Legal framework for the use of artificial intelligence and automated decision-making in public governance

Tombal, Thomas; Willem, Pauline; De Terwangne, Cecile

Published in:
The new digital era governance

Publication date:
2022

Document Version
Publisher's PDF, also known as Version of record

Link to publication

Citation for pulished version (HARVARD):
Tombal, T, Willem, P & De Terwangne, C 2022, Legal framework for the use of artificial intelligence and automated decision-making in public governance. in E Tan & J Crompvoets (eds), *The new digital era governance: how new digital technologies are shaping public governance.* Wageningen Academic Publishers, Wageningen , pp. 141-178.

Chapter 5

Legal framework for the use of artificial intelligence and automated decision-making in public governance

T. Tombal[1,2], P. Willem[3]* and C. De Terwangne[3]

[1]Faculty of Law, University of Namur, Rempart de la Vierge, 5, 5000 Namur, Belgium; [2]Tilburg Institute for Law and Technology, Tilburg University, Prof. Cobbenhagenlaan 221, 5037DE, Tilburg, the Netherlands; [3]Centre de recherche Information, Droit et Société, Namur Digital Institute, University of Namur, Rempart de la Vierge, 5, 5000 Namur, Belgium; pauline.willem@unamur.be

Abstract

European public administration increasingly relies on personal data to deliver their public services, which implies that they must comply with the rules contained in the General Data Protection Regulation. These rules are especially important in the advent of new artificial intelligence (AI) technologies, which increase the public administration's capability to make informed decisions in order to provide public services and to support their decision-making. This chapter presents the legal framework within which these AI technologies can be used, namely data protection rules and core principles of administrative law. Several key takeaways are given. First, data subjects' rights must be respected (right to information, access, erasure, not to be subject to decisions based solely on automated processing). Those rights impact the design of the technologies used by public administrations and have concrete implications. Moreover, the legal framework aims to give more control to the citizens and keep them at the centre. In this regard, it is highly reassuring to see that the European Commission has proposed the adoption of an AI Act. The chapter analyses this proposal, keeping in mind the use of new technologies by public administrations in the fight against fraud.

Keywords: artificial intelligence, GDPR, right not to be subject to automated individual decision-making, administrative law, AI Act Proposal

DOI 10.3920/978-90-8686-930-5_5, © Wageningen Academic Publishers 2022

5.1 Introduction

As outlined in Chapter 4, European public administration increasingly relies on personal data[1] to deliver their public services, which implies that they must comply with the rules contained in the General Data Protection Regulation (hereafter 'GDPR').[2] These rules are especially important with the advent of new artificial intelligence technologies, which increase the public administration's capability to take informed decisions in order to provide public services and to support their decision-making. Indeed, while artificial intelligence (AI) technologies can potentially increase the efficiency of the public administration's decision-making, they can also generate significant impacts on the lives and fundamental rights of the citizens to which automated decision-making systems are applied.

Accordingly, this chapter will aim at presenting the legal framework within which these AI technologies can be used. This legal framework is composed of two categories of rules. First, public administrations willing to rely on AI techniques to improve the delivery of their public services must respect personal data protection rules (Section 5.2). Then, since these AI technologies are used by the administration in the context of the pursuit of their public service missions, they will also have to comply with the core principles of administrative law (Section 5.3). Indeed, these additional legal challenges, distinct from personal data protection challenges, must also be considered when reflecting on the development of algorithmic decision-making tools by the public administration.

Similar to Chapter 4, while most of the following analysis will be equally applicable to any public administration in the European Union, the chapter will focus on the Belgian case, to detail the legal challenges that public administrations face in practice. More specifically, the accent will be put on the use of AI for social security and tax fraud. This is because, in both cases, the use of algorithmic and automated decision-making processes could have a significant impact on citizens' finances and their general well-being.

5.2 Personal data protection and artificial intelligence

AI 'refers to systems that display intelligent behaviour by analysing their environment and taking actions – with some degree of autonomy – to achieve specific goals' (European Commission, 2018). AI thus aims at understanding how human cognition works, in order to reproduce it and to create computer cognition resembling that of humans (Villani *et al.*, 2018). AI methods are, in fact, numerous and diverse (expert systems, neural networks, machine learning, reinforcement learning, deep learning...) and are not new, as many have been developed and refined since the infamous Dartmouth conference of 1956 (Villani *et al.*, 2018).

For the purpose of this contribution, we will simply further elaborate here on two categories of AI, namely *expert systems* and *artificial neural networks*, as the distinction between the two will be relevant for our analysis below. An expert system can be defined as 'a computer system that emulates the decision-making ability of a human expert. Expert systems are designed to solve complex problems by reasoning through bodies of knowledge, represented mainly as 'if-then rules'.[3] In short, expert systems are rules-based AI systems. Based on rules that have been programmed, the AI system will draw inferences, following a form of decision-tree (If A then C, but if B then D, etc.).

On the other hand, artificial neural networks 'are computing systems (..) [that] 'learn' to perform tasks by considering examples, generally without being programmed with task-specific rules. (...) Instead, they automatically generate identifying characteristics from the examples that they process'.[4] Here the rules are not provided in advance to the program. Rather, the program has to understand the rules for itself, through trial and error, and through self-improvement (machine learning, reinforcement learning, deep learning...). The big difference between these two categories of AI is that, while the human programmer knows the rules it has given to an expert system, it will not necessarily understand which rules have been applied by the neural network, because the program has created its own rules, through trial and error. This is sometimes referred to as the 'Black box' (De Streel *et al.*, 2020; Pasquale, 2015).

From a personal data protection point of view, it should be outlined from the start that, as the use of AI algorithms requires a large amount of data, all the developments in Chapter 4 pertaining to big data and personal data protection[5] need to be kept in mind here. Indeed, the collection, linking and processing of large volumes of data to train an algorithm will need to rely on a lawful basis of processing[6] and to respect the purpose limitation and data minimisation principles.[7] That analysis will not be repeated here. Rather, we will focus, in this section, on the specific challenges posed by AI in terms of the GDPR, and these mostly revolve around the need to respect the data subjects' rights. More particularly, we will analyse the interactions between AI and the data subjects' right to information,[8] the data subjects' right of access,[9] the data subjects' right to erasure[10] and the

[1] Personal data are defined as 'any information relating to an identified or identifiable natural person ('data subject'); an identifiable natural person is one who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, an online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person' (Art. 4.1 of the GDPR).

[2] Regulation 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46 (General Data Protection Regulation), OJ [2016] L 199/1.

[3] https://en.wikipedia.org/wiki/Expert_system.

[4] https://en.wikipedia.org/wiki/Artificial_neural_network.

[5] See Chapter 4, Section 4.2.

[6] Art. 6.1 of the GDPR. See Chapter 4, Section 4.2.1.

[7] Arts. 5.1.b) and 5.1.c) of the GDPR. See Chapter 4, Sections 4.2.3 and 4.2.4.

[8] Arts. 12 to 14 of the GDPR.

[9] Art. 15 of the GDPR.

[10] Art. 17 of the GDPR.

data subjects' right not to be subject to automated individual decision-making.[11] However, prior to delving into the analysis of these data subject rights, a word must be said about the importance of considering the risks of AI before taking the decision to rely on such a technology.

5.2.1 A preliminary step: considering the risks of AI use (impact assessments and the proposal for a regulation on AI)

Like any technology, while AI can offer significant benefits, it also presents significant risks, notably in terms of safety and liability, security, bias, discrimination, but also personal data processing (European Commission, 2018).[12] In light of this latter risk, it is important to have Article 35 of the GDPR in mind, which pertains to personal data impact assessments. According to Article 35.1 of the GDPR:

> Where a type of processing in particular using new technologies, and taking into account the nature, scope, context and purposes of the processing, *is likely to result in a high risk to the rights and freedoms of natural persons*, the controller shall, prior to the processing, carry out an assessment of the impact of the envisaged processing operations on the protection of personal data. A single assessment may address a set of similar processing operations that present similar high risks (emphasis added).

Any public administration wishing to rely on AI techniques to support its policy- or decision-making, notably in the field of tax and social security fraud, should thus question whether this will imply a processing of personal data and, if this is the case, whether this is likely to result in a high risk to the rights and freedoms of natural persons (Edwards and Veale, 2017).

In assessing the likelihood of this high risk, public administration should be particularly attentive to Article 35.3 of the GDPR, as, in light of this provision, a public administration will have to perform a data protection impact assessment prior to the use of AI techniques to support its policy- or decision-making, if it implies a systemic and extensive evaluation of personal aspects relating to natural persons, which is based on automated processing[13] and produces legal effects on the person or similarly significantly affects them (Art. 35.3.a). Such an impact assessment would thus be required if a public administration intends to rely on automated processing to detect tax or social security fraud. In fact, the need to conduct such an impact assessment might also be justified, in the context of tax and social security fraud, in the case where a large scale of special categories of personal data (e.g. health data) or of personal data relating to criminal convictions and offences (e.g. data about prior convictions to establish 'suspicious' profiles) are used (Art. 35.3.b) (Edwards and Veale, 2017).[14]

[11] Art. 22 of the GDPR.

[12] For a depiction of some of these risks, see Alston, 2019. Extreme poverty and human rights. Note by the Secretary-General of the United Nations. Available at: https://undocs.org/pdf?symbol=en/A/74/493.

[13] On this point, see Section 5.2.5.

[14] See Chapter 4, Section 4.2.2.

If such a data protection impact assessment has to be performed, Article 35.7 provides that it shall contain, at least:

> (a) a systematic description of the envisaged processing operations and the purposes of the processing, including, where applicable, the legitimate interest pursued by the controller; (b) an assessment of the necessity and proportionality of the processing operations in relation to the purposes; (c) an assessment of the risks to the rights and freedoms of data subjects referred to in paragraph 1; and (d) the measures envisaged to address the risks, including safeguards, security measures and mechanisms to ensure the protection of personal data and to demonstrate compliance with this Regulation taking into account the rights and legitimate interests of data subjects and other persons concerned.

Moreover, the Data Protection Authority will have to be consulted before the process is launched, and will have the power to ban, temporarily or permanently, the use of the system if it is not satisfied with the risk identification and/or the guarantees offered by the public administration to mitigate the risks for the data subjects (Edwards and Veale, 2017).[15]

Finally, it is worth adding that according to a law proposal of 6 April 2021, these impact assessments of the algorithmic tools put in place by the administrations would have to be published in order to increase transparency.[16]

Importantly, the fact that public administration must consider whether an AI application entails 'high risk' to the rights and freedoms of natural persons is not only relevant for their right to personal data protection, but more broadly for all of their human rights (see, for example, Committee of Ministers of the Council of Europe, 2020). In this regard, it should be outlined that the European Commission's White Paper on AI also suggested the adoption of a risk-based approach when it comes to AI applications (European Commission, 2020). For the Commission, the extent of regulatory intervention on AI should be proportionate and should differentiate between categories of AI applications, focussing on 'high risk' applications (European Commission, 2020[17]). In fact, this approach has been confirmed by the Commission in its recent proposal for

[15] Art. 36.1 and 36.2 of the GDPR.

[16] Proposition de loi modifiant la loi relative à la publicité de l'administration du 11 avril 1994 afin d'introduire une plus grande transparence dans l'usage des algorithmes par les administrations, 6 avril 2021, *Doc. parl.*, Chambre, sess. ord., 2020-2021, no 55-1904/001, p. 6 and 8.

[17] In that White Paper, the Commission had provided that an AI application should be considered 'high risk' when it meets the two following cumulative criteria (p. 17):

'First, the AI application is employed in a sector where, given the characteristics of the activities typically undertaken, significant risks can be expected to occur. [These are] areas where, generally speaking, risks are deemed most likely to occur. (...) For instance, healthcare; transport; energy and *parts of the public sector.*

Second, the AI application in the sector in question is, in addition, used in such a manner that significant risks are likely to arise. This second criterion reflects the acknowledgment that not every use of AI in the selected sectors necessarily involves significant risks. (...) The assessment of the level of risk of a given use could be based on the impact on the affected parties. For instance, uses of *AI applications that produce legal or similarly significant effects for the rights of an individual or a company*; that pose risk of injury, death or significant material or immaterial damage; [or] that produce effects that cannot reasonably be avoided by individuals or legal entities' (emphasis added).

an Artificial Intelligence Act (European Commission, 2021).[18] Indeed, the proposal differentiates between 'uses of AI that create: (1) an unacceptable risk; (2) a high risk; and (3) low or minimal risk' (European Commission, 2021).

