Artificial Intelligence and Big Data in Fraud Analytics

Tombal, Thomas; Simonofski, Anthony

Publication date:
2021

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

Artificial Intelligence and Big Data in Fraud Analytics: Identifying the Main Data Protection Challenges for Public Administrations

Thomas Tombal* *(Namur Digital Institute, UNamur, Belgium, thomas.tombal@unamur.be)* **& Anthony Simonofski** *(Namur Digital Institute, UNamur, Belgium; Faculty of Economics and Business, KU Leuven, Belgium)*

Abstract

Fraud Analytics refers to the use of Big Data Analytics to detect fraud. Numerous techniques, from data mining to social network analysis, are applied to detect various types of fraud. While Fraud Analytics offers the promise of more efficiency in fighting fraud, it also raises data protection challenges for public administrations. Indeed, whether they use traditional or advanced techniques, administrations consistently use more and more data to deliver public services. In this regard, they often need to process citizen's personal data. Therefore, administrations have to consider data protection legal requirements. While these legal requirements are well documented, the concrete way in which they have been integrated by public administrations in their Fraud Analytics process remains unexplored. Accordingly, we examine two case studies within the Belgian Federal administration (the detection of tax frauds and of social security infringements), in order to shed light on the main data protection challenges faced by public administrations in this regard.

Keywords **–** Fraud Analytics; Public Administration; Data Protection; Challenges

1 Introduction

The use of Big Data Analytics to detect tax fraud has been examined in previous research (Van Vlasselaer et al., 2017; Yu et al., 2003) and has been labelled as "*Fraud Analytics*" (Baesens et al., 2015). Fraud Analytics refers to a more global approach consisting of using analytics in fraud detection, investigation, confirmation, and ultimately prevention (Baesens et al., 2015; Pencheva et al., 2018). While Fraud Analytics offers the promise of more efficiency in fighting fraud, public administrations face additional constraints, such as the need to be trusted by the citizens and to comply with legal requirements. Indeed, whether they use traditional or advanced techniques, administrations consistently use more and more (big) data to deliver public services. In this regard, they often need to process citizen's personal data, defined by the General Data Protection Regulation (hereafter "GDPR")[1], as "*any information relating to an identified or identifiable natural person*" (Art. 4.1, GDPR). When processing personal data, organisations have to comply with the citizens' fundamental right to personal data protection[2], which derives from their right to privacy.[3]

While these legal requirements are well documented (see 2.1), the concrete way in which they have been integrated in Fraud Analytics practices of public administrations remains unexplored. This is a key issue as the introduction of analytics in organisations without appropriate organisational change can lead to ethical challenges and privacy issues (Gal et al., 2020; Mai, 2016). Therefore, in this paper, we aim to address the following research question: "**What are the main data protection challenges in the Fraud Analytics process?**". To do so, we first provide background information on some core data protection legal requirements, before identifying the research gap we aim to fill (Section 2). Then, we detail our methodology (Section 3) and we present the main data protection challenges that we have identified (Section 4), before concluding (Section 5).

2 Background

2.1 Data protection legal requirements

Legal requirements for Fraud Analytics, which fit in the broader context of the legal requirements that must be considered by public administrations when employing analytics and algorithmic processes, are well documented

[1] 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.
[2] Charter of Fundamental Rights of the European Union, OJ [2012] C 326/391, art. 8.
[3] European Convention for the Protection of Human Rights and Fundamental Freedoms, Rome, 4 November 1950, art. 8; Charter of Fundamental Rights of the European Union, OJ [2012] C 326/391, art. 7; Belgian Constitution, art. 22.

(Basin et al., 2018; Hildebrandt, 2019; Jones and Kaminski, 2020; Kaminski, 2019; van Noordt and Misuraca, 2020). Indeed, as for Big Data Analytics, the opportunities offered by Fraud Analytics must be balanced with the need to protect the citizens' right to privacy and to personal data protection (De Raedt, 2017; Scarcella, 2019).

