

Journal: <https://research-education.com/index.php/ijmeemr/index>

Research Article

Artificial Intelligence in Tax Administration: Legal Limits and Regulatory Risks Automated Risk Scoring, Due Process, and Algorithmic Bias as Challenges to Taxpayer Rights

Authors

Volodymyr Bezditnyi¹, Andriy Matyash²

¹International Tax and Corporate Law Practitioner, Independent Researcher in International Tax Law
Partner of 4B Ukraine United Kingdom

²Expert in International Taxation and Investment Law, Independent Legal Researcher Partner of 4B Ukraine
Switzerland

Abstract

The accelerating deployment of artificial intelligence and machine learning systems in tax administration is reshaping the relationship between fiscal authorities and taxpayers in a manner that existing legal frameworks are ill-equipped to address. Automated risk scoring platforms, predictive audit selection tools, and large-scale data-matching infrastructures now exercise profound influence over enforcement outcomes in jurisdictions across the European Union, the United States, the United Kingdom, and Australia, yet they operate in conditions of structural opacity that are incompatible with foundational requirements of due process, equal treatment, and effective judicial review. This article examines the legal limits and regulatory risks inherent in this transformation, drawing upon comparative constitutional and administrative law, European Union data protection and artificial intelligence regulation, and the emerging scholarship of algorithmic governance. Central to the analysis is the Dutch Toeslagenaffaire - the childcare benefits scandal in which a machine-learning algorithm operated by the Belastingdienst produced discriminatory outcomes affecting approximately 35,000 families, ultimately causing the resignation of the Rutte II cabinet - as a paradigm case that exposes the systemic legal deficits of algorithmic enforcement. The article critically assesses the adequacy of the EU AI Act (Regulation (EU) 2024/1689), the General Data Protection Regulation, and OECD soft-law instruments, argues that all three are structurally deficient as applied to tax administration AI, and advances a framework of de lege ferenda reform proposals grounded in legality, explainability, human oversight, non-discrimination monitoring, and effective remedy.

Keywords: artificial intelligence; tax administration; risk scoring; due process; algorithmic bias; EU AI Act; GDPR; taxpayer rights; comparative tax law; Toeslagenaffaire.

1. Introduction

The use of computational tools by tax authorities to select cases for examination is not a recent innovation. The United States Internal Revenue Service has operated its Discriminant Function System - a statistical scoring mechanism for identifying potentially non-compliant returns - since the 1960s (Coglianese and Lehr, 2017, p. 1155). What is new, however, is the scale, sophistication, and consequential reach of contemporary artificial intelligence deployments in fiscal enforcement, and the profound mismatch between those deployments and the legal frameworks nominally governing them.

Tax authorities in major economies now employ supervised machine learning models to score entire taxpayer populations, natural language processing tools to interrogate commercial databases and corporate filings, and automated data-matching infrastructures that cross-reference financial disclosures against third-party records without the intervention of a trained official at any point in the process. These systems make, or materially determine, decisions that carry substantial legal consequences: the initiation of audits, the issuance of assessments, and the imposition of penalties. Yet the taxpayer who receives an adverse automated determination typically has no visibility into the model that generated it, no access to the variables that drove the outcome, and - in many jurisdictions - no enforceable legal right to a comprehensible explanation (Nathwani, 2025, p. 2; Institute for Fiscal Studies, 2024, pp. 4-8).

The governance deficit is not merely theoretical. The *Toeslagenaffaire* (the Dutch childcare benefits scandal) demonstrated with devastating clarity that an algorithmic enforcement system can produce discriminatory outcomes at scale, target tens of thousands of families on the basis of nationality and ethnic background, operate for years without detection or correction, and resist accountability because its logic is impenetrable to those it injures and to the institutions nominally responsible for oversight (Hadwick and Lan, 2021, pp. 609-614). The Parliamentary Inquiry Committee's characterisation of the affair as unprecedented injustice captures not only the magnitude of the harm but also its novelty: a harm enabled by the structural properties of automated decision-making in a legal environment designed for human judgment.

This article addresses three overarching research questions. First, what legal constraints - constitutional, statutory, and derived from international human rights instruments - govern the deployment of AI-assisted risk scoring systems in tax enforcement? Second, do current implementations of such systems give rise to cognisable violations of equal treatment and non-discrimination norms, and if so, by what doctrinal pathways? Third, are the principal regulatory instruments - including the EU AI Act, the GDPR, and OECD frameworks - adequate to address the identified risks, or is a purpose-specific legal framework for AI in tax administration required?

The article proceeds as follows. Part 2 surveys the taxonomy and comparative deployment of AI systems in tax administration. Part 3 analyses constitutional and human rights constraints, focusing on due process, the right to explanation, non-discrimination, and data protection. Part 4 examines the mechanisms and legal qualification of algorithmic bias. Part 5 assesses existing regulatory frameworks. Part 6 surveys judicial review of AI-assisted tax decisions, with particular attention to the *SyRI* litigation and relevant CJEU jurisprudence. Part 7 advances *de lege ferenda* proposals. Part 8 concludes.

2. Artificial Intelligence in Tax Administration: Taxonomy and Deployment

2.1 A Functional Classification

The term artificial intelligence is applied to a heterogeneous class of computational methods, and legal analysis requires taxonomic precision. For present purposes, three functional categories are most relevant to tax

enforcement. First, *supervised machine learning* - including logistic regression, gradient-boosted trees, and neural networks - constitutes the dominant technology in automated risk scoring. Such systems are trained on labelled historical datasets in which outcomes (compliant / non-compliant) are known, and they learn to identify input patterns statistically associated with the labelled outcome. Second, *unsupervised learning* methods - including clustering and anomaly detection - identify unusual patterns in financial data without prior labelling, enabling authorities to surface compliance risks that are not yet reflected in historical enforcement experience. Third, *natural language processing* tools extract structured intelligence from unstructured sources such as corporate filings and beneficial ownership registers.

