the demands of the individuality of the records, i.e. processing information that is linked to particular individuals, will not be satisfied by any privacy-preserving technique.[930] A solution to overcome these drawbacks could be the use of synthetic data generation in conjunction with differential privacy. Data utility is achieved via synthetic data generation and data protection is privacy is attained through differential privacy.[931] Differentially private synthetic data is one of the strongest privacy-preserving methods for synthetic data.[932] Hence, a combination of synthetic data and the use of differential privacy can be a solution for those cases.

[929] Steven M Bellovin, Preetam K Dutta and Nathan Reitinger, 'Privacy and Synthetic Datasets' (2019) 22 Stanford Technology Law Review 38–39.

[930] ibid 41.

[931] ibid 39.

[932] Joseph Near and David Darais, Differentially Private Synthetic Data (National Institute of Standards and Security blog 03/05/2021) <https://www.nist.gov/blogs/cybersecurity-insights/differentially-private-synthetic-data> accessed 09/05/2022.

CONCLUSIONS

This work tries to explain the main challenges faced by the General Data Protection Regulation to protect individuals against the negative impacts of AI. Both the positive and negative aspects of these new technological developments were addressed and were also considered the risks to the fundamental rights of individuals.

The core of this work was to assess the extent to which the processing of personal data using AI systems complies with the requirements established in the General Data Protection Regulation and how the risks to the fundamental rights of individuals posed by these processing operations can be mitigated. This protection is translated in two ways: by granting strong and effective individual rights and imposing stringent accountability measures on those natural or legal persons who process their personal data. While the data subjects' perspective is recognised throughout this work, the consideration of the accountability mechanisms is of utmost importance to effectively protect the personal data of individuals.

This research is mainly focused on the data protection implications related to the processing of personal data using AI systems, deliberately omitting considerations and risks posed by AI systems in other fields like consumer rights, competition law and intellectual property rights. Additionally, while the territorial scope of the research is mostly focused on the EU legislation and cases, where appropriate, some mentions of international developments were evaluated.

In addition to compiling and updating literature about the data protection implications of AI-assisted processing of personal data, this work makes two significant contributions. First, it attempts to bridge the gap between technical and legal knowledge by explaining the most relevant definitions of AI systems and why understanding the different approaches or techniques employed to build AI systems matters from a data protection perspective. Secondly, it seeks to deliver detailed guidance on how high-level principles or fundamental values required by the relevant legislative framework could be implemented, concretised or translated into practical and operational requirements.

In the following paragraphs, it is provided with a high-level explanation of the main findings of the research conducted, along with the recommendations and proposals to

enhance the protection of individuals where their personal data is processed using AI systems.

The idea of artificial intelligence

Chapter I conceptualizes the notion of artificial intelligence and provides a crucial basic categorization of the different AI techniques. First, the importance of data and its free flow for the development of AI is explained. Even though some ideas and techniques that led to the AI revolution were already available many years ago, the sheer increase in computational power and data availability made the development of current AI systems possible. This chapter also attempts to explain the meaning of 'data' and provides a clear picture of the importance of the free flow of data in the digital economy. However, as personal data is a central piece of this work, the concept of personal data and its importance is explained. For the concept of personal data, the GDPR is the source of the legal definition, and personal data according to this text is any information relating to an identified or potentially identifiable individual. Since the legal definition is open to interpretation, the Court of Justice of the European Union further developed some of the most contentious aspects of the definition. Particularly debated aspects relate to what data elements should be included in the notion of 'any information', how the particular data elements should relate to the individual, and when a person can be potentially identifiable using the information under processing. It is also highlighted that individuals can be more easily identified using AI systems than with traditional processing methods, and information that was previously considered as non-personal data can nowadays allow the identification of individuals, a situation that blurs the boundaries between non-personal and personal data.

The first chapter then moves forward to the conceptualization of AI and evaluates the complexities of this technology. It provides a broad picture of everyday use cases of AI and, most importantly, the risks that the intended or unintended misuses of AI can cause on individuals and society, particularly concerning discrimination, lack of transparency and cybersecurity risks. Furthermore, it attempts to find a definition suitable for the purposes of this work. After surveying some definitions proposed by academics, public institutions and ad-hoc committees, it evaluates in depth the definition proposed by the European Commission in the AI Act draft. In a nutshell, according to this definition, AI is software that: a) is developed using machine, logic

and knowledge-based techniques or statistical approaches; and b) can produce certain outputs for a given set of human-defined objectives. While the production of certain outputs, such as predictions or recommendations, is barely contested, the characterisation of the techniques used to qualify software as an AI system is under scrutiny and little attention was paid to its importance in the literature reviewed. Hence, for a better understanding of the AI systems, it explains the most common models that are included in any of the three techniques or approaches mentioned by the AIA draft. The reason for this explanation is to show, first, that AI is not a single and unified technology and, second, that as the models that form the basis of an AI system differ, the privacy-related problems of these models and the interpretability of their inner workings and outcomes will differ as well. However, it is also noted that the machine learning approaches are the most frequently used systems and their adaptative nature poses particular risks vis-à-vis individuals and society.

The data protection regulation and its relationship with AI

Chapter II studies the data protection legal framework in the EU and how these legal provisions are applied to the processing of personal data using AI systems. It is considered in this work that a basic knowledge of fundamental rights is vital for the correct evaluation of the processing of personal data using AI systems. The chapter first briefly assesses the right to the protection of personal data as a fundamental right in the EU, by explaining the relevant provisions of the European Convention of Human Rights and the interpretation of these provisions by the European Court of Human Rights, and the Treaty of the Functioning of the EU and the Charter of Fundamental Right, both of which integrate the primary legislation of the EU.

After evaluating the core provisions of the fundamental legislation, the most important piece of secondary legislation is evaluated in depth: the General Data Protection Regulation. The principles of data protection are first addressed and the particularities of the processing of personal data using AI systems are explained. Processing personal data using AI systems challenges the principle of purpose limitation since many times establishing concrete predefined objectives to process personal data can be difficult. The purposes to process personal data collected may change throughout an AI project, so this situation should be evaluated accordingly. Moreover, the data minimisation principle is difficult to observe because, in general,

AI systems need to gather substantial amounts of information, including personal data, from the initial stages of development. The more data is available to train the AI algorithms, the more statistically accurate their predictions or classifications tend to be. Furthermore, the principle of accuracy can be problematic to follow, due to the sheer amount of information collected from every data subject and the different sources of collection, which make it problematic to trace the correctness of the information. Additionally, it is usually irrelevant having inaccurate or incomplete data concerning a single natural person, since the overall accuracy of the model remains largely unaffected by a single inaccurate record. Finally, processing personal data using AI systems increases the risks of suffering a malicious attack because AI systems present new opportunities to attackers compared to applications that process personal information. To be more specific, malicious actors can exploit the vulnerabilities of AI systems in three general ways: by exploiting algorithmic design flaws, poisoning the datasets used to produce the prediction or outcome and via adversarial examples.

Another central aspect of the protection of personal data concerns the lawful basis for processing, i.e., whenever a controller processes personal data, they must rely on one or more suitable legal basis for the processing. While the GDPR establishes six legal basis for controllers to process personal data, when the processing of personal data is carried out using AI systems the most important legal basis are the consent of the data subject and the legitimate interests of the controller or a third party. Regarding consent as a lawful basis, obtaining informed consent from the data subjects is particularly challenging where their personal information is processed using AI systems, since the data subject's acceptance may not be based on a freely given, specific and informed indication of their wishes to have their data processed for those purposes**.** The legitimate interests of the controller or a third party can also be a suitable legal basis for processing personal data for the development and deployment of AI systems. The most prominent advantage of this legal basis is that it does not require any active participation or involvement from the data subject (as it is required with the consensual basis). However, this flexibility is not unrestrained, since controllers must carry out a legitimate impact assessment whereby they evaluate whether their legitimate interests override the fundamental rights of the data subjects.

Another issue that should be considered relates to the repurposing of data. If purposes are compatible, personal data can be processed for another different purpose. While a compatibility assessment should be performed before processing personal data for another purpose, controllers may also rely on a presumed compatible purpose, i.e. statistical purpose. In this case, however, controllers must process personal data to find statistical correlations in aggregated data without using the output of such processing activities to take measures or support decisions concerning specific individuals.

Finally, in this chapter it was also highlighted the importance of distinguishing between the two broad stages of the AI lifecycle (development and deployment) since controllers can invoke different purposes in each of these phases. It is generally agreed that at the development stage purposes can be defined more broadly and may encompass research, whereas at the deployment stage more specific purposes must be invoked.

Protecting the rights of individuals when AI systems are used to process personal data

Chapter III evaluates the rights of data subjects whose personal data is processed using AI systems and it provides an overview of the GDPR general accountability mechanisms that have an impact on all of the issues related to the intersection between data protection and artificial intelligence.

