Maastricht University

Faculty of Law

Artificial Intelligence and the Black Box Problem

Can Artificial Intelligence used by Tax Administrations be compliant with EU law?

Student Name: David João Nobre Rito
Student Number: i6211435
Program: LL.M. International and European Tax Law with specialisation in Tax and Technology, Maastricht University
Supervisor: Dr. Marcel Schaper
Date of submission: 20th November 2020
Number of words: 10791 words

Table of contents

1. Introduction

The era of big data brought new possibilities for the decision-maker: allied to his experience, they can make better predictions and smarter decisions.[1] More than trusting their gut, decision-makers had another source of information which is more reliable and leads to better results.[2] In the business context, this is especially true for big companies as they deal with higher amounts of data than the average company.[3]

The proliferation of big data also came to improve the decision-making process in the legal domain. Here too, the use of big data means to have in hand more information than the one observed.[4] In order to process it, Artificial Intelligence (AI) needs to be used. Examples of AI in the legal domain vary from algorithms for predicting recidivism (COMPAS)[5], prediction of legal outcomes[6] or document classification[7]. These and other developments of deployment of AI in law enhanced efficiency both in time spent and in incurred costs by automating routine tasks or supporting decision-making. Thus, the law professional could focus his attention on the core tasks that bring value to his client while the client would have a better service as he would receive advice based on more data than a human could observe alone. Though the general view is that AI cannot provide the same level of analysis and skill as a law professional, and so not possible to automate complex tasks, research argues that as long the results produced by automated processes are at least similar to those performed by humans, AI can replicate complex tasks.[8]

Apart from AI application in businesses, AI is also being used by the public sector, namely by Tax Administrations.[9] Among its advantages, AI allows Tax Administrations to analyse several periods and taxpayers at the same time so they can plan control and risk management and improve service delivery. Moreover, processing big data in Tax Administrations helps to support

---

[1] McAfee, A., & Brynjolfsson, E. (2012). Big data: The management revolution. Harvard Business Review, 90(10), 4, page 4.
[2] Brynjolfsson, E., Hitt, L. M., & Kim, H. H. (2011). Strength in numbers: how does data-driven decisionmaking affect firm performance? Ssrn Electronic Journal, (2011), page 31. https://doi.org/10.2139/ssrn.1819486
[3] Davenport, T. H., & Dyché, J. (2013). Big Data in Big Companies, page 7.
[4] Katz, D. M. (2012). Quantitative Legal Prediction - Or - How I Learned to Stop Worrying and Start Preparing for the Data-Driven Future of the Legal Services Industry. Emory Law Journal, 62, page 928. Retrieved from https://heinonline.org/HOL/Page?handle=hein.journals/emlj62&id=923&div=&collection=
[5] Dressel, J., & Farid, H. (2018). The accuracy, fairness, and limits of predicting recidivism. Science Advances, 4(1), 5580, page 1. https://doi.org/10.1126/sciadv.aao5580
[6] Surden, H. (2014). Machine learning and law. Washington Law Review, 89(1), 87–115, page 102.
[7] Surden, H. (2014). Machine learning and law. Washington Law Review, 89(1), 87–115, page 110.
[8] Surden, H. (2012). Computable Contracts. U.C. Davis Law Review, 46. Retrieved from https://heinonline.org/HOL/Page?handle=hein.journals/davlr46&id=643&div=&collection=
[9] Technologies for Better Tax Administration. (2016). In Technologies for Better Tax Administration, page 36-44. https://doi.org/10.1787/9789264256439-en

decision making of governments by sharing insights and information. Lastly, Tax Administrations can predict taxpayer behaviour which saves time, money and effort during risk profiling.[10]

However, AI also brought challenges that can hinder its full potential. One of which is perpetuation and prejudice reinforcement. When training the model, the selected historical data can be biased due to the circumstances in society when data was retrieved. An example of this challenge is in AI for predictive policing. Defined as "any policing strategy or tactic that develops and uses information and advanced analysis to inform forward-thinking crime prevention"[11], its goal is to improve effectiveness and efficiency for law enforcement decision-making[12] by predicting areas of future crime locations from past crime statistics and other data.[13] A data science tool as PredPol is an example of predictive policing. Used in the United States policing jurisdictions as Los Angeles, Atlanta and Santa Cruz, questions were raised about the accuracy of its prediction, namely about the potential to increase discrimination due to class based profiling.[14] Basing the action of police resources on historical data can become a self-fulfilling prophecy: in typically economically disadvantaged areas more resources will be deployed which results in higher levels of reported crime in these areas.[15]

Another common issue of AI, namely with machine-learning, is the possibility of overgeneralization. Whereas the goal of using models is to digest new data and make accurate predictions after being trained on a training set, overgeneralization happens when parameters of the model are tuned for very high accuracy on your training data set, but do poorly on the unseen examples. In the legal domain, overgeneralization would be relying on a model which is too biased by its training set data (e.g. as individual law firm's past cases) making the rules inferred not useful for predictive purposes.[16] An example of overgeneralization in tax compliance can be Tax Administration training models to finding fraudsters and the training data only contains information from taxpayers who are nationals of Tax Administration's jurisdiction. If the model finds a pattern between nationality and fraud behaviour, when applying the model in real life, all fraudsters who do not hold the nationality will not be signalled by the algorithm restricting its added-value.

---

[10] Technologies for Better Tax Administration. (2016). In Technologies for Better Tax Administration, page 53-54. https://doi.org/10.1787/9789264256439-en

[11] Uchida, C. D. (2009). National Discussion on Predictive Policing: Defining Our Terms and Mapping Successful Implementation Strategies. National Criminal Justice Reference Service. Retrieved from https://www.ncjrs.gov/pdffiles1/nij/grants/230404.pdf%5Cnpapers3://publication/uuid/24AA7AF5-FEDF-496C-AB25-19C28C9BAC4F

[12] Moses, L. B., & Chan, J. (2014). Using big data for legal and law enforcement decisions: testing the new tools. University of New South Wales Law Journal, The, 37(2), 643–678, page 658.

[13] A. G. (2011). Predictive policing and the future of reasonable suspicion. Ssrn Electronic Journal, (2011), page 265. https://doi.org/10.2139/ssrn.1965226

[14] Baldrige, J. (2015). Machine Learning And Human Bias: An Uneasy Pair | TechCrunch. Retrieved June 26, 2020, from https://techcrunch.com/2015/08/02/machine-learning-and-human-bias-an-uneasy-pair/

[15] Kelleher, J. D., & Tierney, B. (2018). Data science (Ser. The mit press essential knowledge series). MIT Press, page 192.

[16] Surden, H. (2014). Machine learning and law. Washington Law Review, 89(1), 87–115, page 106.

Data can only be useful when future cases where decisions are made have pertinent features. Thus, using inductive reasoning to generalize about the future can lead to a new variation or rare cases that the model will not be able to detect.[17] An example that the Public Administration faces is organized crime which as an activity can be very dynamic.[18] Thus, models trained with past data to detect this kind of behaviour will always be limited to the data available and will not be 100% accurate, making the mission of Public Administration to be considered utopian. The best that can be achieved is to reduce the criminal activity in business and catch the businesses which use the strategy (followers) once created by others (leaders).

Apart from above mentioned challenges, one that can encounter more resistance to be accepted and solved is the opacity inherent to some AI models. An opaque model can be described as "relative to a cognitive agent X at time *t* just in case X does not know at *t* all of the epistemically relevant elements of the [system]".[19] One of the questions opacity brings to Public Administration is to understand whether models process data which complies with the legal requirements, namely under Data Protection Regulation (GDPR).

Algorithms often operate within what are known as "black boxes", in which it is known what happens at the end but not what happens inside the intermediate layers, meaning that people and even their creators or developers cannot understand how the algorithms are making decisions.

In this article, focus will be on the opacity as one of the challenges that are found within the AI domain: the consequences that derive from decisions made by AI and that have no explanation.

For humans to work effectively in cooperation with AI-powered robots and systems, trust is essential. But how can trust exist if AI makes a decision that humans do not understand? Other questions arise when black box models come into the picture: who is accountable if things go wrong? Can it be explained why things go wrong? If things are working well, is it possible to know why and how to leverage them further?[20]

Trust will be the heart of the discussion when looking to the usefulness of the models and AI tools by the Tax Administration. Being the models trained with past data and today's society is changing faster than ever[21], how likely is it to use AI to solve today's problems and prevent them in the future? This brings legal concerns attached as the principle of non-discrimination. Data

---

[17] Surden, H. (2014). Machine learning and law. Washington Law Review, 89(1), 87–115, page 105.
[18] Savona, E. U., Riccardi, M., & Berlusconi, G. (Eds.). (2016). Organised crime in European businesses, page 211. Routledge.
[19] Zednik, C. (2019). Solving the black box problem: a normative framework for explainable artificial intelligence. Philosophy and Technology, (2019), page 4. https://doi.org/10.1007/s13347-019-00382-7
[20] Tjoa, E., & Guan, C. (2019). A Survey on Explainable Artificial Intelligence (XAI): Towards Medical XAI, page 1. Retrieved from http://arxiv.org/abs/1907.07374
[21] Bentley, D. (2019). Timeless principles of taxpayer protection: How they adapt to digital disruption. EJournal of Tax Research, 16(3), 679–713, page 288-289.

that is gathered will always be biased to a certain extent. How to make sure data contains only the necessary biases in order to be reliable? Moreover, how likely it is to understand the model so the output can be analysed concerning its reasoning, i.e. output is not bias-driven?

Beyond discrimination, AI stakeholders need also to take into consideration the principle of proportionality. AI and the data used should not go beyond what is necessary to attain the goal for which they are being deployed. An example of the discussion of proportionality between the power of Tax Administration and taxpayer is the decision of the French Constitutional Council with regard to the article 154 of the 2020 “Budget Bill”. Though the Council did not find reason enough to decline the use of the new mechanism proposed regarding the balance between the fundamental rights of the taxpayer and the public interest, a set of accountability rules was delineated to keep the proportionality in place.[22] However, it is arguable to what extent can these conditions be met and how can AI be monitored by the end-users or regulators.[23]

The blurriness behind the requirements to attain the principle of proportionality can lead other fundamental rights to follow the way as the equality of arms. Without clear and practical guidelines to measure proportionality, asymmetry of information will be a natural consequence due to the fact that Tax Administration will be able to gather as much information from the taxpayer without the latter knowing when and how that was processed. Even worse, using the argument of public interest, Tax Administration might disclose neither the data nor the algorithm turning the possibility of the taxpayer to defend himself a mirage as he does not even have access to what led the Tax Administration to highlight his profile. The same situation can happen if the Tax Administration is not capable of explaining what the algorithm did to retrieve a certain output.

