ARTICLE

Ethics in AI Use for Tax Administration: Guidelines for the Future

Tofigh Hasen Nezhad Nisi*, Stephen Daly** & Francien Dechesne***

This paper examines the ethical challenges in automating tax administration, focusing on the integration of Artificial Intelligence (AI) technologies. It critically reviews some existing frameworks, identifying their shortcomings in addressing the risks posed by the use of AI technologies by tax authorities. To mitigate these risks, the paper suggests guidelines for responsible AI adoption by tax authorities, emphasizing human centredness, fairness, transparency, oversight, and constraining the coercive power of the state.

Keywords: Artificial intelligence, ethics, rule of law, tax administration, *Toeslagenaffaire*, ECHR, GDPR, AI Act, Syri, eKasa.

1 Introduction

Across the globe, tax administrations are increasingly turning to Artificial Intelligence (AI) and automation to perform their functions.[1] Most common is the use of AI for risk management.[2] Reductions in staff headcount, coupled with the increasing difficulty in managing tax compliance,[3] have prompted tax authorities to adopt these technologies at an accelerated pace.[4] The Organization for Economic Co-operation and Development (OECD) acknowledges this trend, and envisions a future where automation becomes integral to tax administration, a vision shared by tax authorities eager to keep pace with evolving demands.[5] They term this 'Tax Administration 3.0' and envisage AI bringing about a 'paradigm shift' from the present approach which relies on active, burdensome, voluntary compliance by taxpayers towards seamless and frictionless automation.[6]

However, in the rush to embrace automation, deeper ethical implications of these technological advancements can be overlooked. This is a considerable risk for tax administrations, which cannot afford to adopt an 'after the fact' approach when implementing changes to their internal workings. The use of AI and automation in tax administration not only mechanizes parts of existing and explicitly defined tasks, but implicitly transforms processes, roles and relations within and with the service. Accordingly, there needs to be serious deliberation about the ethical risks to be managed, which can be translated into internal policies and provide the foundations on which tax authorities can further embrace automation.[7]

Notes

* Research assistant at University of Leiden. Email: t.hasen.nezhad.nisi@law.leidenuniv.nl. The authors wish to thank David Hadwick, Andreas Häuselmann, Esther Huiskers, Benita Mathew and two anonymous reviewers for comments on previous iterations of this paper. Thanks also to Katarina Lau for research assistance on the paper. All errors remain those of the authors alone. This work has been supported through research funding provided by the British Academy and Leverhulme Trust Small Research Grants Scheme (grant nr SRG2425\251209).

** Reader in Tax Law, King's College London. Email: stephen.daly@kcl.ac.uk.

*** Associate Prof. of Ethics and Digital Technologies at the Centre for Law and Digital Technologies of the Leiden Law School, University of Leiden. Email: f.dechesne@law.leidenuniv.nl.

1 OECD, *Tax Administration 2022: Comparative Information on OECD and Other Advanced and Emerging Economies* 23–24 (OECD 2022), https://www.oecd.org/en/publications/2022/06/tax-administration-2022_5944859e.html.

2 OECD, *Tax Administration 2024: Comparative Information on OECD and Other Advanced Economies* 91 (OECD 2024), https://www.oecd.org/en/publications/2024/11/tax-administration-2024_5c4606e4.html.

3 For a slightly dated report, *see* European Public Service Union, *The Impact of Austerity on Tax Collection* (Jun. 2020), available at https://www.lrd.org.uk/research/impact-austerity-tax-collection.

4 *See* for instance IRS, *Internal Revenue Service Inflation Reduction Act Strategic Operating Plan: FY2023 – 2031* 3 and 10 (Apr. 2023), available at https://www.irs.gov/about-irs/irs-inflation-reduction-act-strategic-operating-plan.

5 OECD, *Tax Administration 3.0: The Digital Transformation of Tax Administration* (OECD 2020), https://www.oecd.org/en/publications/2020/12/tax-administration-3-0-the-digital-transformation-of-tax-administration_886337a7.html.

6 *Ibid.*, at 3 and 7.

7 *Ibid.*, at 24.

The problem for tax administrations thus is that they must harness the power of AI and automation,[8] while protecting important norms (societal values as well as legal rights) and being mindful of the risks that the technology poses.[9] This paper is prompted by this conundrum and seeks to explore the ethical dimensions of automating tax administration processes. It suggests the need for, and attempts to articulate, a structured framework to guide tax authorities. The 'need' follows from a failure of existing legal frameworks to sufficiently regulate risks inherent in the technology. Such a framework, as we will explain in this article, should emphasize human centredness, fairness, oversight, transparency, and constraining the coercive power of the state. Our framework builds on both the fundamental technological risks and the Rule of Law, which we see as central to tax administration (and upon which *any* framework for regulating AI use by a tax authority must build). Given the scale of risk posed, we ultimately adopt a thicker Rule of Law account for this purpose.

Though this is not the first attempt at developing either a framework for guiding the use of AI by a tax authority[10] nor the first ethical framework developed for guiding the use of AI generally,[11] this is the first attempt at a holistic *ethical* framework for the use of AI by *tax authorities* which responds to fundamental risks with AI and the foundational ideal underpinning tax administration (namely, the Rule of Law).

The paper is broken down into three substantive parts. The first part delves into how the use of AI in tax administration should be governed. In the second part, we focus on how concerns we raise in the first part are not adequately addressed by certain regulatory frameworks, namely the General Data Protection Regulation (GDPR),[12] AI Act[13] and European Convention on Human Rights (ECHR).[14] In the third part, we consolidate our findings into a proposal for guidelines, which we offer as a starting point for a broader discussion. They aim to translate abstract ethical concerns into tangible standards which may serve as safeguards.

The goal of this paper is to analyse issues and propose solutions that are of universal relevance. In doing so, however, we predominantly focus on European regulatory frameworks. Further, though tax authorities are imbued with a range of functions today, owing to the expansion of the state in the twentieth century along with mergers of different tax collecting bodies, this article shall focus on the use of AI by tax authorities in carrying out their primary function of collecting taxes due (which includes all ancillary functions such as providing assistance to taxpayers and cooperating with other tax authorities through the exchange of information or assistance in the collection of taxes).

2 Ethics of AI in tax administration

Before examining the ethical issues that arise where AI is put to use by tax authorities, it is necessary first to articulate the ethical issues which generally arise where AI is used and secondly to explore the ethics that generally govern tax administrations. Thereafter, this part will examine how general AI ethical issues map on to the Rule of Law to reveal the issues arising from the use of AI by tax authorities that ultimately need regulating.

2.1 *An Exploration of AI*

AI is a broad, multidisciplinary field of research and engineering that concerns itself with the construction of machines and systems that perform functions commonly associated with the cognitive capacities of the human mind. While there is no clear consensus on what systems or techniques should be regarded as AI, we take the term broadly to include tasks such as (autonomous) decision making, search, planning, reasoning, image recognition and generation, natural language processing (NLP) and generation. Due to natural cycles in buzzword popularity, we note that 'AI' is also used to describe techniques that are or have been called 'data mining', 'big data', 'business intelligence', 'data science', 'deep learning' and many other things.[15]

Current AI-systems are mostly built using the technical paradigm of machine learning (ML). Such AI-systems

Notes

[8] Nowadays the term Artificial Intelligence (AI) is often used to refer specifically to systems which use machine learning. To avoid the debate on the 'proper' definition of AI, and because we discuss forms of automation broader than machine learning, we will refer to the relevant innovations as 'AI and automation'.

[9] OECD (2020), *supra* n. 5, at 12–24.

[10] *See* for instance IFS Tax Law Review Committee, *Artificial Intelligence in Automated Decision-Making in Tax Administration: The Case for Legal, Justiciable and Enforceable Safeguards* (2024), https://ifs.org.uk/publications/artificial-intelligence-automated-decision-making-tax-administration-case-legal.

[11] OECD, *OECD Framework for the Classification of AI Systems* (OECD 2022), https://www.oecd.org/en/publications/oecd-framework-for-the-classification-of-ai-systems_cb6d9eca-en.html on which *see* Brendt Mittelstadt, *Principles Alone Cannot Guarantee Ethical AI*, 1 Nature Mach. Intelligence 501 (2019), doi: 10.1038/s42256-019-0114-4; *Executive Order 14110: Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence* (30 Oct. 2023).

[12] Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 Apr. 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.

[13] EU Artificial Intelligence Act, *The Act Texts*, https://artificialintelligenceact.eu/the-act/.

[14] Council of Europe, European Convention for the Protection of Human Rights and Fundamental Freedoms, as amended by Protocols Nos. 11 and 14, 4 Nov. 1950.

[15] For the state-of-the-art introduction to the field of Artificial Intelligence, we refer to the standard textbook: Stuart Russell & Peter Norvig, *Artificial Intelligence: A Modern Approach* (Pearson 4th ed. 2020).

convert historical data into a statistical *model*, which in turn is applied to inputs (new data) to produce outputs (results, decisions, predictions or judgments). In traditional AI,[16] expert knowledge and reasoning rules were explicitly represented and coded as 'models' by human programmers. Such an explicit approach has pros and cons, as it both *allows* and *requires* ex ante specification of the instructions for what the system should do. This makes it hard to automate tasks that rely on tacit knowledge or that are otherwise ineffable. ML-algorithms generate or fine-tune models by attempting to optimize a human-provided numerical *objective function* – a function that measures how 'good' an outcome is – based on data it interacts with during a process called *training.* ML allows for the identification of patterns in the data, often unknown to humans and/or hard to represent in a way that adequately corresponds with human understanding. Where data is plentiful, as in the case of tax authorities for instance which acquire huge amounts of data (not just from taxpayers, but also third parties such as employers and banks, as well as other government agencies and foreign tax authorities), the more accurate ML-systems can be.

While the potential of AI is impressive and speaks to the Zeitgeist, especially since the public launch of AI-tools capable of generating text, images and videos,[17] there are also specific challenges that require awareness and attention when implementing AI-techniques in existing practices.[18] First, the way AI systems work does not align with (human) common sense, nor an understanding of context or the concepts they are dealing with. As with computer programs in general, ML-algorithms literally execute computational instructions – in the case of ML these instructions aim at some numerical optimization (the *objective function*), without connection to our non-numerical understanding. It requires both deep mathematical knowledge and human subject matter understanding to map societal goals and concepts onto a fitting mathematical objective function to steer the performance of an AI system towards the intended functionality. If there is a mismatch between the mathematical structures and the nature of the societal problem, this can result in a lack of robustness, unintended and unintelligible outcomes, the inability of the system to deal with different contexts of application, unforeseen circumstances, and harmful decisions.[19] Failing to incorporate humans into the development and operation of AI will leave these problems uncorrected.

Secondly, given that ML works with large sets of numerical data, there are several types of bias that may inadvertently slip into the systems. At its core, a bias is a systemic error, or deviation from a standard, in results or inferences.[20] Certain characteristics or groups of people may be under or overrepresented in the available data (a phenomenon referred to as 'infra-marginality').[21] This is a function of a more general problem inherent in working with numerical data, namely, the focus is shifted to whatever aspects can be measured, quantified and computed at the expense of more qualitative aspects of a decision, and more deliberative decision procedures. The behaviour and reliability of ML-systems depends in large part on the data that is available and selected for training. Such data needs to be accurate, plentiful and meaningfully correlated to the decision problem as well as its context of application. A secondary consequence, as a result, is that proper training of ML-systems requires a lot of computational resources. A key challenge then is ensuring that biases are eliminated (to the extent possible) as this is key to securing fairness by ensuring equal treatment by the systems.

Thirdly, the relation between the outputs and choices made in the training of the system will be hard to capture or explain in a way that makes sense in terms of the decision or other functionality we attribute to the system. This is just one of the ways in which ML tends to produce opaque AI-systems that we cannot map directly onto human reasoning, decision making and understanding.[22] This is particularly the case with predictive models using deep neural networks (DNNs), which contain multiple layers of nodes to process data. These act like neurons in the brain and are referred to as 'black box' systems because it is not clear how the information provided ('inputs') led to the arrival at particular decisions ('outputs'). Those affected by decisions made by opaque systems are denied the opportunity to understand the reasoning behind a

Notes

[16] Often referred to as 'Good Old Fashioned AI' or GOFAI. *See* e.g., Melanie Mitchell, *Artificial Intelligence: A Guide for Thinking Humans* (Farrar Straus and Giroux 2019).

[17] The breakthrough in terms of public awareness of such tools was the launch by OpenAI of ChatGPT in Nov. 2022. The techniques underlying generative AI tools like OpenAI's GPT (text), Dall-E (image) or Sora (video) have been developed over the course of years by researchers at several institutes and companies.

[18] For further and more detailed reading on the issues described in this section, *see* Arvind Narayanan & Sayash Kapoor, *AI Snake Oil: What Artificial Intelligence Can Do, What It Can't, and How to Tell the Difference* (Princeton University Press 2024). Some of these concerns are also shared by the IMF: Joshua Aslett, Stuart Hamilton, Ignacio Gonzalez & Michael A. Hardy, *Understanding Artificial Intelligence in Tax and Customs Administration* 23 (IMF 21 Nov. 2024).

[19] Such issues are articulated as Abstraction Traps in Andrew Selbst, Danah Boyd, Sorelle Friedler, Suresh Venkatasubramanian & Janet Vertesi, *Fairness and Abstraction in Sociotechnical Systems,* ACM Conference on Fairness, Accountability, and Transparency 59 (2019 FAT*), doi: 10.1145/3287560.3287598.

[20] *Cochrane Handbook for Systematic Reviews of Interventions version 5.2.0,* 8.3 (Julian Higgins, Rachel Churchill, Jacqueline Chandler & Miranda Cumpston eds, updated Jun. 2017). Elsewhere, the term is used to positively describe the manipulation of models, such as referring to the use of 'weights' in order to deliberately bias a model towards a particular outcome when training. *See* Stuart Russell & Peter Norvig, *Artificial Intelligence: A Modern Approach* 728 (Pearson 3d ed. 2009).

[21] *See* Arpita Biswas, Siddharth Barman, Amit Deshpande & Amit Sharma, *Quantifying Infra-Marginality and Its Trade-off With Group Fairness* (arXIV 3 Sep. 2019), https://arxiv.org/abs/1909.00982.

