

From Compliance to Intelligence: Integrating AI and Predictive Analytics into U.S. Tax Compliance and Revenue Systems

Issabella Ampofo ^{1,*}, David Amoako ² and Mary Magdalene Yeboah ³

¹ Delta State University, Cleveland, USA.

² Pittsburg, CA, USA.

³ Independent Researcher, Texas, USA.

Magna Scientia Advanced Research and Reviews, 2026, 17(02), 022–029

Publication history: Received on 26 May 2026; revised on 30 June 2026; accepted on 02 July 2026

Article DOI: <https://doi.org/10.30574/msarr.2026.17.2.0119>

Abstract

The gross tax gap in the U.S is over 600 billion a year. AI and predictive analytics have a transformational opportunity to enhance compliance risk scoring, audit selection, revenue forecasting, and fraud detection at the IRS. This critical literature review summarizes peer-reviewed literature (2020-2026) on AI adoption to the U.S. tax compliance and revenue systems, evaluating reported outcomes, methodological conflicts, and governance needs. Organized search in SSRN, Google Scholar, Web of Science, Scopus, and government repositories resulted in 37 sources that satisfy pre-determined inclusion criteria and are rated in three levels of evidence. The ML models decrease audit no-change rates by an estimated 15-20 percentage points and forecasting MAPE by 15-30 percent compared to legacy systems, although equity questions are actualized: ROI-optimal classifiers increase audit load on low-income filers unless fixed through regression-based expected-adjustment targets and fairness limits. Governance alignment with NIST AI RMF 1.0, EO 14110, and 26 U.S.C. § 6103 is critical. Equity-by-design, explainable architecture, federated infrastructure, and modernized statutory framework are all necessary to achieve responsible AI adoption.

Keywords: Artificial Intelligence; Tax Compliance; Predictive Analytics; IRS Modernization; Machine Learning; Algorithmic Accountability; Explainable AI

1. Introduction

The IRS estimates a gross tax gap of more than 600 billion per year, which puts a strain on government investment and financial equity (Mazur, 2025; Makoni et al., 2025). AI and predictive analytics are becoming more and more poised as potentially groundbreaking governance tools that can revolutionize the manner in which the federal government determines and collects tax liabilities. Algorithms in machine learning are able to scale to multi-dimensional taxpayer data at a level that cannot be achieved in traditional approaches, and NLP applications are able to read intricate regulatory texts and financial reports with growing accuracy. (Holtzblatt & Engler, 2022; Pamisetty et al., 2022). The IRS has made investments in high-end analytics infrastructure through the Compliance Data Warehouse and the neural-network-powered return review programs (Bruckner & Coil, 2024; Shaikh, 2025).

However, there are two fundamental methodological issues that have not been tackled. The selective-labels problem: audit outcome data is only available on taxpayers who have been selected to be audited in the past, introducing training biases that are propagated to deployed ML models (Holtzblatt & Engler, 2022). Viewing compliance as binary classification as opposed to a regression objective (amount of adjustment the auditor expects to make) has been demonstrated to focus audits on low-income filers to the detriment of revenue yield and vertical equity (Yang, 2025;

* Corresponding author: Issabella Ampofo

Neuman & Sheu, 2022). The governance schemes in EO 14110, NIST AI RMF 1.0, and OMB M-24-10 offer a boilerplate but are incomplete for the IRS's unique statutory context (Engstrom & Ho, 2020; Siva, 2025).

This is a critical synthesis of evidence that has been conducted on five themes: (1) ML compliance risk modeling and the selective-labels problem; (2) audit selection and revenue forecasting with quantitative benchmarks; (3) fraud and anomaly detection architectures; (4) algorithmic equity, accountability, and governance; and (5) XAI architectures and operational controls. It is speculative in nature, revealing strains and unsettled discussions instead of introducing artificial agreement.