The legal significance of this taxonomy lies principally in the concept of *automation*. Where a supervised learning model generates a risk score that determines, as opposed to merely informing, the decision to initiate an audit or issue an assessment, the resulting administrative act falls within the category of decisions based solely on automated processing within the meaning of Article 22 of the General Data Protection Regulation (Regulation (EU) 2016/679, hereinafter GDPR). As Coglianese and Lehr observe, the distinction between *supportive* and *determinative* algorithmic applications is critical to legal analysis: in the former, the algorithm assists a human decision-maker; in the latter, it supplants her (Coglianese and Lehr, 2017, pp. 1170-1175). In practice, the boundary is frequently blurred because human officials reviewing algorithmic outputs under time pressure and without independent access to underlying data are functionally precluded from exercising genuine independent judgment.

2.2 Comparative Deployment: A Survey

The deployment of AI-assisted risk scoring in tax enforcement is now sufficiently widespread to permit meaningful comparative analysis.

In the *Netherlands*, the Belastingdienst has employed algorithmic risk models for tax audit selection and benefit administration since the 2010s. The most consequential deployment was the Childcare Benefits System administered by the Tax and Customs Administration's Benefit Agency. That system employed a machine-learning model that used nationality and dual citizenship as risk variables, producing a systematic pattern of discrimination against families with a migration background. The Parliamentary Inquiry Committee found that approximately 35,000 families had been wrongly designated as fraudulent and subjected to demands for full repayment of benefits, causing severe financial hardship (Kamerstuk 35 510, nr. 2, 2020, pp. 7-11). Hadwick and Lan analyse the legal and governance failures underlying the scandal in a detailed comparative study, concluding that the combination of algorithmic opacity, absence of meaningful human oversight, and systemic institutional resistance to challenge constituted a multi-layered denial of taxpayer rights (Hadwick and Lan, 2021, pp. 619-632).

In the *United Kingdom*, HM Revenue and Customs deploys a range of AI-assisted compliance tools, including data-matching programmes that cross-reference third-party financial data against self-reported returns, predictive models for VAT and income tax compliance risk, and, more recently, large language model applications for data processing. The Institute for Fiscal Studies' 2024 report, *AI in Automated Decision-Making in Tax Administration: The Case for Safeguards*, identifies the legal framework governing HMRC's use of AI as structurally inadequate: existing legislation does not affirmatively authorise AI-driven administrative decision-making, pre-deployment testing and explanation obligations are absent, and the Taxpayer Charter's general commitments are insufficient to establish legally enforceable procedural safeguards (Institute for Fiscal Studies, 2024, pp. 15-31).

In the *United States*, the Internal Revenue Service has operated its Discriminant Function System since the 1960s and has progressively augmented it with more sophisticated machine learning tools. Coglianesi and Lehr document the IRS's use of algorithmic selection for audit targeting as an early example of governmental machine learning deployment, noting its compatibility with due process doctrine under prevailing *Mathews v. Eldridge* balancing, while acknowledging the significant transparency concerns it raises (Coglianese and Lehr, 2017, pp. 1155-1163). The IRS's Strategic Operating Plan 2023-2031, funded by the Inflation Reduction Act of 2022, contemplates substantial further investment in AI-assisted compliance analytics (Internal Revenue Service, 2023).

Across the *European Union*, virtually all member state tax administrations employ machine learning tools for audit selection, VAT fraud detection, or both, with deployment reportedly beginning as early as 2004 in some jurisdictions. As Hadwick documents in his empirical inventory of EU tax AI systems at taxadmin.ai, AI-driven enforcement tools such as Transaction Network Analysis are available to every EU member state tax administration (Hadwick, 2022, p. 184).

2.3 The Risk Scoring Paradigm and Its Structural Deficits

Automated risk scoring in tax enforcement follows a common logical structure: a model trained on historical enforcement data assigns a probability-weighted score to each taxpayer, and cases above a defined threshold are selected for examination. This paradigm generates three structural legal risks. First, the model inherits the biases of its training data: if historical enforcement has disproportionately targeted particular demographic or economic groups, the model will reproduce and amplify that disparity. Second, the score is an expression of statistical similarity to historically non-compliant cases, not a finding of actual non-compliance, yet it functions in administrative practice as a *de facto* presumption. Third, the opacity of the model precludes meaningful challenge: the taxpayer cannot know what score she received, which variables drove that score, or whether those variables were legitimately tax-relevant.

Nathwani (2025, p. 3) identifies the core legal problem succinctly: the current tax administrative framework is ill-equipped to accommodate AI in discretionary decision-making roles, because moving from human-led interpretation to algorithmic decision-making fundamentally alters the basis of tax authority and discretion. This is not merely a technical observation but a constitutional one: the legitimacy of coercive fiscal power rests on the capacity of those subject to it to understand, contest, and seek review of the decisions that exercise it.



Figure 1. Tax administration AI decision pipeline with rights-sensitive safeguard checkpoints.

End-to-end workflow from data ingestion to enforcement outcomes, highlighting points where legality, transparency, explainability, human oversight, non-discrimination monitoring, and effective remedy should be operationalized.

3. Constitutional and Human Rights Constraints

3.1 Due Process and the Right to a Reasoned Decision

The right to due process – the entitlement to know the case against one, to respond to it meaningfully, and to have the matter resolved by a legally accountable decision-maker – is a foundational norm of liberal constitutional orders and is directly engaged by automated tax enforcement.

In the *European human rights framework*, Article 6(1) of the European Convention on Human Rights guarantees a fair hearing in the determination of civil rights and obligations and, in its criminal limb, applies to tax proceedings attracting penalties of a punitive character. The European Court of Human Rights confirmed in *Jussila v. Finland* (Application No. 73053/01, Grand Chamber, 23 November 2006) that tax surcharges of a punitive character fall within the criminal limb of Article 6. Where an automated risk score forms the evidentiary foundation for such proceedings without disclosure of the model's methodology, the structural requirements of adversarial procedure and equality of arms are *prima facie* violated: the taxpayer cannot challenge what she cannot see.