In terms of data collection, the data must be collected fairly and transparently (Art. 5.1.a, GDPR). According to the purpose limitation principle (Art. 5.1.b, GDPR), personal data shall be collected for specified, explicit and legitimate purposes and not further processed in a manner that is incompatible with those purposes. According to the data minimisation principle (Art. 5.1.c, GDPR), only the adequate, relevant and necessary data for the fulfilment of the specific purpose of processing shall be processed. In terms of data analytics, any Fraud Analytics process must rely on a lawful basis of processing (Art. 6, GDPR). In practice, this will often be a law (Art. 6.1.c, GDPR), but this law needs to meet several requirements, such as being very specific regarding the purposes of processing it allows (Art. 6.3, GDPR).

The GDPR also provides several rights to data subjects, which should be considered when employing data analytics (Art 12-21, GDPR). We highlight here two of these rights that will be further discussed below (see 4.2.3 to 4.2.5). The first is the data subjects' right to information (Art 12-14, GDPR), which mentions that data has to be processed fairly and in a transparent manner. Therefore, the public administrations shall take appropriate measures to provide any information to the data subjects about the data analytics in a concise, transparent, intelligible and easily accessible form, using clear and plain language. The second is the data subjects' right not to be subject to a decision based solely on automated processing (Art. 22, GDPR). While there are exceptions to this right, such as fully automated processing authorised by a law (Art. 22.2, GDPR), safeguards shall be implemented, such as the right to obtain human intervention (Art. 22.3, GDPR).

2.2 Research gap

While these legal requirements are well documented, the concrete way in which they have been integrated by public administrations in their Fraud Analytics process remains unexplored. Indeed, to the best of our knowledge, no paper has taken a more concrete approach on how Fraud Analytics processes have been tailored, in practice, by public administrations to address these legal requirements, nor has identified the main data protection challenges faced by public administrations in doing so.

3 Methodology

In order to shed light on these main challenges, we examined two case studies within the Belgian Federal administration: the detection of tax frauds by the Federal Public Service (FPS) Finances and the detection of social security infringements by the Social Security Institutions (SSIs). We opted for two, rather than one, case studies as this improves the external validity of the research and allows drawing more general conclusions about the contextual factors in Belgium. Data from the cases were extracted through semi-structured interviews. Indeed, this qualitative method is effective when covering a complex topic in detail (Boyce and Neale, 2006). Moreover, this technique is relevant for our research question, as it centres around the expertise of the practitioners, and not around the validation of the knowledge of the researchers. In total, 21 interviews were performed online, from August 2020 to December 2020, with stakeholders from different management levels (strategic, mid-level, operational) and different backgrounds (legal, IT, management). The complete interview guide can be found on the Zenodo platform.[4]

4 Results

4.1 Description of the two cases

Before presenting how these legal requirements have been included by the FPS Finances and the SSIs in their Fraud Analytics processes, the general functioning of these two processes (i.e. the detection of tax frauds and the detection of social security infringements), is briefly presented.

Regarding the tax fraud detection process, data is first extracted from several sources and prepared for analysis. Then, data mining is used to signal potentially fraudulent cases that need to be further examined. These two tasks (in grey) are performed by data miners. Then, at the pre-investigation stage, the signals derived from the data mining tasks are enriched with data from other sources, and it is decided whether a proper investigation should be started. Finally, in the investigation stage, some of the potentially fraudulent cases are examined in-depth, with the support of analytics (e.g. text mining) to explore a large quantity of unstructured data. This stage is also referred to as e-auditing. These inspection tasks are performed by inspectors. Feedback is then given to data miners about the relevance of the signals. It must be noted that cases to be investigated are sometimes also suggested by "Input services" that manually detect cases to be further investigated.

For the social security infringement detection process, it is important to understand that a "Social Security Network"

[4] https://zenodo.org/record/4572708#.YD4POGhKg2w

was created by the law of 15 January 1990[5], in which all the Belgian Federal public SSIs are structured around the "Crossroad Bank for Social Security" (CBSS) (Degrave, 2020). The CBSS acts as the core of the network, and the SSIs are the nodes.[6] While these SSIs remain in control of their authoritative sources of personal social data, the CBSS acts as the central actor for the data sharing between them.[7] The CBSS thus does not itself store any data, but rather acts as a "gatekeeper" that checks that an SSI has the right to access data stored on one of the nodes of the network (another SSI).