The latter case also happens in human behaviour in which many aspects are impossible to explain in detail. Thus, it is fair to say the public may have to be comfortable with the fact that they will not be able to explain all of AI's behaviours.

If it is inevitable to trust in AI's judgment without the ability to analyse the reason behind each decision, it must at least be able to seed AI with values that fit the social norms and rules. But who is responsible if the AI does not meet those standards? Should an AI be responsible for its own decision? Should AI be treated as a human being on a legal level? What if there is bias in the algorithms?

Dependability on machine learning and algorithms has made the problem of biases increasingly important for the developers and users community related to AI initiatives. Part of the problem is

[22] J.M. Calderón Carrero & J.S. Ribeiro, Fighting Tax Fraud through Artificial Intelligence Tools: Will the Fundamental Rights of Taxpayers Survive the Digital Transformation of Tax Administrations?, 60 Eur. Taxn. 6 (2020), Journal Articles & Papers IBFD (accessed 7 July 2020), page 235-236.
[23] J.M. Calderón Carrero & J.S. Ribeiro, Fighting Tax Fraud through Artificial Intelligence Tools: Will the Fundamental Rights of Taxpayers Survive the Digital Transformation of Tax Administrations?, 60 Eur. Taxn. 6 (2020), Journal Articles & Papers IBFD (accessed 7 July 2020), page 236.

that the databases used to train algorithms are not always diverse enough. Those biases can have a tremendous impact if they affect decisions in fields such as health insurance, schooling, or criminal record, as government decisions become more automated. And that is why AI Now - a research institute studying the social implications of AI - argued and warned in a 2017 report that it is vital that government agencies stop using black box algorithms.[24]

Legislation taking into account the realities derived from AI decisions made without human intervention already already exists. In the case of Europe, it is being argued that it is a fundamental legal right to question an AI system about how it arrived at its conclusions. The GDPR, which entered in force in 2018 states for example that an individual may require an explanation of how automated system decisions are made.[25]

But how can that law be enforced when computers have been programmed in a way that it cannot be understood? Even the engineers who built those AI-powered systems cannot fully explain their behaviour. Should the public expect AI to be explained in a way that humans can understand?

To open that black box and build trust, the machine learning models on which these systems are based, should have at least three important characteristics: be explainable - by means of that the reasoning behind each decision can be understood-, be transparent - by means that the model on which the decision-making is based can be fully understood-, and be demonstrable -by means that there is mathematical certainty behind the predictions. The combination of those three characteristics is often referred to as the 'interpretability' of AI models.

The development of technology is much faster than the pace of policymakers to regulate innovation towards protection of the stakeholders involved. As an example, think in the case of technology developments and the interaction between the Tax Administration and the taxpayers (natural or legal persons). To process such a big amount of data, AI needs to use more complex structures which even their developers face difficulties to explain. As mentioned before, AI algorithms that are difficult to explain on how the system has arrived at a decision are called black box models. These models raise questions such as who is accountable by the outcomes of the algorithm, what information is relevant to the outcome, and can an algorithm be trusted to help decision making when the users do not understand how it works, among others. Furthermore, of the foremost importance in the legal domain is, how can a person challenge a decision based on a black box model?

These concerns are crucial to clean the path of development of AI and give technology room to improve our lives as done until today. In the EU, rights of citizens and taxpayers are protected either by Primary EU law or Secondary EU law. The issue here is not the lack of protection

[24] Campolo, A., Sanfilippo, M., Whittaker, M., & Crawford, K. (2017). AI Now 2017 Report. In AI Now Institute, page 1. Retrieved from https://ainowinstitute.org/AI_Now_2017_Report.pdf
[25] See European Parliament and Council of the European Union: The General Data Protection Regulation (GDPR).

given by EU law. It is indeed the challenge to analyse whether technology can fit the expectations of the data controllers (e.g. Tax Administrations) while it respects the legal framework that protects the data-subjects. This paper will focus on GDPR as this is the most comprehensive data protection legislation ever enacted.[26]

This paper aims to contribute to this discussion in two different angles. First, promote use of AI in the tax compliance field to attain the goals of Tax Administrations, both collection of tax revenue and fight tax fraud. At the same time, the author's purpose is to show how the rights of taxpayers can be protected even when Tax Administrations use black box models.

2. Research question

This paper commits to study the use of black box models under EU Law. In that way, the research question this paper will address is: *Can black box models used by Tax Administrations be compliant with EU Law?*

To set up the scene about this topic, explanations need to be given on how a model can be considered as a black box model. Thus, the paper will answer the sub-question: *What are the characteristics of a black box model?*

To focus on the legal concerns, there is a need to address what characteristics of the black box models are relevant for this matter. As this paper will base the legal analysis on GDPR, such description will be defined through the sub-question: *What are the relevant characteristics of the black box model for GDPR purposes?*

In order to explain if it is possible black box models to meet the legal requirements, first is needed to clarify the legal framework through the following sub-question: *What EU legal framework black box models are facing?*

Finally, a look into other fields where black box models are present and how they are being addressed will be given to provide insights on how such models can be used by Tax Administration while being compliant with the legal framework. Thus, the following sub-question will be answered: *What approaches in other fields are being taken to address black box models?*

[26] Bal, A. (2019). Ruled by algorithms: the use of `black box' models in tax law. Tax Notes International, 95(12), 1159–1166, page 1162.

3. Methodology

From a methodological perspective, this paper will be composed of two major analyses: A technological one related to AI, and a legal one related on how AI needs to comply with the legal framework in the EU and the impact and consequences that it has for the purposes of the research question.

From a technology perspective this paper will briefly discuss the main AI models that currently exist, from the less to the more complex ones. After, the problems that arise because of the accuracy vs complexity trade-off, and the impact that such a trade-off creates on the explainability of the AI will be described taking into account what opacity means for the definition of a black box model. Furthermore, the dichotomy of explainability and transparency will be discussed regarding which characteristic is more relevant when deploying AI and namely black box models. Other concepts mentioned in the discussion of black box models as interpretability, understandability or comprehensibility will be treated as interrelated to the concept of explainability[27] and in doing so, no need will arise to mention these concepts in this discussion or in the paper. The output of the discussion will be fundamental to introduce a normative framework suggested by Zednik[28] into the tax domain. An overview of the framework will be done so it can be clarified how Zednik's framework can be implemented in the tax domain. Additionally, it will help to define which approach can be suitable to use when facing black box models, i.e. black box approach. Upon the selected approach, two systems used for targeting detection purposes will be presented: AdFisher and Sunlight. Their purpose and a general overview concerning how they work will be explained to understand the feasibility in theory of its deployment in the tax domain namely by Tax Administrations' use of AI. Though a general overview of black box model concept will be provided, a focus on risk profiling algorithms will be done due to the similarity between the task performed by them and models designed for targeting purposes which the systems previously mentioned argue they can explain. Moreover, the lack of sufficient available data and academic research about other types of black box models used by the Tax Administration would hinder the robustness of the conclusions this paper aims to present if such sources would be included.

Concerning the legal analysis, a descriptive method will be used. First, a review of key provisions of the GDPR will be provided according to the relevant characteristics of automated decision-making as defined in article 22.[29] Though there is research which argues there is no

---

[27] Guidotti, R., Monreale, A., Ruggieri, S., Turini, F., Giannotti, F., & Pedreschi, D. (2018). A survey of methods for explaining black box models. ACM Computing Surveys, 51(5), page 5. https://doi.org/10.1145/3236009
[28] Zednik, C. (2019). Solving the black box problem: a normative framework for explainable artificial intelligence. Philosophy and Technology, (2019). https://doi.org/10.1007/s13347-019-00382-7
[29] WP29 (Article 29 Data Protection Working Party). (2018), page 34. Guidelines on Automated individual decision-making and Profiling for the purposes of Regulation 2016/679

right to explanation in GDPR[30] or such right is the most suitable when explanations from algorithms are needed[31] the author agrees that it is important to mention such right at least as to provide safeguards regarding the behaviour of the algorithm and the possible legal effects that people may suffer from it.[32] Guidelines from the Working Party 29 as well as recitals will be used when no meaning can be found in the text of the Regulation. To better understand how black box AI can face problems with GDPR, a case study regarding Poland's implementation of System Teleinformatyczny Izby Rozliczeniowej (STIR) will be included as well as an analysis of the Taxation Modernization Act implemented by Germany by taking into account its main characteristics that can be relevant under GDPR.

4. Body

4.1. Technology analysis

4.1.1. AI models and machine learning

As shown in Figure 1, among the AI models that currently exist, it is possible to find rule-based models, distance-based models, tree-based models, probabilistic models, linear models, ensemble models and neural network models, among others.

In order to properly understand those models, it is important to first mention the basic characteristics and the application areas of all these algorithms, and in order to fully understand the underlying concepts, it is important to provide a brief description of the concepts related to machine learning.

[30] Wachter, S., Mittelstadt, B., & Floridi, L. (2017). Why a right to explanation of automated decision-making does not exist in the general data protection regulation. International Data Privacy Law, 7(2), 76–99. https://doi.org/10.1093/idpl/ipx005

[31] Edwards, L., & Veale, M. (2017). Slave to the Algorithm? Why a Right to Explanation is Probably Not the Remedy You are Looking for. SSRN Electronic Journal, 16, page 42-43. https://doi.org/10.2139/ssrn.2972855

[32] Roig, A. (2018). Safeguards for the right not to be subject to a decision based solely on automated processing (Article 22 GDPR). European Journal of Law and Technology, 8(3), page 3.