In the *European Union*, Article 41 of the Charter of Fundamental Rights guarantees the right to good administration, including the obligation of the administration to give reasons for its decisions. The Court of Justice of the EU has confirmed that this obligation is general in character, applying to all adverse administrative measures. An automated risk score that triggers an adverse measure without accompanying reasoning is, in this respect, constitutionally deficient as a matter of EU primary law.

In the *United States*, the Due Process Clause of the Fifth Amendment governs federal administrative action. The *Mathews v. Eldridge* balancing framework (424 U.S. 319 (1976)) requires courts to weigh the private interest affected, the risk of erroneous deprivation under existing procedures, and the government's interest. Coglianese and Lehr argue that governmental machine learning can satisfy *Mathews* balancing in principle, because automated systems may in aggregate, produce more accurate outcomes than human review (Coglianese and Lehr, 2017, pp. 1188-1200). However, their analysis presupposes systems that are adequately validated, subject to human oversight, and amenable to case-specific explanation – conditions that, as the IFS report documents, current HMRC and IRS deployments manifestly do not satisfy in full (Institute for Fiscal Studies, 2024, pp. 20-28).

3.2 The Right to Explanation and Algorithmic Transparency

Article 22(1) of the GDPR prohibits data subjects from being subjected to decisions based solely on automated processing, including profiling, which produce legal or similarly significant effects. Where such processing is authorised, Article 22(3) requires the controller to implement suitable measures safeguarding the data subject's rights, including at least the right to obtain human intervention, to express a point of view, and to contest the decision.

Wachter, Mittelstadt, and Russell identify three aims that explanations of automated decisions should serve from the perspective of the affected individual: to inform and help the individual understand why the decision was reached; to provide grounds to contest it if the outcome is undesired; and to clarify what would need to change for a different outcome in the future (Wachter, Mittelstadt, and Russell, 2018, pp. 844-846). They propose that *counterfactual explanations* - describing the minimum change to the relevant facts that would produce a different outcome - can satisfy these three aims without requiring full disclosure of a model's internal logic, thereby accommodating legitimate confidentiality interests while preserving the substance of due process. This proposal is directly relevant to tax enforcement contexts, where the trading of a taxpayer's score

for a specific statement of what facts, if different, would have placed her below the audit threshold would substantially improve the quality of challenge rights without compromising the model's integrity. The GDPR's reach in tax administration is, however, constrained by Article 23, which permits member states to restrict data subject rights where necessary to protect important economic or financial interests, including taxation. Several EU member states have invoked this derogation to shield automated risk scoring systems from GDPR Article 22 obligations. This article argues that such derogations cannot override the minimum requirements of procedural fairness imposed by Article 41 of the EU Charter, which operates at constitutional level above the GDPR.

3.3 Non-Discrimination and Equal Treatment

The principle of equality before the law - of direct constitutional rank in every jurisdiction under examination - prohibits the state from treating similarly situated persons differently on the basis of legally irrelevant characteristics. Applied to tax enforcement, it requires that audit selection criteria be grounded in tax-relevant indicators of non-compliance, not demographic or social proxies.

The *Toeslagenaffaire* is the most extensively documented instance of algorithmic tax enforcement producing discriminatory outcomes at scale. Hadwick and Lan demonstrate that the risk-scoring algorithm operated by the Belastingdienst's Benefit Agency produced a systematic pattern of discrimination rooted in a process of sample bias: the Belastingdienst conducted ethnicity-based enforcement sweeps (auditing all Ghanaian welfare recipients upon receiving information about specific suspected fraudsters), the resulting over-representation of foreign nationals in the non-compliant training dataset caused the machine-learning model to assign disproportionate predictive weight to nationality, and the model's outputs then reinforced the discriminatory enforcement pattern through a feedback loop (Hadwick and Lan, 2021, pp. 614-622). From a doctrinal perspective, the affair illustrates the mechanism of indirect discrimination: a facially neutral criterion that produces disproportionately adverse effects on a protected group. In EU law, the concept of indirect discrimination is established in Article 2(2)(b) of Directive 2000/43/EC (the Race Equality Directive) and Article 2(1)(b) of Directive 2000/78/EC (the Framework Employment Directive). Its application to algorithmic public administration is addressed, in the context of the SyRI system, by the Hague District Court in its 2020 judgment (discussed in Part 6 below). The Hague District Court confirmed that a risk scoring system that correlates with protected characteristics, produces disproportionate adverse effects on vulnerable groups, and lacks sufficient transparency to permit independent assessment of its justification violates Article 8 of the European Convention on Human Rights read in conjunction with the principle of proportionality.

3.4 Data Protection and the Principle of Data Minimisation

Automated tax enforcement is premised on the mass processing of personal financial data assembled from multiple sources - tax returns, bank transaction records, property registers, beneficial ownership declarations, and increasingly open-source intelligence. The aggregation of these data streams into comprehensive financial profiles raises foundational data protection concerns.

The principle of *data minimisation* in Article 5(1)(c) GDPR requires that personal data be adequate, relevant, and limited to what is necessary for the purpose for which they are processed. Mass population-level financial profiling for the purpose of algorithmic risk scoring arguably exceeds what is necessary where less privacy-invasive methods of compliance risk identification are available. The *purpose limitation* principle in Article 5(1)(b) restricts the repurposing of tax-reporting data for algorithmic profile construction without a clear and

specific legal basis. The *Data Protection Impact Assessment* obligation in Article 35 - applicable wherever processing is likely to result in high risk to the rights and freedoms of natural persons, including large-scale automated processing - should be mandatory for all algorithmic risk scoring deployments in tax administration. There is, at present, no public evidence that any major EU member state tax authority has conducted or published a compliant DPIA in respect of its automated risk scoring systems (Nathwani, 2025, p. 5).

4. Algorithmic Bias in Tax Enforcement: Legal Qualification

4.1 Mechanisms of Bias Generation

Algorithmic bias in supervised machine learning is not an aberration but a structurally predictable product of specific design choices made at each stage of model development. The leading technical scholarship identifies four primary mechanisms. *Historical bias* arises where the training data encodes pre-existing societal inequalities or discriminatory institutional practices. *Representation bias* arises where certain groups are systematically overrepresented or underrepresented in the training sample, causing the model to perform differently across groups. *Measurement bias* arises where the observable features used as inputs are measured with differential accuracy across groups. *Feedback loops* arise where a biased model generates biased recommendations, those recommendations produce enforcement actions that confirm the original bias, and the resulting data is fed back into subsequent model iterations, progressively entrenching discriminatory outputs (Wachter, Mittelstadt, and Russell, 2021, pp. 105567-105572).