[22] *See* Jenna Burrell, *How the Machine 'Thinks': Understanding Opacity in Machine Learning Algorithms* 3(1) Big Data & Soc'y 1, 3–5 (2016), doi: 10.1177/2053951715622512.

decision affecting their interests and cannot be confident in the propriety of the decision-making process. Although various approaches have been proposed to improve the explainability of AI systems, significant limitations remain.[23]

Fourthly, there is a need for monitoring, evaluating, updating, managing and maintaining of AI systems as well as the data ecosystem they rely on: data needs to be carefully gathered, sanitized, entered into a database, (often) annotated, evaluated, securely stored, and used to (re)train the ML-system or (re)evaluate predictions. Failure to do this 'pre-processing' can result in inaccuracies, unfair outcomes, and privacy breaches, thus rendering oversight essential. Old systems of accountability thus may be insufficient and instead new, more technical requirements for design and deployment of the technology may be required to oversee the executive branch, such as socio-technical robustness and safety.[24] Definitions of AI systems often stress a certain level of autonomy as an essential characteristic.[25] However, this autonomy is not inherent to an AI system: it is determined by decisions on the extent to which it operates without human oversight or control. Making sure this oversight or control is meaningful is challenging and often at odds with the inherent opacity of ML.

Fifthly, some ML-algorithms are probabilistic in nature. As advanced and as sophisticated as they might become, the outcome is based on probabilities instead of certainties – a prediction is not a promise or proof, and hence predictions should also not be considered decisions.[26] Any ML-system will make 'errors' in the sense of misclassifying new examples. Norvig and Russell use the example of a system classifying emails as either spam or non-spam – some emails will be classified by the algorithm as spam when they are not (so-called *false positives*), and others not spam when they are (*false negatives*).[27] As a result, when designing an ML-system, an error threshold must be chosen – one which is either prone to producing more false negatives, or false positives. Whether false negatives or false positives are preferable will depend on the uses to which the system is being put.[28] In the case of cancer screening for instance, false positives are preferable to false negatives[29] – people who do not have cancer are told that they do, which will obviously cause stress, but further testing will show that they do not; whereas in the converse situation, people with cancer will have been given the all clear! Where AI is used in the exercise of the coercive powers of the state, false positives can result in individuals wrongly subjected to treatment detrimental to their interests (for instance, wrongful imprisonment).

Sixthly, in the specific case of so-called large language models, trained with ML on vast amounts of textual data to produce linguistically correct text, the output reflects what are plausible sentences based on word orders and co-occurrences in training data – but not on the basis of the meanings of the sentences in the texts on which it was trained. Therefore it is to be expected that these models will also produce outputs that express false statements about the world, or clearly immoral or dangerous advice. This phenomenon is called 'hallucination'.[30] It is inherent to the underlying statistical and probabilistic mechanisms, but also the aforementioned problems in data and modelling.[31] Notwithstanding this, it warrants separate mention as a specific risk as it is so notorious.

Once properly appreciated how AI technologies operate, it is obvious that they raise epistemic, moral and political issues. While automation and digitalization of processes may initially have been regarded as ethically neutral, increasing awareness has arisen – especially in the wake of the hype around so-called 'big data'[32] – of ethical (re)evaluation and safeguards required for the introduction of data-driven technologies.[33] At the level

Notes

[23] *See* e.g., Rudresh Dwivedi, Devam Dave, Het Naik, Smiti Singhal, Rana Omer, Pankesh Patel, Bin Qian, Zhenyu Wen, Tejal Shah, Graham Morgan & Rajiv Ranjan, *Explainable AI (XAI): Core Ideas, Techniques, and Solutions. ACM Computing Surveys*, 55(9), ACM Computing Surveys 1 (2023), doi: 10.1145/3561048; Blazej Ku niacki, Marco Almada, Kamil Tylliński, Łukasz Górski, Beata Winogradska & Reza Zeldenrust, *Towards eXplainable Artificial Intelligence (XAI) in Tax Law: The Need for a Minimum Legal Standard*, 14(4) World Tax J. 573 (2022), https://doi.org/10.59403/2yhh9pa.

[24] *See* for a study of requirements for responsible AI in Law Enforcement: Francien Dechesne, Virginia Dignum, Lexo Zardiashvili & Jordi Bieger, *AI Ethics for Law Enforcement: A Study into Requirements for Responsible Use of AI at the Dutch Police*, 2(4) Delphi – Interdisc. Rev. Emerging Technologies 179 (2019), doi: 10.21552/delphi/2019/4/7.

[25] *See* e.g., the Definition provided by the European Commission High-Level Expert Group on AI: European Commission, *Ethics Guidelines for Trustworthy AI* (8 Apr. 2019), https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai.

[26] Reuben Binns, *Human Judgment in algorithmic loops: Individual justice and automated decision-making*, 16 Reg. & Governance 197 (2022), doi: 10.1111/rego.12358.

[27] Norvig & Russell (2020), *supra* n. 15, at 691.

[28] Karen Yeung & Adam Harkens, *How Do 'Technical' Design-Choices Made When Building Algorithmic Decision-making Tools for Criminal Justice Authorities Create Constitutional Dangers? Part II*, Pub. L. 448, 459 (2023), doi: 10.48550/arXiv.2301.04715.

[29] *Ibid.*, at 459.

[30] *See* Cambridge Dictionary definition of 'Hallucinate', https://dictionary.cambridge.org/dictionary/english/hallucinate.

[31] Gary Marcus, *Why DO Large Language Models Hallucinate?* (5 May 2025), https://garymarcus.substack.com/p/why-do-large-language-models-hallucinate; Jim Waldo and Soline Boussard, *GPTs and Hallucination: Why Do Large Language Models Hallucinate?* 22(4) Queue 19 (2024), https://doi.org/10.1145/3688007; IBM, *What are AI Hallucinations?*, https://www.ibm.com/think/topics/ai-hallucinations.

[32] Viktor Mayer-Schonberger & Kenneth Cukier, *Big Data: A Revolution That Will Transform How We Live, Work and Think* (Eamon Dolan/Houghton Mifflin Harcourt 2013).

[33] *See* on this the 'fallacy of equivalence' discussed in Yeung & Harkens (2023), *supra* n. 28, at 469.

of data, the GDPR within the EU attempted to regulate at a general level some of the main ethical issues around the processing of personal data, to promote and protect values of (informational) privacy and autonomy. The GDPR's regulatory strategy is to push accountability obligations on the data controller and processor on the one hand and defines several 'data subject rights' to empower the individual on the other.

It soon became clear, however, that a focus on data alone, without any regard to how it was used, was insufficient, making no clear distinction between use of data by a public authority with intrusive surveillance powers or by a supermarket chain trying to understand consumer trends: AI systems were treated by the GDPR 'as any processing activity or as automated decision-making without differentiation of the purpose of the model, the controllers, and its end consequences'.[34] Technical developments in processing data required further reflection and guidance. For this reason, the European Commission convened a selection of experts in 2018 to form the 'High-Level Expert Group on Artificial Intelligence'.[35] The Group finalized its 'Ethics Guidelines for Trustworthy AI' report in 2019.[36] It lists three high level general principles towards achieving so-called 'Human-Centric' AI-systems, so that human values are central to the way in which AI systems are developed, deployed, used and monitored.[37] These are that the systems should be lawful, ethical and technically robust. The document also specifies seven requirements for the development, implementation and use of 'trustworthy' AI-systems: *(1) human agency (both in terms of empowering humans to make choices and respecting human rights) and oversight; (2) technical robustness and safety; (3) privacy and data governance; (4) transparency; (5) diversity (including stakeholder participation), non-discrimination and fairness; (6) environmental and societal well-being; and (7) accountability.*[38]

Similar principles and requirements were formulated at several levels and by different actors (companies, oversight agencies, engineering societies, governments, the United Nations, and the Council of Europe for instance), with a significant overlap in the principles each converged on[39] – perhaps not surprisingly given the level of generality at which the principles were formulated.

These European Commission High-Level Expert Group principles have also been a stepping stone towards the development of the first globally direct regulation of AI products. Building on the efforts of the High-Level Expert Group, the European Commission published a whitepaper on AI and the 'European approach to Excellence and Trust' in February 2020,[40] followed by a proposal (in April 2021) to regulate AI in the European Union.[41] The final text of the AI Act, a risk-based product regulation, with some reference to human rights (for high-risk systems), was published on 12 July 2024. The formulation of the AI Ethics principles and requirements has been of value in facilitating the debate about the societal risks and impacts on fundamental rights and values of the use of AI technology in general. Nevertheless, the obligation in the AI Act to do broader impact assessments depends on the risk classification of the system (as we will touch on later). We would suggest however that for *all* AI-systems deployed, moral aspirations still need to be operationalized from a level of abstraction to implementable standards.[42]

Coming back to our diagnoses of the six challenges with AI which raise the most pressing ethical issues (common sense, fairness, transparency, error threshold and hallucina, we see commonality, but not complete overlap, with the principles articulated by the High-Level Expert Group. Where we see the lack of human common sense, the High-Level Expert Group *see (1) human agency and oversight*, which demands that decisions be intelligible to human beings. Securing fairness maps on to *(5) diversity, non-discrimination and fairness*, which similarly regards biases as unfair. Counteracting opacity requires *(4) transparency.* Securing oversight relates to *(2) technical robustness and safety*, as both require checks to ensure the system is functioning appropriately, *(3) privacy and data governance*, as data needs to be properly managed, *and (7) accountability*, given the need for affected parties to have the ability to check the propriety of the decision-making process and to secure redress if necessary. Calibrating an appropriate error threshold and

Notes

[34] David Hadwick, *Slipping Through the Cracks, the Carve-outs for AI Tax Enforcement Systems in the EU AI Act*, 9(3) Eur. Papers 936, 945 (2024), doi: 10.15166/2499-8249/793.

[35] European Commission, *High-Level Expert Group on Artificial Intelligence*, https://digital-strategy.ec.europa.eu/en/policies/expert-group-ai.

[36] European Commission, *Ethics Guidelines for Trustworthy AI*, https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai.

[37] High-Level Expert Group, *Ethics Guidelines for Trustworthy AI. Brussels: European Commission* at 37 and 10 (fn. 17) (8 Apr. 2019), https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai.

[38] *Ibid.*, at 15–20.

[39] *See* Jessica Field, Nele Achten, Hannah Hilligoss, Adam Nagy & Madhulika Srikumar, *Principled Artificial Intelligence: Mapping Consensus in Ethical and Rights-based Approaches to Principles for AI* (15 Jan. 2020), https://cyber.harvard.edu/publication/2020/principled-ai.

[40] European Commission, *On Artificial Intelligence – a European Approach to Excellence and Trust* (19 Feb. 2020), https://commission.europa.eu/system/files/2020-02/commission-white-paper-artificial-intelligence-feb2020_en.pdf.

[41] EU Artificial Intelligence Act, *Historic Timeline*, https://artificialintelligenceact.eu/developments/.

[42] *See* e.g., Jeroen van den Hoven, *Ethics for the Digital Age: Where Are the Moral Specs?*, in *Informatics in the Future* 65–76 (Hannes Werthner & Frank van Harmelen eds, Springer 2017).

managing hallucinations engage the fundamental rights relevant to both *(1) human agency and oversight* (which explicitly refers to protection for fundamental rights), as well as *(3) privacy and data governance* (as regards respect for the right to privacy). However, whilst our diagnosis did not pick up on the importance of considering environmental impact of AI which is central to *(6) environmental and societal wellbeing*, as we do not consider that relevant to the uses to which AI *is* put, the potentially destructive uses to which it can be put which affects *societal wellbeing* did underpin our concerns for common sense, fairness, transparency and oversight.

Though we see commonality with the principles enunciated by the High-Level Expert Group, the latter go further than we think strictly necessary. We would suggest, however, that developing an initial ethical framework for the use of AI in tax administration should be grounded principally in the six core concerns that the use of AI raises.

2.2 *The Rule of Law*

The primary function of tax authorities is to collect taxes due. Tax administration ethics is concerned with how tax authorities should perform that function. At the core of these ethical considerations is the Rule of Law. The Rule of Law is an ideal of political morality which has the effect of elevating the importance of law and legal infrastructure in any system of governance.[43] It has two essential components. First, that official action is authorized by law, commonly referred to as the principle of legality.[44] Law 'tempers' power by creating and strengthening it on the one hand whilst constraining it on the other.[45] Second, that laws should have certain qualities in order to perform their functions. Although writers agree as to its importance and on the need for these two basic components, they tend to adopt either thinner or thicker versions of the Rule of Law.

The core function of a 'thin' account of the Rule of Law is that laws should act as guidance[46] – the 'guidance function'. A person should, before committing themselves to any course of action, be able to know the legal consequences that will follow.[47] What advocates of thicker accounts argue is that the Rule of Law makes further, more substantive dictates, such as requiring respect for fundamental human rights[48] and private property.[49] Thicker versions thus consider not just the inner (procedural and formal) morality of laws, but also what laws *should* regulate. Laws to that end should constrain the arbitrary exercise of power.[50] In the context of relationships between the government and its subjects, this means for instance that the 'thick' principle of legality would regard limiting the exercise of government power as important.

The Rule of Law both frames tax authorities' powers and responsibilities *and* informs tax authorities how those powers should be exercised and responsibilities discharged. Tax authorities *should* uphold the Rule of Law. For instance, to enable the guidance function of law, it has been argued that it is desirable for governments to inform people how the rules work through providing published guidance, bespoke rulings and other forms of administrative assistance.[51] Similarly, the Rule of Law's demand for congruity[52] insists that tax authorities should treat taxpayers neither favourably nor unfavourably for extraneous purposes but only in line with the law – demanding no more and conceding no less than the amount of tax that is due under the law. Finally, the Rule of Law implies that there should be accountability through law. Thus, where tax authorities act, there should be an independent body, such as a court, to whom taxpayers can turn to check whether the tax authorities have exercised their powers appropriately. In turn, tax authorities *ought* to act in a transparent manner – making clear how they have arrived at consequential decisions – so that taxpayers are equipped with the information they need to be able to properly challenge the tax authority later before an independent body.

In striving to uphold the Rule of Law, tax administrations touch upon broader ethical considerations such as fairness (by treating taxpayers even-handedly and hence satisfying the Rule of Law's demand for congruity), accountability (through empowering taxpayers to use law to challenge them) and autonomy (through providing taxpayers with the information necessary for them to make decisions about how to comply with the tax code). But where the Rule of Law regards these as

Notes

[43] Jeremy Waldron, *The Rule of Law*, in *The Stanford Encyclopedia of Philosophy* (Edward Zalta ed., Summer 2020 Edition), https://plato.stanford.edu/archives/sum2020/entries/rule-of-law/.