Regarding social security fraud, there is a difference between the types of techniques used to detect fraud committed by beneficiaries of social allocations, on the one hand, and employers, health institutions, independent workers, etc., on the other hand. For the former, SSIs mainly rely on data matching techniques via bilateral cross-checks from other SSIs' databases, aimed at identifying incompatibilities in terms of allocations. These are done either before or after the payment of the allocation. For the latter, social security institutions mainly rely on data mining techniques, through the use of the OASIS data warehouse, where larger quantities of pseudonymised data are compiled. Moreover, one SSI is currently developing a Big Data Analytics Platform to improve the data governance mechanisms between SSIs, notably to tackle social fraud.

4.2 Main data protection challenges

4.2.1 Ensuring reactivity to frauds while respecting purpose limitation

For tax fraud, Article 3 of the Law of 3 August 2012[8] states that the FPS Finances can collect and process personal data to execute its legal missions, and that the data cannot be used for other purposes.[9] Regarding, more specifically, the use of Big Data to fight tax fraud, Article 5.1, which was modified in September 2018[10], provides that the FPS Finances may aggregate data, collected to execute its legal missions, in a "data warehouse" enabling "data mining" and "data matching" operations, including profiling. This can only be done to carry out, in the context of its legal missions, targeted controls on the basis of "risk indicators" and of analyses on data coming from different administrations and/or services of the FPS Finances. Although this Article constitutes the lawful basis for such processing (Art. 6.1.c, GDPR), such law must clearly determine the specific purposes of processing that are allowed (Art. 6.3, GDPR). Yet, the critique formulated by (Degrave and Lachapelle, 2014) regarding the previous version of Article 5, namely that the purposes of data processing were defined too broadly in the Law, as they simply referred to the execution of the FPS Finances' "legal missions", have not been addressed in the 2018 modification, as the same terminology is used. This might thus be problematic in terms of the validity of this Law as lawful basis for the processing, as well as in terms of compliance with the purpose limitation principle (Art. 5.1.b, GDPR).

However, this concern is somewhat alleviated as the data miners have to fill in a DAM (Data Access Management) fiche, which has to be validated by the President of the Executive Committee of the SPF Finances.[11] In this DAM fiche, they have to state the objectives and purposes of the data mining and explain how it fits the organisation's mission. The purpose limitation principle is thus implemented at the process level, but in a way that is not ideal from a democratic perspective (as Parliament does not define the concrete purposes of processing) nor from a legal perspective (as according to Article 8 of the European Convention on Human Rights[12], Article 22 of the Belgian Constitution and Art. 6.3 GDPR, the key elements of personal data processing by public administrations, such as the processing purposes, must be clearly defined by law).

Regarding the social security fraud case study, the use of data matching techniques relying on bilateral cross-checks, aimed at identifying incompatibilities in terms of allocations, must be subject to a data transfer protocol (DTP), as provided in Article 20.1 of the Law of 30 July 2018[13], unless provided otherwise in specific laws (e.g., in Article 15 of the Law of 15 January 1990, as modified in September 2018[14], which requires, in some cases, a prior deliberation of the Information Security Committee (ISC)). The protocol, which must notably contain the purposes of

[5] Loi du 15 janvier 1990 relative à l'institution et à l'organisation d'une Banque-carrefour de la sécurité sociale, *M.B.*, 22 février 1990.
[6] https://www.ksz-bcss.fgov.be
[7] Art. 3 of the Law of 15 January 1990.
[8] Loi du 3 août 2012 portant dispositions relatives aux traitements de données à caractère personnel réalisés par le Service public fédéral Finances dans le cadre de ses missions, *M.B.*, 24 août 2012.
[9] Art. 3, al.1 and 2 of the Law of 3 August 2012.
[10] Modified by art. 71 of the Law of 5 September 2018 (Loi du 5 septembre 2018 instituant le comité de sécurité de l'information et modifiant diverses lois concernant la mise en oeuvre du Règlement (UE) 2016/679 du Parlement européen et du Conseil du 27 avril 2016 relatif à la protection des personnes physiques à l'égard du traitement des données à caractère personnel et à la libre circulation de ces données, et abrogeant la directive 95/46/CE, *M.B.*, 10 septembre 2018).
[11] Art. 4, al.1 of the Law of 3 August 2012, as modified by Art. 70 of the Law of 5 September 2018.
[12] European Convention for the Protection of Human Rights and Fundamental Freedoms, Rome, 4 November 1950.
[13] Loi du 30 juillet 2018 relative à la protection des personnes physiques à l'égard des traitements de données à caractère personnel, *M.B.*, 5 septembre 2018.
[14] Modified by art. 18 of the Law of 5 September 2018.

processing, must be submitted to the Data Protection Officers of the SSIs involved in the sharing.[15] However, they are not subject to a prior validation by the Data Protection Authority, which would bring more certainty in terms of the legitimacy of the purpose of processing. Once this purpose is achieved, the data must be deleted.