In the tax enforcement context, historical bias is particularly acute. Tax authorities that have historically over-audited particular economic sectors, income groups, or demographic categories - whether due to resource constraints, organisational culture, or, as in the Dutch case, intentional discriminatory practice - will produce training datasets that systematically over-represent those groups among the non-compliant observations. A model trained on such data will replicate the historical enforcement pattern with the additional legitimating effect of appearing to be the neutral output of an objective computational process, a phenomenon O'Neil (2016, pp. 28-30) describes as the camouflaging of human bias behind mathematical authority.

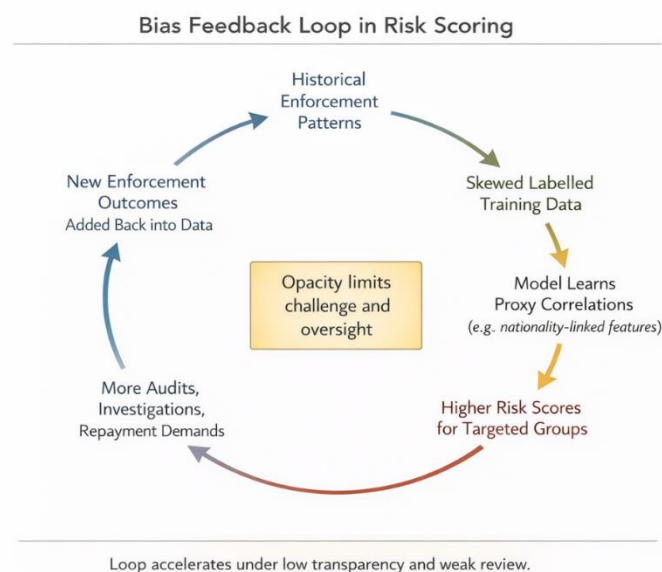


Figure 2. Feedback loop pathway for algorithmic bias in tax and benefits enforcement risk scoring.

Illustrates how historically skewed enforcement data can produce biased model outputs that reinforce subsequent enforcement patterns, entrenching disparate impacts over time.

4.2 Disparate Impact and the Legal Standard of Indirect Discrimination

The translation of the indirect discrimination doctrine to algorithmic tax enforcement is legally viable and increasingly supported by comparative case law. In the EU framework, a provision, criterion, or practice that is apparently neutral but puts persons of a particular ethnic, national, or other protected characteristic at a particular disadvantage constitutes indirect discrimination under Directive 2000/43/EC unless objectively justified by a legitimate aim and the means are appropriate and necessary to achieve it.

Wachter, Mittelstadt, and Russell's 2021 analysis of the relationship between EU non-discrimination law and AI fairness metrics demonstrates that the disparate impact framework applies, in principle, to algorithmic decision-making systems, but that the translation faces significant doctrinal tensions: EU non-discrimination law is calibrated to individual comparative treatment, while algorithmic fairness is typically assessed at a statistical group level (Wachter, Mittelstadt, and Russell, 2021, pp. 105567-105575). For tax enforcement purposes, the relevant doctrinal adaptation is as follows: where a taxpayer can demonstrate that an authority's risk-scoring model assigns disproportionate adverse weight to variables correlated with protected characteristics, the burden should shift to the authority to demonstrate that those variables are tax-relevant, necessary, and proportionate - a burden that the use of nationality in the Dutch system failed to discharge.

4.3 The Evidentiary Problem: Structural Opacity and Burden of Proof

The most acute practical obstacle confronting a taxpayer who suspects algorithmic discrimination is epistemic asymmetry: the taxpayer does not know its score, the variables that drove it, the training data used, or the model's demographic performance metrics. The authority possesses all this information and has compelling institutional incentives to resist disclosure, including the risk of model gaming and the protection of proprietary vendor arrangements.

This asymmetry is not merely an inconvenience but a structural barrier to the effective exercise of legal rights. The European Court of Human Rights recognised in *Dombo Beheer BV v. The Netherlands* (Application No. 14448/88, 27 October 1993) that procedural equality of arms requires that neither party be placed in a substantially disadvantageous position vis-à-vis the other. Applied to algorithmic enforcement, this principle requires, at minimum, that a taxpayer be provided with sufficient information about the model to identify the grounds for challenge - a requirement fundamentally inconsistent with the current regime of comprehensive opacity.

As Coglianese and Lehr argue, the concern that algorithmic transparency enables gaming is real but does not justify comprehensively denying access to decision-making rationales: a distinction can be drawn between disclosing the precise model parameters (which would enable gaming) and disclosing, case-specifically, which category of factors drove an adverse outcome (which enables meaningful challenge without compromising the model's enforcement function) (Coglianese and Lehr, 2017, pp. 1207-1213).

5. Regulatory Frameworks: Adequacy Assessment

5.1 The EU AI Act and Its Application to Tax Risk Scoring

Regulation (EU) 2024/1689 on Artificial Intelligence - the AI Act - entered into force on 1 August 2024 and will become fully applicable in August 2026, with prohibition provisions operative from February 2025. The Act establishes a risk-based regulatory architecture classifying AI systems according to their potential to cause harm. High-risk AI systems - enumerated in Annex III - are subject to mandatory conformity requirements including risk management, data governance, transparency, human oversight, and post-market monitoring obligations.