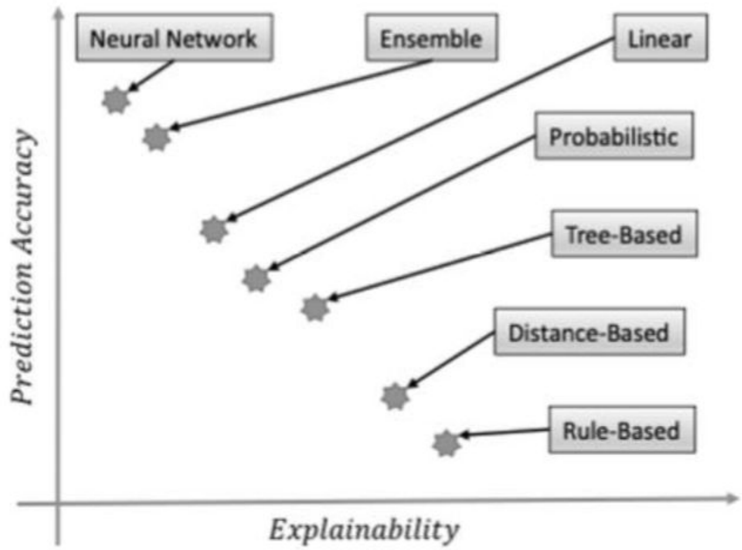

Figure 1: Accuracy and explainability trade-off as described by Hacker et al. (2020. P. 17)

Machine Learning models aim to learn some function (F) that provides an accurate correlation between input values (x) and output values (y). Y = F(x). It is the practice of using algorithms to parse data, learn from it, and then make a determination or prediction about something.

Many of the models and algorithmic systems programmed using machine learning are opaque. This means that it is difficult to know how they work and why they do what they do.

4.1.2. Opacity in machine learning

Opacity is the heart of the black box problem in AI, which is a problem with practical, legal, and theoretical consequences.[33] The black box model problem comes into play when the models and algorithmic systems being developed in AI are opaque.

Opacity can be seen from 3 different perspectives: 1) as intentional corporate or state secrecy 2) as technical illiteracy and 3) the way algorithms operate at the scale of application.[34]

Perceiving opacity as due to technical illiteracy can be understood as agent relative. A computing system is never opaque in and of itself but such opacity will depend on the agent

[33] Zednik, C. (2019). Solving the black box problem: a normative framework for explainable artificial intelligence. Philosophy and Technology, (2019), page 1.
[34] Burrell, J. (2016). How the machine ‘thinks’: Understanding opacity in machine learning algorithms. Big Data and Society, 3(1), page 3-5. https://doi.org/10.1177/2053951715622512

using the computing system.[35] On the other hand, opacity is also an epistemic property, by means that it concerns the lack of a certain kind of knowledge.[36] In this sense, it is not trivial to determine what can be understood by an opaque system, since it can be so for a portion of the population, while it is not for another part.

While it is suggested by literature to overcome this form of opacity by means of widespread educational effort[37], the author takes another perspective. Making everyone knowledgeable about how a code works might be important, but it is not a solution in itself. What individuals need is not knowing what the code is or how the algorithm works but to understand why a certain output was provided by the model taking into account the features of the given input. Thus, rather than transparency, what is requested by the public is to be able to understand the reasoning that led the model to a certain output, i.e. they can interpret and understand why such results appear.

Indeed, education is a way to augment the chances to understand the code as the public will have a higher technical literacy. However, the author argues that a more effective solution is to shift the responsibility to provide explanations to tools which help to increase the explainability of the black box model. Examples such as AdFisher and Sunlight will be further explained under section “Explainability tools to undercover black box models”.

If one thinks for example in the legal domain, where machine learning technology can have a huge impact by automating processes and decisions from legal analytics, and quantitative legal prediction to robot-judges, given the composition of this field, which is composed predominantly of lawyers, who do not have a background in computer science or statistics, it is completely expected that for them many models are black boxes without really being considered as black boxes for computer scientists.

4.1.3. The aim of explainable AI

To open a black box model and build trust to use it, the machine learning models on which these systems are based should have at least two important characteristics: explainability - the reasoning behind each decision can be understood - and transparency - the availability of the model code with its design documentation, parameters and the learning dataset when the model relies on machine learning.[38]

---

[35] Zednik, C. (2019). Solving the black box problem: a normative framework for explainable artificial intelligence. Philosophy and Technology, (2019), page 5.
[36] Humphreys, P. (2009). The philosophical novelty of computer simulation methods. Synthese, 169(3), 615–626.
[37] Burrell, J. (2016). How the machine ‘thinks’: Understanding opacity in machine learning algorithms. Big Data and Society, 3(1), page 4. https://doi.org/10.1177/2053951715622512
[38] Stankovic, M. (2017). Exploring Legal, Ethical and Policy Implications of Artificial Intelligence, page 26.

It was previously mentioned that opacity and consequently transparency is relative to the stakeholder's skills instead of a characteristic by different AI models by itself. Without doubt referring to this, the author believes a more important issue to be addressed is on how to augment the level of explanation a black box model can provide to the different stakeholders.

This position follows the approach of Zednik[39] concerning the types of explanations the agents request and its level of usefulness, i.e. different stakeholders will have different needs to what knowledge they need to derive from the system currently in place.

In a model ecosystem, six kinds of agents, i.e. stakeholders, are suggested as per figure 2.

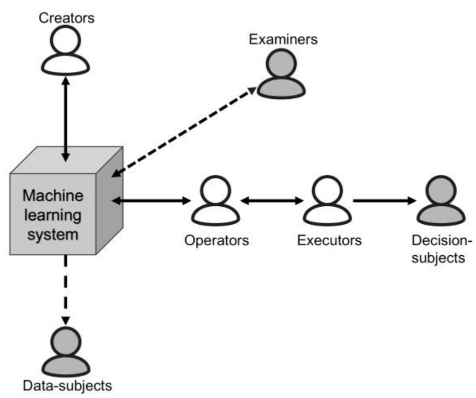

Figure 2: The ML ecosystem. Reproduced from Tomsett et al. (2018)

For the purposes of the current paper, we can split the stakeholders into 3 groups considering the case of risk profiling by Tax Administration: Tax Administration will have creators, operators and executors; taxpayer will be the data-subject and the decision-subject; lastly, the examiners are the courts and attorneys. Tax Administration will develop the computing system (creator), provide inputs to the model and receive the outputs from it (operator) in order to make decisions based on the algorithm (executor). Taxpayer as a data-subject will have his personal data contained in the learning environment and will be a decision-subject if the model will point out such taxpayer to be audited.[40] As the paper aims to study what current models need to improve rather than looking at how to design models which do not face the black box problem, the role of creator will be left aside.

When analysing a computer model, questions should be centred on *what* a system is doing and *why* it does what it does. While a what-question demands a descriptive approach, e.g. the inquirer wants the local properties of the computing system, a why-question requires an

[39] Zednik, C. (2019). Solving the black box problem: a normative framework for explainable artificial intelligence. Philosophy and Technology, (2019), page 5.
[40] Tomsett, R., Braines, D., Harborne, D., Preece, A., & Chakraborty, S. (2018). Interpretable to Whom? A Role-based Model for Analyzing Interpretable Machine Learning Systems, page 9-10. Retrieved from http://arxiv.org/abs/1806.07552

interpretation concerning the behaviour of the model when it comes to recognize a correlation between inputs and outputs.[41]

Concerning the different roles of the agents, the only part where the what-questions are as relevant as the why-questions is to operators that need to know what inputs are mandatory to enter and which outputs are generated. Thus, to fulfil such needs, the system should be rendered as sufficiently transparent to answer those questions. For the executor, more important than knowing the output out of the model, is to interpret the output, i.e. what does it mean, in order to make decisions. Looking from a hierarchical perspective, in Tax Administration the operator will be a person with a low level of responsibility who will provide the model with inputs again and again of the data-subject, i.e. taxpayer, so the model can provide an output that will be analysed at a later stage in the decision-making process. The operator will focus on describing what data from the taxpayer is being entered and what outputs are being generated by the model, e.g. looking at certain features of a taxpayer, his bank account should be blocked. The executor is performed by a person located in a higher part of the hierarchy (most likely the top, e.g. Head of Tax Administration) whose job is to make decisions regarding a taxpayer situation using the output provided by the model. Here interpretation rather than description of the output is of paramount importance to avoid errors and put in place the trust of the public in Tax Administration. Thus, it can be said knowledge and experience needed to provide inputs to the model as requested from the person acting as operator in Tax Administration is less than the ones required from an executor. In addition, as the executor is naturally in a higher place in the hierarchy than the operator it can be considered common ground that the operator tasks are less risky for the final result of Tax Administration decision-making than the executor ones, meaning that for Tax Administration it is more important answer why-questions rather than what-questions.

The same goes to the taxpayer. When personal data is processed it is relevant for the taxpayer to understand how his information is computed and used by the Tax Administration model. Besides that, once a decision is made by the Tax Administration based on the output of the model, the taxpayer wants to know how the outcomes are actually calculated and the factors which contributed to the output so it is possible to object to the decision before the court.

Courts and attorneys have the same needs in terms of why-questions rather than what-questions. The court needs to examine the legitimacy of the assessment made by the algorithm, i.e. interpret the logic of the system based on the law, while the attorney needs to understand the pitfalls of the features used to reach the output so the plaintiff has arguments to object to the decision.

The analysis is not based on whether or not it is possible to have access to see the algorithms, or if there is an asymmetry of information resulting from the presence of industrial or intellectual

[41] Zednik, C. (2019). Solving the black box problem: a normative framework for explainable artificial intelligence. Philosophy and Technology, (2019), page 10-11.

property secrets. The analysis is based on the complexity of the models, algorithms and technologies.

Regarding explainability of AI, as stated in the EU Commission paper[42], related to AI, explainability has different meanings, and the needs vary according to the audience. Not everyone needs the same level and type of explanation. Also, the requirements for explainability vary within Algorithmic Decision Systems (ADS) due to their different levels of abstraction and sophistication, and also because of the impact and relevance that they have regarding the decision-making process that is intended to automate.

Related to AI, transparency and explainability are critical concepts to be understood as to reduce the risks related to the ADS. There are relevant factors in play such as accountability, data protection, security and fairness. However, as mentioned above, explainability has a higher importance in the field currently being reviewed.