SSIs also use data mining techniques. According to Article 5*bis* of the Law of 15 January 1990, which has been inserted in September 2018[16], they may aggregate and process data in a data warehouse, enabling them to carry out data mining operations to prevent, establish, prosecute, and punish offenses against social legislations which fall within their respective powers. This data warehouse is known as OASIS and has existed since 2005. According to (Degrave, 2014, 2020) the purposes of processing in OASIS that are authorised by the law are not clearly defined, which could, here as well, be problematic in terms of the validity of this Law as lawful basis for the processing, as well as in terms of compliance with the purpose limitation principle. However, this concern is somewhat alleviated, although not optimally either from a democratic and legal perspective (see above), in the hypotheses contained in Articles 5*bis*, al.7 and 15 of the Law of 15 January 1990, as the authorisation to process data from the data warehouse must be subject to a prior deliberation by ISC, which will evaluate the purposes of processing. The purpose limitation principle is thus also implemented at the process level in the social security case study, as the purposes of the data matching or data mining operations have to be defined in advance, either in a protocol or in the file to be submitted to the ISC.

The mechanisms mentioned above can be in conflict with the need for reactivity in Fraud Analytics. Indeed, in some cases, such as customs tax fraud detection, administrations have to react very quickly and getting the authorisations is time-consuming. Furthermore, it can be challenging to precisely define the exact type of fraud that they are investigating in advance, as this is sometimes broadly defined at the start and needs to be further refined with time.

4.2.2 Balancing data minimisation with timely access to relevant data sources

For tax fraud, Article 5 of the Law of 3 August 2012 provides that the FPS Finance can use "data collected to execute its legal missions". These are notably data collected from people's and undertakings' tax declarations, from the newspapers, from their own experience, from whistle-blowers and from outputs of investigations. Once again, the critique formulated by (Degrave and Lachapelle, 2014) regarding the previous version of Article 5, namely that the types of data that could be used were defined too broadly, as it provided that the FPS Finance can use, via the data warehouse, any "data collected in order to execute its legal missions", have not been addressed in the 2018 modification either, as the same terminology has been kept. This might be problematic from a data minimisation perspective. However, this concern is somewhat alleviated by the fact that, as outlined above, a DAM fiche must be completed and submitted to the President of the Executive Committee. This constrains the data that data miners can access for a specific project. This is a pragmatic solution, as it would be very difficult for the legislator to pre-define all the types of data that could be processed in this regard. Moreover, the technical access to the data warehouse is built in such a way that the agents of the FPS Finances can only access the electronic records, data or applications that are adequate, relevant and non-excessive in light of the execution of the tasks that fall within their legal missions[17], and this can be checked through access logs.

Regarding the data mining operations conducted in the OASIS database, Art 5*bis* of the Law of 15 January 1990, inserted in 2018, provides that "all the necessary data for the purposes of applying the labour law and social security legislation" can be used. This definition may be too broad as it does not allow the citizens to know exactly which types of data are (or can be) processed. However, this concern is somewhat alleviated by the fact that access to data from the data warehouse must be subject to a DTP or to a prior deliberation of the ISC, in which the necessary and proportionate nature of the accessed data will be controlled (see 4.2.1). The same goes for data matching operations. Moreover, the data minimisation principle is enshrined in the fact that the data warehouse solely contains pseudonymised data and that it can only be accessed by a limited number of data miners/investigators. Importantly, the people who pseudonymise the data to be uploaded in the data warehouse and suggest fraud indicators are not the same as those who use the datawarehouse in order to spot fraudulent patterns based on those indicators.