The critical classificatory question is whether automated risk scoring systems in tax administration fall within Annex III. The AI Act prohibits social scoring by public authorities under Article 5(1)(c) - a prohibition of potential relevance to nationality-based risk profiling of the kind documented in the *Toeslagenaffaire*. However, Annex III does not explicitly enumerate tax risk scoring as a high-risk use case. Point 7 covers AI systems intended to evaluate the creditworthiness of natural persons; Point 8 covers risk assessment in the insurance context. Neither covers tax enforcement. Hadwick has argued that this omission is a deliberate political choice reflecting the lobbying interest of tax authorities in preserving operational freedom and constitutes a significant gap in the regulatory framework (Hadwick, 2022, pp. 184-187). Nathwani characterises the Act's implicit acknowledgment that even ostensibly low-risk AI can produce harmful outcomes in sensitive domains as insufficient to address the structural deficits of algorithmic tax enforcement (Nathwani, 2025, pp. 5-6).

Even under the Act's current scope, tax authorities that deploy AI systems qualifying as high-risk under Annex III must comply with Article 9 risk management requirements, Article 10 data governance obligations (including requirements that training data be free from errors and representative of the populations to which the system will be applied), Article 13 transparency obligations (requiring sufficient transparency to enable deployers to interpret outputs), and Article 14 human oversight requirements (mandating the ability to override automated outputs). Full compliance with these obligations would require a comprehensive review of existing systems and, in many cases, substantial redesign.

	GDPR	EU AI Act	OECD Principles
Legality	✓*	✓	○
Transparency	✓	✓	✓
Explainability	✓	✓	○
Human Oversight		✓	○
Non-Discrimination Monitoring	✓	✓	✓
Effective Remedy	✓		○

 Covered
  Partially Covered
  Not Covered

Figure 3. Regulatory instrument coverage of key risks in tax administration AI.

Comparative mapping of GDPR, EU AI Act, and OECD principles against core risk categories: legality, transparency, explainability, human oversight, non-discrimination monitoring, and effective remedy

5.2 The GDPR: Scope, Derogations, and Enforcement Gaps

The GDPR remains the most immediately applicable instrument governing automated tax enforcement for the foreseeable future. Its Article 22 framework, Article 35 DPIA requirement, Articles 13-15 transparency rights, and Article 5 data quality principles constitute a substantive legal framework of considerable scope. The structural difficulty is the breadth of the Article 23 derogation for taxation and the historically weak enforcement of GDPR obligations against national tax administrations by data protection supervisory authorities.

The European Data Protection Board's Guidelines 03/2017 on Automated Individual Decision-Making and Profiling confirm that tax authorities are not categorically exempt from Article 22 and that Article 23 derogations must satisfy the requirements of necessity and proportionality. Wachter, Mittelstadt, and Russell's analysis of the GDPR's right to explanation concludes that Article 22(3) supports, at minimum, a right to *counterfactual* explanation - that is, a statement of the minimum factual change that would have produced a different decision - even where full model disclosure is resisted (Wachter, Mittelstadt, and Russell, 2018, pp. 871-885). No major EU member state data protection authority has, to date, issued a formal enforcement decision against a national tax authority in respect of automated risk scoring practices.

5.3 OECD Frameworks and the Limits of Soft Law

The OECD's *Tax Administration 3.0: The Digital Transformation of Tax Administration* (2020) documents the scale and trajectory of AI deployment by tax authorities and calls for attention to data governance, transparency, and taxpayer-centric design. The OECD's *Principles on Artificial Intelligence* (2019), endorsed by G20 leaders, articulate five value-based principles - inclusivity, human-centred values and fairness, transparency and explainability, robustness and security, and accountability - that are directly germane to tax risk scoring. The structural limitation of this framework is its voluntarist character: tax authorities are not legally bound, and there is no enforcement mechanism.

Coglianese and Lehr's analysis of administrative law requirements for algorithmic government action in the US context suggests that explainability, in the sense of a reasoned account of the objective function that the algorithm optimises, is legally necessary under prevailing transparency doctrine, even though causal explanation of specific model outputs is not (Coglianese and Lehr, 2017, pp. 1207-1215). This distinction - between systemic transparency (what the algorithm is designed to do and how it is validated) and case-specific transparency (why this taxpayer received this outcome) - is essential to designing proportionate disclosure obligations.

5.4 National Regulatory Responses and Their Limitations

The most significant national regulatory response has been that of the *Netherlands* following the *Toeslagenaffaire*. The *Algoritmeregister* - the Dutch government's publicly accessible registry of algorithms used by national public bodies, launched in 2022 - creates disclosure obligations covering the system's purpose, data inputs, and logic. The *Wet digitale overheid* (Act on Digital Government) imposes transparency obligations on public authorities using algorithmic decision-making systems. These reforms are significant but do not, as yet, extend to the full range of safeguards - mandatory impact assessment, pre-deployment bias testing, independent external audit - that the scale of the failure warrants.

In the *United Kingdom*, the Institute for Fiscal Studies' 2024 report recommends, as a minimum, that enabling legislation affirmatively authorise AI use in tax administrative decision-making, that pre-deployment bias

testing be mandatory, that process-driven explanations be available to taxpayers, that human review of AI-generated determinations be enshrined in statute, and that an HMRC AI Charter with legally binding effect be established (Institute for Fiscal Studies, 2024, pp. 28-37).

6. Judicial Review of Algorithmic Tax Decisions

6.1 Standards of Review and the SyRI Precedent

The most significant judicial precedent in the governance of public sector AI in Europe is the Hague District Court's judgment in *NJCM c.s. v. De Staat der Nederlanden* (Case No. C/09/550982, ECLI:NL:RBDHA:2020:1878, 5 February 2020), commonly referred to as the *SyRI* case. The *SyRI* (System Risk Indication) was a risk-scoring platform operated by the Dutch government to detect social fraud, drawing on data from fourteen government agencies. The court held that the system violated Article 8 ECHR - the right to private life - read in light of the principle of proportionality.

The court's reasoning is of direct relevance to tax risk scoring systems. The violation was found not because risk scoring as such is impermissible, but because the *SyRI* system lacked the minimum transparency required to enable independent assessment of its proportionality: the risk indicators were secret, the weighting methodology was not disclosed, and neither an independent supervisory body nor the individuals affected had any means of evaluating whether the system's adverse effects on privacy and equal treatment were justified by its fraud-detection objectives. The court specifically noted that the absence of public disclosure of the system's logic rendered it impossible for courts to exercise meaningful review (ECLI:NL:RBDHA:2020:1878, para. 6.82-6.95). This reasoning applies with full force to tax risk scoring systems of equivalent or greater opacity.

