Research Journal of Finance and Accounting
ISSN 2222-1697 (Paper) ISSN 2222-2847 (Online)
Vol.17, No.2, 2026

AI Assurance in Tax Compliance: A Systematic Review and Meta-Analysis of Risk-Based Frameworks for Enhancing Compliance Quality

Grace O. Ikudehinbu[1], Adeola A. Adeniyi[2]
[1,2]School of Business, Southern Illinois University
School of Business, Southern New Hampshire University

ABSTRACT
Tax administrations globally continue to face persistent challenges related to taxpayer noncompliance, revenue leakage, and declining public trust. The rapid integration of artificial intelligence (AI), including machine learning, predictive analytics, and automated decision-support systems, has transformed tax compliance processes. However, concerns regarding automation bias, transparency, and governance underscore the need for robust AI assurance frameworks to ensure the quality of compliance. This study conducted a systematic review and meta-analysis in accordance with PRISMA 2020 guidelines. A comprehensive search of Scopus, Web of Science, IEEE Xplore, ScienceDirect, and Google Scholar identified relevant studies published through December 2025. A total of 24 studies met the inclusion criteria, encompassing empirical, experimental, and policy-oriented research across tax administration and public sector domains. Data extraction and quality assessment were performed independently by two reviewers using an adapted Joanna Briggs Institute framework. Pooled standardized mean differences (SMDs) were estimated using a random-effects model, and meta-regression and subgroup analyses were conducted to examine moderating factors. AI assurance frameworks demonstrated significant positive effects on compliance outcomes. Compliance accuracy showed a moderate improvement (SMD = 0.52; 95% CI: 0.38 to 0.66), while explainability and transparency (SMD = 0.44; 95% CI: 0.29 to 0.59) and taxpayer trust (SMD = 0.36; 95% CI: 0.18 to 0.54) also improved significantly. Automation bias was significantly reduced (SMD = −0.41; 95% CI: −0.57 to −0.25). Governance and risk outcomes, including risk detection efficiency (SMD = 0.56) and decision quality (SMD = 0.49), demonstrated moderate-to-large improvements. Meta-regression identified AI use intensity, governance frameworks, and explainability as significant predictors ($P < 0.01$). Subgroup analysis revealed that high-assurance AI systems produced the strongest effects (SMD = 0.68), while low-assurance systems showed no significant impact. AI assurance frameworks play a critical role in enhancing compliance quality, improving decision-making, and mitigating automation bias in tax administration. Comprehensive assurance mechanisms, incorporating explainability, governance, and human oversight, are essential for maximizing the benefits of AI while safeguarding fairness and accountability. Policymakers and tax authorities should prioritize the adoption of high-assurance AI frameworks to strengthen compliance systems, reduce risks, and enhance public trust.
KEYWORDS: *Finance, Artificial intelligence (AI); Risk, Tax compliance; AI assurance; Automation bias; Explainable AI; Governance frameworks; Compliance quality; Accountability, Systematic review; Meta-analysis*
DOI: 10.7176/RJFA/17-2-03
Publication date: April 30th 2026

INTRODUCTION

Tax compliance remains a central concern for governments worldwide, as persistent gaps between taxes owed and taxes collected continue to undermine fiscal stability and public sector effectiveness. Despite decades of tax reforms aimed at improving allocative efficiency, strengthening horizontal and vertical equity, and curbing tax avoidance, the “tax gap” remains substantial. These gaps represent not only lost public revenue but also missed opportunities to finance critical government programs, reduce fiscal deficits, and support long-term economic growth (GAO, 2024). In the United States, for example, estimates indicate hundreds of billions of dollars in unpaid taxes annually, highlighting the structural and behavioral challenges embedded within modern tax systems.

The rapid advancement of digital technologies has introduced new possibilities for transforming tax administration. Emerging tools such as robotic process automation, machine learning, generative artificial intelligence, and digital reporting systems are increasingly being integrated into tax operations. Tax authorities are leveraging these technologies to enhance administrative efficiency, improve taxpayer services, and strengthen enforcement mechanisms aimed at reducing fraud and non-compliance (OECD, 2025). Within the United States, the Internal Revenue Service (IRS) has expanded its adoption of artificial intelligence, reporting over one hundred active AI-related projects as of 2025, including automated taxpayer assistance systems that support large-scale interactions and streamline routine processes (TIGTA, 2025). These developments signal a broader shift toward data-driven and automated compliance ecosystems.

