Institute for Fiscal Studies

TLRC Report

Kunal Nathwani

Tax Law Review Committee

Artificial intelligence in automated decision-making in tax administration: the case for legal, justiciable and enforceable safeguards

Artificial intelligence in automated decision-making in tax administration: the case for legal, justiciable and enforceable safeguards

This discussion paper was written for the Tax Law Review Committee (TLRC) by Kunal Nathwani. The Committee has authorised its publication to inform and promote debate in this area. The views expressed do not necessarily represent the views of the Committee. The Institute for Fiscal Studies has no corporate views.

For a list of the TLRC's sponsors, see https://ifs.org.uk/tax-law-review-committee

Published by **The Institute for Fiscal Studies**

ISBN 978-1-80103-201-8

The Committee is also grateful to Rebecca Williams (Professor of Public Law and Criminal Law, Pembroke College, Oxford) for her invaluable comments and guidance. The paper and its recommendations remain the responsibility of the author.

The Tax Law Review Committee

Chair

Judith Freedman	Emeritus Professor of Tax Law and Policy, University of Oxford

Secretary

David Tipping	Pupil, Field Court Tax Chambers

Members

Paul Aplin	Formerly President, ICAEW and formerly partner, A C Mole
Charlotte Barbour	President, CIOT; formerly Director of Regulatory Authorisations, ICAS
Michael Blackwell	Tribunal Judge; formerly Associate Professor of Law, London School of Economics
Steve Bousher	Barrister, Joseph Hage Aaronson LLP
Tracey Bowler	Tribunal Judge; formerly Research Director, TLRC
Emma Chamberlain	Barrister, Pump Court Tax Chambers; Visiting Professor, University of Oxford and London School of Economics
Dominic de Cogan	Professor of Tax and Public Law and Deputy Director of the Centre for Tax Law, University of Cambridge
Stephen Daly	Reader in Law, King's College London
Chris Davidson	Formerly HMRC; formerly KPMG
Bill Dodwell	Formerly Tax Director, Office of Tax Simplification
Malcolm Gammie KC	Barrister, One Essex Court
Paul Johnson	Director, Institute for Fiscal Studies
Judith Knott	Formerly Director of Large Business, HMRC
Sarah Lane	Partner, Wilson Sonsini Goodrich & Rosati
Amy Lawton	Senior Lecturer in Tax Law, The University of Edinburgh
Sam Mitha	Formerly Head of Central Tax Policy Group, HMRC
Paul Morton	Non-Executive Director, HMRC; formerly Tax Director, Office of Tax Simplification
David Murray	Head of Tax Policy & Sustainability, Anglo American
Kunal Nathwani	Solicitor Associate, Kirkland & Ellis
Dan Neidle	Founder, Tax Policy Associates Ltd; formerly Partner, Clifford Chance LLP
Geoff Pennells	Formerly Tax Policy Director for EMEA, Citigroup
Christopher Sanger	Partner, EY
Heather Self	Consultant, formerly Partner, Blick Rothenberg
Greg Sinfield	President, Tax Chamber of the First-tier Tribunal
Andrew Summers	Associate Professor of Law, London School of Economics
Richard Thomas	Formerly Tribunal Judge; formerly Assistant Director, HMRC
Victoria Todd	Head, Low Incomes Tax Reform Group, CIOT
Edward Troup	Formerly Executive Chair and Permanent Secretary, HMRC

IFS Staff Attendees

Stuart Adam	Senior Economist, Tax Sector, IFS
Helen Miller	Deputy Director and Head of Tax Sector, IFS

Contents

Executive summary

HM Revenue & Customs (HMRC) has a broad range of powers, which can be divided into powers which (i) are administrative or mechanical in nature, and (ii) require the exercise of an element of discretion or subjectivity (including the imposition of penalties).

Automated decision-making (ADM) is any decision or process where the whole or part of the decision or process is made without human intervention (through technology), irrespective of whether the decision or output is *subsequently* reviewed by HMRC. Broadly, the technology underpinning ADM can be divided into:

- artificial intelligence (AI), and
- algorithmic systems developed through conventional programming (i.e. algorithmic systems or rules-based systems).

Although there is no comprehensive list of ADM technology published by HMRC, it is understood that, at present, HMRC employs both types of technology. This paper focuses on the use of AI in ADM in tax administration to make discretionary or subjective decisions.

Although AI has not yet been widely deployed in ADM in tax administration in the UK, it is inevitable that AI will play a more prominent role in such decision-making because it enhances the speed and efficiency of decision-making and helps optimise resources allocated to HMRC and HM Treasury. In addition to operational benefits, AI also has the ability to facilitate detection of previously undetectable or hidden correlations, suspicious activity, trends, indicators of tax loss, etc., which could enable early detection and pre-emptive action, or mitigation of these risks in real time, which would help reduce the tax gap and increase tax collections. Therefore, in considering the tax safeguards that should be introduced with respect to the use of AI in ADM in tax administration, this paper adopts a forward-looking approach with respect to the safeguards that should be introduced in preparation for the more widescale deployment of AI in ADM in tax administration (including to regulate the existing uses of AI by HMRC). Some of the risks presented by the use of AI in ADM in tax administration are set out in the following sections of the paper.

The existing tax administrative framework is not well suited to the use of AI in ADM in tax administration because the use of AI in tax administration (particularly if AI is used to make discretionary decisions) represents a *fundamental shift in the basis of decision-making by HMRC*. Under the current operational framework, generally with respect to decisions made by

HMRC (in particular discretionary decisions), HMRC officers are the primary decision-makers, even though they may be aided by technology to make these decisions. However, once AI is deployed in ADM in tax administration, where a decision is made purely by AI (in particular, machine learning (ML) without human intervention in the decision-making process) this would reflect a decision made *by the system* (and not a decision of an HMRC officer); the decision made by the AI would be based on the model's own interpretation of the data (whether labelled or unlabelled) and correlations drawn by the model. This would reflect a shift in the role of primary decision-maker from the HMRC officer to the AI. Even where an HMRC officer is required to review a decision made by the AI before the decision has an impact on a taxpayer and the HMRC officer provides an explanation of why that decision was arrived at, where black box AI is used, that explanation is just of the officer's *understanding* of why a particular decision was arrived at because of the opacity of black box AI. The HMRC officer's explanation would involve reverse-engineering the logic and basis of the decision generated by the AI, and this reverse engineering inculcates an element of uncertainty and unreliability into the explanations provided by the HMRC officer.

Any such shift in the basis of decision-making should be the product of a conscious transparent policy decision made by the government. Correspondingly, it is recommended that a proactive approach to regulating the use of AI in ADM in tax administration be adopted, rather than a reactive approach, especially given some of the risks (which have manifested themselves in various jurisdictions that have already adopted AI in public administration).

The UK government and international organisations have published various AI codes, standards and frameworks with respect to the use of AI in public administration. However, these codes, standards and frameworks are at this stage non-binding and non-tax-specific, and do not provide taxpayers with any legally enforceable or justiciable rights.

Given the fundamental shift in the way decisions will be made where AI is deployed in ADM in tax administration, it is recommended that to the extent AI is used in ADM in tax administration, taxpayers should be provided with appropriate safeguards that are justiciable, measurable and legally enforceable. There are two alternative solutions that this paper recommends: tax-specific AI legislation or an HMRC AI Charter (which includes some of the key standards and values that HMRC must adhere to) (together, both these measures are referred to as the 'AI Tax Legal Safeguards'). Practically, it is expected that the government may legislate for broader AI legislation (which is not tax-specific) and therefore an HMRC AI Charter may be the alternative preferred by the government.

In addition to introducing AI Tax Legal Safeguards, it is recommended that the government introduce legislation that affirmatively provides HMRC with the power to use AI in ADM in tax

administration to ensure that the use of such AI is legally permissible and not subject to challenge.

The paper recommends the following key safeguards.

Enabling legislation for the use of AI in ADM in tax administration

- Legislation should affirmatively provide for the use of AI in ADM in tax administration.
- AI Tax Legal Safeguards should set out specific circumstances where the use of AI is impermissible.
- AI Tax Legal Safeguards should also require consideration of other relevant legislation including General Data Protection Regulation (GDPR), the Data Protection Act 2018 (DPA) (and processing of data in accordance with GDPR and the DPA), the European Convention on Human Rights (ECHR), Equality Act 2010, the HMRC Charter, etc.
- AI Tax Legal Safeguards should specify how taxpayer risk levels will be determined following the processing of data in risk management systems (e.g. Connect) (although it is noted that, for security and public interest purposes, there may be grounds to limit such disclosure and the exact disclosure should be the subject of public consultation).

Training data and bias

- AI Tax Legal Safeguards should provide broad principles on the use of datasets for the training of AI used in ADM in tax administration to ensure that these datasets (i) are large, diverse, reliable and unbiased, and (ii) represent a wide cross-section of the demographic affected. These principles should apply even where the data used by HMRC is not internal HMRC data but is otherwise bought or procured from third parties or where the system is developed externally.
- AI Tax Legal Safeguards should make clear that the principles on datasets apply at all stages of the development and use of AI, and therefore the datasets used to train the AI should be kept under review, even once the AI has been deployed.
- AI Tax Legal Safeguards should provide for mandatory retraining of AI based on updated datasets, and AI Tax Legal Safeguards should set out the period after which there should be such mandatory retraining once AI has been deployed.
- AI Tax Legal Safeguards should require the development of a tax-specific technical standard (which should be periodically reviewed and updated) to minimise the risk of bias that must be adhered to when developing AI that is used for ADM in tax administration. The technical

standard should be consulted on and should be adopted once it has had wide stakeholder engagement. Once adopted, this technical standard should be publicly disclosed.

Testing and deployment

- AI Tax Legal Safeguards should provide for pre-deployment testing of AI.
- AI Tax Legal Safeguards should require a transition period where post-deployment testing is undertaken in parallel to the deployment of AI.
- AI Tax Legal Safeguards should require annual audits of deployed AI (including impact assessments on under-represented taxpayers).
- Wider government policy should review safeguards with respect to the HMRC information disclosure gateways.

Explainability and transparency

- AI Tax Legal Safeguards should provide that:
    - taxpayers are notified where AI is used for ADM in relation to decisions having a direct impact on them; and
    - taxpayers are provided with outcome-based local rationale explanations with respect to such decisions.
- AI Tax Legal Safeguards should also require process-driven explanations.
- Express policy consideration is also required in respect of transparency, explainability and general policy where AI is being deployed in a manner that does not have a direct impact on taxpayers. In particular, consideration should be given to whether HMRC should be bound by guidance delivered by large language models (LLMs) to taxpayers.

1. Decision-making by HMRC

1. To understand the decisions HMRC makes or is required to make, it serves to first consider what HMRC's powers are. This paper considers decisions in relation to income tax (IT) and capital gains tax (CGT) provided for in the Taxes Management Act 1970 (TMA), and in relation to corporation tax (CT) in Schedule 18 Finance Act 1998 (FA 98), to substantiate the arguments for AI Tax Legal Safeguards (defined below). Penalty decisions in relation to all three taxes are made by HMRC under other legislation and some of these are also considered. There are many other decisions of HMRC that fall to be made under powers provided to HMRC with respect to some of its other functions (e.g. duties) and the same arguments apply to these other powers as well.

2. The TMA, etc., provide HMRC with various specific powers, responsibilities and discretions to enable them to undertake their overall collection and management responsibilities.[1] Inherent in HMRC's collection and management function is the requirement for HMRC to track and detect whether taxpayers have complied with their various obligations under the TMA, etc. (and this requirement to track and detect includes HMRC risk-assessing taxpayers to determine which taxpayers should be subject to enhanced checks). In furtherance of this function, HMRC uses a tool known as 'Connect' that it developed in 2010 with the help of BAE Systems to assist it with 'data matching and risking' by cross-checking positions taken by taxpayers based on data held by HMRC and procured from third parties, for example, identifying hidden relationships between organisations and people (that may have previously been undetected).[2] There is little that has been published by HMRC on the use of Connect but reports suggest that HMRC has access to 55 billion items of data in Connect, which it acquired through taxpayer website browsing records, email records, social media, flight sales and passenger data, DVLA records, tax returns, the Land Registry, online property rental platforms and the UK Border Agency – this has not been verified.[3]

[1] Note that s.1 TMA provides HMRC with its overall responsibility for the collection and management of IT, CGT and CT in the following terms: 'The Commissioners for His Majesty's Revenue and Customs shall be responsible for the collection and management of (a) income tax, (b) corporation tax, and (c) capital gains tax.'