The main difference between transparent AI and explainable AI is the level of sophistication involved in the algorithm. For example, a model can be as transparent as possible, even showing the code of the algorithm, but if it requires a deep knowledge and deep level of abstraction, it will not be understood by most of the people. That is when the concept of explainable AI comes into play. It is important not only to be transparent with the information and processes involved in the algorithm, but also to be able to create AI which is explainable and possible to understand for the audiences and consumers. The black box problem comes into play when an algorithm is difficult enough to understand why a given output is reached taking into account the features the input of the data has inbuilt even if it is transparent with the structure/code of the solution.

Therefore, there is a trade-off between explainability and the sophistication or complexity of the models and algorithms, and as shown in Figure 1, there is consequently, also a trade-off between explainability and prediction accuracy.

4.1.4. Implications of the trade-off between explainability and accuracy

Assuming that opacity or transparency is relative, as mentioned above, and that in turn, is related to explainability, the largest trade-off that is seen in terms of black box models is found in the trade-off between explainability and accuracy, which in turn is given by the complexity of the models used. In general, the more complex the model, the more accurate it is. This leads to a large number of questions, problems, and ethical dilemmas, because depending on the field in which artificial intelligence will be used, it can have great repercussions. Even more so in the

[42] See European Parliament and Council of the European Union: White Paper On Artificial intelligence - A European Approach to Excellence and Trust.

case of fields where there is no symmetry of information or bargaining power between the parties involved, such as any type of interaction between a citizen and a government authority.

For the purposes of this research, the field of tax compliance will be referred, where the two parties involved are the Tax Administration and the taxpayer. The field of taxes is without a doubt a field where the application of AI by the authorities can generate a great dilemma and ethical questions. The complexity of society, globalization and big data demands novel approaches that only AI can offer. In the case of Tax Administrations, AI is of tremendous importance to process all the necessary data to enforce tax compliance. It is fully understandable that a government, through its Tax Administration, seeks adequate compliance with tax regulations, and as a consequence, that each taxpayer pays their fair share of taxes. The problem lies in the eventual use of AI models to achieve this goal, and especially when these models or algorithms fall within the definition of black box models.

The level of information asymmetry and power imbalance does not allow generating trust or a field in which the taxpayer can be certain about whether their rights and their particularities are being considered. Without going any further, the right to an explanation, and the power to understand the decisions behind the models, should be in the foundations of a tax system, which by definition affects the entire population of a country, with its particularities, different levels of education and technological understanding.

It is here where using black box models by the Tax Administrations generates an ethical conflict and sometimes allows questioning even if it is in line with the law. From the ethical side, AI in general and black box models in particular do not have embedded notions regarding human dignity for instance and, once again, the lack of understanding of the model by its users can have severe consequences to the decision-subjects.[43] Moreover, accountability is a key aspect when using black box models, i.e. who is liable to the outputs given by the model?[44] Lastly, another ethical issue raised by black box models deployment concerns privacy related to protection of personal data and how this can be secured.[45] Regarding the legal side, first it must be assessed whether there is legislation which includes the above ethical concerns which is non-exhaustive. If there is, a thorough review of it is requested to make sure the models used by Tax Administration are compliant. A third step is to have a critical attitude to discuss whether the legal framework is the most suitable according to the facts and circumstances.

The problem is far from trivial to solve, since the Tax Administration, in case of deciding that it requires AI to achieve its objective, faces the problem that was mentioned previously referred to

---

[43] Stankovic, M. (2017). Exploring Legal, Ethical and Policy Implications of Artificial Intelligence, page 28.
[44] Castelluccia, C., & Le Métayer, D. (2019). Understanding algorithmic decision-making: opportunities and challenges. In European Parliamentary Research Service (Vol. 15), page 28-29. https://doi.org/10.2861/536131
[45] Bird, E., Fox-Skelly, J., Jenner, N., Larbey, R., Weitkamp, E., & Winfield, A. (2020). The ethics of artificial intelligence: Issues and initiatives. In European Parliamentary Research Service, page 13. https://doi.org/10.2861/6644

the trade-off between explainability and accuracy, and therefore, when it comes to the complexity of the models, their understanding.

4.1.5. Black box approach

Also known as reverse-engineering, the black box approach allows to analyse a model without any knowledge of its code. This approach shows through observations of the algorithm output explanations that can provide insight about how the model works.[46] This approach can be used when the operator is not willing to disclose its code. Furthermore, due to its generality as it does not depend on the model under review, the scope of solutions built under this approach is very broad.[47]

The drawbacks, however, put into perspective the real added-value to open the logic of the model via black box approach. As the tester cannot target specific codes segments or error prone areas, he will not know what is being covered by the model.[48] Furthermore, due to the limited testing time, the test cases might be slow and difficult to write. Another challenge related to the black box approach is the design difficulty of testing a model without having clear functional specifications.[49] To overcome the above mentioned problems, researchers already have proposed algorithms as TREPAN[50] or general frameworks as Local Interpretable Model-agnostic Explanations (LIME)[51].

For the purpose of this paper, models suggested for specific problems such as targeting detection will be used due to the similarities the black box models used in this field have with the risk profiling carried by Tax Administration when fighting tax fraud. Also called multi-segment marketing, targeting is a marketing strategy that involves identifying specific personas or markets for specific content.[52] In the case of web targeting, this will involve the use as an input

---

[46] Castelluccia, C., & Le Métayer, D. (2019). Understanding algorithmic decision-making : opportunities and challenges. In European Parliamentary Research Service (Vol. 15), page 47. https://doi.org/10.2861/536131

[47] Castelluccia, C., & Le Métayer, D. (2019). Understanding algorithmic decision-making : opportunities and challenges. In European Parliamentary Research Service (Vol. 15), page 50-51. https://doi.org/10.2861/536131

[48] Software Testing - Methods - Tutorialspoint. (n.d.). Retrieved June 28, 2020, from https://www.tutorialspoint.com/software_testing/software_testing_methods.htm

[49] Koundinya. (2010). Black Box Testing, Its Advantages and Disadvantages. Retrieved June 28, 2020, from CodeProject website: https://www.codeproject.com/Articles/5579/Black-Box-Testing-Its-Advantages-and-Disadvantages

[50] Craven, M. W., & Shavlik, J. W. (1996). Extracting Thee-Structured Representations of Thained Networks. Advances in Neural Information Processing Systems, 24–30.

[51] Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). “Why should i trust you?” Explaining the predictions of any classifier. Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 13-17-August-2016, 1135–1144. https://doi.org/10.1145/2939672.2939778

[52] What is Targeting: Definition and types of targeting | Snov.io. (n.d.). Retrieved June 28, 2020, from https://snov.io/glossary/targeting/

of a user's personal data to tailor the output to be advertised such as recommendation, a price, search results or news.[53]

4.1.6. Explainability tools to undercover black box models

While in the legal domain at the moment this article is being written there is not a tool for decoding the reasoning behind the patterns found by black box models, in other areas dealing with similar issues Tax Administration faces there is research conducted that shows good indicators on a future of working in a society involved with higher complexity that needs the same level of complexity in the systems deployed to move forward. Following the suggestions provided by Castellucia and Le Metayer[54] derived from the field of targeting, below an overview of two of them will be given: AdFisher and Sunlight.

4.1.6.1. AdFisher

Developed for automating randomized, controlled experiments for studying online tracking, AdFisher can determine the use of information by web advertising algorithms and by personalized ad settings.[55] Though discrimination is inherent to profiling, as its point is to treat people differently, there situations in which the different treatment appears to be inappropriate such as discrimination based on gender in employment ads.[56] This tool creates new user profiles that will have different visit webpages. Then collections of the ads shown will be taken so a statistical analysis is made to understand whether there is discrimination and, if so, which is the feature in which discrimination is found.[57] To infer such conclusions, permutation tests were used to conduct large number of tests and the p-value was selected as the statistic to measure the likelihood of discrimination.[58]

---

[53] Lecuyer, M., Spahn, R., Spiliopoulos, Y., Chaintreau, A., Geambasu, R., & Hsu, D. (2015). Sunlight: Fine-grained targeting detection at scale with statistical confidence. Proceedings of the ACM Conference on Computer and Communications Security, 2015-October, 554–566, page 555. https://doi.org/10.1145/2810103.2813614

[54] Castelluccia, C., & Le Métayer, D. (2019). Understanding algorithmic decision-making : opportunities and challenges. In European Parliamentary Research Service (Vol. 15), page 50. https://doi.org/10.2861/536131

[55] Datta, A., Tschantz, M. C., & Datta, A. (2014). Automated Experiments on Ad Privacy Settings: A Tale of Opacity, Choice, and Discrimination, page 2. Retrieved from http://www.google.com/settings/ads

[56] Datta, A., Tschantz, M. C., & Datta, A. (2014). Automated Experiments on Ad Privacy Settings: A Tale of Opacity, Choice, and Discrimination, page 21. Retrieved from http://www.google.com/settings/ads

[57] Datta, A., Tschantz, M. C., & Datta, A. (2014). Automated Experiments on Ad Privacy Settings: A Tale of Opacity, Choice, and Discrimination, page 10. Retrieved from http://www.google.com/settings/ads

[58] Datta, A., Tschantz, M. C., & Datta, A. (2014). Automated Experiments on Ad Privacy Settings: A Tale of Opacity, Choice, and Discrimination, page 9. Retrieved from http://www.google.com/settings/ads

This tool is argued to have the potential to provide at least the part of the user profile that is used to select ads shown to the person.[59] This is a good example of what a black box approach is: without knowing the code and how it achieved a certain outcome, the tool is capable of presenting the features that led to such output. As opacity is considered agent relative for the purposes of this paper, the fact that the agent knows what personal data led to the algorithm to reach a certain result is more than enough to meet the requirements of explainability.

4.1.6.2. Sunlight

The need of a bigger scalable system to deal with the complexity of the web and how the personal information of the users is processed urged the need to understand problematic effects of data driven web as discriminatory pricing and advertising with solid statistical confidence.[60] Tested with Gmail and web ads, Sunlight can detect targeting of tens of thousands of ads while running simultaneously hundreds of inputs while running on hundreds of virtual machines to process data from targeting experiments with Gmail and web ads.