First, the proposal suggests prohibiting AI uses that create an unacceptable risk, notably because they would contravene EU values and/or would violate fundamental rights (European Commission, 2021).[19] These are:

- AI systems that deploy subliminal techniques beyond a person's consciousness in order to materially distort a person's behaviour in a manner that causes or is likely to cause that person or another person physical or psychological harm (Art. 5.1.a);
- AI systems that exploit any of the vulnerabilities of a specific group of persons due to their age, physical or mental disability, in order to materially distort the behaviour of a person pertaining to that group in a manner that causes or is likely to cause that person or another person physical or psychological harm (Art. 5.1.b);
- AI systems used by public authorities, or on their behalf, for the evaluation or classification of the trustworthiness of natural persons over a certain period of time based on their social behaviour or known or predicted personal or personality characteristics, with the social score leading to either or both of the following: (1) detrimental or unfavourable treatment of certain natural persons or whole groups thereof in social contexts which are unrelated to the contexts in which the data was originally generated or collected; (2) detrimental or unfavourable treatment of certain natural persons or whole groups thereof that is unjustified or disproportionate to their social behaviour or its gravity (Art. 5.1.c);
- 'Real-time' remote biometric identification systems in publicly accessible spaces[20] for the purpose of law enforcement, unless and in as far as such use is strictly necessary for one of the following objectives: (1) the targeted search for specific potential victims of crime, including missing children; (2) the prevention of a specific, substantial and imminent threat to the life or physical safety of natural persons or of a terrorist attack; or (3) the detection, localisation, identification or prosecution of a perpetrator or suspect of a criminal offence referred to in Article 2(2) of Council Framework Decision 2002/584/JHA (Council, 2002) and punishable in the Member State concerned by a custodial sentence or a detention order for a maximum period of at least three years (Art. 5.1.d).[21]

Second, the proposal lays down a risk methodology to define 'high-risk' AI systems that pose significant risks to the health and safety or fundamental rights of persons. The classification of an AI system as high-risk is based on the 'intended purpose'[22] of the AI system, which implies that this classification 'does not only depend on the function performed by the AI system, but also on the specific purpose and modalities for which that system is used'.[23] More precisely, the proposal identifies two main categories of high-risk AI systems:[24]

- AI systems, listed in Annex 2 of the proposal, that are intended to be used as safety component of products, or as products themselves, and that are subject to a third-party *ex ante* conformity assessment in order to be placed on the market (Art. 6.1); and Stand-alone AI systems listed in Annex III (Art. 6.2).

The list of Annex III contains a limited number of AI systems whose risks (mainly in terms of fundamental rights) have already materialised or are likely to materialise in the near future.[25] These high-risk AI systems are divided into eight areas, namely: (1) biometric identification and categorisation of natural persons; (2) management and operation of critical infrastructure; (3) education and vocational training; (4) employment, workers management and access to self-employment; (5) access to and enjoyment of essential private services and public services and benefits; (6) law enforcement; (7) migration, asylum and border control management; and (8) administration of justice and democratic processes.

In the context of tax and social security fraud, it is worth highlighting that 'AI systems intended to be used by public authorities or on behalf of public authorities to evaluate the eligibility of natural persons for public assistance benefits and services, as well as to grant, reduce, revoke, or reclaim such benefits and services'[26] are considered as high-risk AI systems. This is because 'natural persons applying for or receiving public assistance benefits and services from public

[18] European Commission, Proposal for a Regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) and amending certain Union legislative acts, Brussels, 21 April 2021, COM(2021) 206 final, available at https://digital-strategy.ec.europa.eu/en/library/proposal-regulation-european-approach-artificial-intelligence.

[19] See also Article 5.

[20] A 'real-time' remote biometric identification system is 'a remote biometric identification system [i.e. an AI system for the purpose of identifying natural persons at a distance through the comparison of a person's biometric data with the biometric data contained in a reference database, and without prior knowledge of the user of the AI system whether the person will be present and can be identified (Art. 3.36 of the proposal)] whereby the capturing of biometric data, the comparison and the identification all occur without a significant delay. This comprises not only instant identification, but also limited short delays in order to avoid circumvention' (Art. 3.37 of the proposal). Publicly accessible space 'means any physical place accessible to the public, regardless of whether certain conditions for access may apply' (Art. 3.39 of the proposal). On AI applications for the purposes of remote biometric identification (facial recognition), see also Consultative Committee of the Convention for the protection of individuals with regard to automatic processing of personal data, 2021. Moreover, it is worth underlining that Art. 35.3.c of the GDPR provides that a data protection impact assessment will be required in the case of systematic monitoring of a publicly accessible area on a large scale.

[21] In order to apply one of the exceptions of Article 5.1.d), account shall be taken of 'the nature of the situation giving rise to the possible use, in particular the seriousness, probability and scale of the harm caused in the absence of the use of the system' and of 'the consequences of the use of the system for the rights and freedoms of all persons concerned, in particular the seriousness, probability and scale of those consequences', and these uses 'shall comply with necessary and proportionate safeguards and conditions in relation to the use, in particular as regards the temporal, geographic and personal limitations' (Art. 5.2 of the proposal). Moreover, this will be subject to a prior authorisation from a judicial or independent administrative authority, except in duly justified situations of urgency, where the authorisation can be requested during or after the use (Art. 5.3 of the proposal).

[22] 'The use for which an AI system is intended by the provider, including the specific context and conditions of use, as specified in the information supplied by the provider in the instructions for use, promotional or sales materials and statements, as well as in the technical documentation' (Art. 3.12 of the proposal).

[23] Proposal for an Artificial Intelligence Act, p. 13.

[24] See Articles 6 and 7 of the proposal for an Artificial Intelligence Act.

[25] Proposal for an Artificial Intelligence Act, p. 16.

[26] Annex III, point 5.a) of the proposal for an Artificial Intelligence Act.

authorities are typically dependent on those benefits and services and in a vulnerable position in relation to the responsible authorities. If AI systems are used for determining whether such benefits and services should be denied, reduced, revoked or reclaimed by authorities, they may have a significant impact on persons' livelihood and may infringe their fundamental rights, such as the right to social protection, non-discrimination, human dignity or an effective remedy'.[27] This would arguably cover AI systems used to allocate social benefits or to fight against social security fraud.

Moreover, the following are also considered high-risk: 'AI systems intended to be used by law enforcement authorities for predicting the occurrence or reoccurrence of an actual or potential criminal offence based on profiling of natural persons or assessing personality traits and characteristics or past criminal behaviour of natural persons or groups';[28] 'AI systems intended to be used by law enforcement authorities for profiling of natural persons in the course of detection, investigation or prosecution of criminal offences';[29] and 'AI systems intended to be used for crime analytics regarding natural persons, allowing law enforcement authorities to search complex related and unrelated large data sets available in different data sources or in different data formats in order to identify unknown patterns or discover hidden relationships in the data'.[30] Arguably, AI systems used to fight tax fraud, through data matching and data mining, could be considered as falling within these types of high-risk AI systems. However, it should be underlined that Recital 38 of the proposal provides that 'AI systems specifically intended to be used for administrative proceedings by tax and customs authorities should not be considered high-risk AI systems used by law enforcement authorities for the purposes of prevention, detection, investigation and prosecution of criminal offences'. This seems to suggest that AI systems used to fight tax fraud would not be considered as 'high-risk'. Yet, as Recitals are not binding, this creates some uncertainty in this regard.

The Commission will have to conduct an annual assessment of the list of high-risk AI systems,[31] which it can update according to the methodology described in Article 7 of the proposal. This list can notably be expanded to AI systems intended for use in any of the areas listed in Annex III, which 'pose a risk of harm to the health and safety, or a risk of adverse impact on fundamental rights, that is, in respect of its severity and probability of occurrence, equivalent to or greater than the risk of harm or of adverse impact posed by the high-risk AI systems already referred to in Annex III'.[32]

[27] Recital 37 of the proposal for an Artificial Intelligence Act.
[28] Annex III, point 6.e) of the proposal for an Artificial Intelligence Act.
[29] Annex III, point 6.f) of the proposal for an Artificial Intelligence Act.
[30] Annex III, point 6.g) of the proposal for an Artificial Intelligence Act.
[31] Art. 84.1 of the proposal for an Artificial Intelligence Act.
[32] Art. 7.1 of the proposal for an Artificial Intelligence Act.

These high-risk AI systems will have to undergo *ex ante* conformity assessment procedures,[33] and they will have to be registered in a public EU-wide database, operated by the European Commission, to increase public transparency and oversight and strengthen *ex post* supervision by competent authorities.[34] Furthermore, any AI system meeting this 'high-risk' threshold will have to comply with the legal requirements set out in the proposal.[35] According to the Commission, these requirements 'are strictly necessary to mitigate the risks to fundamental rights and safety posed by AI and that are not covered by other existing legal frameworks'.[36] These requirements pertain to the need: (1) to establish, implement, document and maintain a risk management system; (2) to produce technical documentation; (3) to ensure that high-quality data is used; (4) to adopt an appropriate data governance allowing record-keeping – i.e. logs –, transparency and the provision of information to users; (5) to ensure human oversight; and (6) to ensure a certain level of accuracy, robustness, and cybersecurity (European Commission, 2020).[37] Interestingly, several of these requirements contribute to the minimisation of the risks of algorithmic discrimination, namely those pertaining to the design and the quality of datasets, and to the obligations for testing, risk management, documentation and human oversight throughout the AI systems' lifecycle.[38] Finally, the Commission outlined that 'harmonised standards and supporting guidance and compliance tools will assist providers and users in complying with the requirements laid down by the proposal and minimise their costs'.[39]

Third, the proposal suggests imposing transparency obligations for certain low-risk systems, namely those that: '(1) interact with humans; (2) are used to detect emotions[40] or determine association with (social) categories based on biometric data;[41] or (3) generate or manipulate content ('deep fakes')'.[42] Moreover, the proposal establishes a framework for the creation of codes of conduct, in order to encourage low-risk AI systems' providers to voluntarily apply the above-mentioned mandatory requirements for high-risk AI systems.[43]

Finally, it is worth noting that, in order to enforce the above-mentioned rules, the proposal establishes, at the EU level, a 'European Artificial Intelligence Board', composed of representatives from the Member States and the Commission, which will 'facilitate a smooth, effective and harmonised implementation of this regulation by contributing to the effective cooperation of the

[33] Proposal for an Artificial Intelligence Act, p. 3. See also Articles 40 to 50, and Annexes VI and VII.
[34] Proposal for an Artificial Intelligence Act, p. 12-14. See also Articles 51 and 60, and Annex VIII.
[35] Article 8 of the proposal for an Artificial Intelligence Act.
[36] Proposal for an Artificial Intelligence Act, p. 13.
[37] See Articles 9 to 15 and Annex IV of the proposal for an Artificial Intelligence Act.
[38] Proposal for an Artificial Intelligence Act, p. 4.
[39] *Ibid.*, p. 7. See also Articles 40 to 42.
[40] Emotion recognition system: 'an AI system for the purpose of identifying or inferring emotions or intentions of natural persons on the basis of their biometric data' (Art. 3.34 of the proposal).
[41] Biometric categorisation system: 'an AI system for the purpose of assigning natural persons to specific categories, such as sex, age, hair colour, eye colour, tattoos, ethnic origin or sexual or political orientation, on the basis of their biometric data' (Art. 3.35 of the proposal).
[42] Proposal for an Artificial Intelligence Act, p. 14. See Article 52 for more details.
[43] *Ibid.*, p. 16. See also Article 69.

national supervisory authorities and the Commission and providing advice and expertise to the Commission. It will also collect and share best practices among the Member States'.[44] Furthermore, 'Member States will have to designate one or more national competent authorities and, among them, the national supervisory authority, for the purpose of supervising the application and implementation of the regulation'.[45]

5.2.2 AI and the data subject's right to information

As mentioned in Chapter 4, data has to be processed fairly and in a transparent manner.[46] Moreover, if the public administration makes use of individual decision-making, based 'solely' on automated processes, for instance to fight social security infringements and tax fraud, it will have to inform the data subject about the existence of such processes.[47] In that case, the data subject should, at least, receive meaningful information about the logic involved, as well as the significance and the envisaged consequences of such processing for the data subject.[48] This is key for the data subjects' right to informational self-determination, as it allows them to understand what is being done with their data (De Terwangne, 2015). Indeed, the data subject can legitimately wish to know the criteria that have been used, and their respective weight, in the automated decision (De Terwangne, 2015).