The first and most important right is the right not to be subject to a decision based solely on automated processing, including profiling, which produces legal or similarly significant effects on them. While it is allowed under certain circumstances to subject individuals to automated decision-making which significantly affects them (for instance, if they obtain consent from data subjects or rely on a contract), controllers must put in place certain safeguards to protect data subjects. In this section it is considered that the provision should be considered as a general prohibition addressed to controllers rather than a right that must be invoked by the data subject and that it concerns only individual decision-making, leaving aside decisions that affect a group of individuals. Additionally, it is explained that whereas the definition of “decision” is relatively straightforward, there may be some doubts concerning whose decision should be considered (see for instance *Schufa* case debate) and that the phrase

“based solely on automated processing” should be understood as the lack of a person empowered (both technically and organisationally) to alter the decision.

Among the many unresolved issues concerning this right, an important one regards the type of decisions that produce effects similarly significant to legal effects. It is also held that the decision should potentially alter the circumstances, the conduct, and the searches of the person in a substantial manner, produce long-lasting effects on individuals, or, ultimately, exclude or discriminate natural persons without an objective reason. However, after considering some cases recently decided on this matter the only conclusion reached is that the criterion should be decided on a case-by-case basis.

Another important aspect related to this right is that whenever a controller plans to subject an individual to a decision based solely on automated processing or profiling, the controller cannot rely on legitimate interests as a valid legal basis to perform solely automated decision-making that causes legal or similarly significant effects on the data subjects. Instead, they must use the consent of data subjects or a contractual or legal necessity. However, if controllers subject individuals to automated decisions that produce effects significantly similar to legal effects, they must implement appropriate measures to mitigate the risks to data subjects.

Controllers should also provide for human intervention (an individual may request the controller to have the decision reviewed by a human) and the possibility to challenge the decision if data subjects disagree on the merits. This is also related to the information rights of data subjects. Without being properly informed about the system it is impossible to challenge the decision. For this right to be operative, the data subject must be fully informed, both before and after the decision is made, about important and relevant aspects to understand the decision and act accordingly (which includes the right to challenge the decision if needed). Explanations to data subjects do not only relate to the existence of automated decision-making, but also to the need to provide relevant information about the logic involved (which should be sufficient to allow individuals to understand the reasons for the decision and to challenge it) and the significance and the envisaged consequences of such processing (how the processing operations might influence the individuals’ rights and freedoms, but only concerning concrete significant consequences).

Yet this right is not the only one provided for by the GDPR, since this regulation also requires controllers to provide for the right to rectification, erasure, restriction, objection and portability. Honouring data subjects' rights is also more challenging when their personal data is processed using AI systems. The most problematic aspects concerning compliance with data subject requests are the following: the right to rectification can be exercised both at the development of an AI system, for example, if the training dataset contains erroneous data about an individual, but also at the deployment stage when the AI system produces its output. However, in the latter, the individual should be aware that the output may be a statistical prediction, not a statement of fact and, as such, they admit a certain margin of inaccuracy. On the right to erasure, controllers processing personal data using AI systems must, from the early stages of development, ensure that the AI system is capable to honour the deletion of personal data if requested. A complication linked to the right to erasure in AI systems is that to entirely erase the personal data included in an all-encompassing request it is usually necessary to retrain the algorithm with the remaining data or amend the features of the system. Similar issues controllers may face if they have to honour requests to restrict the processing of personal data. Finally, on the right to data portability, it should be noted that it has an important limitation, crucially, concerning the types of personal data that a data subject can transfer to another controller. The personal data at issue must have been provided by the data subject, meaning information intentionally provided by the individual and the information the controller observed from the individual's behaviour or interaction with the service or device, which excludes the data derived and inferred from the information given by the individual (including the outcome of an AI system).

The GDPR does not only grants subjective rights to individuals. The GDPR provides a long list of rights that individuals can exercise, but it also establishes a structure of control that protects the rights granted to data subjects by imposing accountability and oversight obligations to controllers. Even if data subjects do not exercise their rights, data controllers and processors have accountability obligations to comply with, which reinforces the protection afforded to data subjects

The GDPR does not concretely specify which safeguards should be applied in concrete cases, thus giving controllers discretionary powers to decide which particular accountability measure to apply to guarantee the data subjects' rights. The gaps left

by the regulation are filled by guidelines, standards, codes of conduct, best practices and other soft law instruments.

One of the cornerstones of the GDPR is the creation of a register of processing activities. Building a register of processing activities can be burdensome for controllers and processors when the processing of personal data is carried out for the development and deployment of AI systems, since the processing operations carried out by AI systems may be extremely complex, a complication that can be eased with the assistance of an automation tool.

Then, it is highly likely that controllers or processors developing or deploying AI systems or employing big data analytics to carry out online behaviour advertising, tracking individuals across the web or profiling individuals must appoint a data protection officer, and preferably, the person who performs this role should have an adequate understanding of the features and issues related to AI systems.

Additionally, controllers must embed data protection principles into the design of their processing operations and they must preselect processing methods, values and alternatives that have the least data protection impact on individuals. The most well-known strategies to comply with data protection by design are pseudonymisation/anonymisation and encryption. But this obligation is not limited to these techniques. Training of employees, data minimisation (in particular during the AI development), implementation of performance tests at regular intervals and, where necessary, making reasonable adjustments to guarantee fair processing and reduce biases constitute other suitable measures that can be implemented to comply with the principle of privacy by design and by default.

Controllers developing or using AI systems that process personal data will be required to perform a DPIA in the majority of cases. Common challenges that data controllers using AI systems to process personal data may encounter are the need to establish at the outset clear purposes for data processing, the evaluation of the necessity and proportionality of the processing operation considering the stated purposes (explaining the reasons why a particular AI system was employed if they identified a less risky and privacy intrusive method to process personal data and the latter was discarded). When it comes to evaluating the risks for the rights of data subjects, they should consider in particular the risk related to the fairness of the AI outcome which could be produced by errors in the performance of the AI solution and

in particular if individuals whose data is being processed belong to vulnerable groups. Together with this assessment, they must recommend measures to address the risks and demonstrate compliance with the GDPR. Finally, the controller may consult individuals or their representatives concerning the planned processing operations.

Overcoming the weaknesses of the legal framework

Chapter IV evaluates how the limitations of the General Data Protection Regulation can be overcome. While the data protection legislation constitutes a solid framework for the protection of the rights of data subjects, some areas provide still limited protection. This is in particular true concerning the lack of transparency and explainability associated with the processing of personal data using AI systems and the existence of biases in decision-making and the requirement of fair processing and non-discrimination. These topics are addressed in depth in this work and for each of them some proposals are made.

First and foremost, transparency is central to creating social trust in the use of AI systems. Transparent AI solutions allow individuals to know the sources, the reason, and the types of data being processed. Controllers may use non-explainable AI systems and opaque sources of data to train their models and their organisational policies may refrain from disclosing information about the AI systems. While GDPR requires controllers to provide some information to data subjects, it fails to address the problems related to the different information to be provided to different interested parties both before and after the algorithmic decisions are made.

The information to be provided before taking the decisions using AI systems relates to the datasets, the general functioning of the algorithms and the model itself. Not only does it highlight the content of the relevant information to be provided, but also it suggests innovative forms of delivering the information. Concerning the content of the information that should be provided to individuals, it was considered that the ICO guidelines ‘Explaining Decisions Made with AI’ constitute an important starting point. This framework proposes six different rationales to consider when evaluating the kind and depth of information to provide to individuals. According to this framework, explanations should concern the rationale of the decision, who is responsible for the decision, the data used to reach the decision, fairness considerations, measures to ensure safety and performance, and finally, what impact the decision can have on the

individual. In parallel, controllers should also contextualise the information according to the domain in which the decision is taken, the impact of the decision, the data processed, the urgency to deliver the explanation and the audience that will receive the information.

It is also evaluated the adequacy of the AIA draft to promote transparency of AI systems. It is held that while the AIA draft does increases transparency of AI systems since it imposes many transparency obligations to providers and users of AI systems, it also falls short of providing full information to data subjects. In addition to that, this work suggests that the transparency of AI systems could be measured using a non-binding standard published by the Institute of Electrical and Electronics Engineers (IEEE 7001-2021 Standard for Transparency of Autonomous Systems) which considers both the intrinsic features of the AI solution and the kind of stakeholder that demands information from the AI system. Next, this work highlights the importance of the methods to convey the information to end-users of AI systems, which should be redesigned to be adequate for their purposes. This work also explains the initiatives to deliver information concerning datasets used to build AI models (such as Datasheets for datasets or the Dataset Nutrition Label) or documents to explain the models themselves (such as Model Cards for Model Reporting and AI FactSheets).

Along with transparency, fairness and non-discrimination are also important aspects of this work. Algorithmic biases may affect the performance of AI systems and it may result in unfair or discriminatory outcomes. There are different sources of biases in AI models, but they are generally generated in the early stages of AI system development. In particular, biases in AI can occur either in the data collection (for instance, due to the lack of statistical representativity or because of social preconceptions) or during the data preparation (where developers choose the features that the AI system is to evaluate).

This work later explains how algorithmic biases can be addressed. The EU legal framework has many provisions concerning non-discrimination and fairness. But as explained at large in this work, legislation in itself is not enough and for better protection of data subjects, more instruments should be evaluated. To assess and mitigate the risks posed by AI systems when processing personal data, the GDPR requires controllers to conduct a DPIA. However, DPIAs have constraints, in particular, DPIAs are mostly limited to identifying and evaluating only some risks posed by the

processing operations and intervention of data subjects or other stakeholders is not mandatory. This is the reason why impact assessments and audits were evaluated as means to mitigate the risks not covered by the accountability mechanisms mandatory required by the legislation.