Through a four-stage pipeline and with a modular design, Sunlight gives the users the opportunity to adapt the system to their own needs and let them choose the best combination of mechanisms for the problems under research. These features make Sunlight useful for other business cases according to the needs of the user.[61] By generating fictitious profiles as input and analysing the outputs as the advertisements served, Sunlight provides by means of disjunctions simple explanations concerning the features which have a higher impact in the outcome.[62] The research shows that Google uses information about personal traits such as religious interest, race or difficult financial situation, though it is stated as well that the system cannot assign intention of either advertisers or Google for the targeting found.[63] Worth to mention is that though AdFisher has the same goal as to offer statistical confidence in its analysis, it is argued that a drawback is its lack of scalability compared to Sunlight as AdFisher runs all the agents on the same machine to prevent differences based on certain factors (e.g. IP

[59] Datta, A., Tschantz, M. C., & Datta, A. (2014). Automated Experiments on Ad Privacy Settings: A Tale of Opacity, Choice, and Discrimination, page 21. Retrieved from http://www.google.com/settings/ads
[60] Lecuyer, M., Spahn, R., Spiliopoulos, Y., Chaintreau, A., Geambasu, R., & Hsu, D. (2015). Sunlight: Fine-grained targeting detection at scale with statistical confidence. Proceedings of the ACM Conference on Computer and Communications Security, 2015-Octob, 554–566, page 554. https://doi.org/10.1145/2810103.2813614
[61] Lecuyer, M., Spahn, R., Spiliopoulos, Y., Chaintreau, A., Geambasu, R., & Hsu, D. (2015). Sunlight: Fine-grained targeting detection at scale with statistical confidence. Proceedings of the ACM Conference on Computer and Communications Security, 2015-Octob, 554–566, page 558. https://doi.org/10.1145/2810103.2813614
[62] Castelluccia, C., & Le Métayer, D. (2019). Understanding algorithmic decision-making : opportunities and challenges. In European Parliamentary Research Service (Vol. 15), page 50. https://doi.org/10.2861/536131
[63] Lecuyer, M., Spahn, R., Spiliopoulos, Y., Chaintreau, A., Geambasu, R., & Hsu, D. (2015). Sunlight: Fine-grained targeting detection at scale with statistical confidence. Proceedings of the ACM Conference on Computer and Communications Security, 2015-Octob, 554–566, page 561. https://doi.org/10.1145/2810103.2813614

address).[64] This difference can make AdFisher not valuable enough when it is necessary to train a model with a high number of inputs as the effect of each input and each possible combination of inputs needs to be tested separately.[65]

4.1.7. Application of targeting detection tools by Tax Administration

The application of AdFisher and Sunlight will be more useful when algorithms support risk profiling taken by Tax Administration as this is an activity similar to algorithm support targeting, as both are 1) an automated form of processing, carried out on personal data and 2) the objective is to evaluate personal aspects about a person.[66]

Risk profiling is an activity of paramount importance for Tax Administration. Without it, the goal of the revenue body will be hindered, i.e. collection of the tax due will not be fully achieved. Moreover, as due to the limitation of time and resources it is not feasible to audit all taxpayers, looking to those who have a significant risk increases efficiency and resource allocation savings by only focusing on tax returns that need to be audited, which tax issues need to be asked about and to pursue the necessary enquiries. By applying risk profiling, the Tax Administration can gather data about systemic risk to the tax system, e.g. areas where tax law enforcement is producing high compliance costs, and measure the performance in collecting taxes so the process can be reviewed and improved.[67]

Tax Administration thus, can benefit from AdFisher and Sunlight in two different moments regarding the timing in relation to the decision-making process:[68] the first being the moment of decision whether a black box model should be implemented in the process of tax compliance, e.g. prevention or detection of tax fraud and the second moment when a taxpayer requests explanations about why his or her profile triggered the algorithm so the Tax Administration pursued further investigation and to have enough information to object the decision made by the public sector institution in case such taxpayer suffers legal consequences due to the behaviour of the Tax Administration's algorithm.

---

[64] Datta, A., Tschantz, M. C., & Datta, A. (2014). Automated Experiments on Ad Privacy Settings: A Tale of Opacity, Choice, and Discrimination, page 10. Retrieved from http://www.google.com/settings/ads

[65] Lecuyer, M., Spahn, R., Spiliopoulos, Y., Chaintreau, A., Geambasu, R., & Hsu, D. (2015). Sunlight: Fine-grained targeting detection at scale with statistical confidence. Proceedings of the ACM Conference on Computer and Communications Security, 2015-Octob, 554–566, page 554. https://doi.org/10.1145/2810103.2813614

[66] WP29 (Article 29 Data Protection Working Party). (2018). Guidelines on Automated Individual Decision-Making and Profiling for the Purposes of Regulation 2016/679, page 6-7.

[67] Organisation for Economic Co-operation and Development. (2008). Study into the role of tax intermediaries. OECD, page 23. https://doi.org/10.1787/9789264041813-en

[68] Wachter, S., Mittelstadt, B., & Floridi, L. (2017). Why a right to explanation of automated decision-making does not exist in the general data protection regulation. International Data Privacy Law, 7(2), 76–99, page 78. https://doi.org/10.1093/idpl/ipx005

When evaluating whether the algorithm should be accepted in the processes of Tax Administration, using a similar tool capable of being implemented in the tax domain with the same features such as the above explained will provide valuable information regarding the existence of biases in data or in the logic behind the behaviour of the model as discrimination, by creating profiles and providing the features that have the strongest impact on the choice of picking up the taxpayer for further analysis by the Tax inspector. Moreover, when a black box model is used, the availability of a similar tool as AdFisher or Sunlight will allow for the taxpayer who is highlighted by the Tax Administration model to understand why he or she was targeted by the algorithm.

The importance of the application of AI in the development of society is hard to contradict. With more data and related complexity, only using human resources would not be enough to answer the challenges presented to the decision-makers. This is especially true in the public sector, namely in Tax Administration. The dynamics of the economy and the amount of data demand innovative approaches to cope with the pace of the business and taxpayer behaviour. Thus, it was required to go beyond human capital to increase tax compliance, study taxpayer behaviour and look into ways of making the tax compliance process more user friendly to the taxpayer. The complexity inherent in the models used may face the black box problem due to its opacity. This characteristic is argued in this paper to derive from the lack of explanations rather than transparency. Thus, more than having the code at hand, it is necessary to create mechanisms which can give the decision-maker and the decision-subject the information for the performance of their roles in the ecosystem. That is the purpose of the presentation of AdFisher and Sunlight to make it possible in the academia world of tax and technology to discuss similar solutions that can shed light on the black box models without opening them.

So far, black box models were analysed from a technological perspective while reviewing the questions that arise with opacity and applying a normative framework to understand what is more important between transparency and explanation.

However, this analysis is not finished without looking at the legal framework. Thus, it is time now to understand what legal challenges black box models face and what safeguards they need to have in order to be compliant notably when the decision-maker is located in Tax Administration.

4.2. Legal analysis

To balance the safeguard of tax revenues with the protection of fundamental rights of the taxpayer, national laws and regulations are not alone in this matter. At European level, the legal framework is divided into Primary EU law and Secondary EU law. In the former, the European Convention on Human Rights (ECHR) and the Charter of Fundamental Rights of the European Union (Charter) are included together with the EU Treaties. As a secondary EU law, for the purposes of the consequences of AI being used by Tax Administration, GDPR is the one to be referred to.

Concerning the ECHR, the models used by Tax Administration will fall under the principles laid down in the Convention as long as following the Engel criteria[69], the nature of the offence and a severe penalty the taxpayer can incur lead to a dispute that will be considered as a criminal charge.[70] Still under ECHR, a trial is considered fair where each party must be afforded a reasonable opportunity to present his case under conditions that do not place him at a substantial disadvantage vis-à-vis the other party.[71] That means that if information is not disclosed to the taxpayer, it cannot be said there is equality of arms. Additionally, the presumption of innocence is not met if the burden of proof is shifted completely to the taxpayer.[72]

Looking more from a protection of personal data point of view, black box models raise legal concerns concerning the Charter provisions.[73] Though it can be a specified purpose and legitimate basis laid down by law, if taxpayers do not have access to their data it is hard to argue those models can respect personal data from the data-subject and the decision-subject.

Other principles such as proportionality govern EU law. A measure is proportional if it is 1) legitimate, 2) suitable, 3) necessary and 4) reasonable.[74] Though with a legitimate goal, black box models cannot be said to hold the other characteristics: the opacity of the model can mean the taxpayer does not have the information to challenge the decision, no notification is given regarding the decision while it would be possible to have a less restrictive approach such as disclose the grounds on the decision and create clear procedural rules to challenge the decision.[75]

Provided an overview of the legal framework concerning Primary EU law, let us now focus the discussion on GDPR approach to safeguard the data-subjects and decision-subjects rights, namely the right not to be subject to automated decision-making and the right to explanation.

---

[69] Engel and others v. Netherlands, no. 5100/71; 5101/71; 5102/71; 5354/72; 5370/72, CE: ECHR:1976:0608JUD000510071.

[70] Papis-Almansa, M. (2019). The polish clearing house system: a 'stir'ring example of the use of new technologies in ensuring vat compliance in poland and selected legal challenges. Ec Tax Review, 28(1), 43–56, page 51-53.

[71] §142, European Court of Human Rights. (2013). Guide on Article 6: Right to a Fair Trial (Criminal Limb). Retrieved from https://twitter.com/ECHR_CEDH.

[72] §358 European Court of Human Rights. (2013). Guide on Article 6: Right to a Fair Trial (Criminal Limb). Retrieved from https://twitter.com/ECHR_CEDH.

[73] Article 8, Charter of the Fundamental Rights of the European Union [2012] OJ C 326/391.

[74] Craig, P. P., & De Búrca G. (2015). Eu law : text, cases, and materials (Sixth). Oxford University Press.

[75] Papis-Almansa, M. (2019). The polish clearing house system: a 'stir'ring example of the use of new technologies in ensuring vat compliance in poland and selected legal challenges. Ec Tax Review, 28(1), 43–56, page 55.