2. Methodology

This paper follows the expert-led narrative review design with the focus on the depth of the interpretation rather than on a comprehensive list. Structured searches were performed in SSRN, Google Scholar, Web of Science, Scopus, and institutional repositories based on query string families: ("artificial intelligence" OR "machine learning" OR "predictive analytics") AND "tax compliance" OR "IRS" OR "revenue forecasting; (federated learning" OR "differential privacy) AND tax. The 2020–2026 window was applied. Among 312 records found, 37 papers were included, following de-duplication, full-text screening, and quality appraisal. Evidence is graded: Tier 1: peer-reviewed empirical journals (14 sources, 38%); Tier 2: institutional/working papers (13 sources, 35%); Tier 3: preprints/grey literature (10 sources, 27%). The sources in Tier 3 are hedged. The inclusion criteria were AI/ML in tax or very closely related publicly financed subject matter with clear applicability to the U.S., and exclusion was on purely theoretical economic models, non-attributable web content, and redundant reporting.

3. Thematic Analysis

3.1. ML Compliance Risk Modeling and the Selective-Labels Problem

In compliance risk prediction, Yang (2025) discovered that gradient boosting classifiers obtained AUCs of the order of 0.85 -0.87 vs. 0.71 using logistic regression baselines (+14 pp). Neuman and Sheu (2022) demonstrated that ML-selected audit cases had lower no-change rates (~1520 pp) and correspondingly increased dollars measured per audit. Using artificial neural networks, Chan et al. (2025) found 12-18% improvement in audit yield prediction, compared to earlier baselines.

The selective-labels problem is structural: models that are trained on historical audit results are also biased by the historical audit selection and tend to over-flag historical audit populations and under-flag under-audited high-income populations (Holtzblatt & Engler, 2022; Swigart, 2024). Four functional remedies are suggested. To begin with, randomized exploration tranches: allocating 510 percent of audit resources to random selection to create unbiased ground truth, similar to the historical NRP. Second, further NRP-like audits as a statistical reference population on which to calibrate the model. Third, inverse propensity scoring (IPS) and off-policy evaluation: reweighting observed results to assess the performance under alternative audit policies that are not completely implemented. Fourth, regression-target reformulation: forecasting the anticipated adjustment values (continuous) instead of binary misreport probability. The results of NRP-based allocation studies indicate that regression targets raise the expected revenue per audit, and are more consistent with vertical equity, as the average increase in high-income returns is usually an absolute increase despite similar misreport rates (Yang, 2025; Neuman & Sheu, 2022).

3.2. Audit Selection, Revenue Forecasting, and Performance Benchmarks

Swenson (2024) demonstrated that deep-learning LSTM architectures brought a 22-30% reduction in corporate tax payment MAPE compared to ARIMA baselines, and the largest benefit in tail-risk prediction during periods of economic volatility; the most critical situation in fiscal policy. Chung et al. (2022) supported this trend on the state level. Pamisetty et al. (2022) estimated that the audit selection that may be conducted by ML might save the IRS inventory by 20-30 percent of the revenue yield target. Any reduction in revenue forecasting has a direct positive impact on federal budget planning and cash management of the Treasury (Swenson, 2024).

Table 1 Quantitative Performance Benchmarks from Key AI Tax Studies

Domain	Study	AI Approach	Key Metric / Effect Size	Evidence Tier
Compliance risk scoring	Yang (2025)	Gradient boosting	AUC: ~0.85 vs. ~0.71 (+14 pp)	Tier 1 — PLOS ONE
Audit lead selection	Neuman & Sheu (2022)	ML ensemble	No-change rate ↓ ~15–20 pp; higher net revenue/audit	Tier 1 — JATA
Audit yield prediction	Chan et al. (2022)	Artificial neural networks	Yield accuracy improvement: ~12–18%	Tier 1 — PLOS ONE
Revenue forecasting (corporate)	Swenson (2024)	Deep learning / LSTM	MAPE ↓ ~22–30% vs. ARIMA	Tier 1 — Forecasting
Revenue forecasting (state)	Chung et al. (2022)	Ensemble ML	MAPE ↓ ~15–25% in volatile periods	Tier 1 — PPMR
Fraud/anomaly detection	Amoako et al. (2025)	Autoencoders + LSTM	AUC-PR: 0.89–0.93; detection lag: 47 days → <4 hours	Tier 1 — EPRA JEBR

Note: pp = percentage points; MAPE = Mean Absolute Percentage Error; AUC-PR = Area Under Precision-Recall Curve; JATA = Journal of the American Taxation Association; PPMR = Public Performance & Management Review.