An alternative to overcome those shortcomings is to conduct Human Rights Impact Assessments, which is a tool to identify, evaluate and mitigate the risks to human rights derived from the processing activities carried out by AI systems. Additionally, standards have been developed to address the ethical concerns in the design and use of AI systems. Algorithmic audits, in particular, to discover biases in automated decision-making is another tool to consider. However, it should be borne in mind that these mechanisms are voluntary non-binding accountability tools that organisations may implement to address ethical and fundamental rights concerns. The lack of binding effect disincentivises their wider adoption and there is no harmonization of the requirements that developers should abide by or follow. This said, there are currently several initiatives (mostly sectoral but also general) aimed at addressing the problems of processing personal data using AI systems.

The way forward: enhanced protection through better governance mechanisms

Finally, Chapter V acknowledges that more transparency and fairness obligations for controllers are not sufficient to reduce the risks posed by the processing of personal data using AI systems. Therefore, it evaluates some additional important governance and accountability mechanisms that should also be considered since enhanced accountability obligations will pave the way for better development and use of AI systems and reduce the impacts on fundamental rights. To begin with, it was considered the importance of building Public Registers of AI Systems as a method to enhance the transparency of these solutions towards individuals and society. These registers may serve as a public repository of AI systems or AI providers, whose main features can be scrutinised by users of AI systems or the public at large. Then, it was proposed that a specialized person or organization take the role of AI Ethical Officer to operationalize AI-related corporate values and guarantee a trustworthy development and use of AI across the organization. This officer could greatly assist DPOs in carrying out their tasks.

In addition, standardisation and certification of AI systems can be seen as methods to strengthen the protection of individuals and establish a common set of rules to protect personal data. On the one hand, standardisation in AI has been a very active field in the last years and many AI standards were published or are currently under development. These standards can provide a more solid groundwork for privacy practitioners since they specify high-level norms and principles that are oftentimes unclear or need concrete guidance for their full operationalization. However, the practice of regulating through standards is also subject to scrutiny by practitioners and academics, in particular, due to the lack of democratic accountability of standard-setting organisations. On the other hand, certification is a well-known mechanism that allows controllers to provide evidence of compliance with the data protection regulations. However, they do not prove compliance with the regulation by themselves, and still controllers are subject to further demonstration of compliance with the mandatory legal framework. Then, it was evaluated how Codes of Conduct can function as a method to set suitable compliance and ethical rules for institutions working in a specific domain or sector, giving autonomy to controllers to align best practices and find pragmatic solutions to the problems that originated in their industries.

Furthermore, the role of Data Protection Supervisory Authorities was assessed and it was proposed an alternative interpretation of the scope of their functions. Data protection authorities have a wide range of powers to monitor the implementation of the GDPR by controllers and processors, including the imposition of fines and a ban on processing. It was considered that while a literal interpretation of Art. 58 GDPR would not lead to granting power to order the destruction of algorithms developed and trained using tainted personal data (algorithmic disgorgement), since supervisory authorities may request the controller or processor ‘to bring processing operations into compliance’ with the legal framework ‘in a specified manner and within a specified period’, this provision might be used to compel companies to destroy the algorithms developed using tainted data.

Finally, this work evaluated some particular measures to operationalize the principle of Data Protection by Design and By Default, in particular, those concerning the reduction of the identifiability of personal data. While anonymisation/pseudonymisation and encryption are considered standard techniques

to reduce the risks of processing personal data, the use of synthetic data (data created from real data that mimics the statistical proprieties of original datasets) is less common and can greatly contribute to mitigating those risks in certain stages of the AI lifecycle.

As shown, the proposed way forward in this work consists in looking beyond individual rights protected by the GDPR, reinforcing accountability obligations of data controllers and processors and, chiefly, incorporating instruments that are usually considered out of the realm of data protection and privacy for an integral protection of the fundamental rights of individuals against the risks posed by AI systems. Only a combination of methods and solutions will be effective to achieve that objective.

REFERENCES

Bibliography

Aaronson S, ‘Why Trade Agreements Are Not Setting Information Free: The Lost History and Reinvigorated Debate over Cross-Border Data Flows, Human Rights, and National Security’ (2015) 14 World Trade Review 671

Abdelmoula AK, ‘Bank Credit Risk Analysis with K-Nearest-Neighbor Classifier: Case of Tunisian Banks’ (2015) 14 Journal of Accounting and Management Information Systems 79

Ada Lovelace Institute, ‘Examining the Black Box. Tools for Assessing Algorithmic Systems’ (2020)

Agence nationale de la sécurité des systèmes d’information, ‘Avis Scientifique et Technique de l’ANSSI Sur La Migration Vers La Crypographie Post-Quantique’ (2022)

Agencia Española de Protección de Datos, ‘Guía Práctica Para Las Evaluaciones de Impacto En La Protección de Los Datos Sujetas Al RGPD’ (2019)

——, ‘Adecuación Al RGPD de Tratamientos Que Incorporan Inteligencia Artificial. Una Introducción’ (2020)

Anjaria M and Guddeti RMR, ‘A Novel Sentiment Analysis of Social Networks Using Supervised Learning’ (2014) 4 Social Network Analysis and Mining 1

Antonakis AC and Sfakianakis ME, ‘Assessing Naïve Bayes as a Method for Screening Credit Applicants’ (2009) 36 Journal of Applied Statistics 537

Armknecht F and others, ‘General Impossibility of Group Homomorphic Encryption in the Quantum World’ in H Krawczyk (ed), Public-Key Cryptography – PKC 2014. PKC 2014. Lecture Notes in Computer Science (Springer 2014)

Arnold M and others, ‘FactSheets: Increasing Trust in AI Services through Supplier’s Declarations of Conformity’ (2019) 63 IBM Journal of Research and Development 1

Article 29 Data Protection Working Party, ‘Opinion 4/2007 on the Concept of Personal Data’ (2007)

——, ‘Opinion 3/2010 on the Principle of Accountability’ (2010)

——, ‘Opinion 3/2012 on Developments in Biometric Technologies’ (2012)

——, ‘Opinion 03/2013 on Purpose Limitation’ (2013)

——, ‘Opinion 05/2014 on Anonymisation Techniques’ (2014)

——, ‘Opinion 06/2014 on the Notion of Legitimate Interests of the Data Controller under Article 7 of Directive 95/46/EC’ (2014)

——, ‘Guidelines on Data Protection Officers’ (2017)

——, ‘Guidelines on the Right to “Data Portability”’ (2017)

——, ‘Guidelines on Automated Individual Decision-Making and Profiling for the Purposes of Regulation 2016/679’ (2018)

Bandy J, ‘Problematic Machine Behavior: A Systematic Literature Review of Algorithm Audits’, Proceedings of the ACM on Human-Computer Interaction Volume 5 Issue CSCW 1 (2021)

Bellovin SM, Dutta PK and Reitinger N, ‘Privacy and Synthetic Datasets’ (2019) 22 Stanford Technology Law Review

Benolie U and Becher SI, ‘The Duty to Read the Unreadable’ (2019) 60 Boston College Law Review 2256

Bibal A and Frénay B, ‘Interpretability of Machine Learning Models and Representations: An Introduction’, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning (2016)

Biega A and Finck M, ‘Reviving Purpose Limitation and Data Minimisation in Personalisation, Profiling and Decision-Making Systems’ (2021) 21–04

Binns R and Veale M, ‘Is That Your Final Decision? Multi-Stage Profiling, Selective Effects, and Article 22 of the GDPR’ (2021) 11 International Data Privacy Law 319

Bishop C, Pattern Recognition and Machine Learning (Springer 2011)

Bodo B and others, ‘Tackling the Algorithmic Control Crisis –the Technical, Legal, and Ethical Challenges of Research into Algorithmic Agents’ (2019) 19 Yale Journal of Law & Technology 133

Brennan-Marquez K, Levy K and Susser D, ‘Strange Loops: Apparent versus Actual Human Involvement in Automated Decision Making’ (2019) 34 Berkeley Technology Law Journal 745

Brkan M, ‘The Concept of Essence of Fundamental Rights in the EU Legal Order: Peeling the Onion to Its Core’ (2018) 14 European Constitutional Law Review 332

——, ‘Do Algorithms Rule the World? Algorithmic Decision-Making and Data Protection in the Framework of the GDPR and Beyond’ (2019) 27 International Journal of Law and Information Technology 91

Brown S, Davidovic J and Hasan A, ‘The Algorithm Audit: Scoring the Algorithms That Score Us’ (2021) 8 Big Data and Society

Burrell J, ‘How the Machine “Thinks”: Understanding Opacity in Machine Learning Algorithms’ [2016] Big Data and Society 1

Butterworth M, ‘The ICO and Artificial Intelligence: The Role of Fairness in the GDPR Framework’ (2018) 34 Computer Law & Security Review 257