In this regard, it is interesting to mention the SyRI decision of the Court of The Hague,[49] which has been presented in Chapter 4.[50] As a reminder, the Court was asked to assess the compatibility, with Article 8 of the European Convention on Human Rights, of the Dutch government's 'Systeem Risico Indicatie ('SyRI' – System Risk Indication), which is a legal instrument used to detect various forms of fraud, including social benefits, allowances, and tax fraud.[51] In that case, the Court seems to have addressed the obligation to inform the data subject about the use of individual decision-making based 'solely' on automated processes, although not explicitly. Indeed, the Court pointed out that there is a clear transparency problem, because:

> [T]he SyRI legislation does not provide for a duty of disclosure to those whose data are processed in SyRI so that these data subjects can be reasonably assumed to know that their data are or have been used for that processing. The SyRI legislation also does not provide for an obligation to notify the data subjects individually, as appropriate, that a risk report has been submitted. There is only a statutory obligation to announce the start of a SyRI project beforehand by way of publication in the Government Gazette and after the processing access to the register of risk reports upon request. (…). Data subjects are also not informed automatically afterwards. This only occurs if there is a control and investigation in response to a risk report. This does not happen as a matter of course.[52]

Although the Court did not rule specifically on the application of the right to be informed about the logic involved,[53] this right seems not to have been respected in this case, because the data subjects are unaware of the existence of a risk report, while the submission of a risk report has a significant effect on them.[54] Indeed, a data subject whose data were processed in SyRI, but which did not result in a risk report, will not be informed about this processing, and therefore cannot verify that their data was processed on correct grounds.[55]

Additionally, it should be underlined that, depending on the type of AI that is used and on the legal entity that holds the rights to the algorithm, it might be complicated for the public administration to comply with this requirement to provide the data subject with meaningful information about the logic involved.

On the one hand, a difference must be made between expert systems and neural networks. As outlined above, expert systems are rules-based AI systems. Accordingly, the public administration is more likely to be aware of the rules that have been applied by the algorithm and of the logic involved in the decision, as these have been explicitly programmed and dictated to the AI, and should be able, in such cases, to provide the data subject with meaningful information about the logic involved.[56] On the contrary, neural networks create their own rules, through trial and error (e.g. machine learning). Accordingly, the public administration might not necessarily understand which rules have been applied by the neural network, and thus might not be able to provide the data subject with meaningful information about the logic involved.[57]

On the other hand, a difference will have to be made depending on whether the rights to the algorithm are held by the public administration itself, or by a private entity that has been tasked, by means of a public procurement, with developing the algorithm for the public administration. Indeed, in the vast majority of cases, the algorithm will benefit from the copyright protection

[44] *Ibid.*, p. 15. See also Articles 56 to 58.
[45] *Ibid.*, p. 15. See also Article 59.
[46] Arts. 5.1.a) and 12.1 of the GDPR. See Chapter 4, Section 4.2.5.
[47] Arts. 13.1.f) and 14.2.g) of the GDPR.
[48] Art. 15.1.h) of the GDPR.
[49] Rechtbank Den Haag, 5 februari 2020, Zaak n° C-09-550982-HA ZA 18-388, ECLI:NL:RBDHA:2020:865 (ECLI:NL:RBDHA:2020:1878 for the English version).
[50] See Chapter 4, Section 4.2.5.
[51] Wet van 9 oktober 2013 tot wijziging van de Wet structuur uitvoeringsorganisatie werk en inkomen en enige andere wetten in verband met fraudeaanpak door gegevensuitwisselingen en het effectief gebruik van binnen de overheid bekende zijnde gegevens, *Stb.*, 2013, p. 405; Besluit van 1 september 2014 tot wijziging van het Besluit SUWI in verband met regels voor fraudeaanpak door gegevensuitwisselingen en het effectief gebruik van binnen de overheid bekend zijnde gegevens met inzet van SyRI, *Stb.*, 2014, p. 320.

[52] Rechtbank Den Haag, 5 februari 2020, Zaak n° C-09-550982-HA ZA 18-388, point 6.54.
[53] Arts. 13.1.f) and 14.2.g) of the GDPR.
[54] Rechtbank Den Haag, 5 februari 2020, Zaak n° C-09-550982-HA ZA 18-388, point 6.82.
[55] *Ibid.*, point 6.90.
[56] Art. 15.1.h) of the GDPR.
[57] Art. 15.1.h) of the GDPR.

granted to computer programs.[58] Moreover, it might also be protected as a trade secret.[59] In cases where these rights are held by the private entity, this may prevent the public authority from understanding the logic involved behind the algorithm, and therefore from providing this information to their citizens.[60] Indeed, the private entity could invoke its rights to the algorithm in order to refuse to disclose this commercially sensible information to the public administration, which in turn will not be able to communicate it to citizens. It is thus of the utmost importance for public administration that call upon private parties to develop an algorithm, to specify clearly in the public procurement either that it will hold the rights to the algorithm, or that it has the right to receive information about the logic involved behind the algorithm, in order to be able to provide it to the data subjects whose data are being processed.[61] However, an important caveat here is that the private entity itself may not necessarily be able to explain the logic involved in the decision taken by a neural network AI. This is sometimes referred to as the 'black box' (De Streel *et al.*, 2020; Pasquale, 2015).

Finally, it should be added that there have been proposals to include *algorithmic impact assessments* (AIAs) as part of a public administration's procurement procedures pertaining to automated decision-making systems (Misuraca and van Noordt, 2020). The potential benefits of rolling out such AIAs would notably include: 'better communication with the general public; increase of in-house expertise of public agencies; higher levels of accountability of automated decision-making systems and a meaningful way for the public to question them' (Misuraca and van Noordt, 2020).

5.2.3 AI and the data subject's right of access

The data subject's right of access has a double impact for public administrations. On the one hand, it stipulates that the data subject has the right to obtain, from the public administration, the confirmation as to whether or not it processes personal data concerning them.[62] The goal of this right is for the data subject to be aware of, and to verify, the lawfulness of the processing.[63] In the context of AI applications used to fight social security infringements and tax fraud, this means that the data subject has the right to obtain, from the public administration, the confirmation as to whether or not the algorithm processes personal data concerning them. If this is the case, the public administration will have to provide access to the data, as well as to the information listed in points (a) to (h) of Article 15.1 of the GDPR. Since, these elements of information are, in substance, the same as those contained in Articles 13 and 14 of the GDPR, we will simply refer here to what has been said in Chapter 4, Section 4.2.5.

On the other hand, the right of access provides the data subject with the right to obtain a copy of the personal data that is processed by the public administration (Tombal, 2018).[64] In the context of AI applications used to fight social security infringements and tax fraud, this means that the data subject has the right to obtain a copy of the personal data concerning them that is processed by the algorithm. In this regard, simply providing a mass of incomprehensible data for any human being would not be sufficient, as any communication under Articles 15 to 22 of the GDPR must be made in a concise, transparent, intelligible and easily accessible form, using clear and plain language.[65] This means that the data subject should be able to understand the data that she receives. Moreover, and similarly to all of the data subjects' rights, the public administration will have to answer without undue delay and, in any case, within one month of the receipt of the request, and the exercise of the right of access should be free.[66] However, a reasonable fee based on administrative costs may be charged for any further copies requested by the data subject.[67]

Like the right to information, this right of access can be restricted by a Member State law when such a restriction respects the essence of the fundamental rights and freedoms and is a necessary and proportionate measure in a democratic society to safeguard an important objective of general public interest.[68]

5.2.4 AI and the data subject's right to erasure

Article 17 of the GDPR provides that the data subject shall have the right to obtain from the controller the erasure of personal data concerning them, without undue delay, in certain hypotheses.[69] It should, however, be outlined from the outset that this right to erasure shall not apply to the extent that the processing is necessary for the compliance with a legal obligation to which the controller is subject, or for the performance of a task carried out in the public interest or in the exercise of official authority vested in the controller.[70]

That being said, even in cases where such right should apply, doing so in an AI world can turn out to be much more complicated that it seems, as the GDPR does not define the notion of erasure (Fosch Villaronga *et al.*, 2017). As pointed out by Fosch Villaronga *et al.* in their seminal paper,

[58] Directive 2009/24/EC of the European Parliament and of the Council of 23 April 2009 on the legal protection of computer programs, OJ [2009] L 111/16.

[59] Directive (EU) 2016/943 of the European Parliament and of the Council of 8 June 2016 on the protection of undisclosed know-how and business information (trade secrets) against their unlawful acquisition, use and disclosure, OJ [2016] L 157/1. Art. 2.1 defines a trade secret as any 'information which meets all of the following requirements: (a) it is secret in the sense that it is not, as a body or in the precise configuration and assembly of its components, generally known among or readily accessible to persons within the circles that normally deal with the kind of information in question; (b) it has commercial value because it is secret; (c) it has been subject to reasonable steps under the circumstances, by the person lawfully in control of the information, to keep it secret'.

[60] Art. 15.1.h) of the GDPR.

[61] Art. 15.1.h) of the GDPR.

[62] Art. 15.1 of the GDPR.

[63] Recital 63 of the GDPR.

[64] Art. 15.3 of the GDPR. This clarification is important and puts an end to the debate that existed, prior to the GDPR, when Directive 95/46 was applicable, as it was uncertain whether the right of access, as set out in Article 12 of the Directive, implicitly allowed the data subject to obtain a copy of their personal data.

[65] Art. 12.1 of the GDPR.

[66] Arts. 12.3 and 12.5 of the GDPR.

[67] Art. 15.3 of the GDPR.

[68] Art. 23.1 of the GDPR. See Chapter 4, Section 4.2.5.

[69] Art. 17.1 of the GDPR.

[70] Art. 17.3.b) of the GDPR.

Article 17 'seems to push toward the simple deletion of the personal data or the folder containing the personal data from the data controller's system, as if data on a computer was like a physical file that can simply be destroyed' (Fosch Villaronga *et al.*, 2017). Yet, data deletion requirements pose crucial challenges in AI environments, and might actually be practically impossible to satisfy (Fosch Villaronga *et al.*, 2017).

To understand why this is the case, it must be outlined that AI 'minds' do not function exactly as human minds and that, as a consequence, an AI cannot forget data the way humans do, making it much more complex to delete data (Fosch Villaronga *et al.*, 2017). Indeed, it is questionable whether data deletion is, in fact, actually possible in modern AI environments relying on relational database management system (DBMS) (Fosch Villaronga *et al.*, 2017). As outlined by Fosch Villaronga *et al.*:

> [E]very data record added to the database might not only reside at one specific point in the file system, but might be stored at various locations inside internal database mechanisms, as well as across different replicated databases, in logfiles and backups (...) When asking for deletion in a strict sense, these spaces must be identified and overwritten with random information. In several internal mechanisms like the database transaction log, the latter is especially impossible without seriously endangering the consistency of the database, or even simply breaking it altogether (Fosch Villaronga *et al.*, 2017).

Moreover, in most AI environments, when data is 'deleted', it is not directly overwritten with other data, but is only marked as deleted and removed from the search indexes, and it can take a very long time before the space marked as 'deleted' is effectively reused (effectively destroying the old data) (Fosch Villaronga *et al.*, 2017). Accordingly, determining if and when a deletion has occurred, and thus whether the data controller has complied with the data subject's right to erasure, will depend on the interpretation of the words erasure and deletion. Is it the removal from the search index, the overwriting in the file system, the deletion from the log-files and backups or is it the removal from all internal mechanisms? (Fosch Villaronga *et al.*, 2017)

In any case, even if it is assumed that data can be erased from AI systems in a way that complies with Article 17 of the GDPR, such deletion might have an impact on the quality of the AI's results (Fosch Villaronga *et al.*, 2017). As outlined by Fosch Villaronga *et al.*:

> This is especially interesting considering algorithms that use a so-called 'knowledgebase' for calibration, i.e. the algorithm takes the knowledgebase with pre-calculated results as reference data and extracts the common artifacts. It then uses these 'learned' rules on new data, which have to be very close to the training data in terms of data structure and statistical properties. Furthermore, the resulting categorisations are again fed into the knowledge base in order to get even better training data for the next run, thus iteratively extending the knowledge base. (Fosch Villaronga *et al.*, 2017)

In other words, once an algorithm has trained on data to produce a result, it will use these 'learned' results to train on the next batch of data. Accordingly, even if we could assume that the data subject's data can be deleted from the training data, traces of this data will probably still be found in the 'learned' results, which will be used in the training iterations. This may be problematic if it is impossible to erase these traces. In fact, it will likely be impossible to do so, and doing so might actually have large-scale effects on the algorithm's efficiency (Fosch Villaronga *et al.*, 2017). While several approaches have been suggested to solve this issue, none has so far offered satisfactory results (Fosch Villaronga *et al.*, 2017).