The key is thus to be proportionate in the types of data collected and used. Even if administrations could potentially have access to troves of data, a balance must be found with the citizens' data protection. This creates internal discussions about how much data they capture and how much data they may ask for in a timely manner.

[15] Art. 20.2 of the Law of 30 July 2018.
[16] Inserted by art. 12 of the Law of 5 September 2018.
[17] Art. 10.1 of the Law of 3 August 2012.

4.2.3 Facilitating the access to information about Fraud Analytics for citizens

As a rule of thumb, any Fraud Analytics processing must be fair and transparent, and the data subjects must be informed about it. Fairness implies that the laws on which this processing are based must be sufficiently explicit and understandable for the data subjects. They cannot be taken by surprise. For both case studies, citizens are generally informed about the existence of data matching and data mining operations through the laws mentioned above. Yet, according to (De Raedt, 2017; Degrave and Lachapelle, 2014), these laws do not provide sufficiently clear information to the citizens, notably in terms of the concrete processing that will be conducted and in terms of the types of data that will be used (see 4.2.1 and 4.2.2).

To some extent, this lack of transparency is reduced by the fact that these concrete data processing will be subject to a DAM fiche, to a prior deliberation of the ISC or to the conclusion of a DTP, which will provide more specific information. However, citizens do not have access to the DAM fiches. Moreover, while the deliberations of the ISC are published on the website of the CBSS[18], it is hard to obtain information about a specific processing, as the search tool is quite basic. In a similar vein, as the DTPs have to be published on the websites of the relevant data controllers[19], this leads to a diluted publication on a wide variety of websites, whose quality can strongly vary. This makes it almost impossible for citizens to have a good overview of the types of processing conducted with their data, and thus constitutes a major challenge to address.

4.2.4 Ensuring a truly critical human check of quasi-automated decisions

For both case studies, it should be outlined that, as also underlined by (De Raedt, 2017; Scarcella, 2019), even if the machine does not itself decide that a person is a fraudster, the decision to identify a person as "suspicious" could, in and of itself, be qualified as a solely automated decision producing legal effects for this person (i.e. the opening of an investigation). If this interpretation is followed, this would require implementing appropriate safeguards, such as the right to obtain a human intervention (Art. 22.3, GDPR). Moreover, questions could be raised about whether the human intervention remains sufficient, especially if the controllers do not question the fraud inspection suggestions they receive, as they completely rely on the machines to determine the cases to be investigated. For instance, in the specific field of customs frauds, 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 applying what the AI suggests, without any critical thinking. For instance, in the customs frauds example, feedback will be provided by the controllers, which implies that a human will assess the recommendations made by the machine, putting back human control in the process. Yet, looking towards the future, it is possible that, in light of the constant budget cuts and reductions of 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 control quotas, and no longer have time to check the relevance of the indicators suggested by the machine. Such a scenario must be avoided.

4.2.5 Balancing explainability with the need to ensure the confidentiality of Fraud Analytics process

For tax fraud, data miners are able to explain the reasoning behind the detection (indicators, techniques applied, etc.). For social security infringements, automated bilateral *ex-ante* cross-checks relying on data matching 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* cross-checks relying on data matching, 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).

However, it should be outlined that, for both case studies, a person or an undertaking will not be informed that it has been flagged as a potential fraudster following data mining operations if the follow-up investigation did not result in the finding of fraud. Moreover, they will not receive an explanation about why this is the case. This shows that even if administrations can explain their decision, the challenge is to determine when and how they should do it. Indeed, it is complex to find a balance between being fully transparent about the data mining processes and models

[18] https://www.ksz-bcss.fgov.be/fr/deliberations-csi-list?term_node_tid_depth=51

[19] Art. 20.3 of the Law of 30 July 2018.

used, and the need not to disclose their Fraud Analytics processes, as otherwise the fraudsters might adapt and avoid being detected.