Bygrave LA, ‘Minding the Machine: Article 15 of the EC Data Protection Directive and Automated Profiling’ (Elsevier Advanced Technology, 1 January 2001) 17

——, ‘Article 22. Automated Individual Decision-Making, Including Profiling’ in Christopher Kuner and Lee A Bygrave (eds), The EU General Data Protection Regulation (GDPR): A Commentary (OUP 2020)

Bygrave LA and Tosoni L, ‘Article 4(1). Personal Data’ in Christopher Kuner and Lee A Bygrave (eds), The EU General Data Protection Regulation (GDPR): A Commentary (OUP 2020)

Campbell M, Hoane J and Hsu F, ‘Deep Blue’ (2002) 134 Artificial Intelligence 57

Carlini N and others, ‘The Secret Sharer: Evaluating and Testing Unintended Memorization in Neural Networks’, SEC’19: Proceedings of the 28th USENIX Conference on Security Symposium (2019)

Castets-Renard C, ‘6 - Human Rights and Algorithmic Impact Assessment for Predictive Policing’ in Oreste Pollicino and Hans-W Micklitz (eds), Constitutional Challenges in the Algorithmic Society (CUP 2019)

——, ‘Accountability of Algorithms in the GDPR and Beyond: A European Accountability of Algorithms in the GDPR and Beyond: A European Legal Framework on Automated Decision-Making Legal Framework on Automated Decision-Making’ (2019) 30 Fordham Intellectual Property, Media and Entertainment Law 91

Cate FH and Mayer-Schö V, ‘Notice and Consent in a World of Big Data’ (2013) 3 International Data Privacy Law 67

Cavoukian A, Taylor S and Abrams ME, ‘Privacy by Design: Essential for Organizational Accountability and Strong Business Practices’ (2010) 3 Identity in the Information Society 405

Center for Humanitarian Dialogue, ‘Code of Conduct on Artificial Intelligence in Military Systems’ (2021)

Citron DK and Pasquale F, ‘The Scored Society: Due Process for Automated Decisions’ (2014) 89 Washington Law Review 1

Coeckelbergh M, AI Ethics (MIT Press 2020)

Commission nationale de l’informatique et des libertés, ‘Privacy Impact Assessment (PIA) 1: Methodology’ (2018)

Commission Nationale de l’Informatique et des Libertés, ‘Privacy Impact Assessment (PIA). Knowledge Bases’ (2018)

Council of Europe, ‘The Protection of Individuals with Regard to Automatic Processing of Personal Data in the Context of Profiling. Recommendation CM/Rec(2010)13 and Explanatory Memorandum’ (2011)

——, ‘Explanatory Report to the Protocol Amending the Convention for the Protection of Individuals with Regard to Automatic Processing of Personal Data’ (2018)

——, ‘Declaration by the Committee of Ministers on the Manipulative Capabilities of Algorithmic Processes - Decl(13/02/2019)1’ (2019)

Council of Europe - Ad Hoc Committee on Artificial Intelligence (CAHAI), ‘Human Rights, Democracy and Rule of Law Impact Assessment of AI Systems’ (2021)

Council of Europe Commissioner for Human Rights, ‘Unboxing Artificial Intelligence: 10 Steps to Protect Human Rights’ (2019)

Craig P and de Búrca G, EU Law. Text, Cases and Materials (7th edn, OUP 2020)

Dariusz Kloza and others, ‘Data Protection Impact Assessment in the European Union: Developing a Template for a Report from the Assessment Process’ (2020)

Data Protection Commission, ‘Guide to Data Protection Impact Assessments (DPIAs)’ (2019)

Datatilsynet, ‘Artificial Intelligence and Privacy’ (2018)

Davenport T and Kalakota R, ‘The Potential for Artificial Intelligence in Healthcare’ (2019) 6 Future Healthcare Journal 94

De Gregorio G, Digital Constitutionalism in Europe. Reframing Rights and Powers in the Algorithmic Society (CUP 2022)

De Hert P and Papakonstantinou V, ‘The New General Data Protection Regulation: Still a Sound System for the Protection of Individuals?’ (2016) 32 Computer Law & Security Review 179

Development O for EC and, ‘Recommendation of the Council on Artificial Intelligence, C/MIN(2019)3/FINAL’ (2019)

Dhiraj Gurkhe, Niraj Pal and Rishit Bathia, ‘Effective Sentiment Analysis of Social Media Datasets Using Naive Bayesian Classification’ (2014) 99 International Journal of Computer Applications 1

Dignum V, Responsible Artificial Intelligence. How to Develop and Use AI in a Responsible Way (Springer 2019)

Doshi-Velez F and Kortz M, ‘Accountability of AI Under the Law: The Role of Explanation’ (2017)

Drexel J, ‘Legal Challenges of the Changing Role of Personal and Non-Personal Data in the Data Economy’ in Alberto De Franceschi, Reiner Schulze and Oreste Pollicino (eds), Digital Revolution - New Challenges for Law (Nomos 2019)

Ducato R, ‘89. Garanzie e Deroghe Relative Al Trattamento a Fini Di Archiviazione Nel Pubblico Interesse, Di Ricerca Scientifica, o Storica o a Fini Statistici’ in Oreste Pollicino and Roberto D’Orazio (eds), Codice della Privacy e Data Protection (Guiffrè Francis Lefebvre 2021)

Ebers M, ‘Standardizing AI - The Case of the European Commission’s Proposal for an Artificial Intelligence Act’ in Larry Di Matteo, Cristina Poncibò and Michel Cannarsa (eds), The Cambridge Handbook of Artificial Intelligence: Global Perspectives on Law and Ethics (CUP)

——, ‘The European Commission’s Proposal for an Artificial Intelligence Act - Critical Assessment by Members of the Robotics and AI Law Society (RAILS)’ (2021) 4 J 589

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 and Technology Review 18

Emam K El, Mosquera L and Hoptroff R, Practical Synthetic Data Generation. Balancing Privacy and the Broad Availability of Data (O’Reilly 2020)

Engineers I of E and E, ‘IEEE 7000-2021 Standard Model Process for Addressing Ethical Concerns During System Design’ (2021)

Esteva A and others, ‘Dermatologist-Level Classification of Skin Cancer with Deep Learning Neural Networks’ (2017) 542 Nature 115

European Commission's High-Level Expert Group on Artificial Intelligence, 'Ethics Guidelines for Trustworthy AI' (2019)

European Commission, 'The Economics of Ownership, Access and Trade in Digital Data' (2017) 2017–01

European Commission - Joint Research Centre, 'Artificial Intelligence - A European Perspective' (2018)

——, 'AI Watch. Defining Artificial Intelligence. Towards an Operational Definition and Taxonomy of Artificial Intelligence' (2020)

——, 'AI Watch: AI Standardisation Landscape State of Play and Link to the EC Proposal for an AI Regulatory Framework' (2021)

European Commission for the Efficiency of Justice, 'Possible Introduction of a Mechanism for Certifying Artificial Intelligence Tools and Services in the Sphere of Justice and the Judiciary: Feasibility Study' (2020)

European Committee for Standarization, 'CWA 17145-2:2017 (E) - Ethics Assessment for Research and Innovation - Part 2: Ethical Impact Assessment Framework' (2017)

European Data Protection Board, 'Guidelines on the Application and Setting of Administrative Fines for the Purposes of the Regulation 2016/679' (2017)

——, 'Guidelines 1/2018 on Certification and Identifying Certification Criteria in Accordance with Articles 42 and 43 of the Regulation' (2018)

——, 'Guidelines on Data Protection Impact Assessment (DPIA) and Determining Whether Processing Is "Likely to Result in a High Risk" for the Purposes of the GDPR' (2018)

——, 'Guidelines on Transparency under Regulation 2016/679' (2018)

——, 'Guidelines 1/2019 on Codes of Conduct and Monitoring Bodies under Regulation 2016/679' (2019)

——, 'Guidelines 2/2019 on the Processing of Personal Data under Article 6(1)(b) GDPR in the Context of the Provision of Online Services to Data Subjects' (2019)

——, 'Guidelines 05/2020 on Consent under Regulation 2016/679' (2020)

——, 'Guidelines 4/2019 on Article 25 Data Protection by Design and by Default' (2020)

——, 'Guidelines 5/2019 on the Criteria of the Right to Be Forgotten in the Search Engines Cases under the GDPR (Part 1)' (2020)

——, ‘Guidelines 02/2021 on Virtual Voice Assistants’ (2021)

——, ‘Guidelines 8/2020 on the Targeting of Social Media Users’ (2021)

——, ‘Guidelines 01/2022 on Data Subject Rights - Right of Access’ (2022)

European Data Protection Board and European Data Protection Supervisor, ‘Joint Opinion 5/2021 on the Proposal for a Regulation of the European Parliament and of the Council Laying down Harmonised Rules on Artificial Intelligence’

European Data Protection Supervisor, ‘Accountability on the Ground Part I: Records, Registers and When to Do Data Protection Impact Assessments’ (2019)

——, ‘Accountability on the Ground Part II: Data Protection Impact Assessments & Prior Consultation’ (2019)