4.2.1. GDPR: legal framework for black box models

The use of black box models brought legal concerns attached. This was one of the gaps GDPR was intended to close. Designed to harmonise data privacy laws across its Member-States, GDPR gives greater rights to individuals, and allows for large fines to be handed out to organisations in breach of the rules.

Amongst those concerns, GDPR focuses on the explainability criteria decision-making algorithms need to meet. Explainability is fundamental for black box models analysis in the context of public administration processes as Tax Administrations do when checking the compliance of taxpayers. When taking a decision, e.g. due to fraud behaviour, Tax Administrations, the taxpayer and if needed the respective court need to know how the algorithm selected that specific taxpayer. The Tax Administrations need to understand the algorithm output so they are accountable and sure the features which prompted the warning are fair, i.e. absence of undesirable bias. Information needs to be provided to the taxpayer about why he was selected by the algorithm so he can contest the decision of Tax Administrations before the court. Lastly, the court has to have a sufficient level of understanding of the algorithm used by Tax Administrations to decide whether the decision was lawful.

Though GDPR is meant to protect personal data of natural persons via the person's consent, situations involving personal data processing for the performance of a task carried out in the public interest are still lawful.[76] Nevertheless, the public body which processes such data, i.e. data controllers, will be accountable regarding the compliance with the principles relating to processing of personal data.[77]

4.2.1.1. Automated individual decision-making and Profiling

Concerning the use of black box models, Tax Administrations need to comply with article 22 of GDPR when these models take a decision without meaningful human involvement.[78] According to its provisions, any automatic processing with legal effects deserves suitable safeguards, e.g. decision support systems. That is of paramount importance as these systems combine big data with statistical correlations which completely neglect legal theories or legal requirements such

[76] Article 6, Regulation (EU) 2016/679, of the European Parliament and the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation), 2016 O.J. (L 119)

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

[78] WP29 (Article 29 Data Protection Working Party). (2018). Guidelines on Automated individual decision-making and Profiling for the purposes of Regulation 2016/679, page 21.

as GDPR.[79] Thus, there should always be human involvement in such a way the data controller ensures that any oversight of the decision is meaningful rather than just a token gesture. It is also crucial that the controller has the power to change the decision.[80]

However, one must look to the exceptions mentioned in paragraph 2 of article 22, namely 22(2)(b) when automated decision-making could potentially take place if Union or Member-State law authorised its use. Germany used this exception to allow the use of automated decision-making in the context of providing services pursuant to an insurance contract.[81] Less restrictive was the Dutch implementation. In The Netherlands, the right to not be subject to automated individual decision-making, including profiling does not apply if it is necessary to comply with a legal obligation resting on the controller or necessary for the fulfilment of a task of general interest, i.e. public task.[82] Considered as one of the most innovative regulations regarding this topic is French law. Its scope encompasses more than decisions which have legal effects or similarly significant effects: apart from such effects on individuals, French law applies in the judicial field if such processing is intended to evaluate aspects of personality and in administrative decisions in which requirements need to be met to be partially or fully automated, except within the administrative appeal.[83]

Suitable safeguards should be present including “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”.[84]

4.2.1.2. Right to explanation

The provisions outlined in Articles 13–15 specify that data-subjects have the right to access information collected about them, and also requires data processors to ensure data-subjects are

[79] Roig, A. (2018). Safeguards for the right not to be subject to a decision based solely on automated processing (Article 22 GDPR). European Journal of Law and Technology, 8(3), page 6.
[80] WP29 (Article 29 Data Protection Working Party). (2018). Guidelines on Automated individual decision-making and Profiling for the purposes of Regulation 2016/679, page 21.
[81] Malgieri, G. (2019). Automated decision-making in the EU Member States: The right to explanation and other “suitable safeguards” in the national legislations. Computer Law and Security Review, 35(5), 105327, page 7. https://doi.org/10.1016/j.clsr.2019.05.002
[82] Malgieri, G. (2019). Automated decision-making in the EU Member States: The right to explanation and other “suitable safeguards” in the national legislations. Computer Law and Security Review, 35(5), 105327, page 11. https://doi.org/10.1016/j.clsr.2019.05.002
[83] Malgieri, G. (2019). Automated decision-making in the EU Member States: The right to explanation and other “suitable safeguards” in the national legislations. Computer Law and Security Review, 35(5), 105327, page 13. https://doi.org/10.1016/j.clsr.2019.05.002
[84] Recital 71, Regulation (EU) 2016/679, of the European Parliament and the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation), 2016 O.J. (L 119)

notified about the data collected.[85] When it comes to automated decision-making, article 13(2)(f) together with 14(2)(g) settle the information the data controller needs to provide to the data-subject:

- the existence of automated decision-making under Article 22(1) and (4);
- meaningful information about the logic involved;
- significance and envisaged consequences of such processing.[86]

Article 15(h) refers that the data-subject has the right to obtain information about the existence of solely automated decision-making, including profiling.

The right to explanation, namely related to automated decision-making has been discussed between scholars. In the first place it is argued that such a right does not exist.[87] Others argue that, whether the right should be called right to explanation, the difficulty born in what explanation needs to be given.[88] Lastly, there is also the opinion that the explanations developed by scientists are not the same as requested by GDPR.[89] This confusion in scholarship stems from the lack of clarity in the terms used in the provisions. In an attempt to solve it, the Article 29 Data Protection Working Party published guidelines on this topic.[90] Still there is ambiguity to be solved regarding the interpretation of this right.[91]

Jurisprudence is the only hope to clarify the concepts used in the right to explanation provisions of GDPR, especially due to the difference of approaches the Member-States had when implementing it. Up to this point, the Court of Justice of the European Union (CJEU) ruled personal data cases on the basis of the right to be informed about the processing of personal data by public administrative body and its transfer[92] and on the right of access to data and to information on the recipients of data in which there must be a fair balance between the interest of the data-subject in exercising his rights and the burden on the controller to store that

---

[85] Goodman, B., & Flaxman, S. (2017). European union regulations on algorithmic decision making and a ``right to explanation''. Ai Magazine, 38(3), 50–57, page 55.

[86] WP29 (Article 29 Data Protection Working Party). (2018), page 34. Guidelines on Automated individual decision-making and Profiling for the purposes of Regulation 2016/679

[87] Wachter, S., Mittelstadt, B., & Floridi, L. (2017). Why a right to explanation of automated decision-making does not exist in the general data protection regulation. International Data Privacy Law, 7(2), 76–99. https://doi.org/10.1093/idpl/ipx005

[88] Brkan, M. (2019). Do algorithms rule the world? Algorithmic decision-making and data protection in the framework of the GDPR and beyond. International Journal of Law and Information Technology, 27(2), 91–121. https://doi.org/10.1093/ijlit/eay017

[89] Edwards, L., & Veale, M. (2017). Slave to the Algorithm? Why a Right to Explanation is Probably Not the Remedy You are Looking for. SSRN Electronic Journal, 16.

[90] WP29 (Article 29 Data Protection Working Party). (2018). Guidelines on Automated Individual Decision-Making and Profiling for the Purposes of Regulation 2016/679.

[91] Veale, M., & Edwards, L. (2018). Clarity, surprises, and further questions in the Article 29 Working Party draft guidance on automated decision-making and profiling. Computer Law and Security Review, 34(2), 398–404, page 402. https://doi.org/10.1016/j.clsr.2017.12.002

[92] Judgment of 1 October 2015, *Bara and Others*, C-201/14, EU:C:2015:638.

information[93]. However, when it comes to the automated decision-making and what is regarded as a meaningful explanation, it is with expectation that the author awaits the first ruling from the CJEU in this aspect in order to shed light on the meaning of the right to explanation when automated decision-making algorithms are used, namely black box models.

4.2.2. STIR: a Polish approach to fight VAT fraud

As part of a plan to strengthen its VAT system, Poland adopted the System Teleinformatyczny Izby Rozliczeniowej (STIR) in 2017.[94] This system improves the time Polish Tax Administrations need to spot carousel fraud compared with the JPK analyser. [95]

STIR enables risk analysis based on exchange of information between the financial sector, the National Revenue Administration (NRA) and the Central Register of Tax Data. Such risk analysis helps to identify fraudster taxpayers.[96]

The information received by National Clearing House (KIR) on a daily basis from banks and credit unions is used by an algorithm to calculate a risk score that is sent to the Head of NRA. The risk score is then evaluated by the Head of NRA: in case it indicates the qualified person might be actively involved in unlawful activities, his bank account will be blocked up to 72 hours at least. VAT registration can also be refused or cancelled.[97]

4.2.2.1. Relevant characteristics of STIR concerning GDPR

In order to complete our analysis of STIR, the following characteristics are mentioned to provide further an overview about the compliance of the Polish algorithm with GDPR:

- Algorithm used to pursue a public interest such as combat VAT fraud;
- Data can be processed without knowledge and consent of persons to whom the data pertain;
- Neither algorithms developed by KIR nor explanation about how the score is calculated are disclosed to the taxpayer, i.e. black box problem;
- Taxable person is not informed about the blockage of the bank account.

---

[93] Judgment of 7 May 2009, *Rijkeboer*, C-553/07, EU:C:2009:293, paragraph 64.
[94] Act of November 24, 2017, on Preventing the Use of the Financial Sector for VAT Fraud.
[95] Sarnowski, J., & Selera, P. (2019). Reducing the VAT gap: Polish experience and legislative measures introduced in years 2016-2018. International Vat Monitor, Vol. 30 (2019), No. 3 ; P. 121-127, page 125.
[96] Papis-Almansa, M. (2019). The polish clearing house system: a 'stir'ring example of the use of new technologies in ensuring vat compliance in poland and selected legal challenges. Ec Tax Review, 28(1), 43–56, page 44-45.
[97] Papis-Almansa, M. (2019). The polish clearing house system: a 'stir'ring example of the use of new technologies in ensuring vat compliance in poland and selected legal challenges. Ec Tax Review, 28(1), 43–56, page 45-47.