[44] Neil MacCormick, *Institutions of Law: An Essay in Legal Theory* 43 (Oxford University Press 2007).

[45] Martin Krygier, *What's the Point of the Rule of Law*, 67(3) Buff. L. Rev. 743, 782–783 (2019), doi: 10.1007/s40803-019-00094-2.

[46] John Tasioulas, *The Rule of Law*, in *The Cambridge Companion to the Philosophy of Law* 121 (John Tasioulas ed., Cambridge University Press 2020).

[47] *Black-Clawson International Ltd v. Papierwerke Waldhof-Aschaffenberg AG* [1975] A.C. 591 (HL), 638 (Lord Diplock).

[48] Ronald Dworkin, *A Matter of Principle* 11–12 (Harvard University Press 1985).

[49] John Locke, *Two Treatises of Government*, para. 138 (1689).

[50] TRS Allan, *Law, Liberty, and Justice: The Legal Foundations of British Constitutionalism* 23 (Oxford University Press 1994).

[51] Stephen Daly, *Tax Authority Advice and the Public* (Hart Publishing 2020).

[52] Lon Fuller, *The Morality of Law* 39 (Yale University Press revised ed. 1969).

important to the extent that they are instrumental in increasing compliance with the tax laws, other ethical frameworks would regard them as ends in themselves and make broader demands of administration. Fairness and autonomy might require tax authorities to adopt guidelines or codes of conduct to ensure taxpayers are treated with respect. Accountability might demand transparency over spending decisions. To be clear, these are all noble endeavours, but they are not ends pursued *per se* by the Rule of Law which has a slightly narrower focus. We would argue however that whilst these other accounts have value, any account of the ethics of tax administration should build upon the Rule of Law and so we use this as the starting point for our analysis. This is so as tax *is* a creature of law (and indeed many wars have been fought to bring this about) – without law, there is no tax! Those broader values only become relevant because there *is* tax law. How law *should* be used to rule then is foundational to any analysis of the regulation of tax authority conduct.

In short, different accounts of the Rule of Law bring about different concerns when it comes to regulating tax authorities, including, as we will explore, how they use AI. For proponents of thicker accounts of the Rule of Law, the core problem that needs to be managed is the coercive power of the state – the Rule of Law mandates that this power is constrained. With thinner accounts, the demands are less imposing, but would still require tax authorities to use their powers to fulfil the Rule of Law's guidance function and ensure accountability through law.

2.3 *Ethics of AI in Tax Administration*

AI in tax administration involves decision-making by non-human actors which cannot adopt critical human thinking, nor construct a strict pattern of reasoning based on the dataset provided by the programmer. The implementation of AI provides huge opportunities in terms of efficiency and effectiveness.[53] The most common use to which ML has been put by tax authorities is in respect of their compliance risk management systems (RMSs). These systems are central to risk management in advanced tax administrations and predate the use of ML,[54] as taxpayers have long been selected for audit through data analysis of the compliance risk. RMSs function as resource allocation tools which identify the likelihood that particular taxpayers have been non-compliant and hence help tax authorities determine which taxpayers to investigate further. ML-assisted RMSs draw on taxpayer data from myriad sources to build risk profiles of individual taxpayers. They can flag anomalies and inconsistencies in a taxpayer's affairs, as well as highlight where there is a likelihood that the taxpayer has underpaid relative to what would be expected from a similarly placed taxpayer or because that taxpayer has certain characteristics which mean that they are more likely to have underpaid taxes (such as operating in a particular cash-based trade). AI is also used for nudging, chatbots and case law analysis.[55]

Whilst the full capabilities of AI are not currently exploited by tax authorities, it is highly likely that it will become central to how they undertake their duties. This trend is reflected in OECD research, which shows that over half of tax administrations have either implemented AI or are planning to integrate it into their core functions.[56] Moreover, the growing use of AI tools by tax administrations is systematically recorded by several major repositories.[57] There will be teething issues with technology of course. However, it has been repeatedly demonstrated that a legal gap exists concerning the regulation of AI use by tax administrations.[58] As will be shown in the next substantive part of this article, frameworks such as the GDPR and the AI Act primarily focus on consumer products and services. They offer limited guidance for public administration, and even less for the specific context of tax administration. This limited legal coverage demonstrates not only the need for better regulation, but also the importance of developing ethical principles that respond to a new kind of relationship between tax administrations and taxpayers, one shaped by automated, data-driven decision-making. Such systems affect not just consumer safety, but create novel conditions under which individuals exercise their rights, access justice, and organize their lives in relation to the state. It therefore must be appreciated that there exists a fundamental distinction, between the conduct of a public institution adopting AI, and private entities, such as Facebook. While Facebook operates within a framework of defined freedoms and legal constraints, the tax administration

Notes

[53] *See* OECD (2020), *supra* n. 5.

[54] OECD, *Tax Administration in OECD and Selected Non-OECD Countries: Comparative Information Series* (OECD 2006), https://www.oecd.org/en/publications/2006/11/tax-administration-in-oecd-and-selected-non-oecd-countries_g1gh7b25.html.

[55] Hadwick (2024), *supra* n. 34, at 941–942.

[56] OECD (2024), *supra* n. 2, at 91.

[57] *See* e.g., https://www.oecd.org/en/data/datasets/inventory-of-tax-technology-initiatives.html; https://data.rafit.org/?sk=BA91013D-3261-42F8-A931-A829A78CB1EC; https://taxadmin.ai/country-reports/.

[58] *See* e.g., Bruno Peeters, *Editorial: European Law Restrictions on Tax Authorities' Use of Artificial Intelligence Systems: Reflections on Some Recent Developments*, 33(2) EC Tax Rev. 54 (2024), doi: 10.54648/ecta2024006.

exclusively derives its existence and authority from law. A tax administration does not have the luxury to, as Facebooks' motto goes: '*move fast and break things*'. Tax authorities are not players in a game, but instead are regulators tasked with guarding the tax system and enforcing the rules. This imbalance of power and fundamental difference in roles means that it is not acceptable for a tax authority to experiment with AI – to adopt an 'after the fact' approach. Where businesses move fast and break things, they accumulate market share; where public authorities do so, they abuse their public powers.

For this reason, tax administrations should take seriously the ethical issues that are raised by the use of AI (human common sense, fairness, transparency, oversight, error threshold and hallucination challenges), which we explored in the first section of this Part. Meanwhile, though the endeavour to construct a reliable and efficient AI-driven tax system is undeniably challenging, tax administrations should continue to adhere to the Rule of Law. This demands, as explored in the second section to this Part, that laws should guide, public authorities should be accountable through law and laws should constrain public power (to a greater extent in thicker accounts than thinner accounts). When we combine the ethical risks that we raise in section 1 and the demands of the Rule of Law in section 2, we develop a holistic framework for guiding the use of AI by tax authorities, one which emphasizes human centredness, fairness, transparency, oversight and constraining coercive power. In order to explain how we arrive at this framework, four points need to be elaborated.

First, there is a degree of symmetry between the ethical risks we identify and respect for the Rule of Law. Human common sense dictates that the system should be 'human centric' such that reasons for an AI decision should be intelligible to affected taxpayers, which in turn is a precondition of accountability through law. Through providing intelligible reasons, taxpayers are also empowered to make choices as they can better understand the legal consequences that will flow from their actions, as mandated by the guidance function of law.[59] Fairness through the elimination of bias is important as it ensures that all taxpayers are subject to the law in the same way, hence fulfilling the guidance function by ensuring greater consistency between the law as stated and applied. Lack of transparency is problematic because a person needs to be able to understand a system both so that it can guide that person's conduct (the guidance function of law) *and* to ensure that the person can challenge the public authority where it believes that the decision is wrong as a matter of law (accountability through law). Oversight is important so that mistakes are minimized or corrected (through accountability channels, such as courts) and thus there is greater congruity between the law as stated and applied.

Secondly, there are some issues generated by the use of AI which might be said to go beyond the scope of a thin Rule of Law's concerns. This shows that a thicker Rule of Law account, focused more intently on constraining executive power, should be favoured in an AI context. Two points are made to elaborate this argument. One, in a thicker account, but not a thinner account of the Rule of Law, choosing an appropriate error threshold would be particularly pertinent. Consider false positives in a tax audit selection context – taxpayers that have no underpaid taxes would suffer through the stress of an unfounded tax investigation. Investigative tax authority powers are a clear example of the coercive power of the state (intruding upon fundamental rights, including privacy and data rights) that a thick Rule of Law would wish to constrain. Whilst it is correct that a tax authority should aim for false positives as opposed to false negatives (as otherwise there would be systemic non-collection of taxes due), a thicker Rule of Law account would suggest that false positives should be minimized to the greatest extent possible. This is particularly so in a context where false positives can operate at scale given the efficiencies brought on by AI and automation. Two, AI poses a threat as it enhances the capabilities of the executive branch,[60] even becoming an 'Algorithmic Leviathan'.[61] AI has a tendency to increase power for the institution using AI (thereby cutting against the thicker account's principle of legality which serves to both create and constrain power). This is most obvious in respect of the datafication of tax information and the efficiency correspondingly offered by increasingly powerful computational tools to collect tax. Even basic data scraping tools are exponentially quicker than human beings manually collecting and inputting data. It can also be seen in how AI has been used by public authorities to mask policy decisions which shift power from the individual to the state.[62] In amongst the various problems in *Toeslagenaffaire* was that *prima facie* evidence of fraud could result in a benefit claimant's entire annual claims

Notes

[59] Note that the High-Level Expert Group also recognizes the importance both of the Rule of Law to ethical AI systems and the centrality of human dignity (High-Level Expert Group, *supra* n. 37, at 10 (fn. 17) and 37), which underpins the Rule of Law (*see* Daly (2020), *supra* n. 51, Ch. 3).

[60] Gerald Postema, *Law's Rule: The Nature, Value, and Viability of the Rule of Law* 301 (Oxford University Press 2022); Sofia Ranchordás, *Tax and (Digital) Inequalities,* in *Taxation and Inequalities* (Rita de la Feria ed., IBFD 2025).

[61] Nathalie Smuha, *Algorithmic Rule by Law: How Algorithmic Regulation in the Public Sector Erodes the Rule of Law* 4 (Cambridge University Press 2024).

[62] *See* on Robodebt, Jake Goldenfein, *Lost in the Loop – Who is the 'Human' of the Human in the Loop?*, in *Global Governance by Data* (Gavin Sullivan, Fleur Johns & Dimitri Van Den Meerssche eds, Cambridge University Press forthcoming).

being retroactively reclaimed by the public authority.[63] What this underlines is that an AI system is just that – a system. What consequences should be attached to decisions made, or recommended, by the system is a matter of policy. To that end, in line with Smuha, we would support a thick Rule of Law account when it comes to regulating the use of AI[64] given the need to avoid allowing AI to augment the coercive power of the state.

Thirdly, the Rule of Law may also influence the development and deployment of AI for the better (moving beyond simply managing the risks associated with AI use). There are scholars for instance that have written about how law can drive, shape, and enable technological development.[65] AI can empower human choice, something which is germane to the Rule of Law. In the tax administration context, ML-based NLP *could* assist in advancing the Rule of Law's demands for accountability through law and the guidance function of law.[66] AI *could* be harnessed to bring about better explanations as to tax authority actions than is currently the case.[67] These are tasks where tax authorities can overcome the ordinary numerical optimization limitations of AI, as it is possible to generate an objective function which translates the non-numerical problems faced by tax authorities. This idea that the Rule of Law can demand that AI should have an empowering effect on taxpayers mirrors the High-Level Expert Group's aspiration that AI should foster human choice.[68] The development of AI so as to empower human choice also maps on to the High-Level Expert Group's call for the inclusion of stakeholders.[69] From the Rule of Law's perspective, including stakeholders in the development of AI and including humans in decision making would be important to the extent that these human decision-makers and stakeholders would be able to pick up on and prevent legal mistakes being made by AI systems.[70] More generally meanwhile, including stakeholders affected by AI so as to enable the human centric development of AI (as their input can best ensure that AI is developed in line with human sensibilities and values), would be a further nod to the Rule of Law. In other words, the connection to the Rule of Law arises because including stakeholders in this way optimizes human choice. Those choices will then be factored into the development of AI, which in turn will be used by tax authorities through law (as the powers of tax authorities are *all* derived from law, as per the principle of legality). As such, whilst we flagged earlier the issue of the need to have regard to human common sense in the use of AI, the Rule of Law would also suggest that we should have regard to the empowering effect of technology on human choice and the desirability of including human stakeholders in the development of AI. Bringing these together, managing the ethical risks we have identified and respect for the Rule of Law translates to the broader concern for human centric development and use of AI, where 'human centredness' would encapsulate common sense, choice and inclusion.

Fourthly, where AI has been used proactively to assist taxpayers, through chatbots for instance, mistaken advice from the system, such as hallucinations, technically undermines the guidance function of law (as taxpayers are not guided correctly as to their obligations). At the same time, however, it is a policy choice to impose the consequences of the mistake on an innocent taxpayer (in other words that they must pay the taxes due under the law, rather than what would be due on the basis of the chatbot's advice).[71] That policy choice is to favour, and thus further empower, the state over the taxpayer. As such, there is tension - not irreconcilable, but a tension nevertheless - between a demand of the Rule of Law and the desire to constrain the coercive power of the state. In our view, the desire to constrain the coercive power of the state should take precedence, as to prioritize a technical demand of the guidance function of law would be to lose sight of the telos of the Rule of Law in this context which is to constrain coercive power.

2.4 *Summary of the Ethical Use of AI in Tax Administration*

What this part has sought to illuminate are the concerns that should guide the ethical use of AI by tax authorities having regard to the technical challenges posed by the use

Notes

[63] David Hadwick & Shimeng Lan, *Lessons to Be Learned from the Dutch Childcare Allowance Scandal: A Comparative Review of Algorithmic Governance by Tax Administrations in the Netherlands, France and Germany,* 13(4) World Tax J. 609, 613 (2021), 10.59403/27410pa.

[64] *See also* Smuha, *supra* n. 61.

[65] *See* for instance Roger Brownsword & Han Somsen, *Law, Innovation and Technology: Before We Fast Forward – a Forum for Debate*, 1(1) L. Innovation & Tech. 1, 1–3 and 49–53 (2009), doi: 10.1080/17579961.2009.11428364; Julie Cohen, *Between Truth and Power: The Legal Constructions of Informational Capitalism* 16 and Chs 1–3 (Oxford University Press 2019).