European Data Protection Suprevisor, ‘Opinion 4/2020 on the European Commission’s White Paper on Artificial Intelligence. A European Approach to Excellence and Trust’ (2020)

European Parliamentary Research Service, ‘Artificial Intelligence: From Ethics to Policy’ (2020)

——, ‘The Impact of the General Data Protection Regulation (GDPR) on Artificial Intelligence’ (2020)

European Union Agency for Cybersecurity, ‘Artificial Intelligence Cybersecurity Challenges. Threat Landscape for Artificial Intelligence’ (2020)

——, ‘Post-Quantum Cryptography: Current State and Quantum Mitigation’ (2021)

——, ‘Data Protection Engineering. From Theory to Practice’ (2022)

European Union Agency for Fundamental Rights, ‘Preventing Unlawful Profiling Today and in the Future: A Guide’ (2018)

European Union Agency for Fundamental Rights and Council of Europe, Handbook on European Data Protection Law (2018)

European Union Agency For Network And Information Security, ‘Privacy by Design in Big Data. An Overview of Privacy Enhancing Technologies in the Era of Big Data Analytics’ (2015)

Europol, ‘Malicious Uses and Abuses of Artificial Intelligence’ (2020)

Evtimov I and others, ‘Is Tricking a Robot Hacking?’ (2019) 34 Berkeley Technology Law Journal 891

Finck M and Pallas F, ‘They Who Must Not Be Identified — Distinguishing Personal from Non-Personal Data under the GDPR’ (2020) 10 International Data Privacy Law 11

Fjeld J and others, ‘Principled Artificial Intelligence: Mapping Consensus in Ethical and Rights-Based Approaches to Principles for AI’ (2020)

Floridi L, ‘Soft Ethics and the Governance of the Digital’ (2018) 31 Philosophy & Technology 1

Floridi L and Strait A, ‘Ethical Foresight Analysis: What It Is and Why It Is Needed?’ (2020) 30 Minds and Machines 77

Forgó N, Hänold S and Schütze B, ‘The Principle of Purpose Limitation and Big Data’ in Marcelo Corrales, Mark Fenwick and Nikolaus Forgó (eds), New Technology, Big Data and the Law (Springer 2017)

Froomkin M, ‘Big Data: Destroyer of Informed Consent’ (2019) 21 Yale Journal of Law and Technology 27

Fry A and others, ‘Comparison of Sociodemographic and Health-Related Characteristics of UK Biobank Participants with General Population’ (2017) 186 American Journal of Epidemiology 1026

Gebru T and others, ‘Datasheets for Datasets’ (2021) 64 Communications of the ACM 86

German Federal Office for Information Security, ‘Quantum-Safe Cryptography – Fundamentals, Current Developments and Recommendations’ (2021)

German Institute of Standarisation, ‘German Standardization Roadmap on Artificial Intelligence’ (2020)

Ginart A and others, ‘Making AI Forget You: Data Deletion in Machine Learning’, Proceedings of the 33rd International Conference on Neural Information Processing Systems (2019)

Gonzalez Fuster G, The Emergence of Personal Data Protection as a Fundamental Right of the EU (Springer 2014)

——, ‘Article 18. Right to Restriction of Processing’ in Christopher Kuner and Lee A Bygrave (eds), The EU General Data Protection Regulation (GDPR): A Commentary (OUP 2020)

Goodman B and Flaxman S, ‘European Union Regulations on Algorithmic Decision-Making and a “Right to Explanation”’

——, ‘European Union Regulations on Algorithmic Decision-Making and a “Right to Explanation”’ [2017] AI Magazine

Götzmann N, ‘Introduction to the Handbook on Human Rights Impact Assessment: Principles, Methods and Approaches’ in Nora Götzmann (ed), Handbook on Human Rights Impact Assessment. Research Handbooks on Impact Assessment series (Elgar 2019)

Greene T and others, ‘Adjusting to the GDPR: The Impact on Data Scientists and Behavioral Researchers’ (2019) 7 Big Data 140

Gregorio G De and Torino R, ‘Privacy, Protezione Dei Dati Personali e Big Data’ in Vincenzo Franceschelli and Emilio Tosi (eds), Privacy Digitale. Riservatezza e protezione dei dati personali tra GDPR e nuovo Codice Privacy (Guiffrè Francis Lefebvre 2019)

Hacker P, ‘Teaching Fairness to Artificial Intelligence: Existing and Novel Strategies Against Algorithmic Discrimination under EU Law’ (2018) 55 Common Market Law Review 1143

Haenlein M and Kaplan A, ‘A Brief History of Artificial Intelligence: On the Past, Present, and Future of Artificial Intelligence’ (2019) 61 California Management Review 5

Hallinan D and Borgesius FZ, ‘Opinions Can Be Incorrect (in Our Opinion)! On Data Protection Law’s Accuracy Principle’ (2020) 10 International Data Privacy Law 1

Heverly R, ‘The Information Semicommons’ (2003) 18 Berkeley Technology Law Journal

High-Level Expert Group on AI, ‘The Assessment List for Trustworthy Artificial Intellignece (ALTAI)’ (2020)

Human Rights Council, ‘Guiding Principles on Human Rights Impact Assessments of Economic Reforms. A/HRC/40/57’ (2019)

——, ‘Artificial Intelligence and Privacy, and Children’s Privacy. Report of the Special Rapporteur on the Right to Privacy, Joseph A. Cannataci. A/HRC/46/37’ (2021)

Hupperich T and others, ‘An Empirical Study on Online Price Differentiation’, Proceedings of the Eighth ACM Conference on Data and Application Security and Privacy (2018)

Igor H and others, ‘Application of Neural Networks in Computer Security’ (2014) 69 Procedia Engineering 1209

Information Commissioner’s Office, ‘Big Data, Artificial Intelligence, Machine Learning and Data Protection Data Protection Act and General Data Protection Regulation’ (2017)

——, ‘Explaining Decisions Made with AI’ (2020)

——, ‘Guidance on AI and Data Protection’ (2020)

——, ‘The Use of Live Facial Recognition Technology in Public Places’ (2021)

Institute of Electrical and Electronics Engineers, ‘IEEE 7001-2021 Standard for Transparency of Autonomous Systems’ (2022)

Interactive Advertising Bureau Europe, ‘Guidance: GDPR Data Protection Impact Assessment (DPIA) for Digital Advertising under GDPR’ (2020)

International Standard Organisation, ‘ISO/IEC 29100:2011 Information Technology — Security Techniques — Privacy Framework’ (2011)

——, ‘ISO/IEC 17000:2020(En) Conformity Assessment — Vocabulary and General Principles’ (2020)

——, ‘ISO/IEC TR 24028:2020 Information Technology — Artificial Intelligence — Overview of Trustworthiness in Artificial Intelligence’ (2020)

——, ‘ISO/IEC TR 24027:2021 Information Technology — Artificial Intelligence (AI) — Bias in AI Systems and AI Aided Decision Making’ (2021)

Introna LD and Nissenbaum H, ‘Shaping the Web: Why the Politics of Search Engines Matters’ (2000) 16 The Information Society 169

Itoo F, Meenakshi and Singh S, ‘Comparison and Analysis of Logistic Regression, Naïve Bayes and KNN Machine Learning Algorithms for Credit Card Fraud Detection’ (2020) 13 International Journal of Information Technology (Singapore) 1503

Izzo Z and others, ‘Approximate Data Deletion from Machine Learning Models’, Proceedings of The 24th International Conference on Artificial Intelligence and Statistics (2021)

J Lee Riccardi, ‘The German Federal Data Protection Act of 1977: Protecting the Right to Privacy?’ (1983) 6 Boston College International and Comparative Law Review 243

Janssen HL, ‘An Approach for a Fundamental Rights Impact Assessment to Automated Decision-Making’ (2020) 10 International Data Privacy Law 76

Joachims T, ‘Optimizing Search Engines Using Clickthrough Data’, ACM SIGKDD international conference on Knowledge discovery and data mining (2002)

Jordan MI, ‘Artificial Intelligence—The Revolution Hasn’t Happened Yet’ [2019] Harvard Data Science Review

Joy Buolamwini and Gebru T, ‘Gender Shades Intersectional Accuracy Disparities in Commercial Gender Classification’, Conference on Fairness, Accountability and Transparency (2018)

Kaminski ME, ‘The Right to Explanation, Explained’ (2019) 34 Berkeley Technology Law Journal 218

Kaminski ME and Malgieri G, ‘Algorithmic Impact Assessments under the GDPR: Producing Multi-Layered Explanations’ (2021) 11 International Data Privacy Law 125

Kay M, Matuszek C and Munson SA, ‘Unequal Representation and Gender Stereotypes in Image Search Results for Occupations’, Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems (2015)

Kazim E and others, ‘AI Auditing and Impact Assessment: According to the UK Information Commissioner’s Office’ (2021) 1 AI and Ethics 1

——, ‘Systematizing Audit in Algorithmic Recruitment’ (2021) 9 Journal of Intelligence 46

Klimas T and Vaiciukaite J, ‘The Law of Recitals in European Community Legislation ’ (2008) 15 ILSA Journal or International and Comparative Law 61