In light of the above, a public administration wishing to use AI applications in order to fight social security infringements and tax fraud should anticipate these potential erasure requests in the way it builds its AI system. More specifically, it should define, in advance, the moment at which data will be considered as being deleted (removal from the search index, overwriting in the file system, deletion from the log-files or from all internal mechanisms) and justify, in light of the accountability principle,[71] why this complies with Article 17 of the GDPR. Moreover, in light of the iterative way of working of AI systems, which rely on the previous 'learned' results for the next training iterations, it should reflect, from the outset, on the impact that the right of erasure will have on these 'learned' results: is it possible to erase the data subject's data not only from the training data, but also any trace of it in the 'learned' results?

5.2.5 AI and the data subject's right not to be subject to automated individual decision-making

According to Article 22.1 of the GDPR, 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. This right echoes the strong will of the human being not to be entirely subject to the machine, as the former does not accept the idea that a decision may be imposed on it solely on the basis of conclusions reached by that machine (de Terwangne, 2015). Accordingly, it is fundamental for a public administration that wishes to rely on AI to fight social security infringements and tax fraud to comply with this provision of the GDPR.

Automated decision-making based *solely* on automated processing

It is important to highlight from the outset that Article 22 only applies to decisions based *solely* on automated processing. This means that Article 22 could potentially 'be sidestepped relatively easily by inserting human intervention into the process. In other words, once the process is not 'solely' automated, this provision will not apply' (Scarcella, 2019; Zarsky, 2017). Nevertheless, it remains to be seen how courts and regulators would react to the introduction, by data controllers, of fictitious or negligible human intervention in the automated decision process, simply in order to avoid (potentially in bad faith) the application of Article 22 (Zarsky, 2017).

[71] Art. 5.2 of the GDPR.

For instance, if the whole decision-making process is automated, but has to be validated by a human before being effectively applied (e.g. the result of the decision-making appears on the human's screen and she has to validate it by clicking 'Ok' or 'Validate'), can it be said that there has been human intervention in the decision-making process? As coined by Edwards and Veale: 'When does 'nominal' human involvement become no involvement?' (Edwards and Veale, 2017).

The answer will likely differ depending on the leeway that the human has in the automated decision: does he have to follow it or can he divert from it? Moreover, the human's capability to interpret the data and to be sceptical about the result might also have an impact. Indeed, 'human involvement can also be rendered nominal by 'automation bias,' a psychological phenomenon where humans either over or under-rely on decision support systems' (Edwards and Veale, 2017). If the human, whose intervention has been added at the very end of the process, always simply trusts the algorithm's 'suggestion' without ever questioning it (either because it is not able to or because it doesn't want to), this human intervention will have no concrete impact on the decision-making process. For instance, in the specific field of customs fraud, while some fraud indicators result from human knowledge, there is also an automated model that analyses all of the feedback from the controllers on a continuous basis and updates itself every day. Based on these updates, it will produce hundreds of updated selection rules every day to determine which goods/undertakings should be controlled. Therefore, only the feedbacks are provided by humans, not the rules inferred from them. In such cases, it is fundamental to ensure that the inspectors keep collaborating by giving feedback on those newly suggested indicators, rather than simply following what the AI suggests, without any critical thinking. Yet, looking towards the future, it is possible that, in light of the constant budget cuts and reduction in personnel, there is a risk that the few inspectors left will simply end up trusting the machine without any critical thinking, because they have to meet their quotas of controls that they must do, and no longer have time to check the relevance of the indicators suggested by the machine. In such situations, we believe that this should be considered, *de facto*, as a decision based *solely* on automated processing. In this regard, the Article 29 Working Party outlines that:

> The controller cannot avoid the Article 22 provisions by fabricating human involvement. For example, if someone routinely applies automatically generated profiles to individuals without any actual influence on the result, this would still be a decision based solely on automated processing. To qualify as human involvement, the controller must ensure that any oversight of the decision is meaningful, rather than just a token gesture. It should be carried out by someone who has the authority and competence to change the decision. As part of the analysis, they should consider all the relevant data (Article 29 Working Party, 2018).

The key is thus whether a solely automated decision is taken, which produces legal effects concerning the data subject or similarly significantly affects them. In the case of AI applied to social security infringements or tax fraud, there will be a clear difference in whether the AI decides itself that a person has committed a social security infringement or tax fraud, or whether the AI merely warns the civil servants working for the relevant public administration that a specific person has a 'suspicious profile', and that this will lead to a human intervention, by the civil servant, who will look into the case. There is indeed a major difference between an AI 'deciding' itself, and an AI doing a preliminary analysis in the large quantity of data available to the public administration, in order to prioritise the cases on which the civil servants should focus because, by essence, they do not have the time to check every single case.

In this regard, it should be outlined that, at the Belgian FPS Finances, humans play an important role in the pre-investigation and investigation stages. For instance, human controllers have first established a set of typologies (types of suspicious profiles they want to detect) and use analytics to support detection. Indeed, the indicators used to identify these typologies are either proposed by humans or by the machine, which will propose a predictive shortlist of profiles that closely correspond to the typology that the investigators are looking for. To this effect, the machine will identify the most effective factors to detect these typologies, but the final decision to investigate (or not) a profile remains in the hand of the human controller. In fact, the investigators will often not investigate all suspicious cases identified in the pre-investigation. Rather, they will test some of these and will provide feedback on the usefulness of the signals at the end of the investigation. If the signals are relevant, they will investigate more cases from the suggested list. Additionally, for some types of fraud such as those linked to 'direct income taxes', the cases that are investigated following a data-mining recommendation are relatively small compared to the cases that controllers investigate on their own initiative (about 20%). Data mining is however, extremely important for, and well-suited to, other specific types of fraud, such as VAT fraud where about 80% of the cases derive from data mining.

Regarding the social security infringements, a distinction must be made between data-matching operations and the data-mining operations conducted in the OASIS data warehouse. For the former, some forms of *ex ante* data-matching cross-checks (e.g. checks before the social allocation is paid) do not need human intervention and are fully automated. This is because they are used to identifying objective obstacles to the payment of the allowances (e.g. no unemployment benefit if a person has a professional income). It is thus not a matter of interpretation, as there is no flexibility for the machine. This could easily be reviewed by a human, if requested by a data subject. *Ex post* data-matching cross-checks (e.g. checks after the social allocation is paid), on the other hand, always require human verification. Indeed, it is necessary for them to hear the person and ensure the rights of defence before taking a decision. This makes it possible to find cases that have escaped the *ex ante* cross-checks. All these *ex post* cross-checks are justified by a decision in due form with legal and factual justification, which can give rise to complaints to the ombudsman, and appeals. Data-mining operations, on the other hand, merely suggest cases to investigate, while the concrete investigation will always be done by a human. Moreover, the indicators integrated in OASIS have, in fact, been suggested by humans, namely inspectors in the field, who translate their experience of the cases they investigated into indicators. The machine simply looks for those indicators in the large amount of data.

However, even if the machine does not decide on its own that a person is a fraudster, this pre-selection could, in and of itself, be considered as a solely automated decision that significantly affects the data subject, as they are placed on the 'suspect list' and this leads to the opening of an investigation (De Raedt, 2017; Degrave, 2020). Indeed, this may be considered as having an effect on them, as it will entail additional scrutiny from the public administration of their behaviour (see, by analogy, Edwards and Veale, 2017). If this interpretation is followed, this would require implementing appropriate safeguards, such as the right to obtain a human intervention (Section 5.2.5 – Exceptions to the right ...).

Finally, it is interesting to outline that, in the SyRI decision of the Court of the Hague mentioned above,[72] the Court also assessed whether an automated individual decision-making occurred when SyRI was applied.[73] This is because the claimants argued that the submission of a risk report by the Social Affairs and Employment Inspectorate can be considered a decision with legal effect, or at least a decision that affects the data subjects significantly in another way, and that this decision is taken on the basis of automated individual decision-making within the meaning of Article 22 of the GDPR.[74] These claimants added that there is no meaningful human intervention prior to the submission of a risk report, as the mere removal of 'false positives' cannot qualify as such, nor can the assessment of the participating parties after receipt of a risk report.[75] The Dutch State, on the other hand, contested the fact that automated individual decision-making occurred and added that, in any case, the exceptions of Article 22.2 of the GDPR were met and that the amended legislation contained sufficient safeguards to protect privacy, as provided in Article 22.3.[76]

As a starting point, the Court outlined that the SyRI legislation leaves the option open whether predictive analyses, 'deep learning' and data mining can be used in the SyRI infrastructure, but that, at that point in time, no use was made of deep learning and data mining in the implementation of the SyRI legislation.[77] Regarding the potential effects of SyRI on the data subjects, the Court ruled that while the use of SyRI in and of itself is not aimed at having legal effect, a risk report nevertheless does have a similarly significant effect on the private life of the person to whom the risk report pertains.[78] Indeed, according to the Court, '[t]he fact that a risk report does not necessarily always lead to further investigation, or to an administrative or criminal-law sanction, and may also not be used as the sole basis for an enforcement decision, does not alter the significant effect on the private life of the data subject'.[79]

However, and quite surprisingly, the Court decided not to rule on whether this constituted an automated individual decision-making in the GDPR and, insofar as this is the case, on whether one or more of the exceptions to the prohibition in the GDPR had been met, because it deemed this to be irrelevant in the context of the assessment of whether the SyRI legislation meets the requirements of Article 8 of the European Convention on Human Rights.[80] This is quite disappointing, as such an assessment would have provided more clarity for administrations on what constitutes a decision based 'solely' on automated processing, and could also have shed more light on the safeguards that need to be put in place in such cases. Public administration will thus have to be careful when using AI in order to combat social security and tax fraud, as there remains a significant level of uncertainty about what can and cannot be done in this regard.

Exceptions to the right not to be subject to automated individual decision-making, and appropriate safeguards

According to Article 22.2 of the GDPR, individual decision-making based solely on automated processing can nevertheless be used if the decision: '(a) is necessary to enter into, or for the performance of, a contract between the data subject and a data controller; (b) is authorised by a law to which the controller is subject, provided that this law lays down suitable measures to safeguard the data subject's rights and freedoms and legitimate interests; or (c) is based on the data subject's explicit consent'.

In the case where a public administration wishes to rely on AI to fight social security infringements and tax fraud, its only viable option will be to rely on a law. Indeed, it is very unlikely that a contract can be concluded with all the citizens who would likely not be willing to allow the public administration to link their data in order to identify fraudulent behaviour. Moreover, as outlined in Chapter 4,[81] public administrations should avoid relying on the data subject's consent, as there will likely be a clear imbalance between the data subject and the controller, leading to the conclusion that the consent is not freely given (European Data Protection Board, 2020).[82] Accordingly, individual decision-making based solely on automated processing, in order to fight social security infringements and tax fraud, should be authorised by a law.[83] In fact, Recital 71 of the GDPR explicitly states that such automated individual decision-making 'should be allowed where expressly authorised by Union or Member State law to which the controller is subject, *including for fraud and tax-evasion monitoring*' (emphasis added).

This law will, however, have to lay down suitable measures to safeguard the data subject's rights and freedoms and legitimate interests.[84] While Article 22 does not explicitly indicate what such safeguards should be when the automated processing is authorised by a law, it does provide, in its Article 22.3, that in the case of automated processing based on consent or a contract, the data controller should, at least, provide the data subject with the right to obtain human intervention on the part of the controller, to express his or her point of view and to contest the decision. Logically, a law authorising individual decision-making based solely on automated processing should also,

[72] See Section 5.2.2.

[73] Rechtbank Den Haag, 5 februari 2020, Zaak n° C-09-550982-HA ZA 18-388, ECLI:NL:RBDHA:2020:865 (ECLI:NL:RBDHA:2020:1878 for the English version), points 6.55 to 6.60.

[74] *Ibid.*, point 6.57.

[75] *Ibid.*

[76] *Ibid.*, point 6.58.

[77] *Ibid.*, point 6.51.

[78] *Ibid.*, point 6.59.

[79] *Ibid.*

[80] *Ibid.*, point 6.60.

[81] See Chapter 4, Section 4.2.1.

[82] Recital 43 of the GDPR.

[83] Art. 22.2.b) of the GDPR.

[84] Art. 22.2.b) of the GDPR.

at least, provide for these three safeguards. This is supporting by the wording of Recital 71 of the GDPR, which provides that '*in any case*, such processing should be subject to suitable safeguards, which should include specific information to the data subject and the right to obtain human intervention, to express his or her point of view, to obtain an explanation of the decision reached after such assessment and to challenge the decision' (emphasis added). Recital 71 indeed targets all three of the exceptions.