[2] Devereux R. (2016). Letter to Hillier, M. re fraud and error stocktake, 10 June 2016. Available at: https://www.parliament.uk/globalassets/documents/commons-committees/public-accounts/Correspondence/2015-20-Parliament/PAC-Response-final-signed-copy-of-follow-up-letter-to-3rd-party-data.pdf [Accessed: 13 May 2024].

[3] Russel, B. (2023). '55 billion items of taxpayer data now on HMRC's "Connect" AI systems, reveals tax authority'. [Online] *IFA Magazine*. Available at: https://ifamagazine.com/55-billion-items-of-taxpayer-data-now-on-hmrcs-connect-ai-system-reveals-tax-authority/ [Accessed: 13 May 2024].

3. Broadly, HMRC's powers and responsibilities can be split into those which (i) are administrative or mechanical in nature, and (ii) require the exercise of an element of discretion or subjectivity. In the case of discretionary or subjective decisions, the amount of discretion or subjectivity may vary depending on the legislation. Discretion is used in this paper to describe consideration and weighing of various factors to arrive at a decision, which can inherently be described as subjective. Discretion in this paper is not used to describe HMRC's power to enforce a law or not (e.g. extra-statutory concessions), to pursue a violation of the law or to reach compromises or settlements with taxpayers, although similar arguments will indeed be applicable where HMRC chooses to use ADM to operate such powers.

Illustrative HMRC powers

4. To illustrate this, it is useful to consider some examples of the powers provided to officers of HMRC under the TMA.[4] This list is not exhaustive.

    4.1. Examples of administrative or mechanical decisions:

        4.1.1. *tracking* whether persons required to provide notice of chargeability to IT and CGT have provided such notice (s.7 TMA);

        4.1.2. where notice has been provided to a taxpayer to file a personal return (e.g. under s.8 TMA), *tracking* whether such a return has been filed; and

        4.1.3. *amending* a return on the basis that there are obvious errors or omissions in the return (e.g. arithmetic errors) (s.9ZB TMA).

    4.2. Example of decisions involving an element of discretion or subjectivity:

        4.2.1. *amending* a return where an HMRC officer has *reason to believe* that something in the return is incorrect in light of information available to that officer (s.9ZB TMA);

        4.2.2. the *decision* to open an enquiry into a personal tax return filed by a person (s.9A TMA);

[4] 'Officers' are staff appointed by the 'Commissioners' for HMRC to undertake the powers conferred on HMRC (s.2, Commissioners for Revenue & Customs Act 2005). 'HMRC' collectively refers to the 'officers' and 'Commissioners'. Generally legislative provisions in the TMA will confer powers on an officer of HMRC; however, there are exceptions, as discussed below.

4.2.3. issuing a closure notice which requires the notice to state an officer's *opinion* of whether an amendment to a return is required and, if relevant, the amendments to be made (s.28A(2) TMA);

4.2.4. the *decision* to make a discovery assessment or amendment (s.29 TMA); and

4.2.5. the authority to *make* an assessment to tax that is not a self-assessment (s.30A TMA).

Similar provisions in Schedule 18, FA 98 apply for CT.

Penalties

4.3. Provisions on penalties in particular vary with regard to whether the imposition and amount of the penalty is fixed (i.e. fixed penalties where there is no HMRC discretion involved), or the imposition or amount of the penalty involves an element of discretion (including scenarios where the imposition of the penalty is mandatory, but the amount of the penalty is subject to HMRC discretion) (i.e. discretionary penalties):

4.3.1. Paragraphs 1 and 2, Schedule 55, Finance Act 2009 (FA 2009) generally impose fixed penalties where a taxpayer fails to make or deliver a return with respect to IT or CGT, or accounts, statements or documents as required pursuant to s.8(1) TMA (see further discussion in Section 3); the amounts of penalties broadly are (calculated incrementally):

4.3.1.1. an aggregate fixed penalty of £100 (for the first three months after the date of default; i.e. the penalty date) (paragraph 3, Schedule 55, FA 2009);

4.3.1.2. a fixed daily penalty of £10 during the period of 90 days beginning with the end of the three-month period above where the failure continues (paragraph 4, Schedule 55, FA 2009);

4.3.1.3. a fixed penalty of the higher of £300 and 5% of the liability of tax which would have been shown in the return in question where the failure continues after the end of the period of six months beginning with the penalty date (paragraph 5, Schedule 55, FA 2009);

4.3.1.4. a fixed penalty based on the higher of £300 and a statutorily prescribed percentage of liability to tax which would have been shown in the return

in question where the failure continues after the end of 12 months beginning on the penalty date (the percentage determined based on whether the taxpayer deliberately withheld information that would enable HMRC to determine the taxpayer's liability to tax) (paragraph 6, Schedule 55, FA 2009), subject to certain reductions where the taxpayer makes disclosures in certain situations (the amounts of such reductions are discretionary) (paragraph 14, 15 and 15A, Schedule 55, FA 2009);

4.3.2. Paragraphs 1 and 3, Schedule 56, FA 2009 impose a fixed penalty of 5% of the unpaid tax where a taxpayer fails to pay an amount of IT or CGT due pursuant to s.59B TMA with respect to assessments other than simple assessments within a statutorily prescribed amount of time;

Penalties involving HMRC discretion

4.3.3. Paragraphs 1 and 6 of Schedule 41, Finance Act 2008 (FA 2008) impose a fixed penalty based on the potential lost revenue for failure to notify chargeability of IT or CGT under s.7 TMA – however, paragraphs 12, 13 and 13A of Schedule 41, FA 2008 provide that HMRC must mandatorily reduce the amount of the penalty if the relevant taxpayer has made a disclosure to 'one that reflects the quality of the disclosure' up to a minimum percentage provided;

4.3.4. Paragraph 17, Schedule 18, FA 98 imposes a penalty for a company that fails to deliver a company tax return by the filing date of £100, £200, £500 or £1,000 (which is determined based on prescriptive factors of when the return is filed and whether there have been successive failures to file such a return) – however, it is made by way of a determination by an officer setting the penalty at such amount 'as in his opinion, is correct or appropriate' pursuant to s.100 TMA giving the HMRC officer an element of discretion;

4.3.5. Paragraph 18, Schedule 18, FA 98 also imposes a penalty equivalent to 10% or 20% of the unpaid tax in addition to the penalty under paragraph 4.3.4 above depending on the time within which the return is filed (although the actual amount may vary, the percentage is fixed) – however, it is made by way of a determination by an officer setting the penalty at such amount 'as in his opinion, is correct or appropriate' pursuant to s.100 TMA giving the HMRC officer an element of discretion;

4.3.6. S.98(1)(b) TMA (information notices) imposes a penalty *not exceeding* £300 for a failure to deliver certain information requested by HMRC, and a

subsequent daily penalty of up to £600 for each day on which the failure continues after the day on which the initial penalty was imposed;

4.3.7. S.98C TMA (failure to comply with the Disclosure of Tax Avoidance Schemes) imposes a penalty not exceeding a daily maximum of £600 (depending on the provision that is breached) for the initial period and the amount of the daily penalty is to be arrived at 'after taking into account…all relevant considerations, including the desirability of its being set at a level which appears appropriate for deterring the person, or other persons, from similar failures to comply on future occasions having regard to (in particular)' specific factors noted in the legislation; 98C(2ZC) TMA also provides that if the maximum penalty appears inappropriately low after taking account of the above considerations, the penalty is to be of such amount not exceeding £1 million as appears appropriate having regard to those considerations; and

4.3.8. S.109C TMA (companies ceasing to be UK tax resident) imposes a penalty not exceeding the amount of tax which is or will be payable and which has not been paid at that time when a company ceases to be UK tax resident without complying with certain conditions in relation to ensuring that arrangements are made to pay the UK exit tax charge.

4.4. S.103 Finance Act 2020 (FA 2020) provides that 'anything capable of being done by an officer of Revenue and Customs by virtue of a function conferred by or under an enactment relating to taxation may be done by HMRC (whether by means *involving the use of a computer or otherwise*)'. The drafting is expansive but the exact scope of the legislation and what a computer is permitted to do is unclear and untested (for a more detailed summary on this, see 'Comment from the IFS Tax Law Review Committee to the Public Bill Committee: Clause 100 Finance Bill – HMRC: Exercise of Officer Functions'[5]). While there may be some debate as to what decisions are technically permitted to be made by a computer under this legislation (and whether a computer can even have human-like attributes, such as the ability to have an *opinion* or to have *reason to believe or* to have *desirability* for an outcome, which are required for it to undertake some of the discretionary functions listed above), the intention behind the provision seems to be to provide HMRC with wide powers to use computers to wholly or partly

[5] https://ifs.org.uk/publications/comment-ifs-tax-law-review-committee-public-bill-committee

undertake some of its functions.[6] The rest of this paper assumes that this legislation achieves this aim with respect to algorithmic systems developed through conventional programming; however, the following sections discuss legislation required where AI (as defined below) is being used to make decisions instead of officers.

[6] There are some genuine doubts as to what attributes a computer system can have, and some computer scientists argue that a computer is incapable of having an opinion or belief (or any similar human attributes expressing subjectivity). Instead, computer systems can only predict and infer. Therefore, without any further enabling legislation, this puts the legitimacy of the use of AI in ADM in tax administration in doubt and such decisions taken by HMRC may not be beyond challenge where such decisions are made solely using AI. This is an issue that exists beyond just tax administration.

2. Automated decision-making

5. For the purpose of this paper, ADM is defined as any decision or process where the whole or part of the decision or process is made without human intervention (through technology). For the purpose of this definition, it does not matter whether the decision or output of the process is *subsequently* reviewed by HMRC as long as all or part of the initial decision or process is made or undertaken without human intervention.

6. In HMRC's context, ADM can be subdivided into two further broad categories:

   6.1. programs (e.g. conventional decision-trees) developed through conventional programming that do not use ML (also known as algorithmic systems or rules-based systems);[7] and

   6.2. AI (e.g. ML, including through the use of artificial neural networks – sometimes referred to as deep learning – and other techniques, such as random forest, ensemble methods, etc.).

7. At present, HMRC employs both types of technology. However, there are different stakeholders in HMRC that are responsible for the two types of technology and currently there is no overlap in the areas in which such technology is being deployed by HMRC. Further background on both types of technology is provided below.

8. Much has been written and said about the advantages of the automation of decision-making in the public sector, and this paper does not attempt to summarise the relative advantages and disadvantages. It is inevitable that as technology develops ADM will play a more prominent role in decision-making – as it should – to keep pace with developments in the private sector. The main benefits of automation are '[increasing] the speed and efficiency of decision-making as well as [providing] an ability to detect correlations that may be undetectable to the human brain'[8] and ensuring that decisions

[7] Although some taxonomies may include certain algorithmic systems or rules-based systems as AI, for the purposes of this paper, such systems are not treated as AI.

[8] Finck, M. (2020). 'Automated decision-making and administrative law', in P. Cane et al. (eds), *The Oxford Handbook of Comparative Administrative Law*. Oxford: Oxford University Press, pp. 656–676, https://doi.org/10.1093/oxfordhb/9780198799986.013.39.

are 'more systematic, consistent and coherent'.[9] Incorporation (or further incorporation) of ADM in HMRC's processes would have a direct impact on HMRC's efficiency, budgeting and staffing, and could result in tangible cost savings for HMRC and HM Treasury.

9. Recent developments in AI (especially ML) in particular present HMRC with an unprecedented opportunity to incorporate such ADM technology in their processes that have the ability to detect previously undetectable or hidden correlations, suspicious activity, trends, indicators of tax loss, etc. Effective use of such ADM technology could give HMRC the power to mitigate these risks in real time by enabling HMRC to take pre-emptive or defensive measures. If used appropriately, such ADM technology could improve the effectiveness and efficiency of HMRC's monitoring and compliance function and would inevitably help lower the tax gap (subject to a cost–benefit analysis in developing and operating such technology).

10. An example of the benefits of recent developments in AI can be gleaned from the non-tax sector. For example, in the pharmaceuticals field, researchers at Massachusetts Institute of Technology used AI to discover an antibiotic (*Halicin*) that is effective against a strain of bacteria that, prior to the discovery of *Halicin*, had been resistant to all other known antibiotics; the AI managed to detect *Halicin*'s effectiveness by drawing its own correlations (i.e. correlations that scientists had not previously been able to detect).[10] In the absence of AI, *Halicin* may not have been discovered, given the costs involved.[11]

11. This paper does not argue against the use of ADM (or AI) in HMRC's administrative processes. In fact, it is important to acknowledge the potentially significant benefits ADM could bring (and has already brought) to elements of tax administration and to support its integration in tax administration. However, there are significant risks associated with the use of such technology, in particular with the use of AI (especially ML) (given its nascent stage of development), and therefore it is equally important that certain justiciable safeguards are put in place to ensure that the risks of such technology are mitigated and managed.