Kon G, ‘Does Anyone Read Privacy Notices? The Facts’ (Linklaters DigiLink, 2020) <https://www.linklaters.com/en/insights/blogs/digilinks/does-anyone-read-privacy-notices-the-facts> accessed 27 December 2021

Koning ME, ‘The Purpose and Limitations of Purpose Limitation’ (Radboud University - PhD Thesis 2020)

Kosinski M, Stillwell D and Graepel T, ‘Private Traits and Attributes Are Predictable from Digital Records of Human Behavior’ (2013) 110 Proceedings of the National Academy of Sciences of the United States of America 5802

Kranenborg H, ‘Article 17. Right to Erasure ('right to Be Forgotten’)’ in Christopher Kuner, Lee A Bygrave and Christoper Docksey (eds), The EU General Data Protection Regulation (GDPR) (OUP 2020)

Kroll J and others, ‘Accountable Algorithms’ (2017) 165 University of Pennsylvania Law Review 633

Laboratoire National de Métrologie et d'Essais, 'Certification Standard of Processes for AI. Design, Development, Evaluation and Maintenance in Operational Conditions' (2021)

Lambrecht A and Tucker C, 'Algorithmic Bias? An Empirical Study of Apparent Gender-Based Discrimination in the Display of STEM Career Ads' (2019) 65 Management Science 2966

Landers R and Behrend T, 'Auditing the AI Auditors: A Framework for Evaluating Fairness and Bias in High Stakes AI Predictive Models' [2022] American Pshycologist 1

Lauter K, 'Private AI: Machine Learning on Encrypted Data' in Tomás Chacón, Rosa Rebollo and Inmaculada Higueras Donat (eds), Recent Advances in Industrial and Applied Mathematics. SEMA SIMAI Springer Series (Springer 2022)

Lehr D and Ohm P, 'Playing with the Data: What Legal Scholars Should Learn About Machine Learning' (2017) 51 UC Davis Law Review 653

Leslie D, 'Understanding Artificial Intelligence Ethics and Safety. A Guide for the Responsible Design and Implementation of AI Systems in the Public Sector' (2019)

——, 'Human Rights, Democracy, and the Rule of Law Assurance Framework for AI Systems: A Proposal Prepared for the Council of Europe's Ad Hoc Committee on Artificial Intelligence' (2022)

Li T, 'Algorithmic Destruction' (2022) Forthcomin SMU Law Review 1

Li T, Fosch Villaronga E and Kieseberg P, 'Humans Forget, Machines Remember: Artificial Intelligence and the Right to Be Forgotten' (2018) 34 Computer Law & Security Review 308

Liberman A and Rotarius T, 'Pre-Employment Decision Trees: Job Applicant Self-Election.' (2000) 18 Health Care Manager 48

Liu H-W, 'Data Localization and Digital Trade Barriers: ASEAN in Megaregionalism', ASEAN Law in the New Regional Economic Order (Cambridge University Press 2019)

Macdonald DA and Streatfeild CM, 'Personal Data Privacy and the WTO' (2014) 36 Houston Jounal of International Law 625

Malgieri G, 'The Concept of Fairness in the GDPR: A Linguistic and Contextual Interpretation', Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency (2020)

——, ‘5. Principi Applicabili Al Trattamento Di Dati Personali’ in Oreste Pollicino and Roberto D’Orazio (eds), Codice della Privacy e Data Protection (Giufrè Francis Lefebvre 2021)

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

Malgieri G and Niklas J, ‘Vulnerable Data Subjects’ (2020) 37 Computer Law and Security Review 1

Manheim K and Kaplan L, ‘Artificial Intelligence: Risks to Privacy and Democracy’ (2017) 21 Yale Journal of Law & Technology 106

Mantelero A, ‘Personal Data for Decisional Purposes in the Age of Analytics: From an Individual to a Collective Dimension of Data Protection’ (2016) 32 Computer Law & Security Review 238

——, ‘The Common EU Approach to Personal Data and Cybersecurity Regulation’ (2021) 28 International Journal of Law and Information Technology 297

Mantelero A and Esposito MS, ‘An Evidence-Based Methodology for Human Rights Impact Assessment (HRIA) in the Development of AI Data-Intensive Systems’ (2021) 41 Computer Law & Security Review 1

Massey A and Breaux TD, ‘Chapter 7: Interference’ in Travis D Breaux (ed), An Introduction to Privacy for Technology Professionals (IAPP 2020)

Mattioli M, ‘The Data-Pooling Problem’ (2018) 32 Berkeley Technology Law Journal 1

Matwyshyn AM, ‘The Internet of Bodies The Internet of Bodies Repository Citation Repository Citation’ (2019) 61 William & Mary Law Review 77

Mayer-Schönberger V and Padova Y, ‘Regime Change? Enabling Big Data through Europe’s New Data Protection Regulation’ (2016) 17 Science and Technology Law Review 315

McFadden M and others, ‘Harmonising Artificial Intelligence: The Role of Standards in the EU AI Regulation’ (2021)

Medvedeva M, Vols M and Wieling M, ‘Using Machine Learning to Predict Decisions of the European Court of Human Rights’ (2020) 28 Artificial Intelligence and Law 237

Mendoza I and Bygrave LA, ‘The Right Not to Be Subject to Automated Decisions Based on Profiling’ (2017) 2017–20 University of Oslo Faculty of Law Legal Studies Research Paper Series 1

Miller T, ‘Explanation in Artificial Intelligence: Insights from the Social Sciences’ (2019) 267 Artificial Intelligence 1

Mitchell M and others, ‘Model Cards for Model Reporting’, FAT* ’19: Proceedings of the Conference on Fairness, Accountability, and Transparency (2019)

Mitchell T, Machine Learning (McGraw-Hill 1997)

Mökande J and Axente M, ‘Ethics-Based Auditing of Automated Decision-Making Systems: Intervention Points and Policy Implications’ [2021] AI & Society 1

Molnar C, Interpretable Machine Learning. A Guide for Making Black Box Models Explainable (Leanpub 2020)

Moraes TG and others, ‘Open Data on the COVID-19 Pandemic: Anonymisation as a Technical Solution for Transparency, Privacy, and Data Protection’ (2021) 11 International Data Privacy Law 32

Murphy K, Machine Learning. A Probabilistic Perspective (MIT Press 2012)

Narayanan A, ‘21 Fairness Definitions and Their Politics’, Conference on Fairness, Accountability, and Transparency (2018)

National Institute of Standards and Technology, ‘Status Report on the Second Round of the NIST Post-Quantum Cryptography Standardization Process’ (2020)

——, ‘Four Principles of Explainable Artificial Intelligence’ (2021)

——, ‘Psychological Foundations of Explainability and Interpretability in Artificial Intelligence’ (2021)

——, ‘NIST Special Publication 1270. A Proposal for Identifying and Managing Bias in Artificial Intelligence.’ (2022)

Nielsen RØ, Gurzawsk AM and Brey P, ‘Satori Project. Principles and Approaches in Ethics Assessment. Ethical Impact Assessment and Conventional Impact Assessment. Annex 1.A’ (2015)

Norwegian Data Protection Authority, ‘Artificial Intelligence and Privacy’ (2018)

Obar JA and Oeldorf-Hirsch A, ‘The Biggest Lie on the Internet: Ignoring the Privacy Policies and Terms of Service Policies of Social Networking Services’ (2020) 23 Information, Communication & Society 128

Office of the High Commissioner for Human Rights, ‘Data Privacy Guidelines in Context of Artificial Intelligence’ (2020)

Office of the United Nations High Commissioner for Human Rights, ‘Guiding Principles on Business and Human Rights. Implementing the United Nations “Protect, Respect and Remedy” Framework’ (2011)

Ohm P, ‘Chapter 12: Throttling Machine Learning’ in Mireille Hildebrandt and Kieron O’Hara (eds), Life and the Law in the Era of Data-Driven Agency (Elgar 2020)

Oostveen M, Protecting Individuals Against the Negative Impact of Big Data: Potential and Limitations of the Privacy and Data Protection Law (Wolters Kluwer 2018)

Organisation for Economic Co-operation, ‘Guidelines for Multinational Enterprises’ (2011)

Organisation for Economic Co-operation and Development, ‘Digital Trade - Developing a Framework for Analysis’ (2017) 205

——, ‘Trade and Cross-Border Data Flows’ (2019) 220

Orla Lynskey, The Foundations of EU Data Protection Law (OUP 2016)

Party A 29 DPW, ‘Guidelines on Personal Data Breach Notification under Regulation 2016/679’ (2018)

Pellecchia E, ‘Privacy, Decisioni Automatizzate e Algoritmi’ in Vincenzo Franceschelli and Emilio Tosi (eds), Privacy Digitale. Riservatezza e protezione dei dati personali tra GDPR e nuovo Codice PrivacyRiservatezza e protezione dei dati personali tra GDPR e nuovo Codice Privacy (Guiffrè Francis Lefebvre 2019)

Pew Research Center, ‘4. Americans’ Attitudes and Experiences with Privacy Policies and Laws’ (Americans and Privacy: Concerned, Confused, and Feeling lack of Control over their Personal Information, 2019) <https://www.pewresearch.org/internet/2019/11/15/americans-attitudes-and-experiences-with-privacy-policies-and-laws/> accessed 27 December 2021