Regarding Recital 71, it is also interesting to point out that it invites, in its second paragraph, the data controller to take into account the specific circumstances and context in which the personal data are processed, in order to ensure fair and transparent processing in respect of the data subject. In this regard, Recital 71 contains the following recommendations:

> [T]he controller should use appropriate mathematical or statistical procedures for the profiling; implement technical and organisational measures appropriate to ensure, in particular, that factors which result in inaccuracies in personal data are corrected and the risk of errors is minimised; [and] 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, political opinion, religion or beliefs, trade union membership, genetic or health status or sexual orientation, or that result in measures having such an effect.

At first glance, these recommendations appear to have limited binding force, as they are not included in the text of Article 22 of the GDPR. In reality, these recommendations are only the formulation of binding obligations formulated elsewhere in the GDPR, namely the principle of data protection by design[85] (appropriate technical and organisational measures), the requirement of data accuracy[86] (reduction of the risk of errors and correction of errors), and the requirement of data security[87] (secure personal data by taking into account the risks for the data subjects).

Finally, it should be outlined that the use of these exceptions is limited when they lead to the processing of special categories of data.[88] Indeed, Article 22.4 of the GDPR provides that decisions based solely on automated processing shall not be based on special categories of personal data referred to in Article 9.1 of the GDPR, unless point (a) or (g) of Article 9.2 applies and suitable measures to safeguard the data subject's rights and freedoms and legitimate interests are in place. This can have a substantial impact for public administrations wishing to use automated individual decision-making in order to fight social security infringements, as they might want to rely on 'data concerning health',[89] which are listed in the special categories of data.[90] Indeed, such data could be processed if the data subject has given explicit consent to their processing (Art. 9.2.a) of the GDPR). Yet, as we have seen in Chapter 4,[91] public administration should avoid relying on the data subject's consent, which will likely not be deemed as being freely given. Accordingly, the processing of health data by an AI in order to fight social security infringements will have to be necessary for reasons of substantial public interest and will have to be based on a law, which shall be proportionate to the aim pursued, respect the essence of the right to data protection and provide for suitable and specific measures to safeguard the fundamental rights and the interests of the data subject (Art. 9.2.g) of the GDPR).[92] As has been outlined in the SyRI case mentioned above,[93] fighting social security infringements can be considered of substantial public interest, as the Court in The Hague ruled that combatting fraud is key to maintaining citizen support in the social security system, which is one of the pillars of society, that new technological possibilities to prevent and combat fraud should therefore be used, and that the SyRI legislation thus pursues an important objective of general public interest.[94]

[85] Art. 25 of the GDPR.

[86] Art. 5.1.d) of the GDPR.

[87] Art. 5.1.f) of the GDPR.

[88] 'Personal data revealing racial or ethnic origin, political opinions, religious or philosophical beliefs, or trade union membership, (...) genetic data, biometric data, data concerning health or data concerning a natural person's sex life or sexual orientation' (Art. 9.1 of the GDPR).

The right to obtain an explanation of the decision and AI explainability

An attentive reader of the GDPR will have noticed that Recital 71 goes further than Article 22.3 in terms of the appropriate safeguards that must be implemented, as it also mentions the right to obtain an explanation of the decision reached after the automated assessment, which does not appear in Article 22 of the GDPR.

The right to obtain an explanation of the automated decision

According to some authors, this implies that Article 22 of the GDPR does not, in fact, provide a right to obtain an explanation about how the automated decision was reached because, contrary to the Articles of the GDPR, the Recitals are not legally binding (Edwards and Veale, 2017; Wachter *et al.*, 2017). For these authors, this is not a mere omission, but a genuine desire not to include this right in the text of Article 22.3 of the GDPR (Wachter *et al.*, 2017). The European Parliament had indeed proposed to include this right in the article of the GDPR (Committee on Civil Liberties, Justice and Home Affairs, 2013), whereas the Council was of the opinion that this right should be mentioned only in the recitals (Presidency of the Council of the European Union, 2015), which clearly shows, according to them, that the final text of the GDPR is the result of a deliberate choice made during the trialogue negotiations, and not of a mere drafting error (Wachter *et al.*, 2017). However, other authors express more reservations on whether this implies that such a right to explanation does not exist, as 'many issues too controversial for agreement in the main text have

[89] Art. 4.15 of the GDPR.

[90] Art. 9.1 of the GDPR.

[91] See Chapter 4, Section 4.2.1.

[92] Art. 22.4 of the GDPR

[93] See Sections 5.2.2 and 5.2.5 – Automated decision making based solely on automated processing.

[94] Rechtbank Den Haag, 5 februari 2020, Zaak n° C-09-550982-HA ZA 18-388, ECLI:NL:RBDHA:2020:865 (ECLI:NL:RBDHA:2020:1878 for the English version), points 6.3 and 6.4.

been kicked into the long grass of the recitals, throwing up problems of just how binding they are' (Edwards and Veale, 2017). In fact, even the former group of authors admits that it could be argued that 'although it is certainly not explicit in the phrasing of Article 22.3, the right to obtain human intervention, express views or contest a decision is meaningless if the data subject cannot understand how the contested decision was taken' (Wachter *et al.*, 2017).

Thus, this right to obtain an explanation about the decision based solely on automated processing could be implicitly encapsulated in Article 22.3 of the GDPR. This point of view seems to be supported by the explanatory report of the Modernised Convention 108, which outlines that:

> Data subjects should be entitled to know the reasoning underlying the processing of data, including the consequences of such a reasoning, which led to any resulting conclusions, in particular in cases involving the use of algorithms for automated decision-making including profiling. (...) [T]hey should be entitled to know the logic underpinning the processing of their data and resulting (...) decision, and not simply information on the decision itself. Having an understanding of these elements contributes to the effective exercise of other essential safeguards such as the right to object and the right to complain to a competent authority. (Committee of Ministers of the Council of Europe, 2018)

The Article 29 Working Party seems to embrace the same view, as it indicates that: 'the controller should find simple ways to tell the data subject about the rationale behind, or the criteria relied on in reaching the decision' (Article 29 Working Party, 2018) and that the information that is provided should be 'sufficiently comprehensive for the data subject to understand the reasons for the decision' (Article 29 Working Party, 2018). Indeed, according to the Article 29 Working Party, 'the data subject will only be able to challenge a decision or express their view if they fully understand how it has been made and on what basis' (Article 29 Working Party, 2018).

Moreover, we believe that it could be argued that this right to obtain an explanation is also intrinsically included in the data subject's right to information,[95] as well as their right of access.[96] These two rights, which have been presented above,[97] enable them to receive useful information concerning the underlying logic of the processing operation, which should not only enable the data subject to know what is being done with their data, but also to understand the underlying logic of the processing (De Terwangne, 2015). That being said, it would have been preferable, for the sake of clarity, to explicitly include the reference to this right to obtain an explanation in Article 22.3 of the GDPR, rather than only in Recital 71.

[95] Arts. 13.2.f) and 14.2.g) of the GDPR. These articles both stipulate that the data subject should receive information about 'the existence of automated decision-making, including profiling, referred to in Article 22.1 and 22.4 and, at least in those cases, meaningful information about the logic involved, as well as the significance and the envisaged consequences of such processing for the data subject'.

[96] Art. 15.1.h) of the GDPR, which uses the same phrasing as Arts. 13.2.f) and 14.2.g).

[97] See Sections 5.2.2 and 5.2.3.

However, even if, on the basis of the above, it was to be deemed that the data subject has a right to an explanation of the decision, which is controversial, the ability of the public administration to provide explanations about the decisions taken by an AI used to fight social security infringements and tax fraud will depend on the type of AI that is used (expert systems vs. neural networks) and on the legal entity that holds the rights on the algorithm (the public administration itself vs a private entity). Indeed, the above analysis of these factors in the context of the obligation to inform the data subjects about the logic involved in the AI decision-making equally applies to the provision of explanations about the decision.[98]

Yet, it should be underlined that, at this point, public administrations seem to be aware of the importance of explainability. For instance, the Belgian FPS Finances is conscious that it needs to be able to explain why a certain person or company is suspected of tax fraud. For each case, the data miners are able to explain the reasoning behind the detection (indicators, techniques applied, etc.). Even more advanced techniques, such as social network analysis, used to detect more complex fraud types (e.g. 'domino bankruptcies'), are designed by the data miners. However, it should be mentioned that, when investigating a specific case, controllers can rely on AI techniques delivered by private software companies. It can therefore not be excluded that these private companies might hide behind commercial secrecy to refuse to provide explanations about the functioning of their algorithm, and this should be a key point of attention when dealing with those software providers.

Regarding social security infringements, even if the *ex ante* data-matching cross-checks are fully automated (Section 5.2.5 – Automated decision-making based solely on automated processing), they remain explainable because they are used to identify objective obstacles to the payment of the allowances. The machine thus does not have any margin of interpretation. Regarding bilateral *ex post* data-matching cross-checks, their results are also explainable, since they always imply a human verification. Similarly, the results of the data-mining operations conducted in the data warehouse are also explainable, since the indicators that are used to pinpoint suspicious cases have, in fact, been suggested by humans (the data miners). Yet, in the future, as fraud becomes more and more complex, the use of simpler algorithms with explainable business rules may become an issue, especially if public administrations increasingly need to resort to private sector providers.

However, it should be outlined that a person or an undertaking will not be informed that it has been flagged as being a potential (tax or social security) fraudster following data-mining operations conducted in the data warehouse, if the follow-up investigation did not result in the finding of fraud. Consequently, this person/undertaking might be repeatedly flagged as 'suspicious', although erroneously, without being aware of it and without being able to request explanations about why this is the case. This highlights that it is complex for public administrations to find a balance between being fully transparent and explaining the data-mining processes and models used, and the need to protect the confidentiality of their fraud analytics processes, as otherwise the fraudsters will adapt and avoid being detected.

[98] See Section 5.2.2.

To conclude, it is fundamental to emphasise that all of the above depends on how the notion of 'explainability' should be understood. What does it mean for a public administration to explain an algorithmic decision? Arguably, explaining 'the decision' requires more precision than simply explaining the 'logic involved', which could be more general. If we revert to the SyRI case, explaining the 'logic involved' would probably amount to explaining which indicators and which risk model are used,[99] while explaining the 'decision' may require an explanation of how the weight of each indicator has, in the specific case, led to the decision whether the person has a 'suspicious profile' or not.[100] We now briefly turn to the analysis of this notion in the AI context.

AI explainability

In the computer science literature, there are essentially two models of AI explainability (Bibal and Frenay, 2016). On the one hand, there are *interpretable models*, which are 'understandable either because their mathematical expressions are easy to understand (as it is the case with linear models) or can be represented in an easily understandable manner (as it is the case with decision trees)' (De Streel *et al.*, 2020). On the other hand, there are *black box models*, which 'are not easy to understand because their mathematical expression is neither straightforward nor easily representable in an understandable manner. For those models, understanding can be improved through explanations by using methods which are external to the models such as visualisation or approximation with interpretable models (Mittelstadt *et al.*, 2019)' (De Streel *et al.*, 2020).

In law and ethics, however, there is no commonly-agreed definition of AI explainability, although this has become a major concern for policy-makers across the world (De Streel *et al.*, 2020; Pasquale, 2015). At the European level, the requirement of AI explainability is directly highlighted in the 'Ethics Guidelines for Trustworthy AI' of the High-Level Expert Group on AI (High-Level Expert Group on Artificial Intelligence, 2019), and indirectly highlighted in the European Commission's White Paper on AI (European Commission, 2020). (De Streel *et al.*, 2020). Indeed the former lists four ethical principles in the context of AI systems, among which the 'Explicability' principle, which can be understood as a synonym of explainability (High-Level Expert Group on Artificial Intelligence, 2019), and lists seven requirements for trustworthy AI, among which the 'Transparency' requirement which explicitly includes explainability (De Streel *et al.*, 2020; High-Level Expert Group on Artificial Intelligence, 2019). The latter lists six requirements for high-risk AI applications, among which the need for 'Information provision', which does not refer explicitly to explainability, but which refers to the need for transparency and to the need to clearly inform about the functioning of the AI systems (European Commission, 2020).

[99] Rechtbank Den Haag, 5 februari 2020, Zaak n° C-09-550982-HA ZA 18-388, ECLI:NL:RBDHA:2020:865 (ECLI:NL:RBDHA:2020:1878 for the English version), point 4.22. See Sections 5.2.2 and 5.2.5 – Automated decision-making based solely on automated processing.

[100] *Ibid.*, points 4.29 to 4.30.