[9] Williams, R. (2021). 'Rethinking administrative law for algorithmic decision making'. *Oxford Journal of Legal Studies*, 42(2), p. 472. Also see Daly, S. (2024, forthcoming). 'Artificial intelligence, the rule of law and public administration: the case of taxation'. *Cambridge Law Journal*.

[10] Kissinger, H., Schmidt, E. and Huttenlocher, D. (2021). *The Age of AI: And Our Human Future*. London: John Murray Press, p. 9.

[11] Kissinger, H., Schmidt, E. and Huttenlocher, D. (2021). *The Age of AI: And Our Human Future*. London: John Murray Press, p. 10.

12. The next two sections provide a short introduction to both categories of ADM discussed above and explain why this paper focuses on AI (and, within AI, primarily ML).

3. Conventional algorithmic systems

13. Most of the current ADM technology currently in use at HMRC is understood to have been developed through conventional programming (for example, algorithmic decision-trees).

14. Algorithms are 'a set of instructions usually applied to solve a well-defined problem'.[12] The approach to designing (and thereby the use cases for) a set of algorithms for traditional software programming vis-à-vis ML is different.

15. Generally, conventional programming requires specific instructions to be coded with respect to particular inputs and outputs. In other words, traditional programming requires algorithms to provide instructions with respect to each input factor (or combination of input factors) and to provide a particular outcome (or outcomes) based on the input factors. For illustrative purposes, a computer could be programmed with the logic (and it is understood that HMRC already deploys programs with such logic):

Example 1a.

If a taxpayer required to self-assess has not filed a tax return for the tax year ending 5 April 2024 by 20 January 2025, *then* send the taxpayer a reminder via e-mail that they are required to file a tax return for the tax year 2023/2024 by 31 January 2025.

Example 1b.

If a taxpayer required to self-assess has not filed a tax return for the tax year 2023/2024 by (and including) 31 January 2025, *then* impose an aggregate £100 fixed penalty on the taxpayer *and* send the taxpayer notice of this penalty in the form of notice uploaded to the system.

[12] House of Commons: Science and Technology Committee (2018). 'Algorithms in decision-making: Fourth Report of Session 2017–19', HC 351, p.7. Available at: https://publications.parliament.uk/pa/cm201719/cmselect/cmsctech/351/351.pdf [Accessed: 3 June 2024].

16. The above examples are rudimentary (and simplified) examples of the logic of code that could be used for ADM utilising algorithmic systems. These examples seek to illustrate that, generally, conventional programming requires a specific input and output to be provided for in the code. Such ADM technology cannot make *pure* discretionary decisions without human intervention because practically it is impossible to predict every possible input variable and code a corresponding output for such input (taking into account the weighting of such inputs where there are multiple input entries). In other words, for such a program to properly exercise a *pure* discretionary power provided to HMRC (e.g. as provided for in paragraphs 4.2 or 4.3.3 to 4.3.8 above), the code would at least need to specify each input variable to be taken into account in exercising such discretion, each output variable, the effect of each such input on the output and the relative weighting of each input in a particular case. Where there is an input variable that falls outside the input variables that have been coded for, the program does not have the ability to take that input variable into account.

17. HMRC currently already deploys this form of ADM technology in some of its processes and, for example, the use of this technology was discussed in the string of cases *Robert Morgan v HMRC; Keith Donaldson v HMRC* [2013] UKFTT 317 (TC), *HMRC v Keith Donaldson* [2014] UKUT 0536 (TCC), *Nigel Rogers v HMRC* [2018] UKFTT 0312 (TC), *Craig Shaw v HMRC* [2018] UKFTT 0381 (TC) and *HMRC v Rogers and Shaw* [2019] UKUT 406 (TC), amongst others. These cases provided an overview of HMRC's current use of computer programs to impose penalties under the Schedule 55 regime.

    17.1. To make such penalty determinations under Schedule 55, computer programs identify taxpayers meeting the criteria specified in Schedule 55, and thereafter the computer programs automatically determine penalty calculations taking into account various non-discretionary factors (including the type of tax return filed (paper or electronic), the time period of delay, the tax outstanding, etc.) and once calculated, send notices to the affected taxpayers (*Morgan v HMRC* and *HMRC v Donaldson)*.

    17.2. These automated penalties do not involve pure discretion; they are either fixed penalties where (i) HMRC has *no discretion* or (ii) penalties where HMRC only needs to make an institutional decision *on whether* to impose such penalties and once that is done, the amount of the penalties is determined using a fixed formula.

18. There is no published list of the processes for which HMRC uses such ADM technology (whether to make administrative or mechanical decisions, or discretionary decisions). However, it is generally understood that HMRC employs such ADM technology for 'large-scale automated processes to carry out routine tasks such as to give statutory

notice, where making individual decisions on individual cases would be impractical, resource intensive, or simply unnecessary in light of published guidance or underlying legislation.'[13]

[13] HMRC (2019). Automated decisions: technical note October 2019, paragraph 1.1. Available at: https://assets.publishing.service.gov.uk/media/5dbafe0440f0b637a2b05a71/Automation_technical_note_-_final_-_31_Oct__1_.pdf [Accessed: 3 June 2024].

4. Artificial intelligence

19. AI can broadly be defined as a 'set of statistical tools and algorithms that combine to form, in part, intelligent software enabling computers to simulate elements of human behaviour such as learning, reasoning and classification'.[14] The key characteristic of AI (and primary differentiator from ADM software referenced in Section 3) is that it is technology that *simulates* human cognitive behaviour.

20. ML is a subset of AI and can be defined as 'a family of techniques that allow computers to learn directly from examples, data, and experience, finding rules or patterns that a human programmer did not explicitly specify'.[15] Unless specified otherwise, AI is used interchangeably in this paper to reference ML, although it is acknowledged that in practice this is not strictly correct in all circumstances. The next paragraphs summarise some of the key concepts relevant to AI, but these are included mainly to provide a simplistic explanation of AI. AI (and ML) is a complex (and developing) area of technology, and the following paragraphs should not be viewed as attempting to provide a comprehensive explanation of the technology.

21. Unlike traditional programming explained in Section 3, AI does not require the coding of any specific inputs or corresponding outputs; instead, it requires defining the *objective* of the program and the parameters of permitted decision-making. Therefore, unlike algorithmic systems, AI does not require a programmer to code for each input and output variable. This is a key difference between algorithmic systems and AI, and what gives AI its dynamic nature. Fundamentally, if the objective and parameters of an AI system are appropriately defined and the system is properly 'trained', the AI program is capable of dealing with a multitude of inputs and outputs that may or may not have been consciously considered by decision-makers at that time.

22. Training an AI model with data is a key step in developing AI technology. A model is trained by providing it with a large volume of 'training data', which the model uses to learn through a combination of weighting and feedback. There are various learning

[14] House of Commons: Science and Technology Committee (2018). 'Algorithms in decision-making: Fourth Report of Session 2017–19, HC 351, p. 7. Available at: https://publications.parliament.uk/pa/cm201719/cmselect/cmsctech/351/351.pdf [Accessed: 3 June 2024].
[15] House of Commons: Science and Technology Committee (2018). 'Algorithms in decision-making: Fourth Report of Session 2017–19, HC 351, p. 7. Available at: https://publications.parliament.uk/pa/cm201719/cmselect/cmsctech/351/351.pdf [Accessed: 3 June 2024].

techniques that can be used to train an AI model (e.g. supervised learning, unsupervised learning, reinforcement learning, etc.). The technique used to train a model will depend on the objective of the model, the type of data available (and from where such data is procured), availability of resources, etc. With respect to some AI, this training/learning only takes place in the initial stages pre-deployment and, once deployed, there is no additional learning (i.e. the learning is static) – therefore, the training and deployment phases are distinct.[16] In other AI, learning is continuous and therefore the AI continues to learn post-deployment through the analysis of real-time data.

23. Simplistically, in a supervised model, the input and output training data used to train a model are *labelled* (i.e. broadly through human labelling), and the objective of the training is to finesse the algorithm by 'map[ping] input variables…onto desired outputs [and, o]n the basis of these examples, the ML model is able to identify patterns that link inputs to outputs'.[17] When deployed, the objective of the model is to replicate these patterns.

24. In unsupervised learning, the training data is not labelled and the model is trained by analysing the training data and drawing correlations from that data without any guidance through labelling (i.e. devoid of any real human intervention). The correlations drawn by the model from this training phase serve as the foundation of how the model will operate in practice (i.e. the machine learns from the data). Once a model has been trained, it can be applied to new data to carry out the original objective of the program.[18]

25. Therefore, for an AI system to function so that it properly achieves the stated objective and fairly, the quality and volume of the training data is a fundamental aspect that needs to be considered. Further, in supervised models, proper labelling of the data is also critical. Biases in the training data (whether as a result of homogeneous forms of data or inadequate data, or otherwise) can result in the system drawing correlations or developing inferences based on such biased data and 'learning' such biases. If undetected, this could have the effect of introducing or perpetuating biases once the AI ADM technology is deployed.

[16] Kissinger, H., Schmidt, E. and Huttenlocher, D. (2021). *The Age of AI: And Our Human Future.* London: John Murray Press, p. 83.

[17] Information Commissioner's Office and the Alan Turing Institute (2022). 'Explaining decisions made with AI', p.7. Available at: https://ico.org.uk/media/for-organisations/guide-to-data-protection/key-dp-themes/explaining-decisions-made-with-artificial-intelligence-1-0.pdf [Accessed: 8 June 2024].

[18] Reinforcement learning is not explained in this paper, although a helpful summary is provided in: Information Commissioner's Office and the Alan Turing Institute (2022). 'Explaining decisions made with AI', p. 7. Available at: https://ico.org.uk/media/for-organisations/guide-to-data-protection/key-dp-themes/explaining-decisions-made-with-artificial-intelligence-1-0.pdf [Accessed: 8 June 2024].

26. For illustrative purposes only, the development of such an AI model for ADM in tax administration is explained below using a simplified example:

Example 2.

If an AI model is being developed to detect tax evasion, the algorithmic logic will *not* say:

'*If* the taxpayer has done [X], *then* the taxpayer is engaging in tax evasion.'

Given the variety of factors (known and unknown) that could indicate tax evasion, this approach to coding would be too simplistic and it would be impossible to define each factor [X] that might indicate tax evasion. Further, some factors may not always be indicative of tax evasion, and whether they indicate tax evasion may depend on the context (e.g. other factors and circumstances co-existing at the time). Therefore, it is not practically possible to develop a comprehensive system based on conventional programming that properly detects tax evasion.

27. However, hypothetically, AI could be used to develop a system that detects tax evasion (the example below assumes an unsupervised model). In such a situation:

    27.1. Algorithms would be created to define the objective of detecting tax evasion (or an alternative proxy for tax evasion); generally, the algorithms will not specifically code what input factors indicate tax evasion.

    27.2. A learning model would then be developed through which the system would teach itself which factors indicate tax evasion, the circumstances in which those factors indicate tax evasion and any other correlations that it may draw from that data, and this model would be trained on properly selected training data; this training data may hypothetically include inputs from tax returns filed with HMRC and other data points that could be considered proxies for tax evasion (including data that HMRC may purchase or otherwise procure from third parties). Given this example assumes an unsupervised model, the data will not be labelled. However, where a supervised model is used then the data or documents used as training data would need to be labelled.

    27.3. The system would then interpret this data, and use the data to train itself, drawing appropriate correlations as mentioned above. The system may draw correlations based on its own interpretation of the information populated in a tax return and these may be correlations that HMRC may not previously have detected. However, if the system were a supervised model, then the system would draw

correlations or inferences based on the labelling (i.e. which provide the system with guidance on which inputs are indicative of tax evasion).

27.4. Once trained, the system is tested on existing data that has not been used to train the model.

27.5. Once training and testing are completed, HMRC can use this system to detect factors indicative of tax evasion in live cases. The system relies on the correlations and inferences it has drawn based on its training, rather than specifically relying on HMRC's interpretation or an individual HMRC officer's interpretation.[19]

28. The ability to draw correlations that were previously undetectable is one of the main advantages of AI. However, this also serves as the main risk of AI because if a model is trained on biased, unrepresentative, unreliable or insufficient data, the system may draw inferences that are not supported in the real world (or may perpetuate existing or hidden biases). If such AI is widely deployed, such technology could produce far-reaching results that may go largely undetected for a long period of time. Anecdotally, it is understood that HMRC may not currently possess adequate data to automate some of the discretionary penalties in paragraphs 4.3.3 to 4.3.8 using AI.