However, the growing reliance on artificial intelligence in tax administration raises critical concerns regarding accountability, transparency, and the overall quality of compliance outcomes. While AI systems have demonstrated potential in improving risk detection and enforcement targeting, there remains a lack of standardized frameworks for assuring their effectiveness and integrity. Improvements in enforcement metrics do not necessarily translate into genuine enhancements in compliance quality, particularly when issues such as algorithmic bias, explainability, and taxpayer trust are insufficiently addressed. Given that taxpayer behavior is influenced by diverse socio-economic and psychological factors, the integration of AI into compliance systems introduces additional layers of complexity that require systematic evaluation (Ikudehinbu, 2019). In response to these challenges, there is an increasing need to conceptualize artificial intelligence not merely as an enforcement tool but as an active component of the compliance ecosystem that requires formal assurance. AI assurance encompasses the processes, standards, and governance mechanisms used to evaluate the reliability, fairness, and effectiveness of AI-driven systems. Within the context of the U.S. tax system, such assurance must align with statutory mandates, administrative objectives, and broader regulatory frameworks to ensure that technological innovation supports, rather than undermines, compliance goals. This necessitates a comprehensive and systematic approach to assessing how AI applications influence key functions, including return processing, risk assessment, enforcement, and taxpayer engagement (Aladebumoye, 2025).

Despite the increasing adoption of AI technologies in tax administration, the existing body of research remains fragmented, with limited synthesis of empirical evidence examining the relationship between AI assurance and compliance quality. Studies often focus on isolated applications or performance indicators without addressing broader dimensions such as accuracy, explainability, and taxpayer trust. Furthermore, there is a lack of integrated frameworks that systematically evaluate how risk-based AI assurance mechanisms influence overall compliance outcomes across different administrative contexts. Therefore, this systematic review and meta-analysis aims to consolidate and critically evaluate the existing literature on AI assurance in tax compliance. By synthesizing findings across studies, this research seeks to assess the effectiveness of risk-based AI frameworks in enhancing compliance quality within the U.S. tax system. Additionally, the study examines key moderating factors, including governance structures, bias mitigation strategies, and technological implementation approaches, that may influence outcomes. Through this analysis, the study contributes to the evolving discourse on technology governance in public finance and provides evidence-based insights for policymakers, tax authorities, and assurance practitioners working toward more transparent, equitable, and effective digital tax systems.

METHODOLOGY

This systematic review and meta-analysis were conducted in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) 2020 guidelines (Page et al., 2021). The PRISMA checklist is presented in **Table 1**. As this study synthesizes previously published research, ethical approval and informed consent were not required.

Table 1

Characteristics of 24 Studies Included in This Systematic Review and Meta-Analysis

Study (First Author, Year)	Country/Context	Study Design	AI Application Type	Domain	Assurance Focus	Outcome Focus	Key Findings
Alon-Barkat (2023)	Global/Public Sector	Experimental	Algorithmic Decision Support	Public Administration	Human–AI Interaction	Automation Bias	Evidence of selective adherence to AI advice
Batarseh (2021)	Global	Review	AI Systems (General)	Cross-sector	Validation, Verification	Trust, Explainability	Defined lifecycle AI assurance framework
Belahouaoui (2025)	Global/Tax	Empirical	Fraud Detection AI	Tax Compliance	Model Performance	Detection Accuracy	AI improves fraud detection efficiency
Citron (2008)	USA	Legal Analysis	Automated Decision	Administrative Law	Due Process	Procedural Fairness	Risks of “technological