The requirement of AI explainability can also be observed indirectly in the Commission's proposal for an Artificial Intelligence Act, where it is provided that, in order to address 'the opacity that may make certain AI systems incomprehensible to or too complex for natural persons' (European Commission, 2021), high-risk AI systems 'shall be designed and developed in such a way to ensure that their operation is sufficiently transparent to enable users to interpret the system's output and use it appropriately'.[101] Furthermore, Annex IV of that proposal stipulates that the technical documentation pertaining to the AI system should contain: 'the design specifications of the system, namely 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; [and] what the system is designed to optimise for and the relevance of the different parameters'.[102]

Regarding the principle of 'explicability' of AI, the 'Ethics Guidelines for Trustworthy AI' of the High-Level Expert Group on AI stipulate that:

> Explicability is crucial for building and maintaining users' trust in AI systems. This means that processes need to be transparent, the capabilities and purpose of AI systems openly communicated, and decisions – to the extent possible – explainable to those directly and indirectly affected. Without such information, a decision cannot be duly contested. An explanation as to why a model has generated a particular output or decision (and what combination of input factors contributed to that) is not always possible. These cases are referred to as 'black box' algorithms and require special attention. In those circumstances, other explicability measures (e.g. traceability, auditability and transparent communication on system capabilities) may be required, provided that the system as a whole respects fundamental rights. The degree to which explicability is needed is highly dependent on the context and the severity of the consequences if that output is erroneous or otherwise inaccurate (High-Level Expert Group on Artificial Intelligence, 2019).

Regarding the 'explainability' of AI, the 'Ethics Guidelines for Trustworthy AI' of the High-Level Expert Group on AI stipulate that:

> Explainability concerns the ability to explain both the technical processes of an AI system and the related human decisions (e.g. application areas of a system). Technical explainability requires that the decisions made by an AI system can be understood and traced by human beings. Moreover, trade-offs might have to be made between enhancing a system's explainability (which may reduce its accuracy) or increasing its accuracy (at the cost of explainability). Whenever an AI system has a significant impact on people's lives, it should be possible to demand a suitable explanation of the AI system's decision-making process. Such explanation should be timely and adapted to the expertise of the stakeholder concerned (e.g. layperson, regulator or researcher). In addition, explanations of the degree to which

[101] Article 13.1 of the proposal for an Artificial Intelligence Act.

[102] Article 11.1 and Annex IV, point 2.b) of the proposal for an Artificial Intelligence Act.

an AI system influences and shapes the organisational decision-making process, design choices of the system, and the rationale for deploying it, should be available (hence ensuring business model transparency) (High-Level Expert Group on Artificial Intelligence, 2019).

Moreover, additional guidance about the notion of explainability can be found in the 'Explaining decisions made with AI' report of the UK's Information Commissioner's Office and the Alan Turing Institute (De Streel *et al.*, 2020; Information Commissioner's Office and the Alan Turing Institute, 2020). This report first outlines that there are two subcategories of explanation, namely '*process-based explanations* which give you information on the governance of your AI system across its design and deployment; and *outcome-based explanations* which tell you what happened in the case of a particular decision' (Information Commissioner's Office and the Alan Turing Institute, 2020) (emphasis added). If we apply this to our previous developments, 'process-based explanations' would pertain to the explanation of the 'logic involved' behind the algorithm used to fight social security infringements or tax fraud, while 'outcome-based explanations' would pertain to the explanation of the concrete reasons why the algorithm considers that an individual has a 'suspicious profile'. These two subcategories of explanation are sometimes also referred to as 'model-centric' and 'subject-centric' explanations (Edwards and Veale, 2017). While the former focusses more information on the AI model, the training metadata, the performance metrics and the estimated global logics, the latter aims at providing information about a specific query, output or decision (Edwards and Veale, 2017).

Additionally, the 'Explaining decisions made with AI' report identifies six main types of explanation:

1. '*Rationale explanation*: the reasons that led to a decision, delivered in an accessible and non-technical way.
2. *Responsibility explanation*: who is involved in the development, management and implementation of an AI system, and who to contact for a human review of a decision.
3. *Data explanation*: what data has been used in a particular decision and how.
4. *Fairness explanation*: steps taken across the design and implementation of an AI system to ensure that the decisions it supports are generally unbiased and fair, and whether or not an individual has been treated equitably.
5. *Safety and performance explanation*: steps taken across the design and implementation of an AI system to maximise the accuracy, reliability, security and robustness of its decisions and behaviours.
6. *Impact explanation*: steps taken across the design and implementation of an AI system to consider and monitor the impacts that the use of an AI system and its decisions has or may have on an individual, and on wider society' (emphasis in the original text) (Information Commissioner's Office and the Alan Turing Institute, 2020).

Two reservations must, however, be made regarding AI explanations. Firstly, AI systems that rely on a low number of variables are easier to 'explain', yet 'systems with more variables will typically perform better than simpler systems, so we may end up with a trade-off between performance and explicability' (Edwards and Veale, 2017). Moreover, these complex models may be especially

difficult to explain because 'the features that are being fed in might lack any convenient or clear human interpretation in the first place, even if we are creative about it. LinkedIn, for example, claims to have over 100,000 variables held on every user that feed into ML modelling. Many of these will not be clear variables like 'age,' but more abstract ways you interact with the webpage, such as how long you take to click, the time you spend reading, or even text you type in a text box but later delete without posting' (Edwards and Veale, 2017).

Secondly, because AI systems do not function like human minds, it might be impossible, in some cases, to 'explain' in a way that is satisfactory for humans how a complex AI system relying on thousands of variables and correlations has taken a decision, i.e. to provide a 'humanly-understandable decision' (Busuioc, 2020). Accordingly, some authors argue that the focus should instead be set on the human interpretability of a decision (i.e. it can be understood and interpreted by a layman) rather than on its 'explanation' per se (Busuioc, 2020). In this regard, "pedagogical' systems which create explanations around a model rather than from decomposing it may be useful and benefit from not relying on disclosure of proprietary secrets or IP' (Edwards and Veale, 2017).

In light of the above, the public administration using an AI to fight social security infringements and tax fraud must thus be able to explain how the algorithm has, technically, reached its decision in a way that can be understood and verified by humans (High-Level Expert Group on Artificial Intelligence, 2019). This implies the ability to explain not only the logic involved behind the algorithm (process-based explanation) but also the concrete decision of why the algorithm considers that a person has a 'suspicious profile' (outcome-based processing) (Information Commissioner's Office and the Alan Turing Institute, 2020). Similarly, it must be able to explain how human decisions, based on the algorithm's decision, have been taken. This explanation should be adapted to the person to whom it is given, as the person must be able to understand it (a layperson will not have the same level of understanding as a researcher in AI).

In order to be compliant with Article 22 of the GDPR, we believe that – provided that the data subject has a right to an explanation of the AI decision, which is controversial –,[103] the public administration should, at least, be able to explain in an accessible and non-technical way, the reasons that led to the decision (rationale explanation), to identify who is involved in the development, management and implementation of an AI system, and to identify who to contact for a human review of the decision (responsibility explanation) and what data has been used, and how, in the particular decision (data explanation) (Information Commissioner's Office and the Alan Turing Institute, 2020). Indeed, the other three types of explanation outlined by the 'Explaining decisions made with AI' report (fairness explanation, safety and performance explanation, and impact explanation) (Information Commissioner's Office and the Alan Turing Institute, 2020) do not so much focus on the explanation of the decision as such, but more on how the AI system that has taken the decision is built. It is, however, recommended to also provide these types of explanation in order to reinforce the citizens' trust in the use of AI by public administrations.

[103] See previous Section 5.2.5 – The right to obtain an explanation of the automated decision.

As mentioned above,[104] the ability of the public administration to provide such explanations about the decisions taken by an AI used to fight social security infringements and tax fraud will depend on the type of AI that is used and on the legal entity that holds the rights to the algorithm. In those circumstances, 'other explicability measures (e.g. traceability, auditability and transparent communication on system capabilities) may be required, provided that the system, as a whole, respects fundamental rights' (High-Level Expert Group on Artificial Intelligence, 2019). To ensure that it meets these explanation requirements, the public administration could use the 'Explainability checklist' provided in the 'Ethics Guidelines for Trustworthy AI' of the High-Level Expert Group on AI:

> 'Did you assess:
>
> to what extent the decisions and hence the outcome made by the AI system can be understood?
>
> to what degree the system's decision influences the organisation's decision-making processes?
>
> why this particular system was deployed in this specific area?
>
> what the system's business model is (for example, how does it create value for the organisation)?
>
> Did you ensure an explanation as to why the system took a certain choice resulting in a certain outcome that all users can understand?
>
> Did you design the AI system with interpretability in mind from the start?
>
> Did you research and try to use the simplest and most interpretable model possible for the application in question?
>
> Did you assess whether you can analyse your training and testing data? Can you change and update this over time?
>
> Did you assess whether you can examine interpretability after the model's training and development, or whether you have access to the internal workflow of the model?' (High-Level Expert Group on Artificial Intelligence, 2019, 2020).

5.3 Additional legal challenges for the development of algorithmic decision-making tools by the public administration

In Section 5.2, we outlined the main personal data protection rules that must be factored in by public administrations willing to rely on AI techniques to improve the delivery of their public services. In this regard, particular attention was devoted to public policies and decision-making linked to social security infringements and tax fraud. With the same objective in mind, this section will be dedicated to the analysis of additional legal challenges that shall be considered when reflecting on the development of algorithmic decision-making tools by the public administration. These legal challenges revolve around four topics, namely the citizens' right to a human public service, their right to an equal access to public services, the transparency of the public administration and the administrative decision-making through AI systems (Gérard, 2017).

[104] See Section 5.2.2.

5.3.1 Risk of 'dehumanised' public services

Like any new technology, resorting to AI in public services presents both risks and opportunities. The use of AI could first make public services more available to citizens, as they could access it 24 hours a day, 7 days out of 7, and it would arguably speed up services (Gérard, 2017). Moreover, delegating the repetitive and non-complex tasks to an AI would grant more time to the civil servants to focus on more complex cases, which would ultimately benefit the citizens (Gérard, 2017). However, resorting to AI rather than humans might 'dehumanise' the public service (Gérard, 2017). Although there is, as such, no 'right to a human public service', the right to human dignity could come into play (Gérard, 2017). Indeed, this right is recognised as 'the very essence' of the European Convention on Human Rights[105] and is explicitly enshrined in Article 1 of the Charter of Fundamental Rights of the European Union (Gérard, 2017). Accordingly, it should be ensured that machines (AI) do not become the norm, and people the exception (Nevejans, 2016).

5.3.2 Equal access to public services

Resorting to AI, like resorting to any other technology, may lead to a digital divide among citizens, as some of them will not be able to use these technologies, either because they lack the skills to do so, or because they have physical inabilities that prevent them from doing so (Chantillon *et al.*, 2017). In the specific case of AI, the complexity of the technical knowledge required for the effective use of this technology is such as to make it totally or partially inaccessible to part of the population (Gérard, 2017). This may have an impact on two key principles of the public service, namely the principle of equality and the principle of accessibility.

The principle of equality

This principle of equality is enshrined in several supranational Conventions[106] and in Articles 10 and 11 of the Belgian Constitution, and it prevents the legislator from treating differently two equal situations or categories of people, or conversely, from treating equally two different situations or categories of people (Gérard, 2017).[107] Applied to the public service, the principle of equality is defined as the 'law of equality', and constitutes, together with the 'law of mutability' and the 'law of continuity', the three 'laws of the public service' (Gérard, 2017). According to this law, all the users of the public service should be treated equally and should benefit from the same services and advantages (Gérard, 2017).

[105] ECtHR, *Christine Goodwin v. United Kingdom*, 11 July 2002, req. n° 28957/97, point 90; *S.W. v. United Kingdom*, 22 November 1995, req. n° 20166/91, point 44.

[106] Article 14 of the European Convention on Human Rights and Articles 20, 21 and 23 of the Charter of Fundamental Rights of the European Union.

[107] See Belgian Constitutional Court, case n° 21/89, 13 July 1989, point B.4.5.b; and case n° 16/92, 12 March 1992, point B.3.3.

If applied to AI, this means that all the users should be treated equally by the algorithm, which shall not be biased and shall not entail discriminations against some categories of the population. Indeed, even if AI systems are often presented as being deprived of any bias, they might reflect biases that have been integrated, intentionally or not, by their human creators, but also biases that derive from the training data that has been selected to build them (Défenseurs des droits et CNIL, 2020). One such bias is the lack of representative data, which, in the field of facial recognition technologies, led to the finding that the chances of 'false positives' were much greater for woman and for people of colour, because the algorithm had, to a large extent, been trained with images of white men (Buolamwini and Gebru, 2018; Défenseurs des droits et CNIL, 2020). Indeed, even if, at first sight, the AI systems rely on 'neutral criteria', the combination of several of these criteria can lead to biases and, as a matter of consequence, to automatic and systemic discriminations (Défenseurs des droits et CNIL, 2020).