29. AI can be used to undertake active and passive functions.[20] From a tax perspective, some of the key functions that it is envisaged HMRC could use AI for are non-administrative tasks such as risk management (including identification, detection and monitoring), guidance (internal guidance for HMRC and external guidance for taxpayers and agents), and the exercise of discretionary decisions (including those described in Section 1).

30. There is no published list of HMRC uses of AI. However, from discussions with HMRC, it is understood that AI is currently mainly used in compliance risking; for example, risk management to assess taxpayer risk at the point a taxpayer files a tax return or makes an application for repayment, and to assist with selecting cases for compliance investigations. The Risk and Intelligence Service (RIS) team, which is part of the HMRC Compliance team, is understood to be the stakeholder within HMRC responsible for the development of the above technology. The RIS team was assisted by data scientists and analysts employed by HMRC. It is understood that the Chief Digital Information Office – Data Science (CDIO – Data Science) team at HMRC employs data

[19] Kissinger, H., Schmidt, E. and Huttenlocher, D. (2021). *The Age of AI: And Our Human Future.* London: John Murray Press, p. 35.
[20] Kissinger, H., Schmidt, E. and Huttenlocher, D. (2021). *The Age of AI: And Our Human Future.* London: John Murray Press, p. 67.

scientists and other analysts to assist various stakeholders within HMRC with the development of AI solutions.

31. Similarly, it is understood that AI is being used for risk assessment purposes in relation to VAT fraud by detecting patterns of VAT fraud from data collected by HMRC from VAT returns filed by taxpayers.

32. HMRC is also understood to be in the process of developing an LLM. Once this LLM is developed, it is understood that this will be used by HMRC officers to (at the very least) provide HMRC officers with answers to questions that taxpayers may ask in real time (and to refer HMRC officers to the relevant sections of HMRC guidance on which such answers are based). It is unclear at this stage whether the LLM will be taxpayer-facing; however, it seems reasonable to expect that this LLM will at some stage be made available to taxpayers to assist taxpayers with answering questions that would traditionally have been answered through HMRC's manuals or helpline. Some tax authorities already utilise AI to provide guidance (in varying levels of detail) to taxpayers; for example, the Canadian Revenue Authority developed Charlie the Chatbot to provide taxpayers with answers to certain questions.

33. Although, as noted above, HMRC has teams that assist with the development of such technology, it is understood that HMRC also buys technology services from third parties to assist it with developing its systems. However, from discussions with HMRC, it is understood that HMRC maintains overall ownership of the algorithms.

34. In terms of the training data used to train such models, HMRC uses data that it has collected from taxpayers. However, it is understood that in some cases HMRC also procures data from other sources, including data obtained from merchant acquirers. Where HMRC procures such data from other sources, it is unclear whether it relies on labelling (if relevant) from such merchant acquirers or whether it labels this data itself (where it uses such data for supervised learning models).

35. This paper recommends certain taxpayer protections that should be introduced to protect taxpayers where AI is being deployed in ADM. Although the use of AI by HMRC is not particularly widespread currently, the paper approaches these recommendations on a forward-looking basis, taking the position that it is inevitable that the use of AI in tax administration will increase in the near term.

36. Public administrations generally are looking at ways of integrating AI into their systems to assist with ADM with a view to enhancing efficiency and optimising resources (both with respect to tax and non-tax administration). This presents a unique opportunity to

take a proactive approach to regulating AI being used in ADM and establishing taxpayer protections from the outset, rather than adopting a reactive approach and regulating such AI after errors or biases are discovered. Given the relative opacity of AI (in terms of both the explainability of decisions and taxpayers' understanding of the technology), biases or errors in decisions rendered by AI have the propensity to perpetuate for prolonged periods without detection, and any such biases or errors could have pervasive and widespread effects on taxpayers. Therefore, it is imperative that a coherent policy for the use of such AI is developed *before* AI is widely deployed. The taxpayer safeguards recommended in this paper for the use of AI will be equally applicable to the use of algorithmic systems in ADM (although algorithmic systems are not specifically discussed).

5. Government AI standards and the need for AI tax legal safeguards

37. The UK government and international organisations have published various AI codes and frameworks, which set out certain non-binding principles and standards with respect to AI in the public sector. Some of these AI codes are briefly discussed below (although this list is not exhaustive). At the time of writing, there is no tax-specific published code or framework, although HMRC does have internal ethics guidelines that it follows when developing AI solutions.

38. The OECD has developed 'Principles for Responsible Stewardship of Trustworthy AI' (OECD AI Principles, initially published on 22 May 2019),[21] which the UK has signed up to. The OECD put forth five overarching general principles: (i) inclusive growth, sustainable development and well-being; (ii) human-centred values and fairness (such as dignity and autonomy, privacy and data protection, non-discrimination and equality, fairness, social justice, etc.); (iii) transparency and explainability (including enabling those affected by AI systems to challenge its outcome based on information about the facts, and on the logic that led to a particular recommendation, decision or prediction); (iv) robustness security and safety; and (v) accountability. As these are overarching principles, the effectiveness of these policies depends on domestic implementation.

39. The Centre for Digital and Data Office published the 'Generative AI Framework for HM Government' (Generative AI Principles)[22] on 18 January 2024, which HMRC has signed up to. This framework sets out ten common principles directed at UK government departments to guide them in their development and use of generative AI in the public sector: (i) 'you know what generative AI is and what its limitations are'; (ii) 'you use generative AI lawfully, ethically and responsibly' (e.g. AI is used in compliance with data protection, privacy, equality, intellectual property and other rules, and government departments should seek to minimise biases at all stages of the AI life cycle); (iii) 'you know how to keep generative AI tools secure'; (iv) 'you have meaningful human control at the right stage' (which includes having an appropriately trained and qualified person

[21] See https://oecd.ai/en/ai-principles.
[22] See https://www.gov.uk/government/publications/generative-ai-framework-for-hmg.

reviewing outputs produced by AI and validating all decision-making that uses such AI outputs to arrive at a decision); (v) 'you understand how to manage the full generative AI life cycle'; (vi) 'you use the right tool for the job'; (vii) 'you are open and collaborative' (including explicitly identifying where responses to the public are generated through AI and disclosing where AI is being used in official duties); (viii) 'you work with commercial colleagues from the start'; (ix) 'you have the skills and expertise that you need to build and use generative AI'; and (x) 'you use these principles alongside your organisation's policies and have the right assurance in place'.[23]

40. The Generative AI Principles are a step in the right direction. However, from a taxpayer protection perspective these have limited effect because these are not justiciable; these are merely guiding principles and are not legally binding on HMRC. Furthermore, these principles (although expected to be iterative) are currently focused mainly on LLMs[24] (which are only a small subset of the types of applications in which AI (and generative AI) can be (and is already being) used by HMRC, as discussed above).

41. There are other helpful voluntary non-binding standards that have been developed, such as 'Implementing the UK's AI regulatory principles: initial guidance for regulators' (Regulatory Principles)[25] published on 6 February 2024, developed by the Department for Science, Innovation and Technology, which provides regulators with guidance on developing AI for and deploying AI in the public sector. There are also various technical standards that have been published or that are being developed, to provide guidance on issues such as safety, security and robustness, explainability and transparency, fairness, and others. Similar to the point noted on the Generative AI Principles, these are not justiciable. With respect to the technical standards, many of these are understood to be in the process of development and have not yet been published.

42. It is understood that HMRC itself has an internally developed ethics framework (HMRC Ethics Framework), which it uses when developing AI solutions. Details of this framework are not discussed further. From discussions with HMRC, it is understood that the ethics and safeguards with respect to AI and the development of the HMRC Ethics Framework is something that HMRC's Professional Services Committee has been focused on for some time and routinely keeps under review. HMRC is also understood

[23] The Centre for Digital and Data Office (2024). 'Generative AI Framework for HM Government (v 1.0)'. Available at: https://assets.publishing.service.gov.uk/media/65c3b5d628a4a00012d2ba5c/6.8558_CO_Generative_AI_Framework_Report_v7_WEB.pdf [Accessed: 8 June 2024].

[24] The Centre for Digital and Data Office (2024). 'Generative AI Framework for HM Government (v 1.0)', p. 7. Available at: https://assets.publishing.service.gov.uk/media/65c3b5d628a4a00012d2ba5c/6.8558_CO_Generative_AI_Framework_Report_v7_WEB.pdf [Accessed: 8 June 2024].

[25] See https://www.gov.uk/government/publications/implementing-the-uks-ai-regulatory-principles-initial-guidance-for-regulators.

to have consulted (and continues to consult) on managing these risks internally and externally. However, similar to the frameworks and codes discussed above, the HMRC Ethics Framework, while useful as an internal sense-check when developing AI, does not provide taxpayers with any redress, including where AI solutions are deployed that do not properly take into account the factors in the HMRC Ethics Framework or violate some of the standards set out therein, or where the deployment of a model has been signed off despite some of the risks flagged through the HMRC Ethics Framework.

43. It is important that consideration is given as to what regulation of such AI should look like, particularly before widespread deployment of AI in ADM for the purpose of tax administration, as the use of AI in tax administration represents a fundamental shift in the basis of decision-making by HMRC. Although *pure* discretionary decisions are not currently being made by AI (based on information seen by the TLRC), it is inconceivable that AI will not be used in some format to automate discretionary decisions going forward, given the patent advantages of AI as discussed above.

44. Prior to the use of computers in tax administration, a decision made by an HMRC officer purely reflected a decision by a human; if challenged, a human HMRC officer could provide reasons as to why they arrived at that decision. After computers started being used in tax administration, a decision made by an HMRC officer may have been assisted or vetted by the use of computers (or the computers implemented a decision already made by HMRC that had been programmed into the algorithm; e.g. as in the case of the Schedule 55 decisions discussed above). However, even in these cases, the HMRC officer still remained the primary decision-maker, and if challenged, an HMRC officer could still provide reasons as to why a particular decision was arrived at. It is arguable that with respect to AI solutions currently being used by HMRC described above (e.g. risk management technology), the AI still only aids an HMRC officer in making a decision (i.e. the HMRC officer remains the primary decision-maker).

45. However, where decisions are made purely by AI (in particular ML) (without human intervention in the decision-making process), these in effect reflect a decision made by the system based on patterns, correlations or inferences drawn from the data (whether labelled or unlabelled) on which the system was trained. The decision is based on the model's own interpretation. Even in situations where an HMRC officer reviews a decision made by AI before the decision has an impact on a taxpayer and the HMRC officer provides an explanation of why a particular decision was arrived at, where a

black box model[26] has been used, that explanation is just of the officer's *understanding* of why a particular decision was arrived at. This is because decision-making in black box systems is typically opaque, and the process by which such a system arrives at a decision is generally not interpretable. Therefore, to provide an explanation of why a decision was arrived at where a system uses a black box model, an HMRC officer would need to reverse engineer why a decision was arrived at by the system – this process inherently involves an element of uncertainty.[27] This reflects a shift in the role of primary decision-maker from the HMRC officer to the AI system; the HMRC officer no longer explains why they made a certain decision but rather interprets why the AI arrived at a decision.

46. This is a fundamental shift in the approach to tax administration and any such shift should be the product of a conscious policy decision made by the government and HMRC whereby it is accepted and acknowledged that HMRC 'practice' in areas utilising AI for ADM will transition from human-developed practice to a system-developed practice. Any such policy change to tax administration should be clearly disclosed to taxpayers. Incremental creep in the use of AI in ADM in tax administration without taxpayer rights pro-actively being updated to keep pace with the change in status quo would leave a vacuum in taxpayer protections.

47. This change should be facilitated by proactive reconsideration of taxpayer rights with respect to this changing paradigm. Non-binding guidelines and principles do not go far enough to protect taxpayer rights as these are non-justiciable – in other words, there are no consequences for HMRC if these guidelines are not complied with. The legislation described in Section 1 was drafted and catered to a world where HMRC officers (i.e. humans) were the primary decision-makers; for example, at the time the TMA was initially drafted, the use of AI in tax administration was virtually inconceivable. Amendments to the legislation introduced to facilitate the use of computers in tax

[26] A black box model is 'any AI system whose inner workings and rationale are opaque or inaccessible to human understanding' and black box models include artificial neural networks amongst others. This definition is from: Information Commissioner's Office and the Alan Turing Institute (2022). 'Explaining decisions made with AI', p. 69. Available at: https://ico.org.uk/media/for-organisations/guide-to-data-protection/key-dp-themes/explaining-decisions-made-with-artificial-intelligence-1-0.pdf [Accessed: 8 June 2024].