Pollicino O, ‘The Transatlantic Dimension of the Judicial Protection of Fundamental Rights Online’ (2021) 1 The Italian Review of International and Comparative Law 277

Pollicino O and Bassini M, ‘Bridge Is Down, Data Truck Can’t Get Through...A Critical View of the Schrems Judgment in the Context of European Constitutionalism’ in G Ziccardi Capaldo (ed), The Global Community Yearbook of International Law and Jurisprudence 2016 (Oxford University Press 2017)

——, ‘8. Protezione Dei Dati Di Carattere Personale’ in Roberto D’Orazio and others (eds), Codice della Privacy e Data Protection (Giufrè Francis Lefebvre 2021)

Pollicino O, Cannataci J and Falce V, ‘Introduction’ in Oreste Pollicino, Joe Cannataci and Valeria Falce (eds), Legal Challenges of Big Data (Elgar 2020)

Pollicino O and D’Antonio V, ‘The Right to Be Forgotten in Italy’ in Franz Werro (ed), The Right To Be Forgotten. A Comparative Study of the Emergent Right’s Evolution and Application in Europe, the Americas, and Asia (Springer 2020)

Pollicino O and De Giovanni G, ‘A Constitutional-Driven Change of Heart ISP Liability and Artificial Intelligence in the Digital Single Market’ (2019) 18 The Global Community Yearbook of International Law and Jurisprudence

Pollicino O and De Gregorio G, ‘Privacy or Transparency? A New Balancing of Interests for the “Right to Be Forgotten” of Personal Data Published in Public Registers’ [2017] The Italian Law Journal 647

Pollicino O and Nicola FG, ‘The Balkanization of Data Privacy Regulation’ (2020) 123 West Virginia Law Review 115

Porcari A and Mocchio E, ‘Managing Social Impacts and Ethical Issues of Research and Innovation: The CEN/WS 105 Guidelines to Innovate Responsibly’ in Emad Yaghmaei and Ibo van de Poel (eds), Assessment of Responsible Innovation. Methods and Practices (Rutledge 2020)

Rajula HSR and others, ‘Comparison of Conventional Statistical Methods with Machine Learning in Medicine: Diagnosis, Drug Development, and Treatment’ (2020) 56 Medicina 455

Riccio GM and Pezza F, ‘Certifications Mechanism and Liability Rules under the GDPR. When the Harmonisation Becomes Unification’ in Alberto De Franceschi, Reiner Schulze and Oreste Pollicino (eds), Digital Revolution - New Challenges for Law (Beck 2019)

Rich ML, ‘Machine Learning, Automated Suspicion Algorithms, and the Fourth Amendment’ (2016) 164 University of Pennsylvania Law Review 871

Roessler B, ‘Should Personal Data Be a Tradable Good? On the Moral Limits of Markets in Privacy’ in Beate Roessler and Dorota Mokrosinska (eds), Social Dimensions of Privacy: Interdisciplinary Perspectives (CUP 2015)

Roig A, ‘Safeguards for the Right Not to Be Subject to a Decision Based Solely on Automated Processing (Article 22 GDPR)’ (2018) 8 European Journal of Law and Technology 1

Rubinstein IS and Good N, ‘The Trouble with Article 25 (and How to Fix It): The Future of Data Protection by Design and Default’ (2020) 10 International Data Privacy Law 37

Russell S and Norvig P, Artificial Intelligence: A Modern Approach (3rd edn, Pearson 2010)

Sahu M, ‘Plagiarism Detection Using Artificial Intelligence Technique In Multiple Files’ (2016) 5 International Journal of Scientific and Technology Research 111

Schwartz B, Administrative Law (3rd edn, Little Brown & Co 1991)

Schwartz PM, ‘Property, Privacy, and Personal Data’ (2004) 117 Harvard Law Review 2055

Scope Europe, ‘EU Cloud Code of Conduct’ (2021)

Selbst AD and Powles J, ‘Meaningful Information and the Right to Explanation’ (2017) 7 International Data Privacy Law 233

Slaughter R, ‘Algorithms and Economic Justice: A Taxonomy of Harms and a Path Forward for the Federal Trade Commission’ (2021) Special Pu Yale Journal of Law & Technology 1

Standards Council of Canada, ‘CAN/CIOSC 101:2019 - Ethical Design and Use of Automated Decision Systems’ (2019)

Stucke ME and Ezrachi A, ‘How Digital Assistants Can Harm Our Economy, Privacy, and Democracy’ (2018) 32 Berkeley Technology Law Journal 1239

Surden H, ‘Artificial Intelligence and Law: An Overview’ (2019) 35 Georgia State University Law Review 1305

Swedish Data Protection Authority, ‘Supervision Pursuant to the General Data Protection Regulation (EU) 2016/679 – Facial Recognition Used to Monitor the Attendance of Students’ (2019)

Terwangne C de, ‘Article 5. Principles Relating to Processing of Personal Data’ in Christopher Kuner, Lee A Bygrave and Christoper Docksey (eds), The EU General Data Protection Regulation (GDPR): A Commentary (OUP 2020)

Tharani A and others, ‘The COMPASS Self-Check Tool. Enhancing Organizational Learning for Responsible Innovation through Self-Assessment’ in Emad Yaghmaei

and Ibo van de Poel (eds), Assessment of Responsible Innovation. Methods and Practices (Routledge 2020)

Tosoni L, ‘Article 4(5). Pseudonymisation’ in Christopher Kuner, Lee A Bygrave and Christopher Docksey (eds), The EU General Data Protection Regulation (GDPR). A commentary (OUP 2020)

——, ‘The Right to Object to Automated Individual Decisions: Resolving the Ambiguity of Article 22(1) of the General Data Protection Regulation’ (2021) 11 International Data Privacy Law 145

Tucker A and others, ‘Generating High-Fidelity Synthetic Patient Data for Assessing Machine Learning Healthcare Software’ (2020) 3 npj Digital Medicine 1

Tutt A, ‘An FDA For Algorithms’ (2017) 69 Administrative Law Review 84

UNESCO, ‘Recommendation on the Ethics in Artificial Intelligence’ (2021)

van Melle W, ‘MYCIN: A Knowledge-Based Consultation Program for Infectious Disease Diagnosis’ (1978) 10 International Journal of Man-Machine Studies 313

VDE & Bertelsmann Stiftung, ‘From Principles to Practice. An Interdisciplinary Framework to Operationalise AI Ethics’ (2020)

Veale M, Binns R and Ausloos J, ‘When Data Protection by Design and Data Subject Rights Clash’ (2018) 8 International Data Privacy Law 105

Veale M and Borgesius FZ, ‘Demystifying the Draft EU Artificial Intelligence Act’ (2021) 22 Computer Law Review International 97

Venema L, ‘Code of Conduct for Using AI in Healthcare’ (2019) 1 Nature Machine Intelligence 265

Vincent N and others, ‘Measuring the Importance of User-Generated Content to Search Engines’, Proceedings of the International AAAI Conference on Web and Social Media (2019)

Voigt P and von dem Bussche A, EU General Data Protection Regulation (GDPR): A Practical Guide (Springler 2017)

von Grafenstein M, The Principle of Purpose Limitation in Data Protection Laws (Nomos 2018)

Wachter S and Mittelstadt B, ‘A Right to Reasonable Inferences: Re-Thinking Data Protection Law in the Age of Big Data and AI’ (2019) 2 Columbia Business Law Review 494

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

Wachter S, Mittelstadt B and Russell C, ‘Bias Preservation in Machine Learning: The Legality of Fairness Metrics Under EU Non-Discrimination Law’ (2021) 1 West Virginia Law Review 735

Waltraut Kotschy, ‘Article 30. Records of Processing Activities’ in Christopher Kuner and Lee A Bygrave (eds), The EU General Data Protection Regulation (GDPR). A commentary (OUP 2020)

——, ‘Article 6. Lawfulness of Processing’ in Christopher Kuner, Lee A. Bygrave and Christopher Docksey (eds), The EU General Data Protection Regulation (GDPR): A Commentary (OUP 2020)

Warren SD and Brandeis LD, ‘The Right to Privacy’ (1890) 4 Harvard Law Review 193

Wiese Svanberg C, ‘Article 89. Safeguards and Derogations Relating to Processing for Archiving Purposes in the Public Interest, Scientific or Historical Research Purposes or Statistical Purposes’ in Christopher Kuner, Lee A Bygrave and Christopher Docksey (eds), The EU General Data Protection Regulation (GDPR) (OUP 2020)

Willis L, ‘Why Not Privacy by Default?’ (2014) 29 Berkeley Technology Law Journal 61

World Bank Group, ‘IFC Technology Code of Conduct — Progression Matrix — Public Draft’ (2020)

World Economic Forum Global Future Council on Human and Rights 2016-18, ‘How to Prevent Discriminatory Outcomes in Machine Learning’ (2018)

Wright D, ‘The State of the Art in Privacy Impact Assessment’ (2012) 28 Computer Law & Security Review 54

——, ‘Making Privacy Impact Assessment More Effective’ (2013) 29 The Information Society 307