Moreover, these discrimination risks are enhanced by the fact that AI systems often tend to target and control minorities and marginalised social groups (Défenseurs des droits et CNIL, 2020; Eubanks, 2018). This is especially relevant to keep in mind for AI systems aiming at fighting social security infringements and tax fraud. For instance, in the SyRI case brought before the Court in The Hague,[108] the claimants argued that the use of SyRI had a discriminatory and stigmatising effect, because it was allegedly used to 'further investigate neighbourhoods that are known as 'problem areas'. This increases the chances of discovering irregularities in such areas as compared to other neighbourhoods, which in turn confirms the image of a neighbourhood as a problem area, contributes to stereotyping and reinforces a negative image of the occupants of such neighbourhoods, even if no risk reports have been generated about them'.[109] The Court agreed that SyRI had only been applied to 'problem districts' and that there was a risk that this could lead to biases towards people with a lower socio-economic status or an immigration background,[110] but outlined that, because of the lack of transparency of the SyRI legislation, it was unable to assess whether this risk of discrimination had been sufficiently neutralised by the State.[111] The court thus acknowledged the risk of the discrimination but was unable to verify whether this risk materialised *in casu*.

In light of the above, public administrations wishing to rely on AI to fight social security infringements and tax fraud will have to audit their AI systems on a regular basis in order to ensure that they are not biased and that they do not discriminate against some categories of citizens (Défenseurs des droits et CNIL, 2020). Naturally, this implies that the public authorities need to be able to both understand and explain how the AI system works.[112] In this regard, it is relevant to mention that, in the context of tax fraud, data quality checks are performed in the data warehouse, in order to ensure that the data is not biased at the application level and does not lead to discrimination. In fact, several data-mining projects pursued by the data miners solely aim at improving and ensuring data quality. Regarding social security infringements, we can only assume that any risk of inequality is discarded at the stage of the drafting of the data transfer protocol or at the stage of the obtention of the prior authorisation from the Information Security Committee.[113] However, due to the relative opacity in this regard, the existence of inequalities and discrimination cannot be excluded.

The principle of accessibility

The principle of accessibility of public services is enshrined in the Belgian Charter of the public service user.[114] This Charter provides that 'public services must be accessible in the broadest sense of the term, which goes beyond the problems of physical accessibility and proximity (...). It is also a question of the clarity of the texts. Administrative documents and legislation should not be drafted in such a way that the public has great difficulty in understanding them'.[115] In this regard, the application of the principle of accessibility of the public service precludes the introduction of AI that would make this public service excessively complex for users (Gérard, 2017). Therefore, the public administration wishing to rely on AI to fight social security infringements and tax fraud will have to ensure that the functioning of the algorithm is not too complex, and that the citizen can understand, at least at a high level of abstraction, how the algorithm works and why they are, for instance, considered a 'suspicious profile'. This links to the requirement of algorithmic transparency (Section 5.2.2 and 5.3.3).

5.3.3 Transparency of the public administration

The GDPR is not the only legal text that imposes transparency of the public administration for citizens. Indeed, citizens benefit from the fundamental right of administrative publicity, which is enshrined in Article 32 of the Constitution (Gérard, 2017). At the federal level, this constitutional provision is further specified in a Law of 11 April 1994,[116] which provides for two types of administrative publicity, namely active and passive publicity (Gérard, 2017).

In light of the obligation of active publicity, the federal public administration has to actively provide, independently of any request, a clear and objective information to the public about its actions, competences and functioning (Gérard, 2017).[117]

In light of the obligation of passive publicity, the federal public administration should, at the request of a citizen, allow them to consult an administrative document, to obtain explanations about it and to receive a copy of it (Gérard, 2017).[118] In order to broaden the scope of this publicity obligation

[108] Rechtbank Den Haag, 5 februari 2020, Zaak n° C-09-550982-HA ZA 18-388, ECLI:NL:RBDHA:2020:865 (ECLI:NL:RBDHA:2020:1878 for the English version). See Sections 5.2.2 and 5.2.5 – Automated decision-making based solely on automated processing.

[109] *Ibid.*, point 6.92.

[110] *Ibid.*, point 6.93.

[111] *Ibid.*, point 6.94.

[112] See Sections 5.2.5 – The right to obtain an explanation of the decision and AI explainability, and 5.3.4.

[113] See Chapter 4, Section 4.2.3

[114] Charte de l'utilisateur des services publics, 4 décembre 1992, *M.B.*, 22 janvier 1993.

[115] See Chapter II, Section A of the Charter. Author's own translation.

[116] Loi du 11 avril 1994 relative à la publicité de l'administration, *M.B.*, 30 juin 1994.

[117] Art. 2 of the Law of 11 April 1994.

[118] Arts. 4 and 5 of the Law of 11 April 1994.

as much as possible, an administrative document is defined as 'any information, whatever its form, held by a public administration' (Gérard, 2017).[119] For documents of a personal nature – e.g. administrative documents involving an assessment or value judgement relating to a named or easily identifiable natural person, or a description of a conduct the disclosure of which could manifestly cause harm to that person –,[120] the requester must justify an interest to access it.[121]

Regarding, more specifically, the explanations about the use of algorithms for decision-making pertaining to individuals, it is worth mentioning that a law proposal of 6 April 2021 provides that, in order to increase transparency, it should be compulsory for administrations to publish online the rules defining the main algorithmic treatments used in the performance of their tasks when these constitute all or part of the basis for individual decision.[122] Moreover, this proposal provides that, for any administrative document with an individual scope, the administration shall communicate to the person who is the subject of an individual decision taken in whole or in part on the basis of an algorithmic processing, at the latter's request, the characteristics of the algorithm in an intelligible form, provided that this communication does not infringe secrets protected by law.[123] This would cover the degree and type of contribution of the algorithmic processing to the decision-making; the data processed and their sources; the processing parameters and, where appropriate, their weighting applied to the individual's situation; and the operations carried out through the processing.[124] It remains to be seen whether this proposal will be adopted.

It should, however, be noted that Article 6.1.6° of this Law of 11 April 1994 provides that the federal public administration shall refuse a request for consultation, explanation or disclosure in the form of a copy of an administrative document if it is satisfied that the interest in disclosure does not outweigh the protection of a federal economic or financial interest (Degrave and Lachapelle, 2014). Such a decision will have to be formally motivated, as requested by the law of 29 July 1991[125] (Degrave and Lachapelle, 2014).

Public administrations wishing to rely on AI to fight social security infringements and tax fraud will thus have to be transparent about these missions, about the role played by the AI and by humans, and about the logic behind the decisions (active publicity) and will have to conduct a balance of interests in order to determine whether they need to provide a copy of these information to citizens that would request it (passive publicity) (Gérard, 2017). This will allow for a reduction in the opacity of the administration's actions in the eyes of the citizens (Committee of Ministers of the Council of Europe, 2021; Degrave, 2014).

[119] Art. 1.b).2° of the Law of 11 April 1994.

[120] Art. 1.b).3° of the Law of 11 April 1994.

[121] Art. 4 of the Law of 11 April 1994.

[122] Proposition de loi modifiant la loi relative à la publicité de l'administration du 11 avril 1994 afin d'introduire une plus grande transparence dans l'usage des algorithmes par les administrations, 6 avril 2021, *Doc. parl.*, Chambre, sess. ord., 2020-2021, no 55-1904/001, p. 6 and 8.

[123] *Ibid.*

[124] *Ibid.*, p. 8.

[125] Loi du 29 juillet 1991 relative à la motivation formelle des actes administratifs, *M.B.*, 12 septembre 1991.

5.3.4 Administrative decision-making by an AI system

Any use of AI for decision-making, irrespective of whether it is *solely* based on automated processing[126] (i.e. even if there is a human intervention), also generates challenges from an administrative law point of view. Indeed, a difference must be made between decisions where the AI simply applies a precise number of rules that it is bound to follow, and decisions where the AI has more leeway in taking its decision (Gérard, 2017). In the first case, resorting to AI might actually be extremely useful and efficient. For example, in terms of social security, an AI could be extremely efficient in crossing the citizens' data, in order to determine whether they are entitled to a certain subsidy. Here, the conditions to receive this subsidy are known in advance and easily verifiable, e.g. the citizen can receive the subsidy if conditions A, B and C are met, and the AI can easily check this and allow the subsidy based on those rules. In the second case, it might be more problematic to resort to AI decision-making if it has some leeway, as it will not always be possible to verify whether the AI's decision is compatible with the law, notably if the public administration is not able to check nor explain how the AI came to that decision.[127]

Additionally, the Belgian law of 29 July 1991 on the formal motivation of the administrative acts provides that all unilateral legal acts of individual scope emanating from an administrative authority, whose purpose is to produce legal effects in respect of one or more persons under its jurisdiction, have to be 'formerly motivated', which implies that the act must contain the legal and factual conditions that have led to the decision.[128] This formal motivation must be adequate,[129] which means that the person should be able to understand the reasoning that has led to the decision (Gérard, 2017). For decisions taken by an AI, this should be rather easy to do if it simply applies a precise number of rules that it is bound to follow, but it might be more problematic to explain the decision taken by the AI if it has some leeway in doing so (Gérard, 2017). Our developments above regarding the right to obtain an explanation about the AI's decision and about the explainability of an AI decision can be transposed here.[130] Finally, this formal motivation must satisfy the principle of accessibility outlined above, and must thus be written in such a way that the citizen can clearly understand the motives that have led to the decision (Gérard, 2017).[131]

5.4 Conclusions

As outlined throughout this chapter, public administrations are increasingly relying on artificial intelligence technologies and automated decision-making in order to provide public services and support their decision-making. Yet, while this can lead to significant benefits, notably in terms of efficiency, for these administrations, it must not be overlooked that the use of such technologies

[126] See Section 5.2.5 – Automated decision-making based solely on automated processing.

[127] See Section 5.2.5 – The right to obtain an explanation of the decision and AI explainability.

[128] Loi du 29 juillet 1991 relative à la motivation formelle des actes administratifs, *M.B.*, 12 septembre 1991, arts. 1, 2 and 3.

[129] Art. 3 of the Law of 29 July 1991.

[130] See Section 5.2.5 – The right to obtain an explanation of the decision and AI explainability.

[131] See Section 5.3.2 – The principle of accessibility.

can also generate significant impacts on the lives and fundamental rights of the citizens to which automated decision-making systems are applied. Therefore, administrations need to comply with the legal framework that aims at limiting the uses they can make of such technologies, in order to circumscribe these impacts. In this regard, the aim of this chapter was precisely to outline how this could be done in practice.

On that basis, several key takeaways pertaining to the use of artificial intelligence and of automated decision-making technologies by public administrations can be given. First, since data has to be processed fairly and in a transparent manner, public administration making use of individual decision-making based 'solely' on automated processes will have to inform the citizens about the existence of such processes, and will have to provide them, at least with meaningful information about the logic involved, as well as about the significance and the envisaged consequences of such processing.[132] In order to comply with this requirement, it is preferable for public administrations to make use of expert systems rather than neural networks, and to ensure that they hold the rights to the algorithm, rather than to grant those rights to the private entity that has been tasked, through the means of a public procurement, to develop the algorithm for them. Second, a public administration wishing to use AI applications should anticipate the potential data access and data erasure requests that it may receive from citizens, and it should thus make sure that the AI system is designed in a way that allows them to respond in a timely fashion to such requests.[133] Third, it is fundamental for a public administration willing to rely on AI techniques to respect the citizens' right not to be subjected to a decision based 'solely' on automated processing.[134] This right echoes the strong will of the human being not to be entirely subject to the machine, as the former does not accept the idea that a decision may be imposed on it solely on the basis of conclusions reached by that machine. In the same vein, it would not be acceptable for public services, based on AI or automated decision-making technologies, to become 'dehumanised' or no longer equally accessible for all.[135] Coming back to the right not to be subjected to a decision based 'solely' on automated processing, public administrations should refrain from including a fictitious or negligible human intervention in the automated decision process simply in order to avoid (potentially in bad faith) the application of this right.[136] They should also adopt suitable measures to safeguard the citizen's rights and freedoms and legitimate interests, which should, at least, provide the latter with the right to obtain human intervention on the part of the controller, to express his or her point of view and to contest the decision.[137] Last but not least, they should also be able to explain the decisions that have been taken by the algorithm.[138]

[132] See Sections 5.2.2 and 5.3.3.
[133] See Sections 5.2.3 and 5.2.4.
[134] See Section 5.2.5.
[135] See Sections 5.3.1 and 5.3.2.
[136] See Section 5.2.5 – Automated decision-making based solely on automated processing.
[137] See Section 5.2.5 – Exceptions to the right not to be subject to an automated individual decision-making, and appropriate safeguards.
[138] See Sections 5.2.5 – The right to obtain an explanation of the decision and AI explainability and 5.3.4.