3.3. Fraud Detection and Financial Intelligence Architectures

Detection of tax fraud is associated with extreme class imbalance and adversarial adaptation, which renders accuracy measures misguided. Class-balanced measures, AUC-PR and Macro F1, are needed. Amoako et al. (2025) found AUC-PR of 0.89–0.93 on autoencoder architectures compared to 0.72–0.78 on rule-based systems, and detection lag was decreased by 47 days to less than 4 hours on identity theft refund fraud. Nartey et al. (2026) point to graph neural networks (GNNs) as applying to partnership abuse, shell company networks, and similar-party schemes; types of frauds that cannot be discerned using a single-entity feature-based scoring. According to Yeboah et al. (2026), AI-assisted audit systems detect deficiencies 28–35% better than standard practices. Anjarwi (2026) records the transition to real-time API level transaction monitoring, in place of retrospective audit cycles. The problem of data integration is still burning: IRS legacy COBOL systems do not allow even the seamless pipelines that ML fraud detection needs. (Bruckner & Coil, 2024; Ezeife et al., 2022).

3.4. Algorithmic Equity, Accountability, and Legal Frameworks

The evidence of the equity is tangible. Swigart (2024) reports that the audit rates of EITC claimants are 5–7 times higher than that of similar non-EITC filers in the lowest AGI deciles; a product of training data composition, but not a product of normative compliance judgment. The allocation research based on NRP indicates that regression-target reformulation (amounts of adjustments expected) inherently focuses allocation of audit resources in the higher part of the income distribution to minimize the concentration of burden at the lower end and maximize revenue yield (Yang, 2025; Neuman & Sheu, 2022).

Recommended quantifiable equity requirements: (a) Vertical equity monotonicity: audit rate decreases across AGI deciles even with complexity controls; (b) Demographic disparity ratio: audit rate ratio across the covered-class proxies is not more than 1.5:1 without reported justification; (c) Burden proxy measures: audit preparation cost per taxpayer disaggregated by income class; (d) Appeal and resolution rates disaggregated by income class to monitor selection defensibility.

Engstrom and Ho (2020) argue that the current doctrines of administrative law, notice-and-comment rulemaking, arbitrary-and-capricious review, and FOIA are structurally insufficient in terms of algorithmic administrative accountabilities, suggesting model card disclosures, algorithmic impact assessments (AIAs), and adversarial testing protocols. Mazur (2025) recommends changes to IRS statutory authority. According to Bezditnyi & Matyash (2026), one of the U.S. vulnerabilities is the lack of a statute governing AI that applies to the activities of the IRS. Siva (2025) brings up non-delegation doctrine and due process issues to agentic AI enforcement procedures.

3.5. Explainability Architectures, Governance, and Institutional Design

Domain-fit XAI of IRS. The following four architectures should be given priority: (1) Constrained gradient boosting with monotonicity and sparsity controls generates explanations of the form "selected because [Feature A] exceeds [threshold

X]) can be audited by IRS Appeals officers to decide whether a face is reasonable; (3) Disaggregated Model cards by income/complexity class, mandatory performance disclosure checked by TIGTA; (4) Simulation-based stress tests and adversarial probing, test robustness to intentional evasion, and information security (Engstrom & Ho, 2020; Mazur, 2025; Guglyuvatyy, 2025)

Governance alignment: NIST AI RMF 1.0 organizes the risk management over GOVERN, MAP, MEASURE, and MANAGE functions that can be directly applied to the IRS AI deployment lifecycle. EO 14110 also mandates federal agencies to come up with AI governance inventories, safety standards, and red-teaming protocols. The rights-impacting AI systems operationalized by OMB M-24-10 are relevant to IRS audit selection algorithms, which obviously qualify as rights-impacting AI systems. Table 2 cross-tabulates the governance artifacts and these structures and supervision institutions.