5 Conclusion

Our research has enabled us to identify the main data protection challenges faced by public administrations in the Fraud Analytics process. As a final take-away, we suggest ways forward to address some of them. For the first challenge (4.2.1), we argue that to ensure reactivity to frauds while respecting purpose limitation, the data processing authorisation request could be slightly broader at first, and then refined continuously throughout the process, via close collaboration between legal services and data miners following agile analytics principles. Another solution direction would be to anonymise, or at least pseudonymise, the data warehouse data on which the data mining analysis is done, and to only allow the re-identification of the data subjects in the context of a concrete human-led investigation. This would ensure privacy-by-design and by-default (Art. 25, GDPR) and can prevent data processing mistakes. For the third challenge (4.2.3), we believe that in order to facilitate access to information, a solution would be to centralise the publication of all of the DTPs in a single source, such as the Data Protection Authority's website. A good example of this is the city of Amsterdam's "Algorithm register".[20] Moreover, it should be possible to search through this single source, as well as through the Information Security Committee's deliberations on the CBSS website, on the basis of several criteria, such as the types of purposes or of data concerned.

Acknowledgements

This research is supported by the BRAIN-be 2.0 (Belgian Research Action through Interdisciplinary Networks) research grant n° B2/191/P3/DIGI4FED. The authors would like to thank the sponsoring agency, the Belgian Science Policy Office, for their support.

References

Baesens B, Vlasselaer V Van and Verbeke W (2015) *Fraud Analytics Using Descriptive, Predictive, and Social Network Techniques*. DOI: 10.1002/9781119146841.

Basin D, Debois S and Hildebrandt T (2018) On Purpose and by Necessity: Compliance Under the GDPR. In: *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2018. DOI: 10.1007/978-3-662-58387-6_2.

Boyce C and Neale P (2006) Conducting in-depth interviews: A Guide for designing and conducting in-depth interviews. *Evaluation* 2(May): 1–16. DOI: 10.1080/14616730210154225.

De Raedt S (2017) The impact of the GDPR for tax authorities. *R.D.T.I*: 66–67.

Degrave E (2014) *L'E-Gouvernement et la protection de la vie privée. Légalité, transparence et contrôle*,. Bruxelles.

Degrave E (2020) The Use of Secret Algorithms to Combat Social Fraud in Belgium. *European Review of Digital Administration & Law* 1(1–2): 167–177.

Degrave E and Lachapelle A (2014) Le droit d'accès du contribuable à ses données à caractère personnel et la lutte contre la fraude fiscale. *note sous C.C., 27 mars 2014, n°2014/28, R.G.F.C., 2014/5*: 322–335.

Gal U, Jensen TB and Stein MK (2020) Breaking the vicious cycle of algorithmic management: A virtue ethics approach to people analytics. *Information and Organization* 30(2). DOI: 10.1016/j.infoandorg.2020.100301.

Hildebrandt M (2019) Privacy as protection of the incomputable self: From agnostic to agonistic machine learning. *Theoretical Inquiries in Law* 20(1): 83–121. DOI: 10.1515/til-2019-0004.

Jones M and Kaminski ME (2020) An American's Guide to the GDPR. *Denver Law Review, Vol. 98, No. 1, p. 93, 2021*.

Kaminski ME (2019) Binary governance: Lessons from the GDPR'S approach to algorithmic accountability. *Southern California Law Review*. DOI: 10.2139/ssrn.3351404.

Mai JE (2016) Big data privacy: The datafication of personal information. *Information Society* 32(3): 192–199. DOI: 10.1080/01972243.2016.1153010.

Pencheva I, Esteve M and Mikhaylov SJ (2018) Big Data and AI – A transformational shift for government: So, what next for research? *Public Policy and Administration*. DOI: 10.1177/0952076718780537.

Scarcella L (2019) Tax compliance and privacy rights in profiling and automated decision making. *Internet Policy Review* 8(4): 1–19. DOI: 10.14763/2019.4.1422.

van Noordt C and Misuraca G (2020) Exploratory Insights on Artificial Intelligence for Government in Europe. *Social Science Computer Review*. DOI: 10.1177/0894439320980449.

Van Vlasselaer V, Eliassi-Rad T, Akoglu L, et al. (2017) GOTCHA! Network-based fraud detection for social security fraud. *Management Science*. DOI: 10.1287/mnsc.2016.2489.

Yu F, Qin Z and Jia XL (2003) Data mining application issues in fraudulent tax declaration detection. In: *International Conference on Machine Learning and Cybernetics*, 2003. DOI: 10.1109/icmlc.2003.1259872.

[20] https://algoritmeregister.amsterdam.nl/en/ai-register/