[66] Stephen Daly, *Artificial Intelligence, the Rule of Law and Public Administration: The Case of Taxation*, 83(3) Cambridge L. J. 437 (2024), doi: 10.1017/S0008197324000448.

[67] *See* for instance Ku niacki, Almada, Tyliński, Górski, Winogradska & Zeldenrust, *supra* n. 23.

[68] High-Level Expert Group, *supra* n. 37, at 16.

[69] *Ibid.*, at 19.

[70] For an analogous context, *see* Daly (2020), *supra* n. 51, at Ch. 5 analysing consultation with taxpayers on tax authority guidance.

[71] An instructive case on what can be termed 'contra legem legitimate expectations', albeit one that did not involve a chatbot but instead an automated system, is *Pintarich v. Deputy Commissioner of Taxation* [2018] FCAFC 79.

of AI and the central demands of the Rule of Law. For us, this is the best starting point for developing a framework to guide tax authority actions because it is based upon the fundamental issues that any framework should seek to tackle. Our diagnosis is that the following are the key challenges in using AI: operationalizing human common sense, fairness, transparency, oversight; choosing an appropriate error threshold; and managing hallucinations. The Rule of Law meanwhile demands accountability through law, that laws should perform a guidance function (enabling human choice) and that laws should constrain public power, through protecting fundamental rights for instance.

We take the ethical risks of AI use seriously and we also take seriously the importance of the Rule of Law. By virtue of the combination of these concerns, we regard as essential that any framework regulating the use of AI by tax authorities would emphasize human centredness, fairness, transparency, oversight, and constraining the coercive power of the state. Of course, these are not mutually exclusive demands and overlap to a considerable extent. For instance, ensuring fairness through the elimination of bias also helps to constrain the coercive power of the state by protecting people from discrimination at the hands of public authorities.

2.5 *A Word on Trust*

We regard it as paramount that any framework for regulating AI use in tax administration should, as a result, emphasize human centredness, fairness, transparency, oversight, and constraining the coercive power of the state. Doing so will also generate further positive consequences, such as fostering trust amongst taxpayers which in turn may lead to better tax compliance.[72] O'Neill, a foremost authority on trust,[73] suggests that trustworthiness will be partly determined by whether a party lives up to the standards expected of them and whether they are competent in the tasks assigned to them.[74] Both are satisfied in a case where a tax authority manages the risks inherent in AI and upholds the Rule of Law – these are the standards that we should fundamentally expect a tax authority to live up to and how we judge whether they are acting competently! This conclusion is further supported by the work of the High-Level Expert Group. To recall, our concerns overlap considerably with those of the High-Level Expert Group and the principles espoused by them to manage those concerns mirror the demands that we see as central to any framework for regulating the use of AI in tax administration. Just as the High-Level Expert Group argued that satisfying those principles would render the AI 'trustworthy', so we argue the same of satisfying our demands.

3 The gaps

Regulatory frameworks that could affect tax administration policy regarding AI do exist. The GDPR, the AI Act and ECHR, in particular, are developed frameworks which concentrate on ethical and privacy considerations. But a critical problem emerges, as this part shall demonstrate, whereby noteworthy gaps remain in relation to the AI issues in tax administration that we perceive to be most in need of addressing, namely, human centredness, fairness, transparency, oversight, and constraining the coercive power of the state. Whilst there are frameworks that exist as a matter of national law,[75] we focus on these frameworks as they apply to a large number of jurisdictions, and we are yet to come across a framework in domestic law which *significantly* expands upon these frameworks to tackle our core concerns.

3.1 *GDPR*

The GDPR is crafted to safeguard individual privacy and regulate personal data processing.[76] But whilst the GDPR underscores individual rights, transparency, and lawful data processing in line with broader ethical considerations, when the GDPR is applied to AI in tax administration there are a range of exemptions which have the effect of limiting its impact. Critics might flag the legal fallout from *Toeslagenaffaire* as an example of the strength of the GDPR, but a closer look reveals the GDPR's limited utility when faced with the risks we envisage with AI.

Notes

[72] *See* further Erich Kirchler, Erik Hoelzl & Ingrid Wahl, *Enforced Versus Voluntary Tax Compliance: The 'Slippery Slope' Framework*, 29(2) J. Econ. Psychol. 210 (2008), doi: 10.1016/j.joep.2007.05.004.

[73] O'Neill has received numerous awards for her work on trust. *See* for instance Holberg Prisen, *Justification of the Holberg Committee* at https://holbergprisen.no/nb/holbergprisen/prisvinnere/onora-oneill.

[74] Onora O'Neill, *Linking Trust to Trustworthiness*, 26(2) Int'l J. Phil. Stud. 293, 294 (2018), doi: 10.1080/09672559.2018.1454637.

[75] *See* for instance the Canadian Directive on Automated Decision-Making https://www.tbs-sct.canada.ca/pol/doc-eng.aspx?id=32592. Tax administrations conducting risk management decisions are exempted from the requirement for a meaningful explanation, *see* Teresa Scassa, *Administrative Law and the Governance of Automated Decision Making: A Critical Look at Canada's Directive on Automated Decision Making* 54(1) UBC L. Rev. 251, 270 (2021), doi: 10.2139/ssrn.3722192.

[76] *See* for instance GDPR, recitals 1, 2, 4, 6, and 10; Case C-645/19 *Facebook Ireland Ltd and Others v. Gegevensbeschermingsautoriteit* ECLI:EU:C:2021:483, [2021] ECR I-483, para. 45; Case C-319/20 *Meta Platforms Ireland Limited v. Bundesverband der Verbraucherzentralen und Verbraucherverbände – Verbraucherzentrale Bundesverband eV,* ECLI:EU:C:2022:322, [2022] ECR I-322, para. 73; Council of Europe, European Court of Human Rights, European Data Protection Supervisor, European Union Agency for Fundamental Rights, *Handbook on European data protection law* (Publications Office of the European Union 2018).

We do not mean to be flippant about the protection that GDPR offers to taxpayers. It certainly does sanction tax authorities for improper use of taxpayers' data. Sharing taxpayers' data to personal emails for instance, as *Belastingdienst* (the Dutch tax administration) did, will result in breach of GDPR and sanction.[77] But our argument is simply that the protection it affords is insufficient in safeguarding human centredness, fairness, transparency, oversight and constraining the coercive power of the state.

3.1.1 *Legislated Exemptions*

First, consider the way in which taxpayer data is processed. RMSs rely both on the processing of taxpayer data to determine if *that* person has underpaid their taxes or *another* person has underpaid their taxes (either through highlighting anomalies between different sources of information or through comparing risk profiles and benchmarking). Data thus is not solely collected from a taxpayer to determine *that* taxpayer's liability but to investigate other taxpayers. This could be said to be the processing of data for purposes other than the original purpose for which the data was provided. But Article 6(4)(a) allows tax authorities to do this because there is a 'link between the purposes for which the personal data have been collected and the purposes of the intended further processing'.[78]

Moving on to the data-subject rights that are enshrined in the GDPR, consider the right not to be subjected to automated decision-making enshrined in Article 22.[79] There are three cumulative conditions that must be satisfied for this prohibition to apply. First, there must be a decision by the data controller, which, secondly, is automated and, thirdly, produces legal effects or significantly affects the person. A decision will be regarded as automated either where no human is involved at all or where there is a human which ultimately took the decision but that was not a 'meaningful' intervention (as perfunctory as a tick-box exercise).[80] Tax authorities use profiling in their AI-RMSs to assist with the determination of which taxpayers to audit (decisions which certainly affect the taxpayer concerned, thereby satisfying the third condition) – but the decision to instigate an audit is generally undertaken by a human[81] and thus not fully automated.[82] But let's assume either that such a decision is fully automated, or the tax inspector involved did not meaningfully engage with the decision.[83] Even then, Article 22 provides an exemption for tax authorities, enabling them to automate audit selection decisions, where Member State law permits tax authorities to use automation for risk profiling and there are suitable safeguards to protect the data subject's rights, freedoms and legitimate interests. Thus, §88(5) of the German Fiscal Code (*Abgabenordnung*) expressly enables the German tax administration to use automated RMSs and sets out safeguards for their use.

But the exemption in Article 22(2) may be unnecessary given the more general exemption provided by Article 23(1)(e) whereby data subject rights, including the right to an explanation, but also to know (and be given meaningful information)[84] about an automated decision, to rectification,[85] to erasure[86] and so on, may be restricted where this is necessary to safeguard taxation and social security matters (provided the restriction respects the essence of the fundamental rights and freedoms and is a necessary and proportionate measure in a democratic society).[87] This is probably the exemption that tax authorities rely on in order to remain secretive about their use of AI-assisted RMSs. Tax authorities and policymakers could argue that being transparent about the operation of RMSs, including disclosing the risk factors, would jeopardize the collection of taxes as taxpayers could game the system by altering their affairs in order to achieve a low-risk rating.[88] German law expressly legislates for this fear – §88(5) of the German Fiscal Code provides that '[d]etails of the risk management systems may not be published if this could

Notes

[77] *See* WU, Leiden, Tilberg, *Antwerp, Background Paper for AI Governance and Taxpayer Rights in a Digital World* 12–13 (2024).

[78] Some countries such as Germany have specifically legislated to allow the use of taxpayer data other than for the purpose it was collected, *see* AO, §29c.

[79] *See* further Case C-634/21 *OQ v. Land Hessen* ECLI:EU:C:2023:957, para. 52.

[80] Article 29 Data Protection Working Party, *Guidelines on Automated Individual Decision-Making and Profiling for the Purposes of Regulation 2016/679*, 20 (2018a) 17/EN WP251rev.01.

[81] It is mandated in Italy that the inspector will 'autonomously' arrive at a decision whether to audit a taxpayer on the basis of a score produced by the RMS: Ministerial Decree of 26 Jun. 2020, Art. 5. *See* further: Amedeo Rizzo & Giorgio Hassan, *Addressing the Use of AI by EU Tax Authorities: Towards a Common Framework of Taxpayer Protection*, 65(1) Eur. Tax'n 15, 21–23 (2024), doi: 10.59403/1avybfj.

[82] Note that in the UK, such a decision can be undertaken solely by machine – *see* Finance Act 2020, s. 103.

[83] Further guidance on what amounts to a meaningful intervention have been provided by the cases of Case C-202/22 *CK v. Dun & Bradstreet Austria GmbH and Magistrat der Stadt Wien* ECLI:EU:C:2025:117 and Case C-634/21 *OQ*, *supra* n. 79.

[84] *See* GDPR, Art. 15(1)(h) and C-202/22 *Dun, supra* n. 83, para. 57 in particular.

[85] GDPR, Art. 16.

[86] GDPR, Art. 17.

[87] Consider the balance that the Court had to strike in the Case C-202/22 *Dun, supra* n. 83 between on the one hand the right of a data subject and third party data or trade secrets.

[88] On this, various authors argue that this opacity is unjustified. *See* Hadwick & Lan (2021), *supra* n. 63, at 627 and 643; Roman Seer, Sebastian Unger, Emran Sediqi & Philipp Wagner, *Germany*, in *Tax Transparency*, s. 23.2.5 (Funda Basaran Yavaslar & Johanna Hey eds, IBFD 2019).

jeopardise the uniformity and legality of taxation'. As summarized by Ku niacki, Almada, Tyliński, Górski, Winogradska and Zeldenrust, 'current data protection law does not establish a general duty of explanation for AI systems, even when these systems depend on personal data'.[89]

Finally, the GDPR does not apply to non-natural persons,[90] thereby leaving out huge swathes of the taxpaying population, most notably companies, but also trusts, foundations and other non-natural legal persons.[91] This shows a lack of human centredness as it overlooks the fact that humans are very much affected by how the data of their particular legal vehicle is used.

3.1.2 *Toeslagenaffaire*

Notwithstanding these exclusions which can apply in the case of tax authorities, not all provisions of the GDPR can be avoided. Fundamentally, the processing of data under the GDPR is lawful if pursued fairly and for a specified, lawful purpose (Article 5), such as the carrying out of a task in the public interest or exercising official authority (Article 6(1)(e)) such as tax collection. That a lawful purpose can be identified does not give a public authority a carte blanche to process an individual's data – instead, it must still be demonstrated that the processing of data is proportionate to the public task pursued and necessary, in the sense that there is no other less intrusive method of achieving the purpose that could reasonably be used.

To that end, *Toeslagenaffaire* provides a case study on how Articles 5 and 6 have *some* impact on how tax authorities act. The problems in *Toeslagenaffaire* principally stem from the data employed by the algorithm and the policy governing the approach to combat childcare benefit fraud.[92] Over an extended period, the Dutch tax administration utilized ML-algorithms in an effort to combat fraud.[93] It transpired that many of the selected individuals were genuinely eligible citizens seeking state benefits.[94] This misuse of algorithms led to the investigation and blacklisting of thousands of individuals.[95] The discriminatory actions were based on specific data, causing the algorithm to conclude a higher risk of fraud, including factors like dual citizenship, area code, or a foreign last name.[96] The Dutch tax administration, for example, registered the dual citizenship of almost one and a half million Dutch citizens.[97] When these practices came to light, the resulting scandal led to the resignation of the entire Dutch cabinet.[98]

The Dutch data protection authority imposed a fine of EUR 2.75 million on the Dutch tax administration for breach of the principles of lawfulness and fairness found in Articles 5(1)(a) and 6(1)(e) of GDPR (along with their corresponding provisions in Dutch law)[99] by using the erroneous fraud detection and risk classification model.[100] Whilst benefits administration is not a concern of this article, the manner in which the data was used by the Dutch tax administration is similar to how data is used in RMSs for determining which taxpayers to select for audit. As such, the lessons from the episode are relevant to determining how GDPR would apply to tax authorities using RMSs for audit selection.

The Dutch data protection authority found three infractions of the GDPR. First, processing of dual nationality data for the purpose of determining eligibility was unlawful, as only Dutch nationality was relevant to benefits' entitlement.[101] Secondly, using 'nationality' as an indicator of risk in the risk classification model was not lawful, as a less intrusive method of risk assessment could have been used (thus, it was not necessary for achieving the purpose of administering benefits).[102] For instance, the data protection authority suggested that a more finely calibrated risk indicator would be 'possesses Dutch nationality, or EU nationality with registration in a Dutch municipality, or non-EU nationality and a valid

Notes

[89] *See* Ku niacki, Almada, Tyliński, Górski, Winogradska & Zeldenrust (2022), *supra* n. 23, at 606.