[27] There are various supplementary models that have been (and are being) developed to provide explanations of decisions reached by black box AI systems (AI incorporating such tools are also known as explainable AI or 'XAI'). Some examples of these supplementary models are sensitivity analysis, layer-wise propagation (LRP), local interpretable model-agnostic explanations (LIME) and self-explaining and attention-based systems. However, at this stage of development, these models are generally only 'approximations'. See Information Commissioner's Office and the Alan Turing Institute (2022). 'Explaining decisions made with AI', p. 74. Available at: https://ico.org.uk/media/for-organisations/guide-to-data-protection/key-dp-themes/explaining-decisions-made-with-artificial-intelligence-1-0.pdf [Accessed: 8 June 2024]. Even where these supplementary models are incorporated in the AI, this does not do anything to dispel the proposition above that the HMRC officer is *merely reverse engineering why a decision was reached*, even if a supplementary model is used with this reverse engineering process.

administration (e.g. s.103 FA 2020) may intend to provide some legitimacy to the use of AI in tax administration (although see the discussion on *Peter Marano v The Commissioners for His Majesty's Revenue & Customs* [2024] EWCA Civ 876 below), but do not in any way ensure that taxpayer protections are updated to mitigate the risks associated with the use of AI. Taxpayers need to be provided with justiciable rights with respect to the use of AI in ADM in tax administration. Given that the basis of decision-making will change with AI (especially ML) being the primary decision-maker, such rights should provide taxpayer protection with respect to the development, training, deployment and testing of such technology.

48. It is recommended that a proactive approach is adopted because of the widespread and pervasive effects that AI (in particular ML) can have once widely deployed in public administration. As noted above, where black box AI is deployed, it can be extremely difficult to detect and rectify issues in the AI (including biases, hallucinations (in LLMs), etc.) and such AI can have the effect of institutionally entrenching certain issues or biases given their widespread deployment. An argument often raised against specific regulation of AI (particularly in tax administration) is that errors in 'judgement' or discretion by AI are not particularly concerning given that in a non-AI world, where an HMRC officer exercises discretion, the reason for the exercise of such discretion is similarly opaque, subjective and could also in theory be subject to personal biases that are subconscious or uncontrollable. This argument does not appreciate the scale of impact that AI can have; for example, in the Dutch *toeslagenaffaire*, approximately 11,000 parents were subject to audits on a discriminatory basis as a result of AI used by the Dutch tax authorities[28] (see the sections below for a fuller discussion on *toeslagenaffaire)* and in the Australian Robodebt matter AUS $746 million was erroneously recovered from 381,000 people (with AUS $1.751 billion of debt having to be written off).[29] The difference between human decision-making and AI decision-making is that a biased individual making decisions is typically constrained by the number of decisions a human can make, whereas the objective of AI is to *institutionally* replace human decision-making, thereby expanding the scale of any issues in AI decision-making to conceivably all decisions made by the institution. This is the reason why the scale of impact that decisions made by AI can have (as illustrated in the examples above) is materially different from the scale of impact that decisions made by a single human HMRC officer or a group of HMRC officers can have. Even in

[28] Hadwick, D. and Lan, S. (2021). 'Lessons to be learned from the Dutch Childcare Allowance Scandal: a comparative review of the algorithmic governance by tax administration in the Netherlands, France and Germany'. *World Tax Journal*, November 2021, p. 619.

[29] Royal Commissioner (2023). 'Report of the Royal Commission into the Robodebt Scheme', p. xxix. Available at: https://robodebt.royalcommission.gov.au/system/files/2023-09/rrc-accessible-full-report.PDF [Accessed: 9 June 2024].

circumstances where the percentage error margin in decisions made by AI are low, given the role of AI is to *institutionally* replace human decision-making on a go forward basis, the absolute number of taxpayers affected by such errors will be higher than the number of taxpayers affected by an individual biased HMRC officer (or a group of HMRC officers) making decisions.[30] Further, an individual HMRC officer can be questioned (internally or externally) on why a specific decision was made, whereas, for the reasons explained in this paper, at present a decision made by black box AI cannot be definitively explained due to the complexity of the technology and inadequacy of XAI.

49. The recommendation to have a proactive approach to the use of AI in tax administration and providing taxpayers with justiciable rights is not altered even where an HMRC officer is interposed between the decisions being made by the AI and a taxpayer being notified of the decision (i.e. when there is 'a human in the loop'). Where an HMRC officer merely vets decisions (or a sample of decisions) made by the AI (in particular ML), the main issue (especially where black box AI is used) is the ability to understand why a decision was arrived at by the AI (e.g. why an input variable was weighted more than others, the logical flow of a decision, whether a relevant variable was included during the training phase, etc.). Further, it is not inconceivable that given the volume of decisions being made by HMRC and the sophistication of the technology (especially black box AI) that an inertia to overturn decisions made by AI develops (i.e. automation bias). Where the human in the loop involves HMRC officers thoroughly re-reviewing decisions made by the AI (or a large cross-section of decisions made by the AI), this would be a waste of public resources because the AI and HMRC officers would be reviewing the same decision twice, and the objective of the AI is to replace human decision-making. For example, with respect to the Universal Credit advances model discussed in paragraph 64 below, the Public Law Project posits – based on public statements of the Department for Work and Pensions – that the Universal Credit advances model, used for triaging applications for advance payments of Universal Credit, does not provide any explanation of why certain cases are flagged for review by the system, and caseworkers are required to thoroughly re-review each flagged case again.[31] Further, where *every* decision is thoroughly re-reviewed as a matter of course and decisions are eventually made by humans (even where this uses decisions derived from ADM systems using AI), this has the potential to add back human subjectivity in to

[30] Williams, R. (2021). 'Rethinking administrative law for algorithmic decision making'. *Oxford Journal of Legal Studies*, 42(2), p. 486.

[31] Public Law Review (2024). 'Proposed claim for judicia review against the Secretary of State for Work and Pensions in relation to his unlawful use of automation to suspect payment for Universal Credit and/or to triage applications for advanced payment of Universal Credit'. 19 April. [Letter], p. 5. Available at: https://publiclawproject.org.uk/content/uploads/2024/08/Work-Rights-Centre-PAP-For-Publication_Redacted.pdf [Accessed: 24 August 2024].

the making of that decision, and this also would have the effect of eliminating the consistency offered by AI (which is one of the key benefits of AI).

50. The above should not be confused with periodic checks of a cross-section of decisions and system testing. Such periodic checks would be *post-facto* (for example, as part of a yearly audit) and would be undertaken after decisions made by the AI have been notified to taxpayers. The 'human in the loop' concept described in the paragraph above requires checks *before* decisions are notified to taxpayers. Periodic checks of a cross-section of decisions made by the AI are necessary, but these should be part of a wider package of taxpayer safeguards and are not adequate as the only taxpayer safeguard.

51. The risks of miscarriage of justice in relation to tax administration are particularly heightened given that decisions on tax have a direct tangible financial impact on taxpayers and can cause significant hardship to more vulnerable members of society, for example low-income taxpayers (including physical, psychological and other impacts). The need for tax protections and the impact that decisions in relation to tax can have on taxpayers are well documented and are not discussed in further detail.

52. For the reasons above, to the extent that AI is used in ADM for tax administration, taxpayers should be provided with safeguards that are justiciable, measurable and legally enforceable. The most robust way that this can be addressed is through tax-specific AI legislation that includes specific taxpayer protections giving taxpayers the ability to bring a claim where any of the protections or rights set out in the legislation are violated. Practically, however, there may be governmental preference to legislate for broader AI legislation (which is not tax-specific) and this may present some resistance to tax-specific AI legislation.

53. The alternative (less robust) recommendation is for the government to legislate for the development of a *new* HMRC AI Charter, which sets out some of the key standards and values that HMRC must adhere to where AI is used in ADM in tax administration (hereafter, the HMRC AI Charter). The current HMRC Charter on its own does not adequately provide taxpayers with rights that are justiciable, measurable and legally enforceable (even if some of the standards in the HMRC Charter (e.g. 'getting things right', 'treating you fairly', 'being aware of your personal situation' and 'keeping your data secure') could be interpreted to apply to AI used in ADM in tax administration. This is because the current HMRC Charter provides limited redress to taxpayers where the standards in the HMRC Charter are not complied with; the HMRC Charter merely sets out 'the standards of behaviour and values that [HMRC] will *aspire* [to] when dealing with people in the exercise of their function' (s.16A Commissioners for Revenue & Customs Act 2005) [emphasis added]. This does not create any rights that are

challengeable in the tax tribunals.[32] There are arguments that the HMRC Charter may create legitimate expectation as a ground for judicial review, but this is still uncertain and not established. In any event, judicial review is a high standard, and to rely on judicial review as the only form of redress for taxpayers on an ongoing basis does not adequately build in the protections required in the context of AI ADM. Therefore, the enabling legislation implementing any HMRC AI Charter should not be drafted as mere aspirations of HMRC but should be drafted as more affirmative obligations of HMRC. Further, the standards and values themselves should be narrower and more targeted (as opposed to the standards in the HMRC Charter, which are broad), which enable performance against these standards to be properly measurable. The rest of this paper refers to any tax-specific AI legislation and/or HMRC AI Charter as 'AI Tax Legal Safeguards'.

54. Some of the arguments against regulating AI have been that over-regulation of AI could hamper growth, development and investment in this area. While this is acknowledged as a key policy factor in the private sector, the weighting of factors in the public sector are necessarily different. Regulating the use of AI is necessary to ensure that key government functions are consistently and robustly performed to an acceptable standard within the rule of law, without compromising taxpayer protections and the fundamental principles of democratic government.

55. In Section 6 onwards, we discuss some provisions that should be included in any AI Tax Legal Safeguards.

[32] In the HMRC Charter annual report 2022 to 2023 (published on 17 July 2023), it was reported that there were 91,217 new complaints for 2022/2023 and the issues with meeting the standards in the HMRC Charter were discussed. Also referenced in Closs-Davies, S. Burkinshaw, L. and Frecknall-Hughes, J. (2024). 'Is the HMRC Charter fit for purpose: experiences of tax practitioners and vulnerable citizens'. *British Tax Review*, 2024(2), 304–326 (see p. 307).

6. Enabling legislation for the use of AI in ADM in tax administration

56. It may be argued that s.103 FA 2020 can be interpreted to provide HMRC with the power to use AI for ADM in a wide range of circumstances (i.e. in any situation where a function is conferred on an HMRC officer under legislation relating to tax). However, in the Court of Appeal's judgment in *Peter Marano v The Commissioners for His Majesty's Revenue & Customs* [2024] EWCA Civ 876) Asplin LJ (at paragraph [37]) noted with respect to s.103 FA 2020: '[w]ho or what is HMRC for these purposes…HMRC is being referred to the body or department itself, albeit a body comprised of the Commissioners and officers of Revenue and Customs… [o]bviously, the body, which is an emanation of the State can only act through individuals whether they use computers or not. It is common ground that HMRC do not have computers which make decisions themselves. Section 103 is *not intended to authorise the use of artificial intelligence*' [emphasis added].

57. Irrespective of whether tax-specific legislation or an HMRC AI Charter is adopted, a legislative provision should be enacted to affirmatively provide HMRC with the power to use AI in tax administration. Given the points made above, primarily that where AI (in particular ML) is used in ADM in tax administration there is a material risk that the primary decision-maker shifts from the HMRC officer to the AI itself (and the Asplin LJ's judgment above), it is important that the legitimacy of decisions made by HMRC using AI are put beyond doubt, and this can only be done through affirmative legislation.

58. This enabling legislation in relation to the use of AI is also important as a function of the democratic process. Legislating to enable the use of AI (rather than relying on pre-existing powers provided (e.g. those under s.103 FA 2020)) will permit open debate in Parliament on the use of AI in tax administration.

59. It is important that the use of AI in ADM in tax administration is subject to Parliamentary debate (even where AI is being used in risk management systems) because the basis of decision-making will fundamentally change with the use of AI and this should be the product of conscious policymaking. This is given further weight by the Finance Bill Committee debates on s.103 FA 2020, which took place as part of the

Finance Bill 2020; one of the key reasons implied for the introduction of the legislation was to:

> 'put beyond doubt that the law operates in the way Parliament intend it to *and as it has been widely understood to work to date*. It *does not introduce new or additional obligations, and will help to ensure that the tax system applies fairly to all*, while preventing loopholes opening up in tax law that could be exploited by people who do not wish to pay their proper share of tax' [emphasis added].[33]

It is arguable that this premise ceases to exist with the use of AI for the reasons discussed above and there should be new debate on the use of AI in tax administration.