4.2.2.2. Evaluation of STIR

From a technology point of view, STIR can be seen as black box model to the taxpayer, as it not only does not have the information about the decision taken in due time but also no explanation is given about how the output of the algorithm was conceived, i.e. which features were more relevant to the model to reach the outcome.

Though the data is processed without the consent of taxpayers, that is still lawful according to article 6(e). Moreover, article 22 of GDPR does not apply as the Head of NRA has the power to change the decision and ensures that any oversight of the decision is meaningful rather than just a token gesture according to the statistics available.[98]

The case differs when it comes to the right to explanation. The algorithm is a black box model from the taxpayer perspective as he does not get any explanation for those measures. Furthermore, there is no opportunity to contest the decision of blocking the bank account since the taxpayer is not notified.

Hence, it is necessary for STIR to be compliant with GDPR to apply measures that respect article 15 as providing to the taxpayer the reasons why he was selected by the algorithm so he has the chance to contest the decision. This can be supported by a tool similar to AdFisher or Sunlight to reveal the features more important to the model to reach a certain decision. That is in the opinion of the author would be considered under GDPR a suitable safeguard regarding the right to explanation.

4.2.3. Taxation Modernization Act: how the German tax code regulates fully automated taxation procedures

On January 1, 2017 the Taxation Modernization Act was implemented in Germany with the goal of regulating automation-based procedures for tax purposes, namely regarding tax assessments, credits of withheld taxes and prepayments.[99] The automated system is fed with data already available to the German Tax Administration and data provided by the respective taxpayer.[100] Data from third parties can also be regarded as taxpayer data provided that the taxpayer makes statements which contradict such data.

Automated systems are monitored via Risk Assessment Systems (RMS) which allow the German Tax Administration to check which cases have high risk and should be manually

---

[98] Bal, A. (2019). Ruled by algorithms: the use of `black box' models in tax law. Tax Notes International, 95(12), 1159–1166, page 1163.
[99] Tax Code (hereafter AO, German for Abgabenordnung) in the version published on October 1, 2002 (BGBl. I p. 3866; 2003 I p. 61), last amended by Article 6 of the Law of July 18, 2017 (BGBl. I p. 2745).
[100] § 155(4), sentence 1, AO.

checked.[101] To ensure the principle of investigation referred to in section 88, RMS needs to determine a sufficient number of cases to be thoroughly reviewed by the public officials.[102] Other safeguards are included in the article, notably the regular review of the RMS to check whether the goals are being successfully met.[103] In order not to jeopardize taxation compliance, the details about RMS must not be made publicly available.[104]

These systems will be used when there is no need to be reviewed by public officials. Automation also applies to situations such as the "issuance, correction, withdrawal, revocation, cancellation or amendment of administrative acts related to the situations described in the previous paragraph. [105] Lastly, under administrative instructions from the Federal Ministry of Finance or by the supreme financial authorities of a federal state, ancillary provisions to the administrative act can be issued fully on an automated basis.[106]

However, there are exceptions to the full automation. During the process, German Tax Administration public officials can process the tax return by his appointment.[107] Moreover, the RMS may also impose human processing.[108] The taxpayer can also request the human processing into a "qualified free-text field".[109] In this field, the taxpayer can argue why their tax return should be processed by public officials rather than automatically processed.[110] In case the taxpayer includes different data than the one provided by third parties, the tax return is also human processed.

4.2.3.1. Relevant characteristics of German legislation

Looking to the German legal framework from a GDPR perspective, the following are features which are useful when evaluating the compliance with the European regulation:

- Taxpayers are not able to choose freely between fully and partial automated-based assessment;
- The provisions of GDPR related to the right to explanation are included in the legislation; [111]
- As no legal consequences arise to the taxpayer, the taxpayer does not know whether its tax assessment was processed automatically;

[101] Braun Binder, N. (2018). AI and Taxation: Risk Management in Fully Automated Taxation Procedures. SSRN Electronic Journal, page 4. https://doi.org/10.2139/ssrn.3293577.
[102] § 88(5)(1), AO.
[103] § 88(5)(2)-(4), AO.
[104] § 88)(5), sentence 5, AO.
[105] § 155(4)(1), AO.
[106] § 155(4)(2), AO.
[107] § 88(5)(3), AO.
[108] § 88(5)(1)-(2), AO.
[109] § 155(4), sentence 3, AO.
[110] § 150(7), AO.
[111] § 32(a) - (e), AO.

- The risk assessment system which monitors the automated systems is regularly reviewed to determine whether it is fulfilling their objectives;[112]
- The automated procedure is meant to be supportive while processing data of the taxpayer during tax assessments instead of making decisions which may have consequences to the taxpayer;
- Both RMS and the automated systems are not known to the public about the in-built technology, bringing the doubts to the taxpayers concerning its fairness and lack of bias.

4.2.3.2. Evaluation of German legal framework

The German tax code does not include which automated systems are used, namely in terms of complexity (rule-based systems or neural networks). Thus, from a technology point of view, what can be inferred is that in the case complexity of the systems brings a lack of understanding of the outputs given, German Tax Administration needs to find ways to gather interpretations about the results of the automatic processing of tax returns. The same goes to the RMS. Though the RMS is meant to control tax evasion and it is monitored by the public officials, the law is not explicit on how they do it: using the term "sufficient" regarding the number of cases that the RMS needs to determine to be reviewed does not help for the taxpayer to feel safe.

From a GDPR perspective, there is no case of automated decision making, either when processing the data for tax return purposes or other mentioned in the tax code or when the RMS checks whether there is fraud: once RMS signals one return as high risk, is time for the public officials to take a look at it. Thus, these technologies work as a support rather than having power to take decisions on which taxpayers can suffer consequences. Though the taxpayer cannot choose to have the tax return processed manually, there is still the option to provide information in the tax return on why he wants his tax return being processed by a public official.

With regards to the right to explanation, as the provisions of GDPR are included and also extended in the tax code, no issues arise in this field. However, it is important to note that in case neural networks are used in the RMS that due to its complexity it becomes a black box model and the German Tax Administration cannot explain why the RMS signalled that tax return. Therefore, suggestions of using technology to help to interpret black box models as AdFisher and Sunlight may be in the list of options to solve the black box problem.

---

[112] § 88(5)(4), AO.

5. Conclusions

This thesis unfolded whether AI used by Tax Administrations can be compliant with EU law. The discussion was dealt in two different areas: technology and law. Under the former area first it was identified what makes a model become a black box: opacity. The approach taken in this paper is that though opacity can be considered as agent relative, it is more important to understand why a certain output was given by the model concerning the inputs than understanding what the code of the model is. Thus, tools which provide insights about what features were more important to the output should be developed rather than educate people to become knowledgeable about the code under the models used.

Such approach follows Zednik's conclusions in his suggested normative framework when assessing what is requested from a black box model, whether it is transparency and knowing the code which underlines the model or explainability and the ability to understand why a certain output was given by the model. The conclusion is that in a model ecosystem it is more important for the executor (Tax Administration) to answer why questions, i.e. interpret the results and understand the output of the model. From the decision-subject point of view, more important to know what the model does is in fact to understand what were the features of its data that made the model highlight his profile. Thus, the relevant characteristic of black box model to assess its compliance with GDPR is explainability rather than transparency.

When using black box models, attention needs to be given not only to the technology side, but also to the legal framework. Black box models fall under both primary EU law (ECHR and the Charter) and secondary EU law (GDPR), the latter taking special attention since it is the legal source which has more details concerning the use of automated decision-making and what is required under the right to explanation. Whenever AI is used to process data, safeguard measures need to be taken regarding monitoring as well as clarify who has the authority to take the decisions, i.e. the model or a human being. Also, explanations need to be provided according to the audience, being the data-subject whose data is processed or other stakeholders such as judges in the court or lawyers that represent the data -subject.

Following the evaluation of STIR under GDPR, the challenge comes to how can AI provide explanations to taxpayers in case of fraud detection. If the logic of the algorithm becomes completely public, the potential fraudsters would structure their activities accordingly and what could be considered an effective black box model would become a useless white box model. However, as the taxpayer is subject to sanctions, explanations regarding those measures need to be provided.[113] This analysis makes the mission of Tax Administrations very hard to accomplish.

---

[113] Bal, A. (2019). Ruled by algorithms: the use of `black box' models in tax law. Tax Notes International, 95(12), 1159–1166, page 1163.

On the other hand, Germany implemented a legal framework to be able to process tax returns automatically. The approach is a good example on how to use AI for tax purposes: technology should be a support to human decision rather than the decision-maker. Moreover, the GDPR criteria are included which provides a sufficient level of safety to the taxpayer.

Tax Administrations use AI in order to prevent lack of tax compliance or to find ex post taxpayers who did not follow the law. This means Tax Administrations take actions according to the classification of the taxpayer, i.e. if a taxpayer takes the correct behaviour, no legal consequences will arise, whereas if law is not complied with, the taxpayer will suffer the impact of such actions as long as Tax Administrations find out.

The classification of the taxpayer goes in line with what is done by marketing and pricing departments with AI: profiling their target. Accordingly, when research about what approaches in other fields are being taken to address black box models, the paper showed that tools such as AdFisher and Sunlight are used to shed light on the practices of profiling and micro-targeting are suggested in this paper due to the similar issues that black box models contain when used by Tax Administrations. Using these tools can in theory solve the issue of the right of explanation.

Thus, black box models used by Tax Administrations can be compliant with EU law as long as 1) there is a legal basis which allows its use and maintains the rights of the taxpayer (as the German case shows) and 2) there are tools implemented to guarantee the explainability of such models (AdFisher and Sunlight).

Further research will be needed to actually understand whether these tools can be deployed in the tax domain and what changes are necessary from its original design to be able to comply with GDPR concerns. If the suggestion is successfully accomplished, Tax Administrations can continue to have their decision-making process supported by black box AI without hindering EU Law. Also, such achievement will promote the focus on developing AI based on the results provided rather than the process to achieve such an outcome. Hence, Tax Administrations will have available more choices related to the accuracy of AI regarding the expected output of its analysis which will lead to an improvement in decision-making and, at the end, more effective tax compliance as well as a decrease in tax fraud.