Hadwick and Lan draw directly on the *SyRI* judgment in their comparative analysis of the *Toeslagenaffaire*, arguing that the legal requirements established by the Hague District Court - transparency, independent oversight, accessible challenge mechanisms - constitute a minimum standard applicable to all algorithmic public enforcement systems in the Netherlands and, through the operation of Article 8 ECHR as interpreted by the court, across the Council of Europe's member states more broadly (Hadwick and Lan, 2021, pp. 635-641).

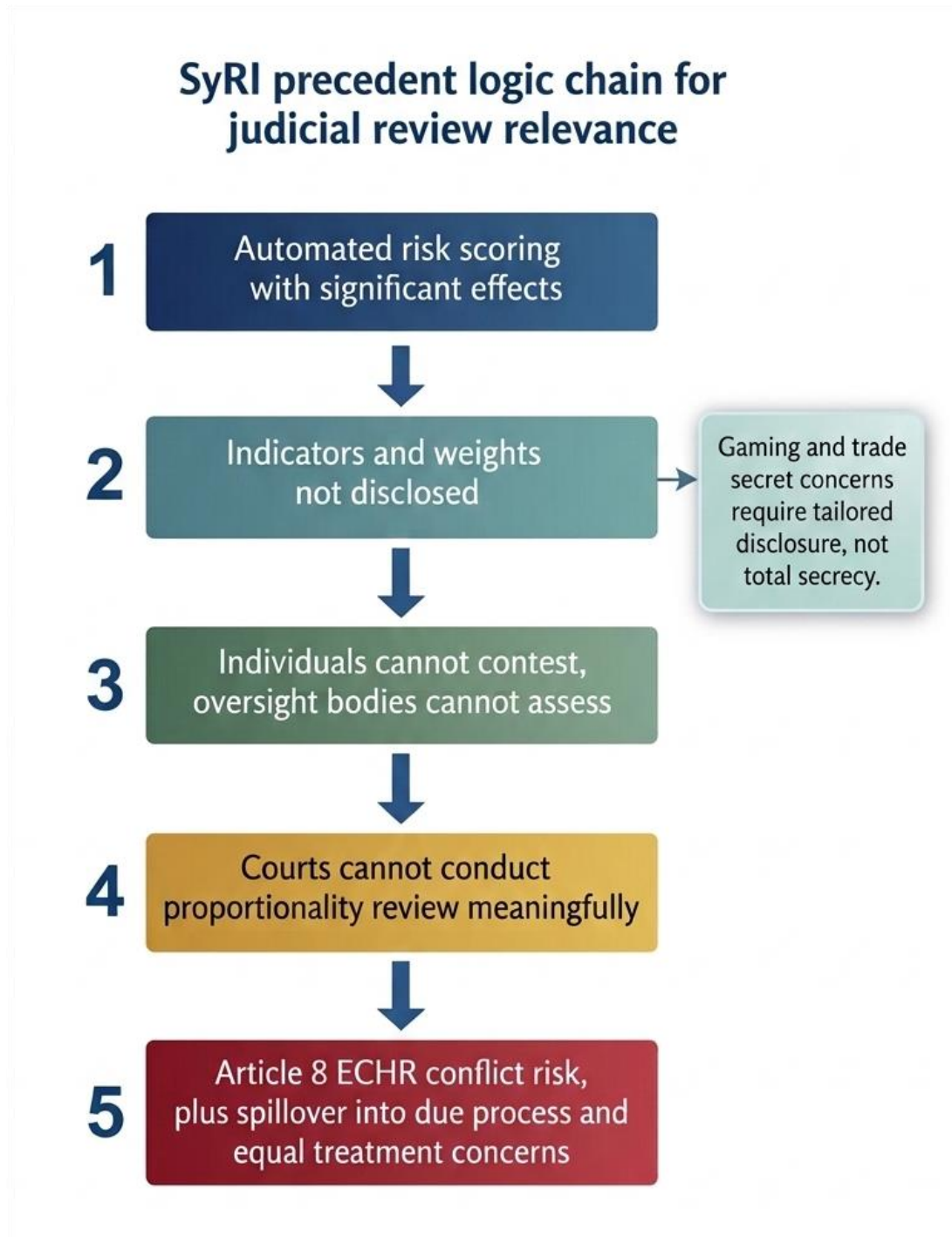


Figure 4. SyRI judicial review logic chain applied to tax risk scoring systems.

Demonstrates how opacity undermines proportionality assessment and meaningful judicial review, making rights compliance non-verifiable.

6.2 CJEU Jurisprudence and the Principle of Effective Judicial Protection

The Court of Justice of the EU's approach to judicial review of algorithmic administrative measures is shaped by the principle of effective judicial protection enshrined in Article 47 of the EU Charter. In *Ligue des droits*

humains v. Conseil des ministres (Case C-817/19, EU:C:2022:491, 21 June 2022), addressing the EU's Passenger Name Record Directive, the Grand Chamber confirmed that automated profiling systems must be subject to genuine judicial scrutiny - not merely deference to the authority's assessment of necessity - and that the absence of disclosed risk indicators precludes such scrutiny.

The CJEU's analysis of automated decision-making in *Schrems II* (Data Protection Commissioner v. Facebook Ireland Limited, Case C-311/18, EU:C:2020:559) establishes the broader principle that large-scale automated processing of personal data by public authorities is subject to heightened scrutiny under the EU Charter's privacy and data protection provisions, and that generalised derogations from data subject rights are incompatible with the principle of proportionality where they do not provide minimum safeguards against abuse.

6.3 Disclosure of Algorithmic Models in Litigation: The Trade Secret Problem

Tax authorities confronted with demands for disclosure of risk scoring models in litigation have asserted three principal grounds for resistance: *state secrets* (in US federal court proceedings); *trade secrecy* of commercial AI vendors; and the *gaming* risk - the concern that model disclosure would enable sophisticated taxpayers to structure their affairs to fall below the detection threshold.