60. AI Tax Legal Safeguards should then provide for limitations on the use of AI in ADM in tax administration. Various limitations on the use of AI have been recommended across the various standards noted in Section 5 above. Further, in the guidance published by the Information Commissioner's Office and the Alan Turing Institute,[34] various recommendations have been made on the limitations on the use of AI. These limitations should be consulted on through industry-wide consultations to ensure that appropriate safeguards are built to protect taxpayers.

61. As an example, Germany has adopted a s.88(5) in its Fiscal Code (*Abgabenordnung*), whereby it specifically provides for the use of risk management systems to determine whether 'further investigations and reviews are necessary to ensure the consistent and lawful assessment of taxes and tax rebates and the consistent and lawful crediting of withheld taxes and prepayment'.[35] This provision further sets out certain minimum basic requirements for the use of such systems and reviews. Although it is considered that this enabling legislation is narrow in scope, it serves as a useful example of countries specifically legislating for the use of AI.

62. In addition, there are other restrictions on the use of AI used in ADM in tax administration that HMRC needs to consider before the deployment of any such technology to ensure that AI is not subject to challenge, and consideration of these factors should be noted in the AI Tax Legal Safeguards. For example, there are

[33] HC Deb (18 June 2020). Jesse Norman. Available at: https://hansard.parliament.uk/Commons/2020-06-18/debates/27174a3a-c211-4174-897a-5d4f3f5a3986/FinanceBill(NinthSitting)?highlight=tax%20law%20review%20committee#contribution-E6670301-54E4-4447-8101-9BCB245D658B [Accessed: 4 August 2024].

[34] Information Commissioner's Office and the Alan Turing Institute (2022). 'Explaining decisions made with AI'. Available at: https://ico.org.uk/media/for-organisations/guide-to-data-protection/key-dp-themes/explaining-decisions-made-with-artificial-intelligence-1-0.pdf [Accessed: 8 June 2024].

[35] An English translation of the Fiscal Code of Germany is available at: https://www.bundesfinanzministerium.de/Content/EN/Downloads/Resources/Laws/2018-03-26-fiscal-code.pdf?__blob=publicationFile&v=4 [Accessed: 15 June 2024]

questions on the compatibility of the use of AI under GDPR and the DPA, the ECHR, the Equality Act 2010 (including the public sector equality duty (PSED)) and the HMRC Charter, etc., and any AI Tax Legal Safeguards should explicitly require HMRC to consider or evaluate these factors before AI is deployed in ADM in tax administration. There are also questions around the compatibility of the use of AI to make completely automated decisions without any enabling legislation with GDPR and the DPA, which HMRC will need to engage with.[36]

63. Some of these issues have presented themselves in other jurisdictions. For example, in the Dutch case *NCJM et al. and FNV v State of Netherlands (*District Court of the Hague, 6 March 2020, ECLI:NL:RBDHA:2020:865), the legislation relating to, and the use of, System Risk Indicator (SyRI), a social welfare fraud risk assessment technology, was found not to comply with Article 8(2) of the ECHR.[37] SyRI utilised various datasets that it procured from various government agencies (including various unknown datasets) to create risk profiles for individuals that indicated an assessment of the fraud risk; in particular, the various assessees were not notified of any risk assessments and there was no transparency on what factors contributed to a particular risk assessment.[38] In reaching its decision, the Hague District Court drew heavily (amongst others) on principles in EU GDPR.[39] Although the Hague District Court found that SyRI did not utilise deep learning,[40] the principles still apply. Similarly human rights, data protection and legality issues[41] were found by the Slovak Supreme Constitutional Court in the *eKasa* case (492 finding of the Constitutional Court of the Slovak Republic, PL. ÚS 25 /2019-117) with the Slovak tax authority's use of automatic processing of certain data (and related ML algorithms) used for the purpose of risk assessing taxpayers for VAT fraud based on data collected from sellers.[42] A detailed analysis of the ECHR and GDPR consequences of the use of AI in ADM in tax administration is beyond the scope of this paper.

[36] Williams, R. (2021). 'Rethinking administrative law for algorithmic decision making'. *Oxford Journal of Legal Studies*, 42(2), p. 472.
[37] Rachovista, A. and Johann, N. (2022). 'The human rights implications of the use of AI in the digital welfare state: lessons learned from the Dutch *SyRI* Case'. *Human Rights Law Review*, 22, ngac010, p. 1.
[38] Rachovista, A. and Johann, N. (2022). 'The human rights implications of the use of AI in the digital welfare state: lessons learned from the Dutch *SyRI* Case'. *Human Rights Law Review*, 22, ngac010, p. 3.
[39] Rachovista, A. and Johann, N. (2022). 'The human rights implications of the use of AI in the digital welfare state: lessons learned from the Dutch *SyRI* Case'. *Human Rights Law Review*, 22, ngac010.
[40] Rachovista, A. and Johann, N. (2022). 'The human rights implications of the use of AI in the digital welfare state: lessons learned from the Dutch *SyRI* Case'. *Human Rights Law Review*, 22, ngac010, p. 5.
[41] University of Antwerp. The eKasa case, *AI TaxAdmin.EU*. Available at: https://www.uantwerpen.be/en/projects/aitax/publications/ekasa/#:~:text=The%20eKasa%20case%20concerns%20proceedings,risk%2Dscoring%20algorithms%20%E2%80%93%20and%20the [Accessed: 15 June 2024].
[42] Kuźniacki, B. and Hadwick, D. (2023). '(Non)natural born killers of XAI in tax law: trade secrecy, tax secrecy and how to kill the killers'. *Kluwer International Tax Blog*. Available at: https://kluwertaxblog.com/2023/09/12/nonnatural-born-killers-of-xai-in-tax-law-trade-secrecy-tax-secrecy-and-how-to-kill-the-killers/ [Accessed: 9 June 2024].

64. Challenges to the use of AI in public administration are already starting to present themselves and it is expected that as the use of AI in public administration continues (or existing use of such AI becomes more publicised), this will continue. For example, on 19 April 2024, Work Rights Centre (represented by the Public Law Project) delivered a letter to the Secretary of State for Work and Pensions in accordance with the Pre-Action Protocol for Judicial Review proposing to bring a claim with respect to the use by the Department for Work and Pensions of automated systems (suspected of utilising ML) to identify and suspend payments with respect to fraudulent Universal Credit claims and to triage applications for advance payment of Universal Credit (in the latter case using a system known as the Universal Credit advances model). The validity of the claims have not been vetted and no opinion is offered on the strength of these claims.[43]

Recommendations

- **Legislation should affirmatively provide for the use of AI in ADM in tax administration.**
- **AI Tax Legal Safeguards should set out specific circumstances where the use of AI is impermissible.**
- **AI Tax Legal Safeguards should also require consideration of other relevant legislation including GDPR, the DPA (and processing of data in accordance with GDPR and the DPA), the ECHR, Equality Act 2010, the HMRC Charter, etc.**
- **AI Tax Legal Safeguards should specify how taxpayer risk levels will be determined following the processing of data in risk management systems (e.g. Connect) (although it is noted that, for security and public interest purposes, there may be grounds to limit such disclosure, and the exact disclosure should be the subject of public consultation).**

[43] Public Law Review (2024). 'Proposed claim for judicia review against the Secretary of State for Work and Pensions in relation to his unlawful use of automation to suspect payment for Universal Credit and/or to triage applications for advanced payment of Universal Credit'. 19 April. [Letter]. Available at: https://publiclawproject.org.uk/content/uploads/2024/08/Work-Rights-Centre-PAP-For-Publication_Redacted.pdf [Accessed: 24 August 2024].

7. Training data and bias

65. Training is a fundamental step in developing an AI model (as discussed above). Where an AI model (especially an ML model) is being deployed to make decisions in a public sector context, the model must be trained on datasets that are large, diverse, reliable and unbiased. This has been acknowledged and discussed in the Generative AI Principles. The main advantage of using AI in ADM in tax administration would be because AI has the capability to distinguish between taxpayers based on their individual circumstances and to weigh factors based on such differences in arriving at a decision – similar to human cognition.[44] Therefore, where a model is not trained on the right datasets, in simple terms, it impacts the ability of the model to correctly distinguish between taxpayers and weigh factors appropriately, creating discriminatory effects or bias.[45]

66. The House of Commons Science and Technology Committee identify four causes of bias or discrimination with respect to AI: (i) the use of inappropriate data; (ii) lack of data; (iii) correlation disguised as causation; and (iv) unrepresentative development teams.[46,47] The first three factors are briefly discussed below. The fourth factor is self-explanatory.

Inappropriate data

67. This risk has been widely documented and arises as a consequence of the training data importing human bias into the system. As explained earlier, the model will draw correlations and inferences based on the data that it has been trained on (whether labelled or unlabelled) and so if the data is biased and inherently discriminatory, the model will be biased and discriminatory. In the context of ADM, bias could be

[44] House of Commons: Science and Technology Committee (2018). 'Algorithms in decision-making: Fourth Report of Session 2017–19, HC 351, p. 7. Available at: https://publications.parliament.uk/pa/cm201719/cmselect/cmsctech/351/351.pdf [Accessed: 3 June 2024].

[45] House of Commons: Science and Technology Committee (2018). 'Algorithms in decision-making: Fourth Report of Session 2017–19', HC 351, pp. 18–19. Available at: https://publications.parliament.uk/pa/cm201719/cmselect/cmsctech/351/351.pdf [Accessed: 3 June 2024].

[46] House of Commons: Science and Technology Committee (2018). 'Algorithms in decision-making: Fourth Report of Session 2017–19', HC 351, p. 19. Available at: https://publications.parliament.uk/pa/cm201719/cmselect/cmsctech/351/351.pdf [Accessed: 3 June 2024].

[47] There are various other forms of bias that may exist and some of these may be seen as subcategories of the forms of bias listed above. These are not explored herein; however, a helpful summary is provided in Mehrabi, N., Morstatter, F., Saxena, N., Lerman, K. and Galstyan, A. (2022). 'A survey of bias and fairness in machine learning' [v3], arXiv:1908.09635v3. Available at: https://arxiv.org/abs/1908.09635 [Accessed: 4 August 2024].

introduced by something as simple as an overemphasis on a particular factor.[48] This issue arose in the Netherlands as part of the *toeslagenaffaire*, where an AI risk management software that was used to detect fraud in childcare allowance claims erroneously flagged families that had foreign backgrounds or non-Dutch nationalities as fraud risks, and these families were subject to additional audits by the tax authority.[49] This was because, as part of various input factors used, (i) the risk management software used foreign backgrounds and Dutch/non-Dutch nationality as input factors, (ii) the Dutch tax authorities disproportionately opened audits with respect to parents belonging to certain nationalities (for example, because they detected 120 to 150 fraudulent requests from an IP address linked to a Ghanian institution, the Dutch tax authorities opened an investigation into all Ghanian welfare recipients in the Netherlands (approximately 6,047 recipients), (iii) the Dutch tax authorities overemphasised the importance of nationality in opening these audits, and (iv) the system used this increased incidence of audits into parents of certain nationalities to train itself (and thereby entrenching bias); this resulted in a third of the parents flagged by the AI system (approximately 11,000 parents) being flagged because of their non-Dutch nationality.[50]

Example 3.

Where a model is being developed to determine a discretionary penalty for delay in paying a tax, and the dataset used to train the model overemphasises the importance of the number of days of delay as opposed to other factors, then the model will also overemphasise the importance of the time by which the tax payment is delayed, irrespective of any other mitigating factors.

Insufficient data

68. A dataset used to train a model needs to be large (so that there are adequate data points for appropriate correlations or inferences to be drawn, including where there are slight variations in the facts) and diverse (so that the system has data points on a vast cross-section of the relevant demographic, including under-represented groups).[51] Where the

[48] Kissinger, H., Schmidt, E. and Huttenlocher, D. (2021). *The Age of AI: And Our Human Future.* London: John Murray Press, p. 80.
[49] Hadwick, D. and Lan, S. (2021). 'Lessons to be learned from the Dutch childcare allowance scandal: a comparative review of the algorithmic governance by tax administration in the Netherlands, France and Germany'. *World Tax Journal*, November 2021, pp. 619–623.
[50] Hadwick, D. and Lan, S. (2021). 'Lessons to be learned from the Dutch childcare allowance scandal: a comparative review of the algorithmic governance by tax administration in the Netherlands, France and Germany'. *World Tax Journal*, November 2021, pp. 619–623.
[51] Kissinger, H., Schmidt, E. and Huttenlocher, D. (2021). *The Age of AI: And Our Human Future.* London: John Murray Press, p. 65.

dataset does not meet either of these criteria, the correlations or inferences drawn may be inappropriate and could discriminate against certain taxpayers.