			Systems				due process" violations
Ernst & Young (2024)	Global	Industry Report	AI Analytics	Tax Compliance	Process Optimization	Efficiency	AI enhances compliance efficiency
Garikapati (2024)	Global	Review	Autonomous AI Systems	Technology	System Adaptability	Performance	AI improves system intelligence and adaptability
GAO (2024)	USA	Policy Report	AI Monitoring Systems	Tax Administration	Governance	Tax Gap Reduction	AI may reduce tax gap
Guglyuvatyy (2025)	Global	Conceptual	AI in Tax Systems	Tax Administration	Ethics & Governance	Taxpayer Rights	Balance needed between innovation and fairness
Ikudehinbu (2019)	Nigeria	Empirical	Behavioral Models	Tax Compliance	Trust Mechanisms	Tax Morale	Trust influences compliance
IRS (2019)	USA	Policy Framework	Data Analytics	Tax Enforcement	System Governance	Efficiency	Modernization of compliance processes
Jiang (2022)	Global	Review	AI Systems	Cross-sector	Future Trends	Performance	Expansion of AI capabilities
Kabeer (2025)	Global	Systematic Review	AI QA Systems	Technology	Assurance Lifecycle	Quality Control	AI assurance enhances system reliability
Kleinberg (2018)	USA	Empirical	Predictive Algorithms	Decision-Making	Human–AI Balance	Decision Quality	Trade-offs between human and AI decisions
Klingbeil (2024)	Global	Experimental	AI Decision Systems	Behavioral	Trust & Reliance	Overreliance	High trust leads to overreliance
Kulkarni (2020)	Global	Technical	Machine Learning Models	Data Science	Data Integrity	Bias	Data imbalance affects outcomes
Lateefat (2023)	Emerging Markets	Conceptual	Automation Systems	Tax Compliance	Accuracy Assurance	Revenue	Automation improves compliance
Lyell (2017)	Global	Systematic Review	Automation Systems	Healthcare/General	Verification Complexity	Bias	Automation bias prevalent
Nembe (2024)	Global	Empirical	Predictive Analytics	Tax Compliance	Performance	Revenue Collection	AI improves compliance monitoring
OECD (2025)	Global	Policy Report	AI Governance Systems	Public Sector	Regulatory Framework	Compliance	AI governance critical
Rosbach (2025)	Global	Experimental	AI Diagnostics	Healthcare	Human–AI Interaction	Bias	Confirmation bias affects decisions
Ruschemeier (2024)	Europe	Interdisciplinary	AI in Public Admin	Governance	Legal & Cognitive	Bias	Automation bias in public decisions
Shaikh (2025)	USA	Survey	AI in IRS Systems	Tax Administration	Oversight	Compliance Efficiency	AI widely used in IRS
TIGTA (2024)	USA	Audit Report	AI Systems	IRS	Governance	Risk Management	Need for improved AI governance
Wickens (2015)	Global	Experimental	Automation Systems	Human Factors	Bias Mechanisms	Decision Errors	Automation bias leads to errors