Table 2 Governance Artifact Mapping for IRS AI Systems

Artifact	NIST AI RMF Function	EO 14110 / OMB M-24-10	IRS / Treasury Oversight Body
Model card (disaggregated performance)	MEASURE	M-24-10: Rights-impacting AI minimum practice	TIGTA; IRS Office of Chief Counsel
Algorithmic impact assessment (AIA)	MAP: Risk identification	EO 14110 Sec. 5(b): Agency AI safety	Treasury AI Safety Board; OMB OIRA
Red-team / adversarial testing	MANAGE: Risk response	EO 14110 Sec. 4(a): High-impact AI red-teaming	TIGTA; IRS Cybersecurity Division
Vertical equity monotonicity audit	MEASURE: Fairness metrics	M-24-10: Civil rights impact assessment	IRS EDI Office; TIGTA
Continuous MLOps monitoring dashboard	GOVERN: Ongoing accountability	M-24-10: Ongoing risk management	IRS IT Modernization; Treasury CIO

Note: EO = Executive Order; OMB = Office of Management and Budget; TIGTA = Treasury Inspector General for Tax Administration; OIRA = Office of Information and Regulatory Affairs; EDI = Equity, Diversity & Inclusion.

Agentic AI taxonomy. Three levels of IRS AI autonomy with varied HITL requirements: (A) Advisor: AI comes up with ranked recommendations that are reviewed by people, no action is taken without human decision (present IRS baseline). (B) Conditional: AI triggers low-stakes automated actions (CP2000 notices, mathematical corrections) in statutory types of correction; human review is triggered by taxpayer contest. (C) Agentic: AI programs run multi-step enforcement processes on their own (lien filings, penalty initiation). Tier C poses the most serious non-delegation and due process questions (Siva, 2025) and must be deemed out of bounds unless expressly authorized by Congress.

Federated learning under 26 U.S.C. § 6103. Section 6103 forbids the publication of taxpayer information beyond legitimate mediums, limiting the cross-agency data exchange to build ML. Federated learning: Training models on distributed silos with gradients instead of raw record sharing: This is a technically promising architecture that can prevent the disclosure violations of § 6103. Feasibility is governed by two unresolved conditions: (1) whether gradient sharing is considered as disclosure under § 6103 (must be OLC-guided); (2) whether the differential privacy noise budgets that are consistent with the privacy standard of section 6103 (recommended $\epsilon \leq 1.0$) can be zero-calibrated without unacceptable degradation of utility. Proposed minimal viable pilot test: a federated IRSSSA identity theft fraud detection model (using the current exception of § 6103(d) SSA disclosure), which will have to be independently reviewed by the IRS Chief Counsel and DOJ Tax Division (Nartey et al., 2026).

Dual-use LLM risks. Compliance tools that are powered by LLM can help decrease the burden on taxpayers and automate the review of unstructured documents, yet the same can be used to create new avoidance strategies (Shaikh, 2025; Bruckner & Coil, 2024). Recommended mitigations: (a) curated authority-grounded corpora restricted to IRS and court documents; (b) retrieval-augmented generation (RAG) with binding authority citations instead of synthesizing new strategies; (c) refusal policies to structure queries in avoidance form; (d) tiered access control between taxpayer-service and internal analytics interfaces; (e) independent legal review before any advisory.

4. Applications Across the Tax Compliance Lifecycle

Table 3 AI Applications — IRS-Specific Evidence vs. Cross-Domain Analogies

Domain	Methodology	IRS-Specific Evidence	Cross-Domain Analogy	Tier
Compliance risk scoring	Gradient boosting, neural nets	Neuman & Sheu (2022); Yang (2025)	Banking credit risk	Tier 1
Audit lead selection	ANN; ensemble methods	Chan et al. (2022); Holtzblatt & Engler (2022)	OCC regulatory exams	Tier 1–2
Revenue forecasting	Deep learning; LSTM	Swenson (2024); Chung et al. (2022)	State revenue authorities	Tier 1
Fraud & anomaly detection	Autoencoders; GNN; blockchain	Amoako et al. (2025) — banking AML	FinCEN SAR analytics	Tier 1 (cross-domain)
NLP / document processing	Transformers; RAG	Bruckner & Coil (2024 — IRS pilots)	SEC EDGAR processing	Tier 2
Agentic enforcement workflows	Multi-agent; LLM orchestration	Siva (2025) — conceptual/legal	SEC robo-advice	Tier 3

Note: Cross-domain analogies are explicitly labeled. Findings should not be generalized to IRS contexts without direct empirical validation. ANN = Artificial neural network; GNN = Graph neural network; OCC = Office of the Comptroller of the Currency.