The gaming concern has some merit, but is frequently overstated. Wachter, Mittelstadt, and Russell's counterfactual explanation framework addresses it directly: case-specific disclosure of the minimum factual change that would have altered the outcome enables meaningful challenge without disclosing the model's full parameter architecture (Wachter, Mittelstadt, and Russell, 2018, pp. 874-880). The Institute for Fiscal Studies proposes *in camera* judicial inspection - in which a court, rather than the taxpayer, reviews the model and rules on its legal compliance without disclosing proprietary parameters - as a proportionate solution to the tension between transparency and legitimate confidentiality (Institute for Fiscal Studies, 2024, pp. 32-34).

7. Towards a Legal Framework for AI in Tax Enforcement: De Lege Ferenda

7.1 Six Foundational Principles

Drawing upon the foregoing analysis, this article proposes six foundational legal principles to govern the deployment of AI systems in tax enforcement. These principles are grounded in existing legal norms - constitutional, statutory, and international - and constitute a framework capable of legislative implementation.

- ❖ **Legality.** The authority to deploy AI-assisted decision-making in tax enforcement must rest upon an express, accessible, and sufficiently precise legal basis, specifying the categories of data that may be processed, the types of decisions that may be made or supported by automated means, and the procedural safeguards applicable. General statutory mandates to collect taxes do not constitute adequate authority for AI-driven automated enforcement.
- ❖ **Transparency.** Tax authorities must publicly disclose the existence, purpose, and general methodology of their risk scoring systems, including the categories of data used as inputs, the general logic of the scoring mechanism, and the governance arrangements for model maintenance and validation. Full parameter-level disclosure may be legitimately withheld, but the general architecture must be accessible to affected individuals and supervisory bodies.
- ❖ **Explainability.** Any adverse decision materially based on algorithmic output - audit selection, assessment, penalty imposition - must be accompanied by a case-specific explanation that enables the taxpayer to understand the grounds for the decision and to mount a meaningful challenge. Following

Wachter, Mittelstadt, and Russell (2018), counterfactual explanations satisfy this requirement proportionately.

- ❖ **Human oversight.** Significant enforcement decisions must not be based solely on algorithmic output. A trained official with independent access to underlying data and with genuine practical capacity to override the algorithmic recommendation must authorise each such decision. Nominal human review that merely rubber-stamps algorithmic outputs does not satisfy this requirement.
- ❖ **Non-discrimination monitoring.** Tax authorities must conduct regular, independent, and publicly disclosed audits of their AI systems' disparate impacts across demographic groups. Where adverse impact is identified, the authority must either modify the system to eliminate the disparity or demonstrate objective, necessary, and proportionate tax-relevant justification for its continuation. The monitoring obligation must be external, not self-assessed.
- ❖ **Effective remedy.** Taxpayers must have access to a legally enforceable mechanism for challenging adverse algorithmic decisions, including the right to invoke a competent supervisory body or court with authority to order disclosure and, where violations are established, to award adequate redress. The *SyRI* litigation and the *Toeslagenaffaire* parliamentary inquiry demonstrate that informal complaint mechanisms and internal review processes are structurally insufficient.

7.2 Algorithmic Impact Assessments for Tax Systems

The GDPR's Data Protection Impact Assessment framework provides a template for a mandatory *Algorithmic Impact Assessment* (AIA) applicable to all tax administration AI systems making or supporting decisions with significant effects on taxpayers. An AIA should include: the system's purpose and legal basis; a description of the data processing and model architecture; an assessment of risks to fundamental rights including non-discrimination and due process; identification of mitigation measures; and mandatory consultation with the competent data protection authority and, where parliamentary financial oversight mechanisms exist, with the relevant committee. No AI system in tax enforcement should be deployed before an AIA has been completed and made public.

This proposal builds on the Institute for Fiscal Studies' recommendation of mandatory pre-deployment testing (Institute for Fiscal Studies, 2024, pp. 28-30), the OECD's endorsement of algorithmic transparency and accountability as governance principles (OECD, 2019; OECD, 2020), and the AI Act's conformity assessment requirements for high-risk systems. It extends these frameworks by requiring *public disclosure* of AIA outcomes - a requirement that existing frameworks do not consistently impose.

7.3 International Coordination

The multinational character of significant tax compliance risks - base erosion and profit shifting, offshore asset concealment, cross-border VAT fraud - creates structural arguments for international coordination in the governance of AI tax enforcement tools. Where multiple jurisdictions' tax authorities exchange data and coordinate enforcement activity through instruments such as the OECD's Common Reporting Standard, the governance of the AI systems processing that data should be subject to coordinated oversight. The OECD's Forum on Tax Administration is well placed to develop, as a minimum, peer-reviewed standards for AI governance in tax administration, addressing data quality requirements, bias testing protocols, transparency obligations, human oversight requirements, and taxpayer remedy mechanisms. Elevation of these standards to

binding commitments, analogous to the CRS framework's legal architecture, should be the long-term objective.

8. Conclusion

The deployment of artificial intelligence in tax administration has crossed the threshold from pilot programme to structural feature of fiscal governance. Tax authorities across the EU, the US, and the UK use machine learning daily to select taxpayers for examination, identify compliance risks, and in some jurisdictions and some contexts, to make or materially determine legally consequential decisions. The potential efficiency gains are genuine. But the legal infrastructure that would render this transformation compatible with the constitutional values of the legal orders deploying it has not kept pace.

The *Toeslagenaffaire* is the definitive illustration of what happens in the absence of that infrastructure. A system that used nationality as a discriminatory proxy, operated for years without transparency, resisted challenge through structural opacity, and caused devastating harm to 35,000 families was not the product of exceptional institutional failure. It was the product of ordinary institutional incentives operating in the absence of legally enforceable constraints on algorithmic enforcement. Hadwick and Lan's forensic analysis of the affair demonstrates that the legal deficits it exposed - absence of legal basis, opacity, no meaningful human oversight, no effective remedy - are systemic and are replicated, to varying degrees, wherever AI-driven tax enforcement operates without adequate governance (Hadwick and Lan, 2021, pp. 641-645).