Reference List

- Act of November 24, 2017, on Preventing the Use of the Financial Sector for VAT Fraud.
- Bal, A. (2019). Ruled by algorithms: the use of `black box' models in tax law. Tax Notes International, 95(12), 1159–1166.
- Baldrige, J. (2015). Machine Learning And Human Bias: An Uneasy Pair | TechCrunch. Retrieved June 26, 2020, from https://techcrunch.com/2015/08/02/machine-learning-and-human-bias-an-uneasy-pair/
- Bentley, D. (2019). Timeless principles of taxpayer protection: How they adapt to digital disruption. EJournal of Tax Research, 16(3), 679–713.
- Bird, E., Fox-Skelly, J., Jenner, N., Larbey, R., Weitkamp, E., & Winfield, A. (2020). The ethics of artificial intelligence: Issues and initiatives. In European Parliamentary Research Service. https://doi.org/10.2861/6644
- Braun Binder, N. (2018). AI and Taxation: Risk Management in Fully Automated Taxation Procedures. SSRN Electronic Journal. https://doi.org/10.2139/ssrn.3293577
- Brkan, M. (2019). Do algorithms rule the world? Algorithmic decision-making and data protection in the framework of the GDPR and beyond. International Journal of Law and Information Technology, 27(2), 91–121. https://doi.org/10.1093/ijlit/eay017
- Brynjolfsson, E., Hitt, L. M., & Kim, H. H. (2011). Strength in numbers: how does data-driven decisionmaking affect firm performance? Ssrn Electronic Journal, (2011). https://doi.org/10.2139/ssrn.1819486
- Burrell, J. (2016). How the machine 'thinks': Understanding opacity in machine learning algorithms. Big Data and Society, 3(1). https://doi.org/10.1177/2053951715622512
- Campolo, A., Sanfilippo, M., Whittaker, M., & Crawford, K. (2017). AI Now 2017 Report. In AI Now Institute. Retrieved from https://ainowinstitute.org/AI_Now_2017_Report.pdf
- Castelluccia, C., & Le Métayer, D. (2019). Understanding algorithmic decision-making : opportunities and challenges. In European Parliamentary Research Service (Vol. 15). https://doi.org/10.2861/536131
- Charter of the Fundamental Rights of the European Union [2012] OJ C 326/391.
- Convention for the Protection of Human Rights and Fundamental Freedoms, Rome, 4 Nov. 1950
- Craig, P. P., & De Búrca G. (2015). Eu law : text, cases, and materials (Sixth). Oxford University Press.
- Craven, M. W., & Shavlik, J. W. (1996). Extracting Thee-Structured Representations of Thained Networks. Advances in Neural Information Processing Systems, 24–30.
- Davenport, T. H., & Dyché, J. (2013). Big Data in Big Companies.
- Datta, A., Tschantz, M. C., & Datta, A. (2014). Automated Experiments on Ad Privacy Settings: A Tale of Opacity, Choice, and Discrimination. Retrieved from http://www.google.com/settings/ads
- Dressel, J., & Farid, H. (2018). The accuracy, fairness, and limits of predicting recidivism. Science Advances, 4(1), 5580. https://doi.org/10.1126/sciadv.aao5580

- Edwards, L., & Veale, M. (2017). Slave to the Algorithm? Why a Right to Explanation is Probably Not the Remedy You are Looking for. SSRN Electronic Journal, 16.
- Engel and others v. Netherlands, no. 5100/71; 5101/71; 5102/71; 5354/72; 5370/72, CE: ECHR:1976:0608JUD000510071.
- European Parliament and Council of the European Union: White Paper On Artificial intelligence - A European Approach to Excellence and Trust.
- Ferguson, A. G. (2011). Predictive policing and the future of reasonable suspicion. Ssrn Electronic Journal, (2011). https://doi.org/10.2139/ssrn.1965226
- Goodman, B., & Flaxman, S. (2017). European union regulations on algorithmic decision making and a ``right to explanation''. Ai Magazine, 38(3), 50–57.
- Guidotti, R., Monreale, A., Ruggieri, S., Turini, F., Giannotti, F., & Pedreschi, D. (2018). A survey of methods for explaining black box models. ACM Computing Surveys, 51(5). https://doi.org/10.1145/3236009
- Humphreys, P. (2009). The philosophical novelty of computer simulation methods. Synthese, 169(3), 615–626.
- J.M. Calderón Carrero & J.S. Ribeiro, Fighting Tax Fraud through Artificial Intelligence Tools: Will the Fundamental Rights of Taxpayers Survive the Digital Transformation of Tax Administrations?, 60 Eur. Taxn. 6 (2020), Journal Articles & Papers IBFD (accessed 7 July 2020).
- Katz, D. M. (2012). Quantitative Legal Prediction - Or - How I Learned to Stop Worrying and Start Preparing for the Data-Driven Future of the Legal Services Industry. Emory Law Journal, 62. Retrieved from https://heinonline.org/HOL/Page?handle=hein.journals/emlj62&id=923&div=&collection=
- Kelleher, J. D., & Tierney, B. (2018). Data science (Ser. The mit press essential knowledge series). MIT Press.
- Koundinya. (2010). Black Box Testing, Its Advantages and Disadvantages. Retrieved June 28, 2020, from CodeProject website: https://www.codeproject.com/Articles/5579/Black-Box-Testing-Its-Advantages-and-Disadvantages
- Lecuyer, M., Spahn, R., Spiliopoulos, Y., Chaintreau, A., Geambasu, R., & Hsu, D. (2015). Sunlight: Fine-grained targeting detection at scale with statistical confidence. Proceedings of the ACM Conference on Computer and Communications Security, 2015-October, 554–566. https://doi.org/10.1145/2810103.2813614
- Malgieri, G. (2019). Automated decision-making in the EU Member States: The right to explanation and other “suitable safeguards” in the national legislations. Computer Law and Security Review, 35(5), 105327. https://doi.org/10.1016/j.clsr.2019.05.002
- Marr, D. (2010). Vision : a computational investigation into the human representation and processing of visual information. MIT Press. https://web-b-ebscohost-com.ezproxy.ub.unimaas.nl/ehost/ebookviewer/ebook/bmxlYmtfXzMyNDcwOF9fQU41?sid=dcd4fa36-c631-40e9-b024-23c34547208c@pdc-v-sessmgr06&vid=0&format=EB&rid=1
- McAfee, A., & Brynjolfsson, E. (2012). Big data: The management revolution. Harvard Business Review, 90(10), 4.

- Moses, L. B., & Chan, J. (2014). Using big data for legal and law enforcement decisions: testing the new tools. University of New South Wales Law Journal, The, 37(2), 643–678.
- Organisation for Economic Co-operation and Development. (2008). Study into the role of tax intermediaries. OECD. https://doi.org/10.1787/9789264041813-en
- Papis-Almansa, M. (2019). The polish clearing house system: a 'stir'ring example of the use of new technologies in ensuring vat compliance in poland and selected legal challenges. Ec Tax Review, 28(1), 43–56.
- Regulation (EU) 2016/679, of the European Parliament and the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation), 2016 O.J. (L 119).
- Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). "Why should i trust you?" Explaining the predictions of any classifier. Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 13-17-August-2016, 1135–1144. https://doi.org/10.1145/2939672.2939778
- Roig, A. (2018). Safeguards for the right not to be subject to a decision based solely on automated processing (Article 22 GDPR). European Journal of Law and Technology, 8(3).
- Sarnowski, J., & Selera, P. (2019). Reducing the VAT gap: Polish experience and legislative measures introduced in years 2016-2018. International Vat Monitor, Vol. 30 (2019), No. 3 ; P. 121-127.
- Savona, E. U., Riccardi, M., & Berlusconi, G. (Eds.). (2016). Organised crime in European businesses. Routledge.
- Software Testing - Methods - Tutorialspoint. (n.d.). Retrieved June 28, 2020, from https://www.tutorialspoint.com/software_testing/software_testing_methods.htm
- Stankovic, M. (2017). Exploring Legal, Ethical and Policy Implications of Artificial Intelligence.
- Surden, H. (2012). Computable Contracts. U.C. Davis Law Review, 46. Retrieved from https://heinonline.org/HOL/Page?handle=hein.journals/davlr46&id=643&div=&collection=
- Surden, H. (2014). Machine learning and law. Washington Law Review, 89(1), 87–115.
- Technologies for Better Tax Administration. (2016). In Technologies for Better Tax Administration. https://doi.org/10.1787/9789264256439-en
- Tjoa, E., & Guan, C. (2019). A Survey on Explainable Artificial Intelligence (XAI): Towards Medical XAI. Retrieved from http://arxiv.org/abs/1907.07374
- Tomsett, R., Braines, D., Harborne, D., Preece, A., & Chakraborty, S. (2018). Interpretable to Whom? A Role-based Model for Analyzing Interpretable Machine Learning Systems. Retrieved from http://arxiv.org/abs/1806.07552
- Uchida, C. D. (2009). National Discussion on Predictive Policing: Defining Our Terms and Mapping Successful Implementation Strategies. National Criminal Justice Reference Service. Retrieved from https://www.ncjrs.gov/pdffiles1/nij/grants/230404.pdf%5Cnpapers3://publication/uuid/24AA7AF5-FEDF-496C-AB25-19C28C9BAC4F

- Veale, M., & Edwards, L. (2018). Clarity, surprises, and further questions in the Article 29 Working Party draft guidance on automated decision-making and profiling. Computer Law and Security Review, 34(2), 398–404. https://doi.org/10.1016/j.clsr.2017.12.002
- Wachter, S., Mittelstadt, B., & Floridi, L. (2017). Why a right to explanation of automated decision-making does not exist in the general data protection regulation. International Data Privacy Law, 7(2), 76–99. https://doi.org/10.1093/idpl/ipx005
- What is Targeting: Definition and types of targeting | Snov.io. (n.d.). Retrieved June 28, 2020, from https://snov.io/glossary/targeting/
- WP29 (Article 29 Data Protection Working Party). (2018). Guidelines on Automated Individual Decision-Making and Profiling for the Purposes of Regulation 2016/679.
- Zednik, C. (2019). Solving the black box problem: a normative framework for explainable artificial intelligence. Philosophy and Technology, (2019). https://doi.org/10.1007/s13347-019-00382-7