[90] GDPR, Art. 4(1).

[91] *See* in this respect Esther Huiskers-Stoop, Almut Breuer & Mark Nieuweboer, *Exchange of Information, Tax Confidentiality, Privacy and Data Protection from an EU Perspective,* Erasmus L. Rev. 86, 97–99 (2022), doi: 10.5553/ELR.000225.

[92] Sjoerd Zijlstra, *Lessen uit de toeslagenaffaire. Nederlands Tijdschrift voor Bestuursrecht*, 34(3) JournalNederlands Tijdschrift voor Bestuursrecht 125 (2021), doi: 10.54195/ntm.23181.

[93] Amnesty International, *Xenophobic Machines: Discrimination Through Unregulated Use of Algorithms in the Dutch Childcare Benefits Scandal* 15–17 (25 Oct. 2021), https://www.amnesty.org/en/documents/eur35/4686/2021/en/.

[94] *Ibid.*, at 10–14.

[95] *Ibid.*, at 5.

[96] *Ibid.*, at 31–33.

[97] Corien Prins, *Discriminerende algoritmes* (2021) 96(20) Nederlands Juristenblad 1454.

[98] Zijlstra (2021), *supra* n. 92.

[99] Wet bescherming persoonsgegevens (Personal Data Protection Act), Arts 6 and 8.

[100] Authoriteit Persoonsgegevens, *Boete Belastingdienst kinderopvangtoeslag* (2021).

[101] *Ibid.*, paras 8 and 40–44.

[102] *Ibid.*, paras 9 and 45–50.

residence permit'.[103] The use of nationality as a risk indicator also breached the GDPR safeguard on fairness, as this unjustly distinguished individuals based on nationality without an objective justification.[104] Again, a risk indicator that did not unlawfully discriminate would be: 'possess[es] Dutch nationality, or an EU nationality with registration in a Dutch municipality, or a non-EU nationality and a valid residence permit'.[105] Thirdly, processing of nationality for the detection of organized fraud was unlawful as it was not necessary for the purpose of detecting fraud.[106] Rather than mapping *all* nationalities, as the tax authority had done, for the purpose of detecting fraud, a less intrusive means of doing so would have been to focus only on *certain* nationalities *if* a substantial increase in fraud could be identified.[107] It was not proportionate because it was not clear there was any link in fact between likelihood of organized fraud and nationality.[108] The use of nationality as an indicator of organized fraud was also 'unfair' because it was discriminatory.[109] The heightened scrutiny afforded to individuals based on non-Dutch nationality was not driven by, nor correlated with, actual evidence of a heightened risk.[110]

This episode, however, does not showcase the strength of the GDPR in regulating the use of AI assisted RMSs. If we take a step back, we see that what actually occurred is that the GDPR was engaged because the RMS, firstly, was highly inaccurate (creating as many as 94% false positives[111]) and hence the processing was not necessary and proportionate, and secondly, discriminatory because it disproportionately selected people for audit on the basis of nationality (had the system not been so inaccurate then it could have been plausible to argue it was actual tax non compliance risk which droe the investigations as opposed to nationality). The episode is extreme and an outlier. RMSs are not contrary to GDPR nor does the Regulation insist that they are accurate, but rather the episode highlights that *highly* inaccurate and discriminatory systems are contrary to GDPR. Indeed, RMSs continue to be used in the Netherlands![112]

To recall, we suggest that a framework governing the use of AI by tax authorities should emphasize human centredness, fairness, transparency, oversight and constraining coercive power. By prohibiting highly inaccurate and discriminatory AI systems, the GDPR does regulate basic fairness and does, to an extent, impose oversight requirements (in that the data used must not produce highly inaccurate predictions of non-compliance). But it says little about human centredness and how this would emphasize common sense, choice and inclusion (for instance through mandating the inclusion of humans in the development of AI, or in promoting the use of AI to advance human understanding of tax rules). The GDPR very clearly, through its exemptions, explicitly does not promote transparency in the use of AI systems by tax authorities. Finally, it has nothing to say more generally about how accurate an RMS must be (as opposed to how *inaccurate* it may not be) and as such does not reign in the coercive power of the state by insisting for instance that RMSs produce as few false positives as possible. Moreover, RMSs can seriously intrude, lawfully, into the personal sphere by using highly sensitive personal data, such as sex and race, so long as proportionality and necessity are demonstrated.[113]

3.2 *AI ACT*

The AI Act, while often compared to the GDPR, diverges significantly in its scope and objectives. Whereas the GDPR primarily focuses on safeguarding personal data, the AI Act is centred around regulating the safe and responsible use of AI systems. Rooted in Article 114 of the Treaty on the Functioning of the European Union, the AI Act places a strong emphasis on product safety and the societal impact of AI technologies. Its focus is on the use of AI products in various domains, classified according to societal risk. For the purposes of the regulation, 'risk' means the combination of the probability of an occurrence of harm and the severity of that harm (Article 3(2)). The regulation specifies which types of systems are considered to pose unacceptable risk (Article 5), and in which domains the use of certain AI systems is considered to be 'high-risk' (Article 6).[114]

Notes

[103] *Ibid.*, para. 48.

[104] *Ibid.*, paras 11 and 64–71.

[105] *Ibid.*, para. 69.

[106] *Ibid.*, paras 10 and 51–63.

[107] *Ibid.*, para. 59.

[108] *Ibid.*, paras 60–61.

[109] *Ibid.*, paras 11, 64–68 and 72–76.

[110] *Ibid.*, paras 73–76 in particular.

[111] Tweede Kamer der Staten-Generaal, Eindverslag Parlementaire ondervragingscomissie Kinderopvantoeslag, *Ongekend Onrecht* 22–23 (17 Dec. 2020) cited in David Hadwick, *'Error 404 – Match not found': Tax Enforcement and Law Enforcement in the EU Artificial Intelligence Act*, 1 EU Crim 55, 57 (2023), doi: 10.30709/eucrim-2023-005.

[112] Anne-Marie Noordenbos, *Kabinetsreactie artikel 'Belastingdienst blijft wet overtreden met mogelijk discriminerende fraude-algoritme* (Taxence 18 Mar. 2024), https://www.taxence.nl/nieuws/kabinetsreactie-artikel-belastingdienst-blijft-wet-overtreden-met-mogelijk-discriminerende-fraude-algoritmen/.

[113] *See* GDPR, Art. 9(2)(g).

[114] *See* further European Commission, *Commission Guidelines on Prohibited Artificial Intelligence Practices Established by Regulation (EU) 2024/1689 (AI Act)* (4 Feb. 2025), https://digital-strategy.ec.europa.eu/en/library/commission-publishes-guidelines-prohibited-artificial-intelligence-ai-practices-defined-ai-act.

While the obligations around systems classified as low-risk are limited transparency obligations, the European Commission can make amendments to Annex III (per Article7(1)) to expand the high-risk category. This makes it advisable for both developers and deployers of low-risk systems to anticipate such possible reclassification and actively adopt ethical considerations even if the AI Act (currently) does not require it. Where a system is regarded as high-risk, more stringent obligations apply (Articles 8–17) in relation for instance to managing the risks the system poses, data governance, and oversight.

As it stands, however, the use of AI systems by tax authorities in their performance of administrative tasks, such as in choosing taxpayers to audit, is not regarded as high risk. As explained in Recital 59, AI systems specifically intended to be used for administrative proceedings by tax and customs authorities should not be classified as high-risk AI systems.[115] The use of AI by tax authorities will only be regarded as high risk if used in relation to criminal investigations.[116] Technically, an AI-RMS may detect tax non-compliance which ultimately leads to a criminal prosecution. But that does not seem to be what the recital is driving at. The pertinent sentence in Recital 59 speaks of 'profiling in the course of detection, investigation or prosecution of criminal offences'. Whilst RMSs *do* profile taxpayers, it is only in the course of detecting tax non-compliance. The purpose is not to detect a criminal offence for the simple reason that tax risk profiling *cannot* detect a criminal offence. As Hadwick has astutely pointed out, an RMS will only detect the potential act of underpayment (the 'actus reus'), whereas to determine if a tax crime has taken place, 'intention' (or 'mens rea') is an additional and necessary condition.[117] An RMS simply does not care about intent – its job is merely to flag potential underpayment – and intent can only be established *after* a taxpayer has been investigated.

The alternative interpretation is that an AI-RMS as used by the tax authority *is* high-risk because it cannot make any distinction *ex ante* between detection of non-compliance which is or is not of a criminal nature – any time an RMS is used there is a chance that it may detect non-compliance, which upon later investigation, turns out to be criminal. Because of this inability to distinguish actions *ex ante*, the RMS cannot be partly high-risk: it either is or is not high-risk in its entirety. But the interpretation that a tax RMS is high-risk cannot hold given the very explicit explanation in the Recital that AI systems specifically intended to be used for administrative proceedings by tax and customs authorities should not be classified as high-risk AI systems.

3.3 *ECHR*

The ECHR provides a framework for protecting human rights in various contexts. However, its application to AI in tax administration encounters challenges similar to those found with the GDPR and AI Act. The Convention outlines broad principles such as the right to privacy (Article 8), prohibition of discrimination (Article 14), and the right to an effective remedy (Article 13). The Convention's broad strokes are not finely tuned to the detailed processes and potential biases introduced by AI algorithms in the tax context. This is unsurprising given that the Convention dictates minimum standards for jurisdictions[118] – leaving it to those jurisdictions to determine if they wish to impose more stringent restrictions on public power. Keen observers of this area might counter that the Syri[119] and eKasa[120] cases are examples of the ECHR's strength in regulating the use of AI in tax administration. But we would disagree with this conclusion. In this section, we will explain that the technologies *could* have been made ECHR compliant, but there would still be ethical failings. The UK is also used as a further case study to support this argument.

3.3.1 *Syri, eKasa and the UK*

SyRI was an ML-based decision-support tool used by several Dutch public authorities to help detect tax and social security fraud.[121] SyRI used data from a wide range of government agencies and looked to identify inconsistencies, anomalies and asymmetries in the data in order to build a risk profile for individuals.[122] It was passed into law in 2014[123] and was in use until it was declared unlawful by the Hague District Court (which, in the Dutch judicial system, is two tiers below the Supreme Court) in 2020.[124] The Court found that the legislation

Notes

[115] *See also* Peeters, *supra* n. 58, at 56–57. Though note that where a tax administration has been entrusted with administering welfare benefits, not considered in this article, an AI system used for this purpose *would* be considered high risk: AI Act, recital 58; Hadwick (2024), *supra* n. 34, at 953–954.

[116] *See* AI Act, Annex III, point 6.

[117] *See* Hadwick (2023), *supra* n. 111.

[118] For instance, Eva Brems, *Human Rights: Minimum and Maximum Perspectives,* 9 Hum. Rts. L. Rev. 349 (2009), doi: 10.1093/hrlr/ngp016.

[119] *NJCM et al. v. The Dutch State* (2020) The Hague District Court ECLI: NL: RBDHA:2020:1878.

[120] Finding of the Constitutional Court of the Slovak Republic no. k. PL. ÚS 25/2019-117 of 10 Nov. 2021.

[121] Yeung & Harkens (2023), *supra* n. 28, at 453.

[122] *NJCM* (2020) *supra* n. 119, paras 6.45–6.54.

[123] *See* Wet structuur uitvoeringsorganisatie werk en incomen, Arts 64 and 65.

[124] Joanna Redden, Jessica Brand, Ina Sander & Harry Warne, *Automating Public Services: Learning from Cancelled Systems* 23 (Data Justice Lab 21 Sep. 2022).

infringed the right to private and family life as enshrined in Article 8 of the ECHR and that such infringement was not 'necessary in a democratic society' because it was disproportionate to the (legitimate) aims pursued.[125] This was due to lack of specificity and safeguards. As to the former, the SyRI legislation did not make clear the causal link between the data used by SyRI and the level of risk which was then attributed to the data subject.[126] As for the latter, the SyRI legislation was silent on the risk model used and on the risk analysis method used.[127] By consequence, data subjects were unable to defend themselves 'against the fact that a risk report ha[d] been submitted' about them.[128] Further, individuals not subjected to a risk report would have no way of knowing if their data had been 'processed on correct grounds'.[129] The silence also meant it was impossible to verify if the system produced discriminatory effects.[130]

eKasa is a system which transmits information from electronic cash registers to the Slovakian tax authority. That information is also processed by a risk-scoring algorithm, in turn used to determine which taxpayers to audit.[131] The Slovakian Supreme Court found the use of eKasa to be unlawful. The Court accepted that there was a basis in law for the provision of information to the tax authority. But the Court found that this law did not permit the *processing* of that data through *algorithmic means.* The processing interfered with the constitutional right to informational self-determination found in Articles 19(3) and 22(1) of the Slovakian constitution.[132] Though the reasoning was grounded principally in the requirements of Slovakian law,[133] the Court opined that the ECHR supported this assessment.[134] The critical point was that legislation permitting an interference with these rights must 'clearly provide the citizen with an appropriate indication of the circumstances and conditions under which the public authority is authorized to encroach on his right'.[135] But the manner in which the system operated was not known to taxpayers as the legislation left open how taxpayers' data would be processed.[136]

We disagree with the contention that these instances show the power of the ECHR in regulating the use of AI in tax administration because it was entirely possible for the Dutch and Slovakian legislatures to have amended the law to render the systems ECHR compliant. All that would have been required was for the legislation to make clear how the taxpayers' data was used and how the systems operated, as well as having in place a method for checking for discriminatory effects. But whilst the ECHR helped ensure fairness and basic transparency in these cases, it does not satisfy many of the demands that we consider important. First, the ECHR pays no regard to the need for human centredness in respect of common sense, choice and inclusion. Secondly, neither in the Netherlands nor in Slovakia was meaningful oversight regarded as necessary for ECHR compliance (rather, it was only required by the Slovakian constitution).[137] The focus was on basic transparency. Thirdly, the ECHR does not insist that AI systems choose an appropriate error threshold – rather the focus is only on stopping an AI system from discriminating. Countering discrimination tackles the coercive power of the state to an extent, but still leaves open shifting power to tax authorities through policy decisions (like shifting the burden of proof to taxpayers on the basis of *prima facie* evidence of underpayment) and the substantial (but indiscriminate) auditing of innocent taxpayers.[138]

The ECHR in reality only marginally constrains tax authorities in their use of AI. This point is brought out particularly by looking at the use of automation and AI by

Notes

[125] *NJCM* (2020) *supra* n. 119, para. 6.72.