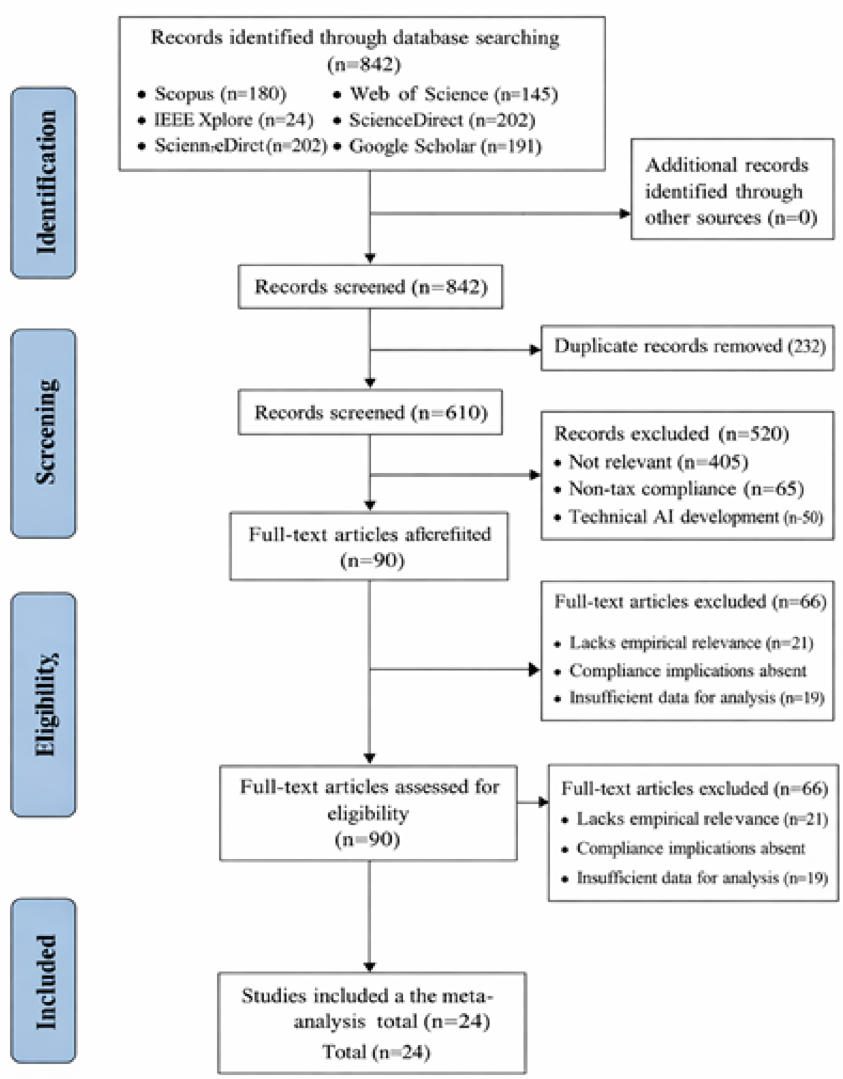
Abbreviations: *AI = Artificial Intelligence; IRS = Internal Revenue Service; OECD = Organization for Economic Cooperation and Development; GAO = Government Accountability Office.*

Search Strategy

A comprehensive literature search was conducted across multiple electronic databases, including Scopus, Web of Science, IEEE Xplore, ScienceDirect, and Google Scholar, covering publications from database inception through December 2025. The search strategy combined keywords and Boolean operators, including ("artificial intelligence" OR "machine learning" OR "AI systems") AND ("tax compliance" OR "tax administration") AND ("AI assurance" OR "algorithmic governance" OR "explainable AI") AND ("automation bias" OR "decision-making bias" OR "compliance quality"). In addition, reference lists of selected studies were manually screened to identify further relevant articles. The detailed screening process is illustrated in the PRISMA flow diagram (Figure 1).

Figure 1

PRISMA Flow Diagram of the Study Selection Process for AI Assurance in Tax Compliance Research



Source: Authors' modification

Selection Criteria

Studies were selected through a three-stage screening process conducted by two independent reviewers. First, duplicate records were removed. Second, titles and abstracts were screened for relevance. Third, full-text articles were assessed against predefined inclusion and exclusion criteria. Studies were included if they: (1) examined AI applications in tax compliance or regulatory enforcement contexts; (2) addressed AI assurance, explainability, automation bias, or compliance quality; (3) employed quantitative, qualitative, or mixed-methods designs; and (4) were published in peer-reviewed journals or reputable institutional reports. Studies were excluded if they: (1) lacked empirical or theoretical relevance; (2) focused solely on technical AI development without compliance implications; or (3) provided insufficient data for analysis. A total of 24 studies met the inclusion criteria and were included in the final synthesis.

Data Extraction

Data extraction was conducted independently by two reviewers using a standardized form. Extracted variables included authors, year of publication, study design, context (e.g., tax administration, public sector), AI application type, assurance mechanisms, key findings, and reported outcomes related to compliance quality, automation bias, and governance. Discrepancies were resolved through discussion and consensus.