Example 4.

Where a model is being developed to risk assess individual taxpayers, and is trained on a small dataset, there is a risk that the dataset lacks the data needed by the model to draw the right correlations or inferences and that decisions made by the model will be misleading or inaccurate. Alternatively, if there is a large dataset used, but the dataset contains little or no data on low-income taxpayers, or taxpayers in the informal sector, or taxpayers with disabilities, the model may still produce decisions that are misleading or incorrect; this is because the model does not have enough data on such taxpayers to draw the right correlations or inferences with respect to such persons.

Another example of this could be where AI is developed to determine penalties payable under s.109C TMA. This is because it is understood that there are limited instances of HMRC having used its power under s.109C to impose a penalty on a taxpayer.

Correlation disguised as causation

69. As discussed above, the trained model will make decisions based on *correlations* or *inferences* that it has drawn from the data. Therefore, it is possible that although the dataset is not inherently discriminatory, there are data points in the training data that are proxies for other data points which may be discriminatory; therefore, when the model learns from that data it draws correlations that generate bias in the model.[52] While developing the model, parameters can be set to mitigate this risk; however, this does not entirely eliminate the risk of bias creeping in through this form of discrimination. This form of discrimination became apparent with respect to the Correctional Offender Management Profiling for Alternative Sanctions (COMPAS) risk management system used in the United States, which provided offenders with risk scores assessing the likelihood of these offenders committing future crimes (i.e. recidivism); these risk scores were then used in making decisions on parole for such individuals.[53] Bias was reported in relation to the system on the basis that the false positive rate for Black offenders was higher than that for White offenders (i.e. a greater number of Black Americans who would not have reoffended were erroneously predicted to be at risk of being

[52] House of Commons: Science and Technology Committee (2018). 'Algorithms in decision-making: Fourth Report of Session 2017–19, HC 351, p. 21. Available at: https://publications.parliament.uk/pa/cm201719/cmselect/cmsctech/351/351.pdf [Accessed: 3 June 2024].

[53] Angwin, J., Larson, J. Mattu, S. and Kirchner, L. (2016). 'Machine bias'. *ProPublica*, 23 May. Available at: https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing [Accessed: 11 June 2024].

rearrested),[54] and this bias arose even though race was not understood to be an input data point included in the dataset.[55] Therefore, there are expected to have been other data points in the dataset that were proxies for race that resulted in a biased model (even though race was not explicitly used as a data point). For example, certain research has shown that there may be different approaches to law enforcement with respect to Black and White Americans in the US in certain circumstances (e.g. Black Americans have been arrested for marijuana offences more than White Americans, although both use marijuana at approximately equal rates), and where the number of arrests may have been taken into account by COMPAS in determining recidivism, this may have had a disproportionate effect on Black Americans because of the attitude to enforcement historically taken in the US.[56]

70. From the above, it is clear that AI (particularly ML) developed using inappropriate datasets can result in discrimination where this is deployed in ADM. The opacity of ML (and other AI technology) makes it imperative that *appropriate technical standards to mitigate the risk of bias* are followed at the *outset* while developing such AI. If such standards are not adhered to, any discrimination imported through the datasets could be pervasive and could affect large groups of people before detection, given that such technology is usually employed in areas where large volumes of decisions are made in short timespans. The opacity of the technology arises because of the black box nature of the technology (which makes flaws in the correlations or inferences drawn by the models generally difficult or impossible to detect); the flaws in the inferences or correlations drawn by the technology often only come to light once the technology has been deployed and has had a significant impact on large groups of people. It is imperative that the AI Tax Legal Safeguards provide for the use of appropriate technical standards to mitigate the risk of bias and that where such standards are not used, taxpayers are provided with rights. Various non-governmental standards for mitigating bias have been developed and these standards have been referenced in the Regulatory Principles. However, as noted above, the Regulatory Principles and use of these non-governmental standards are non-binding and not specific to tax. Further, by providing the standards that must be adhered to in the AI Tax Legal Safeguards, this builds in an element of transparency on how decisions using AI (and in particular ML) are made.

[54] Mayson, S. G. (2019). 'Bias in, bias out'. *The Yale Law Journal*, 128, p. 2218. Available at: https://scholarship.law.upenn.edu/cgi/viewcontent.cgi?article=3396&context=faculty_scholarship [Accessed: 4 August 2024].

[55] Angwin, J., Larson, J. Mattu, S. and Kirchner, L. (2016). 'Machine bias'. *ProPublica*, 23 May. Available at: https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing [Accessed: 11 June 2024].

[56] Mayson, S. G. (2019). 'Bias in, bias out'. *The Yale Law Journal*, 128, p. 2218. Available at: https://scholarship.law.upenn.edu/cgi/viewcontent.cgi?article=3396&context=faculty_scholarship [Accessed: 4 August 2024], p. 2255.

71. The requirement for the technology and the datasets not to be discriminatory extends beyond the scope of the PSED provided for in section 149(1) of the Equality Act 2010 and the protected characteristics enshrined in the Equality Act 2010. As shown in some of the examples, discrimination may arise from categories of information that would not be considered protected characteristics or from proxies for such information. Therefore, it is important that when developing the technical standard discussed above to mitigate the risk of bias, there is wide stakeholder consultation to ensure that HMRC takes into account that the technical standard protects a wide cross-section of the demographic from discrimination through biased data. This positive obligation on HMRC should exist even where the data used to train a model is not HMRC's internal data (and is procured from a third party) or when HMRC purchases technology from a third-party provider. A similar issue was discussed in *R (on the application of Edward Bridges) v The Chief Constable of South Wales Police* [2020] EWCA Civ 1058, where the Court of Appeal found that the Chief Constable of New South Wales Police had not properly discharged its PSED with respect to the use of live automated facial recognition software that it had purchased from a third-party supplier because it did not seek to 'satisfy themselves, either directly or by way of independent verification, that the software program in this case does not have an unacceptable bias on ground of race or sex' with respect to the data that was used to train the technology.[57] As noted above, HMRC's obligation should go beyond the PSED, and there will need to be a positive obligation on HMRC to satisfy itself that there will be no bias in the training data or technology used in ADM in tax administration.

Recommendations

- **AI Tax Legal Safeguards should provide broad principles on the use of datasets for the training of AI used in ADM in tax administration to ensure that these datasets are (i) large, diverse, reliable and unbiased, and (ii) represent a wide cross-section of the demographic affected. These principles should apply even where the data used by HMRC is not internal HMRC data but is otherwise bought or procured from third parties or where the system is developed externally.**

- **AI Tax Legal Safeguards should make clear that the principles on datasets apply at all stages of the development and use of AI, and therefore the datasets used to train the AI should be kept under review even once the AI has been deployed.**

[57] *R (on the application of Edward Bridges) v The Chief Constable of South Wales Police* [2020] EWCA Civ 1058, paragraph 199.

- **AI Tax Legal Safeguards should provide for mandatory retraining of AI based on updated datasets, and AI Tax Legal Safeguards should set out the period after which there should be such mandatory retraining once AI has been deployed.**

- **AI Tax Legal Safeguards should require the development of a tax-specific technical standard (which should be periodically reviewed and updated) to minimise the risk of bias that must be adhered to when developing AI that is used for ADM in tax administration. The technical standard should be consulted on and should be adopted once it has had wide stakeholder engagement. Once adopted, this technical standard should be publicly disclosed.**

8. Testing and deployment

72. Given the risks with black box AI (especially ML), particularly as the deployment of AI in the public sector is still at a nascent phase, safeguards in relation to testing and deployment are important.

73. At present, it is unclear the extent to which HMRC tests new AI technology before deployment in tax administration. In general, deployment of AI without adequate testing pre-deployment and, instead, testing the AI post-deployment poses significant risks for the reasons stated above, i.e. detection can be difficult, and often detection would be after AI has had a material impact on taxpayers. AI systems should be rigorously tested *before* deployment. This is generally not contentious and is not developed further. However, a point to note in relation to this is that the metrics on which such technology is tested pre-deployment is important. For example, with respect to the COMPAS technology discussed above, supporters of the technology have argued that the technology achieves predictive parity (i.e. Black and White defendants who were classified as high risk were rearrested at the same rate – therefore, there was no discrimination), but detractors of the technology have argued that the technology was biased because there were greater false positives with respect to Black defendants than White defendants (i.e. the technology inaccurately labelled more Black defendants as high risk than it did White defendants – therefore, this unfairly discriminated against Black defendants).[58] Therefore, any pre-deployment testing and the metrics against which such testing is undertaken should be properly considered and consulted on.

74. It is also recommended that the AI Tax Legal Safeguards provide for a transition period with respect to any AI being deployed in ADM in tax administration (this is developed in the rest of this section). It is recommended that this approach should be adopted for at least the next few years while the use of AI in tax administration is still relatively new and while HMRC and the government look to refine their approach on the use and regulation of AI in the public sector. In the absence of such testing, there is a risk that, once deployed, AI does not operate in the manner intended; for example, in Australia in

[58] Mayson, S. G. (2019). 'Bias in, bias out'. *The Yale Law Journal*, 128, p. 2218 (see p. 2234). Available at: https://scholarship.law.upenn.edu/cgi/viewcontent.cgi?article=3396&context=faculty_scholarship [Accessed: 4 August 2024].

the Robodebt scandal, technology developed with the involvement of the Australian Tax Office entirely misapplied the law on identifying debtors.[59]

75. The transition period proposal set out below focuses on the deployment of AI in ADM where the AI is being used to make discretionary decisions, but similar rules could be applied for the use of AI in other areas of tax administration. The exact transition period and error ranges may vary based on the exact use case of the AI, but the AI Tax Legal Safeguards should provide a framework of the transition period.

The framework

76. Where AI is to be used to make automated discretionary decisions, once initial testing of the AI has been completed and AI has been deployed in live cases, the AI Tax Legal Safeguards should broadly provide for the following ongoing testing framework.

First post-deployment testing period

76.1. HMRC officers should in parallel review a certain percentage (for example 50%) of the decisions made by the AI before these are notified to taxpayers. This should be done for an initial period (the 'first post-deployment testing period').

76.1.1. The exact percentage of decisions reviewed by HMRC in parallel and the length of the first post-deployment testing period can vary depending on the AI being deployed, and should be determined by HMRC as a matter of policy based on a variety of factors (including the use, complexity, risk, etc., of the AI).

76.1.2. Review in this context refers to an HMRC officer conducting an independent review based on the same information available to the AI model and arriving at a decision unaided by the decision made by the AI. This is conceptually different to an HMRC officer merely reviewing and signing off on a decision already made by the AI. In the latter case, there is a risk of rubberstamping by the reviewing HMRC officer due to inertia to overturn a decision made by technology (i.e. automation bias). An independent review by an HMRC officer is necessary to ensure the efficacy of ongoing testing.

[59] Bentley, D. (2022). 'Tax Officer 2030: the exercise of discretion and artificial intelligence'. *eJournal of Tax Research*, 20(1), pp. 72–100 (see p. 80). Available at: https://www.unsw.edu.au/content/dam/pdfs/business/acct-audit-tax/research-reports/ejournal-of-tax-research/2022-volume-20-number-1/2022-11-Volume20-No1-P72.pdf [Accessed: 13 June 2024].

76.2. There should be a proper feedback loop put in place, which ensures that the decisions made by the AI and the decisions made by the HMRC officer in parallel are properly tracked and that any variances in decisions are appropriately evaluated. An acceptable error range should be determined by HMRC based on a variety of factors (including the use, complexity, risk, etc., of the AI).

76.3. If, at the end of the first post-deployment testing period, decisions made by the AI over the course of the first post-deployment testing period, do not fall within the defined error range, then a decision should be made by the Commissioners whether to retrain the AI, to discontinue the use of the AI or to continue the use of the AI.