The EU AI Act is a significant step, but it contains a critical gap: tax risk scoring is not explicitly classified as high-risk, giving tax authorities a plausible basis for resisting the Act's most demanding obligations. The GDPR provides important but partial protection, limited by broad public interest derogations and historically weak enforcement. OECD soft law establishes valuable normative expectations but lacks the binding force and enforcement architecture necessary to generate consistent compliance.

The six-principle framework and the mandatory Algorithmic Impact Assessment requirement advanced in this article constitute a minimum agenda for reform. They are grounded in existing legal norms, are technically implementable, and are economically proportionate to the risks they address. They require, above all, a recognition by legislators, courts, and tax authorities themselves that the efficiency of fiscal enforcement is a legitimate objective, but not one that can be pursued at the cost of the rule of law.

References

I. Legislative and Regulatory Instruments

1. Charter of Fundamental Rights of the European Union [2012] OJ C 326/391.
2. Convention for the Protection of Human Rights and Fundamental Freedoms (European Convention on Human Rights, as amended) (Council of Europe, 1950) ETS No. 005.
3. Directive 2000/43/EC of the European Parliament and of the Council of 29 June 2000 implementing the principle of equal treatment between persons irrespective of racial or ethnic origin [2000] OJ L 180/22.
4. Directive 2000/78/EC of the European Parliament and of the Council of 27 November 2000 establishing a general framework for equal treatment in employment and occupation [2000] OJ L 303/16.

5. Internal Revenue Service (2023) IRS Strategic Operating Plan FY2023-2031. Washington DC: US Department of the Treasury. Available at: <https://www.irs.gov/pub/irs-pdf/p3744.pdf> (Accessed: 15 January 2025).
6. Kamerstuk 35 510, nr. 2 (2020) Ongekend onrecht: Verslag van de parlementaire ondervragingscommissie kinderopvangtoeslag. The Hague: Tweede Kamer der Staten-Generaal.
7. OECD (2019) Recommendation of the Council on Artificial Intelligence, OECD/LEGAL/0449. Paris: OECD Publishing. Available at: <https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449> (Accessed: 10 November 2024).
8. OECD (2020) Tax Administration 3.0: The Digital Transformation of Tax Administration. Paris: OECD Publishing. Available at: <https://www.oecd.org/tax/forum-on-tax-administration/publications-and-products/tax-administration-3-0-the-digital-transformation-of-tax-administration-6d9862f6-en.htm> (Accessed: 8 October 2024).
9. Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data (General Data Protection Regulation) [2016] OJ L 119/1.
10. Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) [2024] OJ L 1689.
11. United States Constitution, Amendment V.

II. Judicial Decisions

12. Data Protection Commissioner v Facebook Ireland Limited and Maximillian Schrems (Case C-311/18) EU:C:2020:559 (CJEU Grand Chamber, 16 July 2020).
13. Dombo Beheer BV v The Netherlands (Application No. 14448/88) (1993) 18 EHRR 213 (European Court of Human Rights, 27 October 1993).
14. Jussila v Finland (Application No. 73053/01) (2009) 48 EHRR 32 (European Court of Human Rights, Grand Chamber, 23 November 2006).
15. Ligue des droits humains v Conseil des ministres (Case C-817/19) EU:C:2022:491 (CJEU Grand Chamber, 21 June 2022).
16. Mathews v Eldridge 424 US 319 (1976) (United States Supreme Court).
17. NJCM c.s. v De Staat der Nederlanden (Case No. C/09/550982) ECLI:NL:RBDHA:2020:1878 (Hague District Court, 5 February 2020) [SyRI case].

III. Scholarly Literature and Reports

18. Coglianese, C. and Lehr, D. (2017) 'Regulating by Robot: Administrative Decision Making in the Machine-Learning Era', Georgetown Law Journal, 105(5), pp. 1147-1223. Available at: https://scholarship.law.upenn.edu/faculty_scholarship/1734 (Accessed: 5 December 2024).
19. Hadwick, D. (2022) 'Behind the One-Way Mirror: Reviewing the Legality of EU Tax Algorithmic Governance', EC Tax Review, 31(4), pp. 184-195.
20. Hadwick, D. and Lan, S. (2021) 'Lessons to Be Learned from the Dutch Childcare Allowance Scandal: A Comparative Review of Algorithmic Governance by Tax Administrations in the Netherlands, France and Germany', World Tax Journal, 13(4), pp. 609-645. Available at: <https://ssrn.com/abstract=4282704> (Accessed: 12 November 2024).

21. Institute for Fiscal Studies (2024) AI in Automated Decision-Making in Tax Administration: The Case for Safeguards. London: Institute for Fiscal Studies. Available at: <https://ifs.org.uk/sites/default/files/2024-09/Artificial-intelligence-in-automated-decision-making-in-tax-administration.pdf> (Accessed: 3 October 2024).
22. Nathwani, T. (2025) 'Balancing Innovation and Integrity: AI in Tax Administration and Taxpayer Rights', Humanities and Social Sciences Communications, 12, Article 1818. <https://doi.org/10.1057/s41599-025-06099-7>.
23. O'Neil, C. (2016) Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy. New York: Crown Publishers.
24. Wachter, S., Mittelstadt, B. and Russell, C. (2018) 'Counterfactual Explanations Without Opening the Black Box: Automated Decisions and the GDPR', Harvard Journal of Law and Technology, 31(2), pp. 841-887. Available at: <https://jolt.law.harvard.edu/assets/articlePDFs/v31/Counterfactual-Explanations-without-Opening-the-Black-Box-Sandra-Wachter-et-al.pdf> (Accessed: 20 November 2024).
25. Wachter, S., Mittelstadt, B. and Russell, C. (2021) 'Why Fairness Cannot Be Automated: Bridging the Gap Between EU Non-Discrimination Law and AI', Computer Law and Security Review, 41, 105567. <https://doi.org/10.1016/j.clsr.2021.105567>.



Research Education
JOURNALS