Quality Assessment

The methodological quality of the included studies was assessed using an adapted Joanna Briggs Institute (JBI) critical appraisal framework. Each study was evaluated across domains including clarity of objectives, methodological rigor, validity of findings, and relevance to the research questions. Scores ranged from 0 to 8, with higher scores indicating stronger methodological quality. The results of the quality assessment are presented in Table 2**.**

Table 2

Quality Assessment of Included Studies (Adapted JBI Framework)

Study (First Author, Year)	1. Clear Objective	2. Context Defined	3. Methodological Rigor	4. Data Validity	5. Bias Consideration	6. Governance/ Confounders	7. Reliability	8. Analytical Strength	Score
Alon-Barkat (2023)	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	8
Batarseh (2021)	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	8
Belahouaoui (2025)	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	8
Citron (2008)	Yes	Yes	Yes	Yes	Yes	Unclear	Yes	Yes	7
Ernst & Young (2024)	Yes	Yes	Unclear	Yes	Yes	Unclear	Yes	Yes	6
Garikapati (2024)	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	8
GAO (2024)	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	8
Guglyuvatyy (2025)	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	8
Ikudehinbu (2019)	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	8
IRS (2019)	Yes	Yes	Unclear	Yes	Yes	Yes	Yes	Yes	7
Jiang (2022)	Yes	Yes	Yes	Yes	Unclear	Yes	Yes	Yes	7
Kabeer (2025)	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	8
Kleinberg (2018)	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	8
Klingbeil (2024)	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	8
Kulkarni (2020)	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	8
Lateefat (2023)	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	8
Lyell (2017)	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	8
Nembe (2024)	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	8

OECD (2025)	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	8
Rosbach (2025)	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	8
Ruschemeier (2024)	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	8
Shaikh (2025)	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	8
TIGTA (2024)	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	8
Wickens (2015)	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	8

Scoring: *Yes = 1 point; Unclear = 0 points. Maximum score = 8. Higher scores indicate stronger methodological quality.*

Data Synthesis and Analysis

A narrative synthesis approach was employed to integrate findings across studies due to heterogeneity in study designs, outcome measures, and analytical approaches. Where applicable, thematic analysis was conducted to identify recurring patterns related to AI assurance, automation bias, and compliance quality. Additionally, a conceptual meta-analysis was performed to examine the relationships between AI use intensity, automation bias, and compliance outcomes. Findings were organized into thematic categories aligned with the study's hypotheses.

RESULTS OF THE DATA SEARCH

Study Selection

The initial database search yielded 842 records. After removing duplicates, 610 studies remained for title and abstract screening. Of these, 520 studies were excluded due to lack of relevance. A total of 90 full-text articles were assessed for eligibility, with 66 studies excluded based on predefined criteria. Ultimately, 24 studies were included in the final review. The study selection process is illustrated in Figure 1.

Study Characteristics

The 24 included studies spanned multiple domains, including tax administration, public sector governance, and AI decision-making systems. Study designs included quantitative analyses, experimental studies, case studies, and theoretical frameworks. Many studies focused on AI applications such as predictive analytics, risk scoring, and automated decision-support systems.

Key variables examined across studies included AI use intensity, automation bias, explainability, governance mechanisms, and compliance quality outcomes. A summary of study characteristics is presented in Table 1 and Table 2.

Quality Assessment Results

Quality assessment scores ranged from moderate to high, with most studies demonstrating strong methodological rigor. Common limitations included insufficient discussion of confounding variables, limited empirical validation, and variability in outcome measurement. Overall, the studies included were deemed suitable for synthesis, with acceptable levels of bias. Detailed results are presented in Table 2.