Second post-deployment testing phase

76.4. If, at the end of the first deployment testing period, decisions made by the AI over the course of the first post-deployment testing period fall within the acceptable error range or if the Commissioners decide to continue with the use of the AI as discussed above, there should be a second transition period (the 'second post-deployment testing phase') where the percentage of decisions independently reviewed by HMRC officers in parallel (as explained above) should be brought down to a percentage determined by HMRC based on a variety of factors (including the use, complexity, risk, etc., of the AI) (for example 25%). If, at the end of the second post-deployment testing period, decisions made by the AI over the course of the second post-deployment testing period fall within the acceptable error range, then no further post-deployment parallel independent testing by HMRC officers is required. However, if decisions made by the AI do not fall within the acceptable error range, then the Commissioners would need to decide whether to retrain the AI, to discontinue the use of the AI or to continue with the use of the AI, and whether to subject the AI to further post-deployment testing periods.

76.5. *Prior* to deployment of the AI, HMRC should publish (i) the acceptable error ranges for the first and second post-deployment testing phases, (ii) the periods of testing for the first and second post-deployment testing phases and (iii) the percentage of decisions subject to parallel independent review by HMRC officers in the first and second post-deployment testing phases. HMRC should also publish the results of testing and decisions (if any) made by the Commissioners at the end of each post-deployment testing phase.

77. The AI Tax Legal Safeguards should provide for annual audits of the technology to ensure that the AI continues to operate as expected and within the acceptable error

range. The German Fiscal Code provides a similar provision in the context of the risk management systems discussed above at s.88(5)(4) of the German Fiscal Code (*Abgabenordnung*).[60] This annual testing should be rigorous and the impact of the AI on under-represented taxpayers should specifically be considered. Determining if there are under-represented taxpayers with respect to specific decisions being made by the AI will vary on a case-by-case basis, and this should be determined by HMRC.

78. Finally, where HMRC through the post-deployment testing or through annual audits, or otherwise, discovers errors in specific decisions made by the AI, or systemic errors in decisions by the AI, then where these decisions or outcomes deriving from such decisions have been shared with other governmental departments through the various information gateways and this information has been or can be used as a basis for such departments to make decisions (i.e. non-tax decisions) affecting the relevant taxpayers, there should be a means to rectify such information sent to the other departments, and for the other departments to rectify decisions they have made in reliance on the shared information. In the absence of this, taxpayers would indirectly suffer consequences as a result of technology deployed by HMRC, without avenues for easy redressal. This is not specifically something that can be dealt with in the AI Tax Legal Safeguards but will need to be part of wider government policy.

Recommendations

- **AI Tax Legal Safeguards should provide for testing prior to the deployment of AI.**
- **AI Tax Legal Safeguards should require a transition period where post-deployment testing is undertaken in parallel to the deployment of AI.**
- **AI Tax Legal Safeguards should require annual audits of deployed AI (including impact assessments on under-represented taxpayers).**
- **Wider government policy should review safeguards with respect to the HMRC information disclosure gateways.**

[60] An English translation of the Fiscal Code of Germany is available at: https://www.bundesfinanzministerium.de/Content/EN/Downloads/Resources/Laws/2018-03-26-fiscal-code.pdf?__blob=publicationFile&v=4 [Accessed: 15 June 2024].

9. Explainability and transparency

79. One of the main conclusions from the examples above (including *toeslagenaffaire* and COMPAS) is that public authorities should ultimately be accountable for their use of AI in administration. This extends to HMRC's use of AI in ADM in tax administration.

80. The obligation for HMRC to provide explanations and justify decisions forms a fundamental part of how taxpayers hold HMRC accountable,[61] especially given the fact that decisions made by HMRC tend to have tangible financial impacts on taxpayers. Although, based on recent jurisprudence, there has been a tendency for courts to find (based on various justifications) that public authorities are under a duty to provide reasons, this position is not entirely free from doubt under constitutional law, and the base view seems to be that currently there is no general duty to provide reasons under common law.[62] Similarly, although domestic legislation (mainly GDPR and the DPA) imposes obligations to provide explanations and reasons under certain circumstances,[63] such legislation was not developed specifically to protect taxpayer rights and the scope of when such explanations need to be provided can be limited, and what such explanation should exactly provide is unclear.[64]

81. For these reasons, it is important that the AI Tax Legal Safeguards provide that (i) taxpayers are notified where AI is used for ADM in relation to decisions having a direct impact on them (which includes instances where the decision would have a direct financial impact on the taxpayer and instances where there would not be a direct financial impact but there would be a legal impact on the taxpayer – for example, the TMA provisions described in Section 1) and (ii) taxpayers are provided with explanations of why that decision was made (in other words, taxpayers should be provided with an outcome-based local rationale explanation – based on the

[61] Turpin, C. and Tomkins, A. (2007). *British Government and the Constitution*, 6th edn. Cambridge: Cambridge University Press, p. 132.
[62] Mountfield KC, H. (2024). 'Duty to give reasons'. *Practical Law*. Available at: https://uk.practicallaw.thomsonreuters.com/Document/Ib5556dd8e83211e398db8b09b4f043e0/View/FullText.html?transitionType=SearchItem&contextData=(sc.Search)#co_anchor_a1036564 [Accessed: 14 April 2024].
[63] Information Commissioner's Office and the Alan Turing Institute (2022). 'Explaining decisions made with AI'. Available at: https://ico.org.uk/media/for-organisations/guide-to-data-protection/key-dp-themes/explaining-decisions-made-with-artificial-intelligence-1-0.pdf [Accessed: 8 June 2024].
[64] Tomlinson, J., Sheridan, K. and Harkens, A. (2020). Judicial review evidence in the era of the digital state, p. 20. Available at SSRN: http://dx.doi.org/10.2139/ssrn.3615312 [Accessed: 14 June 2024].

nomenclature used by the Information Commissioner's Office and the Alan Turing Institute).[65] There may be a case to exclude risk ratings assigned to taxpayers from the scope of these provisions included in the AI Tax Legal Safeguards, given some of the sensitivities around this and the fact that these ratings may not have a direct impact on the taxpayers, so long as the fact that AI is being used for risk rating taxpayers is publicly disclosed as per this paragraph. The exclusion of risk ratings from these provisions is subject to further analysis being done on how these risk ratings are being used by other government departments to make decisions directly affecting persons.

82. A distinction is drawn here between an explanation of *why* a decision was arrived at, and *why it is expected or understood that* a decision was arrived at. It is important that explanations provided by HMRC are able to provide the former. The difficulties with explainability of AI, especially black box AI, have been discussed in detail above. This will likely mean that, at the outset, the models that can be used and deployed may be limited to some of the simpler non-black box models (sometimes referred to as glass box models) until new models are developed or supplementary models are developed that can reliably provide material explanations on why particular decisions have been arrived at.[66]

83. Even in circumstances where an independent statutory review may be available to a decision made by AI, it is not adequate that the initial explanation does not properly explain *why* a decision was made. As much as the explanation is a taxpayer safeguard, it also serves as an ongoing method to test the veracity of ADM by the AI. Explanations of why it is expected or understood that a particular decision was arrived at is not as robust.

84. There is also a need for the AI Tax Legal Safeguards to require public transparency in relation to the AI being deployed in ADM in tax administration (including risk management technology being used in tax administration, such as AI being used in CONNECT), the model and training method being used for such AI, its training data, whether supplementary models have been incorporated, etc. (i.e. process-driven explanations, as described by the Information Commissioner's Office and the Alan Turing Institute[67]). This is not covered further in this paper as this is something that is already expected to be implemented by the government by phasing in, over the course of

[65] Information Commissioner's Office and the Alan Turing Institute (2022). 'Explaining decisions made with AI'. Available at: https://ico.org.uk/media/for-organisations/guide-to-data-protection/key-dp-themes/explaining-decisions-made-with-artificial-intelligence-1-0.pdf [Accessed: 8 June 2024].

[66] Information Commissioner's Office and the Alan Turing Institute (2022). 'Explaining decisions made with AI'. Available at: https://ico.org.uk/media/for-organisations/guide-to-data-protection/key-dp-themes/explaining-decisions-made-with-artificial-intelligence-1-0.pdf [Accessed: 8 June 2024].

[67] Information Commissioner's Office and the Alan Turing Institute (2022). 'Explaining decisions made with AI'. Available at: https://ico.org.uk/media/for-organisations/guide-to-data-protection/key-dp-themes/explaining-decisions-made-with-artificial-intelligence-1-0.pdf [Accessed: 8 June 2024].

2024, the mandatory requirement for all central government departments to use the Algorithmic Transparency Recording Standard (ATRS); it is expected that this will require government departments (such as HMRC) to disclose the use of AI in decision-making.[68,69] The ATRS will require government departments to disclose information in relation to the models that are used in the AI, the data used to train models, etc. The actual effectiveness of the ATRS will depend on how this is implemented in practice, and how much information is actually disclosed or withheld from the public. On the date of writing (4 August 2024) no disclosures have been made by HMRC.[70]

85. Relatedly, there are questions around transparency where AI is deployed in a manner that does not have a direct impact on a taxpayer, for example the use of LLMs to provide taxpayers with automated guidance (and, as mentioned above, LLMs are a use case that HMRC is already working on). These points are material and warrant an analysis in themselves, but equally dovetail into wider points of HMRC policy. For example, some of the questions that would need to be considered before deployment of such technology are the following. Can guidance provided by an LLM to a taxpayer be *relied* on by a taxpayer (and what happens when the LLM hallucinates)? Should there be a mitigation of penalties where a taxpayer relied on advice generated by an LLM? Where an HMRC officer relies on automated guidance generated by an LLM to provide a taxpayer with advice and such advice is incorrect, can the taxpayer rely on that advice (and what happens when the LLM hallucinates)? How and for how long should such LLMs store data and conversations from an evidentiary perspective, should this evidence be made available to the taxpayer in case of a dispute with HMRC, and what technical safeguards should be built in to mitigate the risk of hallucination?[71,72] As has been discussed above,

[68] Department for Science, Innovation and Technology (2024). 'A pro innovation approach to AI regulation: government response', CP 1019, paragraphs 44 and 93. Available at: https://assets.publishing.service.gov.uk/media/65c1e399c43191000d1a45f4/a-pro-innovation-approach-to-ai-regulation-amended-governement-response-web-ready.pdf [Accessed: 15 June 2024].

[69] It will need to be confirmed if the new Labour government will continue to implement this.

[70] The algorithmic transparency records can be found at: https://www.gov.uk/algorithmic-transparency-records.

[71] Although there are new techniques being developed to mitigate hallucinations in LLMs (for example, retrieval-augmented generation), these techniques are still largely unproven and material questions exist on whether these entirely mitigate the risk of hallucination. See, for example, Magesh, V. Surani, F., Dahl, M., Suzgun, M., Manning, C. D. and Ho, D. E. (2024). 'Hallucination-free? Assessing the reliability of leading AI legal research tools'. [Manuscript in review]. Available at: https://dho.stanford.edu/wp-content/uploads/Legal_RAG_Hallucinations.pdf [Accessed: 2 September 2024].

[72] In the non-tax sector as well, this is a developing area. For example, in *Moffatt v. Air Canada*, 2024 BCCRT 149, the Civil Resolution Tribunal found that erroneous advice provided by Air Canada's customer support chatbot amounted to negligent misrepresentation and entitled the applicant to damages from Air Canada. The Civil Resolution Tribunal rejected the argument that the chatbot was a 'separate legal entity that is responsible for its own actions'. Interestingly, the Civil Resolution Tribunal also stated that 'it makes no difference whether the information comes from a static page or a chatbot'. Although beyond the scope of this paper, it is noted that in the context of HMRC (i.e. public sector) where an LLM is being specifically deployed to provide taxpayers with guidance and advice on their rights and obligations (either directly or through HMRC officers), the dynamic nature of the advice and guidance provided by an LLM (where a response is provided to a specific question asked by a taxpayer) ought to be given greater weight with respect to reliance (in comparison to the static nature of HMRC's guidance pages wherein a taxpayer is required to self-determine its tax consequences).

if such LLMs are deployed, these will represent a material shift in the way advice or guidance is analysed and delivered to taxpayers and needs to be considered as point of active policy.

Recommendations

- **AI Tax Legal Safeguards should provide that:**
  - **taxpayers are notified where AI is used for ADM in relation to decisions having a direct impact on them; and**
  - **taxpayers are provided with outcome-based local rationale explanations with respect to such decisions.**
- **AI Tax Legal Safeguards should also require process-driven explanations.**
- **Express policy consideration is also required in respect of transparency, explainability and general policy where AI is being deployed in a manner that does not have a direct impact on taxpayers. In particular, consideration should be given to whether HMRC should be bound by guidance delivered by LLMs to taxpayers.**