5. Benefits, Limitations, and Ethical Considerations

5.1. Benefits

The cost-effectiveness of the audit selection process in AI-assisted audit selection is enhanced by a factor of around 1520 pp over legacy DIF systems, improving the cost-efficiency of a long-term resource-limited enforcement strategy (Neuman & Sheu, 2022). Deep learning revenue forecasting lessens MAPE by 15-30% under economic circumstances, significantly enhancing accuracy of federal budget planning (Swenson, 2024; Chung et al., 2022). Real-time fraud detection minimizes a 47 days detection lag to less than 4 hours (Amoako et al., 2025). AI tools can be used to reduce compliance load and enhance data quality by facing taxpayers (Islam et al., 2025). The effects of algorithmic deterrence, the improvement of voluntary compliance out of the perceived monitoring, could contribute 10-15 percent on top of the gains through the direct detection (Islam et al., 2025).

5.2. Limitations

ML data pipelines are limited by legacy COBOL-based IRS infrastructure (Bruckner & Coil, 2024; Mazur, 2025). The selective-labels issue creates structural training bias that must be addressed actively. Model drift requires continuous retraining regimes and MLOps infrastructure for monitoring data drift (PSI, KS tests) and calibration (ECE, Brier scores). Deep neural nets, high-performance architectures are functionally opaque, which demands legal exposure under doctrines of administrative law enforcing reasoned explanation (Engstrom & Ho, 2020; Guglyuvatyy, 2025).

5.3. Ethical Considerations

The rate of audits of EITC claimants is 5-7 times higher than the rate of audits of similar non-EITC filers; a training data artifact, rather than a compliance artifact (Swigart, 2024). Regression-target reformulation is empirically viable and technically viable, and should be considered a design requirement. The issue of privacy and data minimization is that the broader the range of data gathered by third parties, the more fiscal surveillance expands on constitutionally unresolved Fourth Amendment and other standards of the Fourth Amendment and § 6103 (Bezditnyi & Matyash, 2026; Siva, 2025). The mitigation architecture in Section 2.5 is needed to address dual-use LLM risks. The asymmetry in burden of AI enforcement; legal expenses, documentation expenses places heavy burdens on taxpayers who lack professional representation (Soled & Thomas, 2023; Alarie et al., 2021).

5.4. Organizational Capacity and MLOps

The IRS skills profile of the workforce; that of rule-based administration needs to be systematically upskilled in the basics of ML, statistical analysis, and algorithmic responsibility. Standard procurement requirements must state compulsory XAI requirements, fairness auditing procedures and MLOps requirements as contractual. Such continuous monitoring infrastructure must comprise: automated data drift detection by PSI and KS tests; calibration monitoring by ECE and Brier scores; and official incident reporting channels to TIGTA and Congressional oversight committees. Sajid et al. (2025) and Ezeife (2021) give complementary organization readiness models that focus on the institutional aspects that can either make AI investments return expected benefits.

Future Directions and Research Gaps

This synthesis gives rise to six priority research gaps. To obtain credible causal estimates of equity effects, first, causal equity studies that use temporal variation in the intensity of IRS AI deployment or randomized exploration tranches as suggested above will be used to obtain such estimates. Second, empirical comparison of classification and regression audit targets using actual IRS NRP data and multi-objective optimization frameworks that both specify revenue, equity, cost and burden targets. Third, XAI sufficiency testing of legal requirements: simulated IRS Appeals that address how constrained gradient boosting descriptions and counterfactual productions meet administrative law reasoned-explanation criteria. Fourth, the feasibility of federated learning as disclosed in § 6103: OLC gradient sharing as disclosure, DP calibration experiments, and the IRS-SSA pilot study outlined above. Fifth, longitudinal behavioral designs to assess deterrence, chilling, and gaming of AI-enhanced enforcement visibility among taxpayer groups. Sixth, AI compliance software tailored to gig workers, crypto holders and small businesses as the groups of taxpayers who are at high risk of non-compliance and face the highest challenges to compliance support.