[126] *Ibid.*, para. 6.87.

[127] *Ibid.*, para. 6.89.

[128] *Ibid.*, para. 6.90.

[129] *Ibid.*, para. 6.90.

[130] *Ibid.*, paras 6.91 and 6.93.

[131] Slovakian Constitutional Court (2021), *supra* n. 120, paras 118 and 130.

[132] *Ibid.*, para. 120.

[133] *Ibid.*, para. 148.

[134] *Ibid.*, para. 148.

[135] *Ibid.*, para. 122.

[136] *Ibid.*, paras 123–124.

[137] *Ibid.*, para. 148.

[138] Space does not allow us to elaborate precisely upon how the ECHR does or does not regulate RMSs. We examine only the cases that *have* arisen here, but a theoretical argument could be founded on 'status' protection under the ECHR. Tax authorities must not discriminate against a taxpayer on the basis of that taxpayer's status. This would look beyond race or nationality for instance and look at sub groupings, such as profession for instance. So a group of taxpayers from a particular profession that are selected for audit by an RMS could claim that they have been discriminated against because of their membership of a particular profession. There are two challenges for such persons. The first is convincing a court that the group does indeed qualify for status protection under the ECHR. The second is convincing a court that discriminating against that group is unjustified. A run through the case law at the level of the European Court of Human Rights would reveal that both elements of such an argument would struggle. For further reading, *see* European Court of Human Rights, *Guide on Article 14 of the European Convention on Human Rights and on Article 1 of Protocol No. 12 to the Convention* 38–48 (31 Aug. 2024).

HMRC, which is not strictly controlled by legislation. Since 2010,[139] HMRC has used a system known as CONNECT, which is a data matching and risk profiling tool.[140] As of 2023, it held approximately fifty-five billion data items[141] and apparently prompts 90% of HMRC investigations.[142] This tool engages in web-scraping and the processing of data for the purpose not just for detecting unreported or under-reported earnings, but also for the purpose of performing population wide risk assessments. However, HMRC's legal power to use this technology does not derive from any specific legislative mandate. Instead, the legal power to do so derives from HMRC's general managerial discretion.[143]

In terms of automation, the UK enacted section 103 of the Finance Act 2020, which provides that *any* responsibility imposed on HMRC can be executed by a computer, including the commencement of an audit.[144] This updates provisions which previously referred to tasks assigned to an 'officer' of the tax authority.[145] But this provision did not precede the commencement of automated operations by HMRC. What prompted the legislative change was that there had been conflicting tribunal determinations on the automation of certain HMRC processes. One tribunal determined that HMRC could use automation on the basis of a purposive interpretation of the word 'officer' in the relevant legislation.[146] In other words, that an 'officer' need not be a human. Another tribunal came to the opposite conclusion and determined that HMRC *could not* use automation unless allowed expressly by legislation. By the time section 103 had actually been introduced, the matter had been resolved. The Upper Tribunal in *HMRC v. Rogers and Shaw*[147] found that the old legislation did indeed permit the taking of decisions by computers despite the legislation not specifically providing for this.[148] In short, the Tribunal had found that HMRC did not need specific authorization to use automation, as such power was said to already exist as inferred from a purposive interpretation of its powers which existed *before* largescale automation was even possible.

How HMRC checks to ensure the accuracy of its RMS is entirely a matter for HMRC. It is not compelled to routinely test CONNECT, nor are there any instructions on how such testing should work. Meanwhile, with automation, it is envisaged that highly consequential decisions can be taken, such as the commencement of audits into a taxpayer's affairs up to twenty years in the past.[149]

We would suggest that despite all this – the fact that HMRC was not regarded as requiring specific legislative authorization to use automation, that it does not have specific legislation permitting the use of AI in its RMS, that there are not requirements regarding testing of CONNECT, that taxpayers are not told that they have been flagged for an audit by an AI RMS, that taxpayers cannot know if the system is justified in arriving at the results that it does – HMRC's use of CONNECT does not breach Article 8. CONNECT serves a sufficiently compelling and legitimate purpose. The Hague Court in *Syri* similarly found that the AI RMS there also served a legitimate purpose[150] which was sufficiently compelling to justify interference.[151] In that case it was combatting tax fraud, given the scale in fiscal terms of the problem. HMRC could equally claim that reducing the tax gap is also a legitimate, sufficiently compelling purpose to which CONNECT can be put, thus securing the payment of taxes which are necessary for economic wellbeing. The lack of specificity and safeguards in the legislation meanwhile does not *per se* render its operation disproportionate to the objective. As the Hague Court noted in *Syri*, the determination of what is 'necessary in a democratic society' is subject to a margin of appreciation.[152] For Council of Europe Member States, there is a range of legitimate interpretations as to how public authorities should be allowed to manage pressing social issues. The UK would presumably point to the reduction in the tax

Notes

[139] Angharad Carrick, *Taxman is Snooping on Emails and Social Media – and Now Holds 55 BILLION Items of Our Data on Its AI System in a Bid to Tackle Tax Evasion* (This is Money 16 May 2023), https://www.thisismoney.co.uk/money/bills/article-12089927/HMRC-Connect-holds-55bn-data-items-bid-tackle-tax-evasion.html.

[140] Letter From the DWP Permanent Secretary Sir Robert Devereux & HMRC Chief Executive Jon Thompson to Meg Hillier, para. 8 (10 Jun. 2016).

[141] Carrick, *supra* n. 139.

[142] Jay Sanghrajka, *Degrees of Connection*, 186(4752) Tax'n 8 (2020), doi: 10.1061/40826(186)54.

[143] Commissioners for Revenue and Customs Act 2005, ss 5 and 9.

[144] *See* Finance Act 2020, s. 103(2).

[145] *See* Finance Act 2020, s. 103(1).

[146] HMRC, *Automated Decisions: Technical note October 2019*, para. 1.2 (2019), https://www.gov.uk/government/publications/securing-the-tax-base-affirming-the-legislative-framework-for-hmrc-to-use-automated-processes.

[147] *HMRC v. Rogers and Shaw* [2019] UKUT 406 (TCC), [2020] 4 W.L.R. 23.

[148] *Ibid.*, para. 32.

[149] *See* Finance Act 2020, s. 103(2) combined with Taxes Management Act 1970, ss 30A, 29 and 36(1A).

[150] *NJCM* (2020) *supra* n. 119, para. 6.4.

[151] *Ibid.*, para. 6.74.

[152] *Ibid.*, para. 6.73.

gap owing to the use of CONNECT, therefore showing the connection between the objective and the means, and argue that there are internal safeguards to ensure that the system operates appropriately. Presumably this is the same reasoning that many other European countries would take – very few tax authorities are given a specific mandate to use AI nor ensure that it is operating appropriately.[153]

3.4 *Concluding Thoughts on Existing Frameworks*

While the GDPR, AI Act and ECHR establish foundations for ethical AI, their application to AI in tax administration reveals gaps that necessitate addressing. None, as a result, properly address human centredness, fairness, transparency, oversight, and constraining the coercive power of the state. A nuanced regulatory approach is needed, one that recognizes the specific challenges presented by AI-driven decision-making in tax processes.

4 BEYOND LEGAL MANDATES: GUIDELINES FOR THE USE OF AI IN TAX ADMINISTRATION

It is important that tax authorities do not fall foul of existing legal frameworks regulating the use of AI in tax administration. Nevertheless, they are far from comprehensive in satisfying the demands that *we would expect* from a framework governing the use of AI by tax authorities. The purpose of this part is to set out such a framework.

Before setting out our proposed framework, there are several matters which we regard as important to the human centredness and oversight challenges that we include in the framework but have not had the opportunity to articulate yet in this article. We take that opportunity now.

4.1 *Human Centredness*

4.1.1 *Stakeholders*

Our demands for 'human centredness' would encapsulate common sense, choice and inclusion. Earlier in this article we highlighted the need to have 'human' values and knowledge translated into computational terms, thus underscoring the need, when developing, operating and overseeing AI, to include humans coming from both the public and private sector.[154]

Politicians and policymakers play a role in delineating the operational boundaries of tax administrations. Their decisions and policies influence the framework governing AI and will eventually determine the power of this tool for tax administration and how it should be checked. Thus, the political task is to leverage AI for efficiency, while also ensuring alignment with societal values and other important norms. Technologists meanwhile should have an important voice given that they will be tasked with translating these values into operable systems. Citizens and businesses are also integral components of the tax ecosystem as they are the parties that will ultimately be subjected to the application of AI in tax administration – providing them with opportunities to be included in the development process and to exercise oversight can help with the determination as to whether AI is acting appropriately or inappropriately; in other words, in determining whether AI is assisting with taxpayer needs, producing justifiable results or operating in line with the values and norms that the community would consider important. If AI does not, then it is not helping tax authorities fulfil their primary function of collecting taxes due in a competent manner.

Genuine consultation engenders transparency and can ultimately have a bearing on trust between tax authorities and taxpayers. This sort of transparency generates evidence that the tax authorities are using AI in a trustworthy manner, insofar as taxpayers can see that tax authorities using AI for instance to competently fulfil their primary function of collecting taxes due. Thus, rather than perceiving transparency as an added burden, tax administrations should recognize it as an opportunity.

4.1.2 *Explainability*

AI could be harnessed to empower taxpayers and increase taxpayer choice. For instance, as it currently stands, there is no meaningful explanation provided to taxpayers by tax authorities as to why those taxpayers have been chosen for audit. But with AI today, there is an opportunity to provide taxpayers with such explanations. Providing clear explanations of how AI algorithms contribute to tax assessments and outlining the rights of taxpayers when their data is processed also has secondary benefits for tax administration through encouraging compliance. As per Kirchler, Hoelzl and Wahl, tax compliance is a function of power and trust – power (coercive methods of collecting taxes) is required for some taxpayers, building a trusting relationship works for others (for instance through the provision of support, services and information).[155]

Notes

[153] *See* Taxadmin.AI, *AI for Revenue Agencies – EU Overview*, https://taxadmin.ai/ai-for-revenue-agencies-eu-summary/.

[154] *See also* WU, Leiden, Tilburg, *supra* n. 77, at 20–21.

[155] *See* further Kirchler, Hoelzl & Wahl, *supra* n. 72.

A modern method to empower taxpayers and increase choice is Explainable AI (XAI), which refers to methods aimed at making the decision-making processes of AI systems more interpretable to humans. While the concept is not new, with early work on rule-based expert systems dating back several decades, it has gained renewed attention in the context of modern deep learning models. Traditional AI systems, such as decision trees, are usually interpretable (though depending on their length, even these can become opaque), as their decision paths can be traced and understood. However, this is not the case for DNNs, where XAI potentially offers a solution. There arguably exists a trade-off between model accuracy and explainability[156]: highly accurate models, such as DNNs, often lack transparency, while simpler models like decision trees offer greater interpretability, but at the cost of predictive power. XAI takes on this challenge, addressing the tension between the inherent opacity of ML-models on the one hand and the need for oversight on the other. It does so through two main approaches: 'global' and 'local' explanations. Global explanations consider broadly how black box ML-systems operate, for instance how 'input' factors, in general, affect outputs.[157] Local explanations try to establish how a system arrived at a particular decision, for instance by testing the system against counterfactuals in order to narrow down how the system arrived at a particular outcome.[158] In a recent study of XAI in the tax administration context, Górski et al., examined a fraud detection model developed in conjunction with the Buenos Aires tax authority and evaluated the ability of different XAI methods to explain the outcomes from the model.[159] They concluded that whilst work is still to be done to ensure that these methods meet the minimum standard required for use in tax decisions, a combination of different methods can be used to develop an XAI model to the requisite standard.[160]

Critics might say that post-hoc explanations of XAI are insufficient as they do not reflect the real reasoning behind an AI generated decision. This is true. But the challenge is misconceived.[161] Without AI, we have humans that make decisions. But we also do not know, truly, how humans arrive at the decisions they do. Black box AI systems operate just like neural networks in the brain. Where we consider there to be potential impropriety in the *human* decision-making process, we rely upon post-hoc evidence gathering techniques: cross-examination in the court and analysis of contemporaneous written or oral evidence for instance. These techniques strive to do the same thing that XAI does – they seek to logically determine what must (on the balance of probabilities, or other evidentiary threshold) have influenced the decision-making process. XAI then has the capability to be no worse than the techniques we currently rely on to understand the decision-making process. In short, whilst XAI techniques are a form of reverse engineering, so too, in truth, are the techniques we use in law to understand human reasoning.

4.2 *Oversight: Testing*

The integration of AI into tax administration requires a robust framework in order to systematically and continuously assess its effectiveness and influence. Incorporating regular audits and performance evaluations into the policy framework, ensures continuous optimization of the AI tools. For instance, consider the introduction of a predictive analytics tool aimed at optimizing tax compliance. Periodic assessments not only validate its alignment with rapidly evolving tax laws, but also provide insights for necessary adaptations, enhancing the overall efficacy of the tool. Thorough testing, evaluation, validation of the system and datasets, and a commitment to adaptation, should not only precede full-scale adoption. They should also be applied throughout the system lifecycle.

Cross-disciplinary teams, equipped with expertise in both technology and law, must play a critical role.[162] For instance, in the implementation of AI algorithms for risk assessment in tax audits, cross-disciplinary managers should provide oversight, ensuring adherence to legal standards while leveraging their insights to enhance the efficiency and fairness of the algorithm.

Notes

[156] Andrew Bell, Ian Solano-Kamaiko, Oded Nov & Julia Stoyanovich, *It's Just Not That Simple: An Empirical Study of the Accuracy-Explainability Trade-off in Machine Learning for Public Policy*, ACM Conference Fairness, Accountability, & Transparency (FAccT '22) (Seoul, Republic of Korea 21–24, Jun. 2022), doi: 10.1145/3531146.3533090.

[157] *See* David Rumelhart, James McClelland and the PDP Research Group, *Parallel Distributed Processing: Explorations in the Microstructures of Cognition, Volume 1: Foundations* (MIT Press 1986) on parallel distributed processing.

[158] *See* Christopher K. Wikle, Abhirup Datta, Bhava Vyasa Hari, Edward L. Boone, Indranil Sahoo, Indulekha Kavila, Stefano Castruccio, Susan J. Simmons, Wesley S. Burr & Won Chang, *An Illustration of Model Agnostic Explainability Methods Applied to Environmental Data*, 34 Environmetrics 1 (2023), doi: 10.1002/env.2772.