An important common point between the personal data protection rules and the core principles of administrative law that have been presented in this chapter is that they both aim to give more control to the citizens, through increased transparency and understanding of the concrete uses of these technologies made by administrations. Furthermore, they aim to ensure that the human, rather than the machine, always remains at the centre. Indeed, it is fundamental for these public administration not to be blinded by the (efficiency) advantages that such technologies offer, and to always keep in mind that the primary goal of any public policy should be to improve the lives of their citizens.

In this regard, it is highly reassuring to see that the European Commission has proposed the adoption of an Artificial Intelligence Act (European Commission, 2021), which emphasises that any actor wishing to rely on AI techniques (including public administration using AI to support their policy- or decision-making) should question whether this entails a 'high risk' to people's rights and freedoms.[139] This risk-based approach is to be welcomed, as regulatory intervention on AI should be proportionate and should differentiate between categories of AI applications, focussing on those that generate the greatest risks for people. More concretely, the proposal first suggests to prohibit AI uses that create an unacceptable risk, notably because they would contravene EU values and/or would violate fundamental rights.[140] Second, the proposal lays down a risk methodology to define 'high-risk' AI systems that pose significant risks to the health and safety or fundamental rights of persons. The classification of an AI system as high-risk is based on the 'intended purpose' of the AI system, which implies that this classification 'does not only depend on the function performed by the AI system, but also on the specific purpose and modalities for which that system is used'.[141] These high-risk AI systems will have to undergo *ex ante* conformity assessment procedures,[142] and they will have to be registered in a public EU-wide database to increase public transparency and oversight and strengthen *ex post* supervision by competent authorities.[143] Furthermore, any AI system meeting this 'high-risk' threshold will have to comply with the legal requirements set out in the proposal, which pertain to the need: (1) to establish, implement, document and maintain a risk management system; (2) to produce technical documentation; (3) to ensure that high quality data is used; (4) to adopt an appropriate data governance allowing record-keeping – i.e. logs –, transparency and the provision of information to users; (5) to ensure human oversight; and (6) to ensure a certain level of accuracy, robustness, and cybersecurity.[144] Finally, the proposal suggests imposing transparency obligations for certain low-risk systems as well.[145]

[139] See Section 5.2.1.
[140] See Article 5 of the Proposal for an Artificial Intelligence Act.
[141] *Ibid.*, p. 13.
[142] *Ibid.*, p. 3. See also Articles 40 to 50, and Annexes VI and VII.
[143] *Ibid.*, p. 12-14. See also Articles 51 and 60, and Annex VIII.
[144] *Ibid.*, Articles 8 to 15 and Annex IV.
[145] *Ibid.*, p. 14. See Article 52 for more details.

References

Alston, P. 2019. Extreme poverty and human rights. Note by the Secretary-General of the United Nations. Available at: https://undocs.org/pdf?symbol=en/A/74/493.

Article 29 Working Party. 2018. Guidelines on Automated individual decision-making and Profiling for the purposes of Regulation 2016/679. Available at: https://ec.europa.eu/newsroom/article29/item-detail.cfm?item_id=612053.

Bibal, A., and B. Frenay. 2016. Interpretability of machine learning models and representations: an introduction. In M. Verleysen (ed.) *24th European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning: ESANN 2016*. 77-82. CIACO, Bruges, Belgium.

Buolamwini, J., and T. Gebru. 2018. Gender shades: intersectional accuracy disparities in commercial gender classification. In S.A. Friedler, and C. Wilson (eds.) *Proceedings of the 1st Conference on Fairness, Accountability and Transparency, Proceedings of Machine Learning Research*. 1-15. New York University, New York, NY, USA.

Busuioc, M. 2020. Accountable artificial intelligence: holding algorithms to account. *Public Administration Review* 81: 825-836. https://doi.org/10.1111/puar.13293

Chantillon, M., R. Kruk, A. Simonofski, T. Tombal, J. Crompvoets, C. de Terwangne, N. Habra, M. Snoeck, and B. Vanderose. 2017. FLEXPUB Public e-Service Strategy – Work package 2 – Baseline Measurement. KU Leuven Public Governance Institute, Leuven, Belgium.

Committee of Ministers of the Council of Europe. 2018. Explanatory report of the Protocol amending the Convention for the Protection of Individuals with regard to Automatic Processing of Personal Data (ETS No. 108) – Modernised Convention 108, CM(2018)2-addfinal. Available at: https://ccdcoe.org/uploads/2019/09/CoE-180518-Explanatory-Report-to-the-Protocol-amending-the-Convention-for-the-Protection-of-Individuals-with-regard-to-Automatic-Processing-of-Personal-Data.pdf.

Committee of Ministers of the Council of Europe. 2020. Recommendation CM/Rec(2020)1 of the Committee of Ministers of the Council of Europe to member States on the human rights impacts of algorithmic systems. Available at: https://search.coe.int/cm/pages/result_details.aspx?objectid=09000016809e1154.

Committee of Ministers of the Council of Europe. 2021. Declaration by the Committee of Ministers of the Council of Europe on the risks of computer-assisted or artificial-intelligence-enabled decision making in the field of the social safety net, Decl(17/03/2021)2. Available at: https://search.coe.int/cm/Pages/result_details.aspx?ObjectID=0900001680a1cb98.

Committee on Civil Liberties, Justice and Home Affairs. 2013. Report on the proposal for a regulation of the European Parliament and of the Council on the protection of individuals with regard to the processing of personal data and on the free movement of such data (General Data Protection Regulation), A7-0402/2013 – 2012/0011(COD). Available at: https://www.europarl.europa.eu/sides/getDoc.do?pubRef=-//EP//TEXT+REPORT+A7-2013-0402+0+DOC+XML+V0//EN.

Consultative Committee of the Convention for the protection of individuals with regard to automatic processing of personal data. 2021. Guidelines on Facial Recognition, T-PD(2020)03rev4. Available at: https://rm.coe.int/guidelines-on-facial-recognition/1680a134f3.

Council. 2002. Council Framework Decision 2002/584/JHA of 13 June 2002 on the European arrest warrant and the surrender procedures between Member States. Available at: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A32002F0584.

De Raedt, S. 2017. The impact of the GDPR for tax authorities. *Revue du Droit des Technologies de l'Information* 66-67: 129-143.

De Streel, A., A. Bibal, B. Frenay, and M. Lognoul, 2020. Explaining the black box – when law controls AI. CERRE Issue Paper. Available at: https://www.cerre.eu/publications/explaining-black-box-when-law-controls-ai.

De Terwangne, C. 2015. La réforme de la Convention 108 du Conseil de l'Europe pour la protection des personnes à l'égard du traitement automatisé des données à caractère personnel. In: C. Castets-Renard (ed.) *Quelle protection des données personnelles en Europe?* 81-120. Larcier, Brussels, Belgium.

Défenseurs des droits et CNIL. 2020. Algorithmes: prévenir l'automatisation des discriminations. Available at: https://www.defenseurdesdroits.fr/sites/default/files/atoms/files/synth-algos-num2-29.05.20.pdf.

Degrave, E. 2014. *L'E-Gouvernement et la protection de la vie privée. Légalité, transparence et contrôle*. Larcier, Brussels, Belgium.

Degrave, E, 2020. The use of secret algorithms to combat social fraud in belgium. *European Review of Digital Administration & Law* 1: 167-177.

Degrave, E., and A. Lachapelle. 2014. Le droit d'accès du contribuable à ses données à caractère personnel et la lutte contre la fraude fiscal. *Revue Générale du Contentieux Fiscal* 28: 322-335.

Edwards, L., and M. Veale. 2017. Slave to the algorithm? Why a 'right to an explanation' is probably not the remedy you are looking for. *Duke Law & Technology Review* 16: 77-79.

Eubanks, V. 2018. Automating inequalities: how high-tech tools profile, police, and punish the poor. *Law, Technology and Humans* 1: 162-164. https://doi.org/10.5204/lthj.v1i0.1386

European Commission. 2018. Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions. Artificial Intelligence for Europe. COM(2018) 237 final. Available at: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=COM%3A2018%3A237%3AFIN.

European Commission. 2020. White Paper on Artificial Intelligence – A European approach to excellence and trust, COM(2020) 65 final. Available at: https://ec.europa.eu/info/sites/default/files/commission-white-paper-artificial-intelligence-feb2020_en.pdf.

European Commission. 2021. Proposal for a Regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) and amending certain Union legislative acts, Brussels, COM(2021) 206 final. Available at: https://digital-strategy.ec.europa.eu/en/library/proposal-regulation-european-approach-artificial-intelligence.

European Data Protection Board. 2020. Guidelines 05/2020 on consent under Regulation 2016/679 (V.1.1). Available at: https://edpb.europa.eu/sites/edpb/files/files/file1/edpb_guidelines_202005_consent_en.pdf.

Fosch Villaronga, E., P. Kieseberg, and T. Li. 2018. Humans forget, machines remember: Artificial Intelligence and the Right to be Forgotten. *Computer Law and Security Review* 34: 304-313. https://doi.org/10.1016/j.clsr.2017.08.007

Gérard, L. 2017. Robotisation des services publics: l'intelligence artificielle peut-elle s'immiscer sans heurt dans nos administrations. In: H. Jacquemin, and A. De Streel, A. (eds.) *L'Intelligence Artificielle et le Droit*. 413-436. Larcier, Brussels, Belgium.

High-Level Expert Group on Artificial Intelligence. 2019. Ethics guidelines for trustworthy AI. Available at: https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai.

High-Level Expert Group on Artificial Intelligence. 2020. The assessment list for trustworthy artificial intelligence (ALTAI). Available at: https://ec.europa.eu/digital-single-market/en/news/assessment-list-trustworthy-artificial-intelligence-altai-self-assessment.

Information Commissioner's Office and the Alan Turing Institute. 2020. Explaining decisions made with AI (v. 1.0.11). Available at: https://ico.org.uk/for-organisations/guide-to-data-protection/key-data-protection-themes/explaining-decisions-made-with-artificial-intelligence/.

Misuraca, G., and C. van Noordt. 2020. Overview of the use and impact of AI in public services in the EU. Publications Office of the European Union, Luxembourg, Luxembourg, https://doi.org/10.2760/039619

Mittelstadt, B., C. Russell, and S. Wachter. 2019. Explaining explanations in AI. In: *Proceedings of the conference on fairness, accountability, and transparency (FAT)*. 279-288. Association for Computing Machinery, New York, NY, USA.

Nevejans, N. 2016. Study on the European Civil Law rules in robotics. Commissioned by the European Parliament's Legal Affairs Committee. Available at: https://www.europarl.europa.eu/RegData/etudes/STUD/2016/571379/IPOL_STU(2016)571379_EN.pdf.

Pasquale, F. 2015. *Black box society. The secret algorithms that control money and information*. Harvard University Press, London, UK.

Presidency of the Council of the European Union, 2015. Proposal for a Regulation of the European Parliament and of the Council on the protection of individuals with regard to the processing of personal data and on the free movement of such data (General Data Protection Regulation) – Preparation of a general approach, 9565/15. Available at: https://data.consilium.europa.eu/doc/document/ST-9565-2015-INIT/en/pdf.

Scarcella, L. 2019. Tax compliance and privacy rights in profiling and automated decision making. *Internet Policy Review* 8: 1-19.

Tombal, T., 2018. Les droits de la personne concernée dans le RGPD. In: C. De Terwangne, and K. Rosier (eds.) *Le Règlement Général sur la Protection des Données (RGPD/GDPR): analyse approfondie*. 407-557. Larcier, Brussels, Belgium.

Villani, C., M. Schoenauer, Y. Bonnet, C. Berthet, A-C. Cornut, F. Levin, and B. Rondepierre. 2018. Donner un sens à l'intelligence artificielle: Pour une stratégie nationale et européenne. Available at: https://hal.inria.fr/hal-01967551/document.

Wachter, S., B. Mittelstadt, and L. Floridi. 2017. Why a right to explanation of automated decision-making does not exist in the general data protection regulation. *International Data Privacy Law* 7: 76-99.

Zarsky, T., 2017. Incompatible: the GDPR in the age of big data. *Seton Hall Law Review* 47: 995-1020.