6. Conclusion

The AUC score improvements of approximately 14 pp in compliance risk scoring, no-change rate improvements of 15-20 pp in audit selection and MAPE reduction of 15-30% in revenue forecasting, and improvement of detection lag from 47 days to less than 4 hours, all demonstrate that AI can substantially enhance IRS operational effectiveness. The problem of selective-labels, the regression-target reformulation imperative and the equity-amplification risk reported in the literature all prove just as much that naive implementation in the absence of structural solutions will reinforce the status quo inequalities.

The governance structure needed is now more clearly defined: NIST AI RMF 1.0, EO 14110, and OMB M-24-10 offer regulatory scaffolding; model cards, AIAs, adversarial testing, and counterfactual explanation protocols offer operational tools; TIGTA offers accountability with respect to institutional oversight. The three-level model of agentic AI taxonomy offered here Advisory, Conditional, and Agentic offers an effective classification system to IRS procurement and Congressional oversight. To researchers: causal equity-studies, regression-target empirical comparison, XAI adequacy testing, and federated learning feasibility. To the administrators of the IRS: equity-by-design procurement standards, workforce development and MLOps infrastructure. To the policymakers: legislative modernization whereby the IRS AI governance powers, entitlement to explain to taxpayers, and authority to audit systems deployed to the congress.

Compliance with ethical standards

Disclosure of conflict of interest

No conflict of interest to be disclosed.

References

- [1] Akinrinola, O., Addy, W. A., et al. (2024). Application of machine learning in tax prediction: A review with practical approaches. *Global Journal of Multidisciplinary Research*.
- [2] Alarie, B., Niblett, A., & Yoon, A. (2021). Data analytics and tax law. In *Research Handbook on Big Data Law*. Edward Elgar Publishing.
- [3] Amoako, D., Boboye, C. O., Boateng, V., & Laryea, J. E. N.-L. (2025). Artificial intelligence and machine learning for fraud detection in the U.S. banking industry. *EPRA JEBR*. <https://doi.org/10.36713/epra23647>