[159] Łukasz Górski, Bła ej Ku niacki, Marco Almada, Kamil Tyliński, Madalena Calvo, Pablo Matias Asnaghi, Luciano Almada, Hilario Iñiguez, Fernando Rubianes, Octavio Pera & Juan Ignacio Nigrelli, *Exploring Explainable AI in the Tax Domain*, Artificial Intelligence & L. (2024), doi: 10.1007/s10506-024-09395-w.

[160] On the work still to be done between current XAI methods and post hoc interpretability, *see* more generally Brendt Mittelstadt, Chris Russell & Sandra Wachter, *Explaining Explanations in AI, Proceedings of FAT*'19: Conference on Fairness, Accountability, and Transparency,* Conference on Fairness, Accountability, and Transparency. ACM, 279 (2019), doi: 10.1145/3287560.3287574.

[161] *See also* Daly (2024), *supra* n. 66, at 447.

[162] Carlos Silvani and Katherine Baer, *Designing a Tax Administration Reform Strategy: Experiences and Guidelines* 16 (1 Mar. 1997), https://www.imf.org/en/Publications/WP/Issues/2016/12/30/Designing-a-Tax-Administration-Reform-Strategy-Experiences-and-Guidelines-2134.

4.3 *Guidelines*

Our claim is that a basic ethical framework that should guide tax authorities when using AI would emphasize human centredness, fairness, transparency, oversight and constraining the coercive power of the state. We suggest that, in the first instance, the framework should be a set of guidelines as opposed to legal rules. This is for three reasons. First, some of the demands do not easily translate into enforceable legal obligations, such as encouraging tax authorities to positively embrace AI in order to uphold the Rule of Law. Secondly, our framework is more comprehensive than the existing legal frameworks that we have analysed, meaning that, by consequence, *some* of the demands of our framework are already legally enforceable (in relation to fairness, imposed by both the ECHR and GDPR, (basic) transparency imposed by the ECHR, and (basic) oversight imposed by the GDPR). Thirdly, we do not think our set of guidelines should be the final say and as technology develops, this framework should be capable of becoming more sophisticated and more granular. As a result, we offer it here as a starting point for a broader discussion and think it would be premature to legally enshrine our demands as a result.

Many of these demands, as already noted, overlap with each other. However, for the purposes of annotation and simplicity here, we have used just one label (even though it would be more accurate to use several). Here are our guidelines:

- **Human centredness**: Tax authorities must engage stakeholders, including politicians, technologists, citizens, and businesses, to ensure alignment with societal values, important norms and legal standards, while also considering the practical implications of AI integration. Additionally, over time, this process will establish precedents clarifying the boundaries for tax administration in AI implementation.
- **Human centredness**: Ensure that every decision made through automation and AI is understandable to the individual affected by it. To clearly explain these decisions is crucial for human choice, as well as holding tax authorities to account through relevant legal appeal avenues. XAI presents an opportunity for taxpayers to be better informed about tax authority decision-making than is currently the case.
- **Human centredness**: Empowering taxpayers requires tax authorities to use their available technologies to actively assist taxpayers in complying with their tax obligations, such as through the use of chatbots.
- **Fairness**: There needs to be continuous monitoring for any potential biases or assumptions that may arise from overreliance on automated systems, and proactive measures must be taken to mitigate these risks.
- **Fairness**: Always classify the AI and assess its associated risk of bias and the harm that could potentially be caused (for instance because of the system's ability to operate at scale). Recognize that not all AI applications demand equal levels of preparation and safeguarding. By accurately classifying the AI and delineating its specific task to ascertain the requisite level of care needed to guard against bias.
- **Transparency**: To properly understand its automated decisions, the tax administration must ensure that the system itself, as opposed to merely the decisions it arrives at, is transparent. This entails visually projecting the key factors influencing each decision, allowing tax officials and taxpayers to understand the rationale behind the decision, and enabling them to scrutinize it effectively, and ask, if necessary, an independent legal adjudicator to determine if the tax authority has followed and correctly applied the law.
- **Oversight**: There should always be human oversight integrated into the system, not only during its implementation but also throughout its usage. This includes ongoing monitoring of the system's functionality and outcomes to ensure accountability, transparency, and the ability to intervene if necessary. In contested automated decisions or instances where an evaluative human judgment would previously have been required, a second opinion, based on human judgment, becomes necessary to address difficult cases that may not be suitable for AI processing alone. Cases of inequality require special consideration (with access to an independent adjudicator available).
- **Oversight**: A continuous and periodic testing process, coupled with data collection on the functionality, effectiveness, and outcomes of the AI must be in place. This ensures that the AI undergoes thorough evaluation and can be appropriately adjusted as necessary to maintain proper performance as well as being compliant with the relevant legal standards.
- **Constraining the coercive power of the state**: Whilst false positives, and thus the auditing of innocent taxpayers, are unavoidable in tax audits, prior to implementing AI technologies, it is important to assess compliance with applicable laws and regulations (including the GDPR and ECHR). This minimizes the risk of crossing legal boundaries when exercising tax administration powers, ensuring adherence to the principle of legality.
- **Constraining the coercive power of the state**: Beyond simply ensuring compliance with existing legal frameworks, an appropriate error threshold should be adopted for predictive technologies which ensures the least incursion practicable into the private lives of taxpayers. With respect to audit selection, false positives are acceptable, but care needs to be taken to ensure that the error rate is as low as feasible. As the error rate follows from modelling, routine testing of the model, and updating, if necessary, is essential.
- **Constraining the coercive power of the state**: AI systems should not be used to mask policy decisions which shift power from taxpayers to the state. Care needs to be taken as to how AI systems are integrated

into the legal system. AI systems do (and will, given their probabilistic nature) err and so the legal system should not attach significant consequences to AI decisions alone. Rather, where an administrator's decision is supported by an AI system, that administrator should be required to fully understand and explain in intelligible reasons why a negative decision should be imposed on that taxpayer. Meanwhile, the consequences of errors on the part of taxpayer assistants, such as chatbots, should not be imposed on innocent taxpayers.

5 Conclusion

In this article, we aimed to take a first step towards developing a practical and implementable set of guidelines for tax administrations on their journey to integrating AI into their processes. We highlighted why AI, due to its ability to operate not solely based on predefined rules but by learning from past data, offers a unique advantage in the automation of tax administration. Given the data-intensive nature of tax administrations, this technology is an ideal match – the vast amounts of available tax data can be used effectively to train AI systems.

However, there are challenges inherent in using AI systems as regards human common sense, fairness, transparency, oversight, determining an appropriate error threshold and hallucinations. In a tax administration context, these challenges must be understood in light of the demands of the Rule of Law, which is foundational in any account of what tax authorities can and *ought* to do. A thicker Rule of Law will aim to constrain the state's coercive power, while a thinner account will focus more on reducing non-compliance with the tax code and ensuring that taxpayers are adequately informed. A thinner account is insufficient in regulating AI use by tax administration as it has little to say about the imperative to constrain the coercive power of the state. Thus, we regard as central to any framework regulating the use of AI by tax authorities the need for human centredness, fairness, transparency, oversight, and controlling the coercive power of the state. Existing legal frameworks go some of the way to satisfying these demands, but do not go far enough. Meanwhile, incidentally, satisfying these demands will also provide evidence that tax authorities are trustworthy.

To that end, we proposed a foundational set of guidelines for the future implementation of AI systems within tax administration. But this point cannot be overemphasized: the demands of human-centredness, fairness, oversight, transparency and countering the coercive power of the state serve as the starting point of our ethical discussion to define the boundaries for AI implementation and should not be compromised. However, building upon these guidelines in the future would be desirable. This is important, because we must not underestimate the significance of tax systems as a vital thread in the fabric of our society. Therefore, substantive changes to tax administration must be carefully assessed, especially when these changes involve unforeseen technologies that are novel to developers, implementers, users, and those subject to the technology. We are very much at the beginning.

6 List of bibliography

Literature

T. R. S. Allan, *Law, Liberty, and Justice: The Legal Foundations of British Constitutionalism* (Oxford University Press 1994).

A. Bell, I. Solano-Kamaiko, O. Nov & J. Stoyanovich, *It's Just Not That Simple: An Empirical Study of the Accuracy-Explainability Trade-off in Machine Learning for Public Policy*, ACM Conference Fairness, Accountability & Transparency (FAccT '22) 248–266 (Seoul, Republic of Korea 21–24, Jun. 2022), doi: 10.1145/3531146.3533090.

A. Biswas, S. Barman, A. Deshpande & A. Sharma, *Quantifying Infra-Marginality and Its Trade-off with Group Fairness* 1–26 (arXIV 3 Sep. 2019), https://arxiv.org/abs/1909.00982.

R. Binns, *Human Judgment in Algorithmic Loops: Individual Justice and Automated Decision-Making*, 16 Reg. & Governance 197–211 (2022), doi: 10.1111/rego.12358.

E. Brems, *Human Rights: Minimum and Maximum Perspectives,* 9 Hum. Rts. L. Rev. 349–372 (2009), doi: 10.1093/hrlr/ngp016.

R. Brownsword & H. Somsen, *Law, Innovation and Technology: Before We Fast Forward – a Forum for Debate*, 1(1) L. Innovation & Tech. 1–73 (2009), doi: 10.1080/17579961.2009.11428364.

J. Burrell, *How the Machine 'Thinks': Understanding Opacity in Machine Learning Algorithms* 3(1) Big Data & Soc'y 1–12 (2016), doi: 10.1177/2053951715622512.

J. Cohen, *Between Truth and Power: The Legal Constructions of Informational Capitalism* (Oxford University Press 2019).

S. Daly, *Artificial Intelligence, the Rule of Law and Public Administration: The Case of Taxation*, 83(3) Cambridge L. J. 437–464 (2024), doi: 10.1017/S0008197324000448.

S. Daly, *Tax Authority Advice and the Public* (Hart Publishing 2020).

F. Dechesne, V. Dignum, L. Zardiashvili & J. Bieger, *AI Ethics for Law Enforcement: A Study into Requirements for Responsible Use of AI at the Dutch Police*, 2(4) Delphi – Interdisc. Rev. Emerging Technologies 179–185 (2019), doi: 10.21552/delphi/2019/4/7.

R. Dwivedi, D. Dave, H. Naik, S. Singhal, R. Omer, P. Patel, B. Qian, Z. Wen, T. Shah, G. Morgan & R. Ranjan, *Explainable AI (XAI): Core Ideas, Techniques, and Solutions. ACM Computing Surveys*, 55(9), ACM Computing Surveys Arts 1–33 (2023), doi: 10.1145/3561048;

R. Dworkin, *A Matter of Principle* (Harvard University Press 1985).

L. Fuller, *The Morality of Law* 39 (Yale University Press revised ed. 1969)

J. Goldenfein, *Lost in the Loop – Who is the 'Human' of the Human in the Loop?*, in *Global Governance by Data* (G. Sullivan, F. Johns & D. Van Den Meerssche eds, Cambridge University Press forthcoming).

Ł. Górski, B. Ku niacki, M. Almada, K. Tyliński, M. Calvo, P. Matias Asnaghi, L. Almada, H. Iñiguez, F. Rubianes, O. Pera & J. Nigrelli, *Exploring explainable AI in the Tax domain*, Artificial Intelligence & L. (2024), doi: 10.1007/s10506-024-09395-w.

D. Hadwick, *'Error 404 – Match not found': Tax Enforcement and Law Enforcement in the EU Artificial Intelligence Act*, 1 EU Crim 55–60 (2023), doi: 10.30709/eucrim-2023-005.

D. Hadwick, *Slipping Through the Cracks, the Carve-outs for AI Tax Enforcement Systems in the EU AI Act*, 9(3) Eur. Papers 936–955 (2024), doi: 10.15166/2499-8249/793.

D. Hadwick & S. Lan, *Lessons to Be Learned from the Dutch Childcare Allowance Scandal: A Comparative Review of Algorithmic Governance by Tax Administrations in the Netherlands, France and Germany,* 13(4) World Tax J. 609–645 (2021), doi: 10.59403/27410pa.

Cochrane Handbook for Systematic Reviews of Interventions Version 5.2.0 (J. Higgins, R. Churchill, J. Chandler & M. Cumpston eds, updated Jun. 2017).

J. van den Hoven, *Ethics for the Digital Age: Where Are the Moral Specs?*, in *Informatics in the Future* 65–76 (H. Werthner & F. van Harmelen eds, Springer 2017).

E. Huiskers-Stoop, A. Breuer & M. Nieuweboer, *Exchange of Information, Tax Confidentiality, Privacy and Data Protection from an EU Perspective,* Erasmus L. Rev. 86–99 (2022), doi: 10.5553/ELR.000225.

E. Kirchler, E. Hoelzl & I. Wahl, *Enforced Versus Voluntary Tax Compliance: The 'Slippery Slope' Framework*, 29(2) J. Econ. Psychol. 210–225 (2008), doi: 10.1016/j.joep.2007.05.004.

M. Krygier, *What's the Point of the Rule of Law*, 67(3) Buff. L. Rev. 743–791 (2019), doi: 10.1007/s40803-019-00094-2.

B. Ku niacki, M. Almada, K. Tyliński, Ł. Górski, B. Winogradska & R. Zeldenrust, *Towards eXplainable Artificial Intelligence (XAI) in Tax Law: The Need for a Minimum Legal Standard*, 14(4) World Tax J. 573–616 (2022), doi: 10.59403/2yhh9pa.

J. Locke, *Two Treatises of Government* (1689).

N. MacCormick, *Institutions of Law: An Essay in Legal Theory* (Oxford University Press 2007).

V. Mayer-Schonberger & K. Cukier, *Big Data: A Revolution That Will Transform How We Live, Work and Think* (Eamon Dolan/Houghton Mifflin Harcourt 2013).

M. Mitchell, *Artificial Intelligence: A Guide for Thinking Humans* (Farrar Straus and Giroux 2019).

B. Mittelstadt, *Principles Alone Cannot Guarantee Ethical AI*, 1 Nature Mach. Intelligence 501–507 (2019), doi: 10.1038/s42256-019-0114-4.

B. Mittelstadt, C. Russell & S. Wachter, *Explaining Explanations in AI, Proceedings of FAT*'19: Conference on Fairness, Accountability, and Transparency,* Conference Fairness, Accountability & Transparency. ACM 279–288 (2019), doi: 10.1145/3287560.3287574.