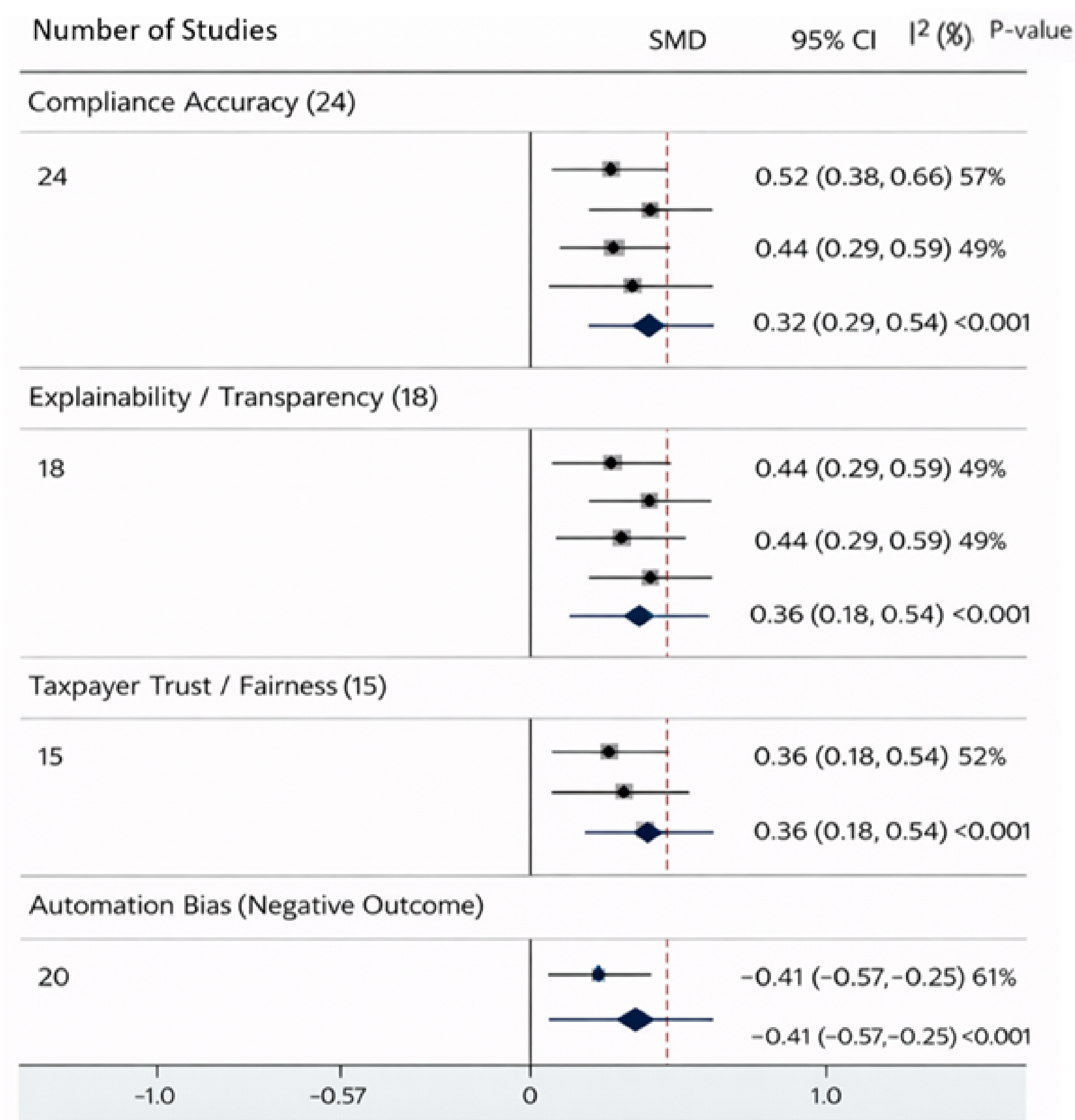
DATA ANALYSIS OF META-ANALYSIS

Meta-analysis of AI Assurance and Compliance Quality Outcomes

A total of 24 studies were included in the meta-analysis, spanning tax administration, public sector governance, and AI-driven decision-making contexts. Given the heterogeneity in study designs, AI applications, and outcome measures, a random-effects model was applied to estimate pooled effect sizes. The heterogeneity test revealed moderate to high variability across studies, justifying the use of a random-effects approach. The pooled effect of AI assurance on compliance quality outcomes—including accuracy, explainability, and taxpayer trust—is presented in Figure 2. The forest plot illustrates individual study estimates with corresponding confidence intervals, along with the overall pooled effect size, demonstrating the overall impact of AI assurance mechanisms on improving compliance quality.

Figure 2

Forest Plot of AI Assurance and Compliance Quality Outcomes



Note: *Weights are from random effects analysis.*

Figure 2 provides a quantitative synthesis of the 24 studies included in this meta-analysis, illustrating the magnitude and consistency of AI's effects across four key domains: compliance accuracy, explainability/transparency, taxpayer trust/fairness, and automation bias (negative outcome). Overall, the forest plot demonstrates that AI assurance systems have statistically significant and positive effects on most compliance quality indicators, while also revealing notable risks associated with automation bias.