- [4] Amoakoh, C., Nwinyi, I. P., & Ampofo, I. (2026). AI-enabled assurance frameworks: Intelligent systems enhancing accuracy and independence in U.S. financial auditing. *EPRA JEBMS*, 154. <https://doi.org/10.36713/epra26251>
- [5] Anjarwi, A. W. (2026). The digital transformation of tax audits: AI, big data, blockchain, and advanced analytics in tax evasion detection. *Journal of Business Analytics*.
- [6] Bezditnyi, V., & Matyash, A. (2026). Artificial intelligence in tax administration: Legal limits and regulatory risks. *International Journal of Modern Studies*.
- [7] Bruckner, C., & Coil, C. (2024). AI and the modern tax agency. IBM Center for The Business of Government.
- [8] Chan, T., Tan, C. E., & Tagkopoulou, I. (2022). Audit lead selection and yield prediction from historical tax data using artificial neural networks. *PLOS ONE*.
- [9] Chung, I. H., Williams, D. W., & Do, M. R. (2022). For better or worse? Revenue forecasting with machine learning approaches. *Public Performance & Management Review*.
- [10] Engstrom, D. F., Ho, D. E., Sharkey, C. M., & Cuéllar, M. F. (2020). Government by algorithm: AI in federal administrative agencies. *NYU School of Law*.
- [11] Engstrom, D. F., & Ho, D. E. (2020). Algorithmic accountability in the administrative state. *Yale Journal on Regulation*.
- [12] Enuma, E., Eseoghene, K., & Princess, E. O. (2025). Agile tax technology development in the US: A conceptual framework. *Gulf Journal of Management*.
- [13] Ezeife, E. (2021). AI-driven tax technology in the United States: A business analytics framework. *International Journal of Multidisciplinary Studies*.
- [14] Ezeife, E., Kokogho, E., Odio, P. E., et al. (2022). Managed services in the US tax system: A theoretical model for scalable tax transformation. *International Journal of Research*.
- [15] Guglyuvatyy, E. (2025). Balancing innovation and integrity: AI in tax administration and taxpayer rights. *Humanities and Social Sciences Communications*.
- [16] Hayek, A. F., & Noordin, N. A. (2024). *Tax data analytics*. Springer.
- [17] Hoffer, S. (2020). What if tax law's future is now? *Ohio State Technology Law Journal*.
- [18] Holtzblatt, J., & Engler, A. (2022). Machine learning and tax enforcement. Urban Institute & Tax Policy Center.
- [19] Islam, M. I., Nisa, K. U., Mufti, S., & Ansarullah, S. I. (2025). Artificial intelligence in tax compliance: Transforming taxpayer behavior and system efficiency. *IGI Global*.
- [20] Kaulwar, P. K. (2025). Reimagining tax and advisory services: Intelligent systems, security, and data-driven decision making. Springer.
- [21] Makoni, C. K., & Laryea, J. E. N.-L. (2025). Consequences of non-tax compliance: Revenue leakage and its impact on federal infrastructure. *EPRA JEBR*, 13(9). <https://doi.org/10.36713/epra24069>
- [22] Martinez, A. L. (2025). Artificial intelligence in tax administration: Enhancing compliance, transparency, and ethical governance. *SSRN*.
- [23] Mazur, O. (2025). Modernizing the IRS in the age of AI. *Virginia Tax Review*.
- [24] Mitchell, M., Wu, S., Zaldivar, A., et al. (2019). Model cards for model reporting. *Proceedings of the ACM FAT* Conference*. <https://doi.org/10.1145/3287560.3287596>
- [25] Nartey, O. L., Agyei, E., Anim, B., & Yeboah, M. M. (2026). AI and blockchain in forensic auditing: Tools for investigating financial crimes in the U.S. *EPRA JEBR*.
- [26] Neuman, E. L., & Sheu, R. J. (2022). Big data analytics in IRS audit procedures and its effects on tax compliance. *Journal of the American Taxation Association*.
- [27] Nwinyi, I. P., & Oman-Amoako, M. (2026). AI-driven business analytics frameworks for government financial management. *EPRA JEBR*. <https://doi.org/10.36713/epra26112>
- [28] Okeke, P. T., Okumu, F., Obodai, T., Adams, I., & Buabin, W. K. A. (2025). Predictive analytics in financial forecasting: A comparative study. *EPRA JEBR*.
- [29] Pamisetty, V. (2023). AI for strategic decision-making in tax administration and policy design. *SSRN*.

- [30] Pamisetty, V., Pandiri, L., Singreddy, S., et al. (2022). Leveraging AI, ML, and big data for tax compliance, fraud detection, and predictive analytics. SSRN.
- [31] Sajid, M. H., Ali, A., & Jadoon, A. K. (2025). Smart tax systems and AI: Transforming compliance and enforcement in the digital era. MPRA Paper.
- [32] Shaikh, S. (2025). Balancing innovation and oversight: AI in the US Treasury and IRS. arXiv.
- [33] Siva, P. K. (2025). Legal and regulatory implications of agentic AI in tax administration. Tax Notes International.
- [34] Soled, J. A., & Thomas, K. D. L. (2023). Predictive analytics and the tax code. Florida State University Law Review.
- [35] Swenson, C. (2024). Using deep learning AI to improve forecasting of tax payments for corporations. Forecasting.
- [36] Swigart, N. (2024). A new age of equity: Machine learning and equity in tax law. Texas Review of Law & Politics.
- [37] Yang, L. (2025). Predictive modeling of tax compliance risks: A comparative study of machine learning approaches. PLOS ONE.
- [38] Yeboah, M. M., Agyei, E., Anim, B., Nwinyi, I. P., & Nartey, O. L. (2026). AI in financial auditing: Addressing U.S. audit failures and strengthening PCAOB/SEC compliance. EPRA JEBR.