A. Narayanan & S. Kapoor, *AI Snake Oil: What Artificial Intelligence Can Do, What It Can't, and How to Tell the Difference* (Princeton University Press 2024).

O. O'Neill, *Linking Trust to Trustworthiness*, 26(2) Int'l J. Phil. Stud. 293–300 (2018), doi: 10.1080/09672559.2018.1454637.

B. Peeters, *Editorial: European Law Restrictions on Tax Authorities' Use of Artificial Intelligence Systems: Reflections on Some Recent Developments*, 33(2) EC Tax Rev. 54–57 (2024), doi: 10.54648/ecta2024006.

G. Postema, *Law's Rule: The Nature, Value, and Viability of the Rule of Law* (Oxford University Press 2022).

C. Prins, *Discriminerende algoritmes* (2021) 96(20) Nederlands Juristenblad 1454.

S. Ranchordás, *Tax and (Digital) Inequalities,* in *Taxation and Inequalities* (R. de la Feria ed., IBFD 2025).

A. Rizzo & G. Hassan, *Addressing the Use of AI by EU Tax Authorities: Towards a Common Framework of Taxpayer Protection* 65(1) Eur. Tax'n 15–29 (2024), doi: 10.59403/1avybfj.

D. Rumelhart, J. McClelland and the PDP Research Group, *Parallel Distributed Processing: Explorations in the Microstructures of Cognition, Volume 1: Foundations* (MIT Press 1986).

S. Russell & P. Norvig, *Artificial Intelligence: A Modern Approach* (Pearson 3d ed. 2009)

S. Russell & P. Norvig, *Artificial Intelligence: A Modern Approach* (Pearson 4th ed. 2020)

T. Scassa, *Administrative Law and the Governance of Automated Decision Making: A Critical Look at Canada's Directive on Automated Decision Making* 54(1) UBC L. Rev. 251–298 (2021), doi: 10.2139/ssrn.3722192.

R. Seer, S. Unger, E. Sediqi & P. Wagner, *Germany*, in *Tax Transparency,* 549–574 (F. Basaran Yavaslar & J. Hey eds, IBFD 2019).

A. Selbst, D. Boyd, S. Friedler, S. Venkatasubramanian & J. Vertesi, *Fairness and Abstraction in Sociotechnical Systems,* ACM Conference on Fairness, Accountability, and Transparency 59–68 (2019 FAT*), doi: 10.1145/3287560.3287598.

N. Smuha, *Algorithmic Rule by Law: How Algorithmic Regulation in the Public Sector Erodes the Rule of Law* (Cambridge University Press 2024).

J. Tasioulas, *The Rule of Law*, in *The Cambridge Companion to the Philosophy of Law* 117–134 (J. Tasioulas ed., Cambridge University Press 2020).

J. Waldo & S. Boussard, *GPTs and Hallucination: Why do Large Language Models Hallucinate?* 22(4) Queue 19–33 (2024), doi: 10.1145/3688007.

J. Waldron, *The Rule of Law*, in *The Stanford Encyclopedia of Philosophy* (E. Zalta ed., Summer 2020 Edition), https://plato.stanford.edu/archives/sum2020/entries/rule-of-law/ (accessed 18 Sep. 2025).

C. Wikle, A. Datta, B. Vyasa Hari, E. Boone, I. Sahoo, I. Kavila, S. Castruccio, S. Simmons, W. Burr & W. Chang, *An Illustration of Model Agnostic Explainability Methods Applied to Environmental Data*, 34 Environmetrics 1–26 (2023), doi: 10.1002/env.2772.

K. Yeung & A. Harkens, *How Do 'Technical' Design-Choices Made When Building Algorithmic Decision-Making Tools for Criminal Justice Authorities Create Constitutional Dangers? Part II*, Pub. L. 448–472 (2023), doi: 10.48550/arXiv.2301.04715.

S. Zijlstra, *Lessen uit de toeslagenaffaire. Nederlands Tijdschrift voor Bestuursrecht*, 34(3) JournalNederlands Tijdschrift voor Bestuursrecht 125–136 (2021), doi: 10.54195/ntm.23181.

Cases

Black-Clawson International Ltd v. Papierwerke Waldhof-Aschaffenberg AG [1975] A.C. 591.

Case C-202/22 *CK v. Dun & Bradstreet Austria GmbH and Magistrat der Stadt Wien* ECLI:EU:C:2025:117.

Case C-319/20 *Meta Platforms Ireland Limited v. Bundesverband der Verbraucherzentralen und Verbraucherverbände – Verbraucherzentrale Bundesverband eV,* ECLI:EU:C:2022:322, [2022] ECR I-322.

Case C-634/21 *OQ v. Land Hessen* ECLI:EU:C:2023:957.

Case C-645/19 *Facebook Ireland Ltd and Others v. Gegevensbeschermingsautoriteit* ECLI:EU:C:2021:483, [2021] ECR I-483.

Finding of the Constitutional Court of the Slovak Republic no. k. PL. ÚS 25/2019-117 of 10 Nov. 2021.

HMRC v. Rogers and Shaw [2019] UKUT 406 (TCC), [2020] 4 W.L.R. 23.

NJCM et al. v. The Dutch State (2020) The Hague District Court ECLI: NL: RBDHA:2020:1878.

Pintarich v. Deputy Commissioner of Taxation [2018] FCAFC 79.

Governmental documents

Authoriteit Persoonsgegevens, *Boete Belastingdienst kinderopvangtoeslag* (2021).

Executive Order 14110: Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence (30 Oct. 2023).

HMRC, *Automated Decisions: Technical note October 2019* (2019), https://www.gov.uk/government/publications/securing-the-tax-base-affirming-the-legislative-framework-for-hmrc-to-use-automated-processes (accessed 18 Sep. 2025).

IRS, *Internal Revenue Service Inflation Reduction Act Strategic Operating Plan: FY2023 – 2031* 3 and 10 (Apr. 2023), available at https://www.irs.gov/about-irs/irs-inflation-reduction-act-strategic-operating-plan (accessed 18 Sep. 2025).

Letter From the DWP Permanent Secretary Sir Robert Devereux & HMRC Chief Executive Jon Thompson to Meg Hillier (10 Jun. 2016).

Tweede Kamer der Staten-Generaal, Eindverslag Parlementaire ondervragingscomissie Kinderopvantoeslag, *Ongekend Onrecht* 22–23 (17 Dec. 2020).

International Organizations' Documents

J. Aslett, S. Hamilton, I. Gonzalez & M. Hardy, *Understanding Artificial Intelligence in Tax and Customs Administration* 23 (IMF 21 Nov. 2024).

Amnesty International, *Xenophobic Machines: Discrimination Through Unregulated Use of Algorithms in the Dutch Childcare Benefits Scandal* (25 Oct. 2021), https://www.amnesty.org/en/documents/eur35/4686/2021/en/ (accessed 18 Sep. 2025).

Article 29 Data Protection Working Party, *Guidelines on Automated individual decision-making and Profiling for the purposes of Regulation 2016/679*, 20 (2018a) 17/EN WP251rev.01.

C. Silvani & K. Baer, *Designing a Tax Administration Reform Strategy: Experiences and Guidelines* (1 Mar. 1997), https://www.imf.org/en/Publications/WP/Issues/2016/12/30/Designing-a-Tax-Administration-Reform-Strategy-Experiences-and-Guidelines-2134 (accessed 18 Sep. 2025).

Council of Europe, European Court of Human Rights, European Data Protection Supervisor, European Union Agency for Fundamental Rights, *Handbook on European data protection law* (Publications Office of the European Union 2018).

EU Artificial Intelligence Act, *Historic Timeline*, https://artificialintelligenceact.eu/developments/ (accessed 18 Sep. 2025).

European Commission, *Commission Guidelines on Prohibited Artificial Intelligence Practices Established by Regulation (EU) 2024/1689 (AI Act)* (4 Feb. 2025), https://digital-strategy.ec.europa.eu/en/library/commission-publishes-guidelines-prohibited-artificial-intelligence-ai-practices-defined-ai-act (accessed 18 Sep. 2025).

European Commission, *Ethics Guidelines for Trustworthy AI*, https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai (accessed 18 Sep. 2025).

European Commission High-Level Expert Group on AI: European Commission, *Ethics guidelines for trustworthy AI* (8 Apr. 2019), https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai (accessed 18 Sep. 2025)

European Commission, *High-Level Expert Group on Artificial Intelligence*, https://digital-strategy.ec.europa.eu/en/policies/expert-group-ai (accessed 18 Sep. 2025).

European Commission, *On Artificial Intelligence – a European approach to excellence and trust* (19 Feb. 2020), https://commission.europa.eu/system/files/2020-02/commission-white-paper-artificial-intelligence-feb2020_en.pdf (accessed 18 Sep. 2025).

European Court of Human Rights, *Guide on Article 14 of the European Convention on Human Rights and on Article 1 of Protocol No. 12 to the Convention* 38–48 (31 Aug. 2024).

European Public Service Union, *The Impact of Austerity on Tax Collection* (Jun. 2020), available at https://www.lrd.org.uk/research/impact-austerity-tax-collection (accessed 18 Sep. 2025).

High-Level Expert Group, *Ethics Guidelines for Trustworthy AI. Brussels: European Commission* (8 Apr. 2019), https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai (accessed 18 Sep. 2025).

OECD, *OECD Framework for the Classification of AI systems* (OECD 2022), https://www.oecd.org/en/publications/oecd-framework-for-the-classification-of-ai-systems_cb6d9eca-en.html (accessed 18 Sep. 2025)

OECD, *Tax Administration in OECD and Selected Non-OECD Countries: Comparative Information Series* (OECD 2006), https://www.oecd.org/en/publications/2006/11/tax-administration-in-oecd-and-selected-non-oecd-countries_g1gh7b25.html (accessed 18 Sep. 2025).

OECD, *Tax Administration 2022: Comparative Information on OECD and other Advanced and Emerging Economies* (OECD 2022), https://www.oecd.org/en/publications/2022/06/tax-administration-2022_5944859e.html (accessed 18 Sep. 2025).

OECD, *Tax Administration 2024: Comparative Information on OECD and Other Advanced Economies* (OECD 2024), https://www.oecd.org/en/publications/2024/11/tax-administration-2024_5c4606e4.html (accessed 18 Sep. 2025).

OECD, *Tax Administration 3.0: The Digital Transformation of Tax Administration* (OECD 2020), https://www.oecd.org/en/publications/2020/12/tax-administration-3-0-the-digital-transformation-of-tax-administration_886337a7.html (accessed 18 Sep. 2025).

Legislation

Commissioners for Revenue and Customs Act 2005.

Council of Europe, European Convention for the Protection of Human Rights and Fundamental Freedoms, as amended by Protocols Nos. 11 and 14, 4 Nov. 1950.

EU Artificial Intelligence Act, *The Act Texts*, https://artificialintelligenceact.eu/the-act/ (accessed 18 Sep. 2025).

Finance Act 2020.

Ministerial Decree of 26 Jun. 2020.

Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 Apr. 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.

Taxes Management Act 1970.

Wet bescherming persoonsgegevens.

Wet structuur uitvoeringsorganisatie werk en incomen.

Weblinks and Web documents

A. Carrick, *Taxman is Snooping on Emails and Social Media – and Now Holds 55 BILLION Items of Our Data on its AI System in a Bid to Tackle Tax Evasion* (This is Money 16 May 2023), https://www.thisismoney.co.uk/money/bills/article-12089927/HMRC-Connect-holds-55bn-data-items-bid-tackle-tax-evasion.html (accessed 18 Sep. 2025).

J. Field, N. Achten, H. Hilligoss, A. Nagy & M. Srikumar, *Principled Artificial Intelligence: Mapping Consensus in Ethical and Rights-based Approaches to Principles for AI* (15 Jan. 2020), https://cyber.harvard.edu/publication/2020/principled-ai (accessed 18 Sep. 2025).

IBM, *What are AI hallucinations?*, https://www.ibm.com/think/topics/ai-hallucinations (accessed 18 Sep. 2025)

A. Noordenbos, *Kabinetsreactie artikel 'Belastingdienst blijft wet overtreden met mogelijk discriminerende fraude-algoritme* (Taxence 18 Mar. 2024), https://www.taxence.nl/nieuws/kabinetsreactie-artikel-belastingdienst-blijft-wet-overtreden-met-mogelijk-discriminerende-fraude-algoritmen/ (accessed 18 Sep. 2025).

G. Marcus, *Why DO Large Language Models Hallucinate?* (5 May 2025), https://garymarcus.substack.com/p/why-do-large-language-models-hallucinate (accessed 18 Sep. 2025).

Holberg Prisen, *Justification of the Holberg Committee* at https://holbergprisen.no/nb/holbergprisen/prisvinnere/onora-oneill (accessed 18 Sep. 2025).

https://dictionary.cambridge.org/dictionary/english/hallucinate (accessed 18 Sep. 2025).

https://www.oecd.org/en/data/datasets/inventory-of-tax-technology-initiatives.html (accessed 18 Sep. 2025).

https://data.rafit.org/?sk=BA91013D-3261-42F8-A931-A829A78CB1EC (accessed 18 Sep. 2025).

https://taxadmin.ai/country-reports/ (accessed 18 Sep. 2025).

https://www.tbs-sct.canada.ca/pol/doc-eng.aspx?id=32592 (accessed 18 Sep. 2025).

IFS Tax Law Review Committee, *Artificial Intelligence in Automated Decision-making in Tax Administration: The Case for Legal, Justiciable and Enforceable Safeguards* (2024), https://ifs.org.uk/publications/artificial-intelligence-automated-decision-making-tax-administration-case-legal (accessed 18 Sep. 2025).

J. Redden, J. Brand, I. Sander & H. Warne, *Automating Public Services: Learning from Cancelled Systems* 23 (Data Justice Lab 21 Sep. 2022), https://carnegieuk.org/publication/automating-public-services-learning-from-cancelled-systems/ (accessed 18 Sep. 2025).

J. Sanghrajka, *Degrees of Connection* ,186(4752) Tax'n 8 (2020), doi: 10.1061/40826(186)54.

https://www.taxation.co.uk/articles/hmrc-s-connect-computer-and-investigations (accessed 18 Sep. 2025).

Taxadmin.AI, *AI for Revenue Agencies – EU Overview*, https://taxadmin.ai/ai-for-revenue-agencies-eu-summary/ (accessed 18 Sep. 2025)

WU, Leiden, Tilberg, *Antwerp, Background Paper for AI Governance and Taxpayer Rights in a Digital World* (2024).