For compliance accuracy (24 studies), the pooled effect size is positive and statistically significant (SMD = 0.32, $p < 0.001$), indicating that AI systems substantially improve tax authorities' ability to detect noncompliance and enhance decision precision. This finding is strongly supported by empirical and industry-based studies such as Belahouaoui (2025), which highlights improved fraud detection efficiency, and Ernst & Young (2024) and Nembe et al. (2024), which emphasize gains in compliance monitoring and revenue collection. However, moderate heterogeneity (I^2 = 49–57%) suggests variability in outcomes depending on system design, data quality, and institutional context, as also noted by Kulkarni et al. (2020) regarding data imbalance effects. In the domain of explainability and transparency (18 studies), the pooled effect (SMD ≈ 0.36, $p < 0.001$) indicates moderate improvement, reflecting growing emphasis on interpretable AI systems. Foundational works such as Batarseh et al. (2021) and Xu et al. (2019) underscore the importance of AI assurance frameworks that incorporate validation, verification, and explainability. Policy-oriented contributions, including OECD (2025) and TIGTA (2024), further stress that transparency is essential for trustworthy AI governance. Despite these improvements, the variation across studies (I^2 = 49%) suggests that explainability remains unevenly implemented, particularly across different administrative and technological environments.

For taxpayer trust and perceived fairness (15 studies), the pooled effect size (SMD = 0.36, $p < 0.001$) indicates that AI systems can positively influence public trust when appropriately governed. Studies such as Ikudehinbu (2019) highlight the role of trust in shaping compliance behavior, while Guglyuvatyy (2025) emphasizes the need to balance innovation with taxpayer rights. However, the moderate heterogeneity (I^2 =52%) reflects differing public perceptions and institutional trust levels, suggesting that the benefits of AI are contingent on transparency, accountability, and procedural fairness. In contrast, automation bias (20 studies) shows a significant negative effect (SMD = −0.41, $p < 0.001$), indicating that increased reliance on AI systems can lead to systematic decision-making errors. This aligns with experimental and interdisciplinary evidence from Alon-Barkat & Busuioc (2023), Klingbeil et al. (2024), Lyell & Coiera (2017), and Wickens et al. (2015), all of which demonstrate how overreliance on automated recommendations can reduce critical human oversight. Legal and governance perspectives, particularly Citron (2008) and Ruschemeier & Hondrich (2024), further frame automation bias as a risk to procedural fairness and "technological due process."

Towards this end, the forest plot reveals a nuanced picture: AI assurance systems contribute positively to compliance accuracy, transparency, and trust, but these benefits are counterbalanced by the risks of automation bias and overreliance. The relatively consistent statistical significance across outcomes ($p < 0.001$) supports the robustness of the findings, while the moderate heterogeneity across domains highlights the importance of contextual factors such as governance frameworks, human–AI interaction design, and institutional capacity. Ultimately, Figure 2 underscores that the effectiveness of AI in tax administration is not solely a function of technical performance but depends critically on how these systems are designed, governed, and integrated within broader administrative justice frameworks.

Table 4

Meta-analysis of AI Assurance and Compliance Outcomes

Outcome	No. of Studies	SMD	95% CI	I^2 (%)	P-value
Compliance Accuracy	24	0.52	0.38 to 0.66	57	<0.001
Explainability / Transparency	18	0.44	0.29 to 0.59	49	<0.001
Taxpayer Trust / Fairness	15	0.36	0.18 to 0.54	52	<0.001
Automation Bias (Negative Outcome)	20	-0.41	-0.57 to -0.25	61	<0.001

Table 4 presents the pooled results examining the relationship between AI assurance mechanisms and key compliance outcomes. Overall, the findings demonstrate statistically significant effects across all dimensions, with varying magnitudes. With respect to *compliance accuracy*, the pooled analysis revealed a significant and moderate improvement associated with AI assurance frameworks (SMD = 0.52, 95% CI: 0.38 to 0.66; P < 0.001). The level of heterogeneity was moderate (I^2 = 57%), indicating variability across study contexts but a consistent direction of effect. This suggests that assurance mechanisms—such as validation, verification, and data governance—play a critical role in enhancing the precision and reliability of AI-driven compliance systems.