Wright D and Friedewald M, ‘Integrating Privacy and Ethical Impact Assessments’ (2013) 40 Science and Public Policy 755

Xu G and others, ‘A User Behavior Prediction Model Based on Parallel Neural Network and K-Nearest Neighbor Algorithms’ (2017) 20 Cluster Computing 1703

Xu X, Zhou C and Wang Z, ‘Credit Scoring Algorithm Based on Link Analysis Ranking with Support Vector Machine’ (2009) 36 Expert Systems with Applications 2625

Yen SJ and others, ‘A Support Vector Machine-Based Context-Ranking Model for Question Answering’ (2013) 224 Information Sciences 77

Zahra Azizi and others, ‘Can Synthetic Data Be a Proxy for Real Clinical Trial Data? A Validation Study’ (2021) 11 BMJ Open 1

Zanfir-Fortuna G, ‘Article 13 Information to Be Provided Where Personal Data Are Collected from Teh Data Subject’ in Christopher Kuner, Lee A Bygrave and Christopher Docksey (eds), The EU General Data Protection Regulation (GDPR) (OUP 2020)

——, ‘Article 21. Right to Object’ in Christopher Kuner and Lee A Bygrave (eds), The EU General Data Protection Regulation (GDPR). A commentary (OUP 2020)

Zarsky T, ‘Incompatible: The GDPR in the Age of Big Data’ (2016) 47 Seton Hall Law Review 995

Zech H, ‘Data as a Tradeable Commodity’ in Alberto De Franceschi (ed), European Contract Law and the Digital Single Market. The Implications of the Digital Revolution (Intersentia 2016)

Zou J and Schiebinger L, ‘AI Can Be Sexist and Racist — It’s Time to Make It Fair’ (2018) 559 Nature 324

Cases

Court of Justice of the European Union

Case C-582/14 Patrick Breyer v Bundesrepublik Deutschland. [2016] ECLI:EU:C:2016:779.

Case C-673/17, Bundesverband der Verbraucherzentralen und Verbraucherverbände v Planet49 GmbH. [2019] ECLI:EU:C:2019:801.

Case 215/88, Casa Fleischhandels-GmbH v Bundesanstalt für landwirtschaftliche Marktordnung [1989] ECLI:EU:C:1989:331.

Case C-543/09, Deutsche Telekom AG v Bundesrepublik Deutschland. [2011] ECLI:EU:C:2011:279.

Joined Cases C-293/12 and C-594/12 Digital Rights Ireland Ltd v Minister for Communications. [2014] ECLI:EU:C:2014:238.

Case C-40/17, Fashion ID GmbH & CoKG v Verbraucherzentrale NRW eV. [2019] ECLI:EU:C:2019:629.

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

Case C-524/06 Heinz Huber v Bundesrepublik Deutschland. [2008] ECLI:EU:C:2008:724.

Case C-613/14, James Elliott Construction Limited v Irish Asphalt Limited. [2016] ECLI:EU:C:2016:821.

Case C-101/2001 Bodil Lindqvist v Åklagarkammaren i Jönköping. [2003] ECLI:EU:C:2003:596.

Case C-434/16 Peter Nowak v Data Protection Commissioner. [2017] EU:C:2017:582.

Case C-184/20 OT v Vyriausioji tarnybinės etikos komisija. [2022] ECLI:EU:C:2022:601.

Case C-634/21, OQ v SCHUFA Holding AG and Land Hesse. Request for a preliminary ruling from the Verwaltungsgericht Wiesbaden (Germany) lodged on 15 October 2021.

Joined cases C-465/00, C138-/01, and C-139/01 Rechnungshof v Österreichischer Rundfunk. [2003] ECLI:EU:C:2003:294.

Case C-212/13 František Ryneš v Úřad pro ochranu osobních údajů. [2014] ECLI:EU:C:2014:2428.

Case C-201/14, Smaranda Bara and Others v Casa Naţională de Asigurări de Sănătate and Others. [2015] ECLI:EU:C:2015:638.

Case C-362/14, Maximillian Schrems v Data Protection Commissioner. [2015] ECLI:EU:C:2015:650.

Case C-291/12 Michael Schwarz v Stadt Bochum. [2013] ECLI:EU:C:2013:670.

Case C-536/15, Tele2 (Netherlands) BV and Others v Autoriteit Consument en Markt (ACM). [2017] ECLI:EU:C:2017:214.

Case C-355/95, Textilwerke Deggendorf GmbH v Commission of the European Communities and Federal Republic of Germany, [1997] ECLI:EU:C:1997:24.

Case C-141/12 YS v Minister voor Immigratie [2014] ECLI:EU:C:2014:2081.

European Court of Human Rights

Antović and Mirković v Montenegro App no. 70838/13 (ECtHR 28 November 2017).

Bărbulescu v Romania App no. 61496/08 (ECtHR 5 September 2017).

Ben Faiza v France App No 31446/12 (ECtHR 8 February (2018).

Breyer v Germany App no. 50001/12 (ECtHR 30 January 2020).

Dragan Petrović v Serbia App no. 75229/10 (ECtHR 14 April 2020).

Gaskin v the UK App no. 10454/83 (ECtHR 7 July 1989).

Haralambie v Romania App no. 21737/03 (ECtHR 27 October 2009).

L.H. v Latvia App no. 52019/07 (ECtHR 29 April 2014).

L.L. v France App no. 7508/02 (ECtHR 10 October 2006).

Leander v Sweden App no. 9248/81 (ECtHR 26 March 1987).

Libert v France App no. 588/13 (ECtHR 22 February 2018).

López Ribalda and Others v Spain App no.1874/13 and 8567/13 (ECtHR 17 October 2019).

M.S. v Sweden App no. 20837/92 (ECtHR 27 August 1997).

Malone v the UK App no. 8691/79 (ECtHR 2 August 1984).

Malone v the UK App no. 8691/79 (ECtHR 2 August 1984).

Mehmedovic v Switzerland App no. 17331/11 (ECtHR 11 December 2018).

Mockutė v Lithuania App no. 66490/09 (ECtHR 27 February 2018).

P.G. and J.H. v the UK App no. 44787/98 (ECtHR 25 September 2001).

Peck v the UK App no. 44647/98 (ECtHR 28 January 2003).

Rotaru v Romania App No. 28341/95 (ECtHR 4 May 2000).

Rotaru v Romania App No. 28341/95 (ECtHR 4 May 2000).

S. and Marper v the UK Apps nos. 30562/04 and 30566/04 (ECtHR 4 December 2008).

Satamedia v Finland App no. 931/13 (ECtHR 27 June 2017).

Uzun v Germany App No. 35623/05 (ECtHR 2 September 2010).

Vetter v France App no. 59842/00 (ECtHR 31 May 2005).

Vukota-Bojić v Switzerland App no. 61838/10 (ECtHR 18 October 2016).

Z. v Finland App no. 22009/93 (ECtHR 25 February 1997).

National Courts

Austrian Federal Administrative Court (BVwG), Public Employment Service Austria (2020) W256 22353.

Constitutional Court of the Slovak Republic, Judgment no k PL ÚS 25 / 2019-117 - 492/2021 Coll

Italian Court of Cassation, Garante per la Protezione dei Dati Personali v Associazione Mevaluate Onlu (2021) case 14381.

District Court of Amsterdam, Ola Netherlands BV (2021) C/13/68970.

District Court of Amsterdam, Uber BV 1 (2021) C/13/68731.

District Court of Amsterdam, Uber BV 2 (2021) C/13/69200.

English and Wales High Court, R (Bridges) v Chief Constable of South Wales Police and other. [2019]

Data Protection Authorities

Agencia Española de Protección de Datos, Procedimiento No: PS/00185/2020.

Commission Nationale de l'Informatique et des Libertés, Clearview AI [2021].

Garante per la Protezione dei Dati Personali, Ordinanza ingiunzione nei confronti di Clearview AI. [2022] 9751362

Garante per la Protezione dei Dati Personali, Ordinanza ingiunzione nei confronti di Deliveroo Italy SRL (2021) 9685994.

Garante per la Protezione dei Dati Personali, Ordinanza ingiunzione nei confronti di Foodinho SRL (2021) 9675440.

Garante per la Protezione dei Dati Personali, Parere sul sistema SARI Real Time (2021) 9575877

Hamburgische Beauftragte für Datenschutz und Informationsfreiheit, Clearview AI Inc. [2020]

Information Commissioner’s Office, Clearview AI Inc. [2022]

United States of America

Supreme Court of Wisconsin, State v Loomis. [2016] No. 2015AP157–CR.

Federal Trade Commission, Cambridge Analytica LLC. Federal Trade Commission File No. 1823107 (July 24, 2019).

Federal Trade Commission, Everalbum Inc. Federal Trade Commission File No. 1923172 (January 8, 2021).

Federal Trade Commission, Facebook Inc. Federal Trade Commission File No. 1823109 (July 24, 2019).

Federal Trade Commission, Google LLC and YouTube LLC. Federal Trade Commission File No. 1723083 (September 4, 2019).

Federal Trade Commission, Kurbo Inc & WW International Inc. Federal Trade Commission File No. 1923228 (March 4, 2022).