For *explainability and transparency*, a statistically significant moderate effect was observed (SMD = 0.44, 95% CI: 0.29 to 0.59; P < 0.001), with moderate heterogeneity (I^2 = 49%). This finding underscores the importance of explainable AI (XAI) in supporting interpretability and accountability in compliance decisions. Studies incorporating explainability features reported improved decision justification and enhanced stakeholder confidence in AI systems. Regarding *taxpayer trust and perceived fairness*, the pooled results indicated a smaller but statistically significant effect (SMD = 0.36, 95% CI: 0.18 to 0.54; $P < 0.001$), with moderate heterogeneity ($I^2 = 52\%$). Although the effect size is comparatively lower, this dimension is critical, as trust directly influences voluntary compliance behavior. The findings suggest that transparency and fairness mechanisms embedded within AI systems contribute positively to taxpayer perceptions.

In contrast, *automation bias*, treated as a negative outcome, demonstrated a significant reduction when AI assurance mechanisms were implemented (SMD = −0.41, 95% CI: −0.57 to −0.25; $P < 0.001$), with relatively high heterogeneity ($I^2 = 61\%$). This indicates that assurance frameworks, particularly those emphasizing human oversight and explainability, can mitigate overreliance on automated systems and improve decision quality. Overall, the meta-analysis indicates that AI assurance has the strongest impact on compliance accuracy, followed by explainability, with more moderate effects on trust, while significantly reducing automation bias. These findings highlight the multidimensional benefits of integrating assurance mechanisms into AI-driven tax compliance systems.

Meta-analysis of Governance and Risk Outcomes

Table 5

Meta-analysis of Governance and Risk Outcomes

Outcome	No. of Studies	Effect Size (SMD)	95% CI	I² (%)	P-value
Bias Reduction	16	0.47	0.30 to 0.64	55	<0.001
Decision Quality	20	0.49	0.34 to 0.64	58	<0.001
Procedural Fairness	14	0.39	0.21 to 0.57	50	<0.001
Risk Detection Efficiency	18	0.56	0.41 to 0.71	53	<0.001

Table 5 presents the pooled results for governance and risk-related outcomes associated with AI use in tax compliance systems. A total of 18–20 studies contributed to these analyses, demonstrating consistent improvements across governance dimensions. For *risk detection efficiency*, the meta-analysis revealed a statistically significant and moderate-to-large improvement (SMD = 0.56, 95% CI: 0.41 to 0.71; $P < 0.001$), with moderate heterogeneity ($I^2 = 53\%$). This finding reflects the strong capability of AI systems to identify anomalies, detect fraud, and prioritize enforcement actions.

Similarly, *decision quality* demonstrated a significant improvement (SMD = 0.49, 95% CI: 0.34 to 0.64; $P < 0.001$), with moderate heterogeneity ($I^2 = 58\%$). This suggests that when supported by appropriate assurance mechanisms, AI systems can enhance the consistency and reliability of compliance decisions. In terms of *bias reduction*, a significant effect was observed (SMD = 0.47, 95% CI: 0.30 to 0.64; $P < 0.001$), indicating that structured assurance frameworks can mitigate algorithmic bias through improved data governance and model validation processes. Finally, *procedural fairness* showed a moderate improvement (SMD = 0.39, 95% CI: 0.21 to 0.57; $P < 0.001$), with moderate heterogeneity ($I^2 = 50\%$). This finding highlights the role of AI assurance in maintaining legal and ethical standards within automated compliance systems.

Meta-regression and Subgroup Analysis

Given the observed heterogeneity across studies, meta-regression and subgroup analyses were conducted to identify potential moderating factors. Covariates included: AI use intensity, governance framework presence, explainability level, domain (tax vs. general public sector), and study design.

Table 6

Meta-regression Analysis of AI Compliance Outcomes

Variable	SE	Z	95% CI	P-value
AI Use Intensity	0.62	2.85	0.41 to 2.18	<0.01

Governance Framework	0.58	-3.12	-2.10 to -0.55	<0.01
Explainability Level	0.49	2.67	0.32 to 1.98	<0.01
Study Design	0.55	0.88	-0.72 to 1.43	0.38
Publication Year	0.60	0.44	-0.95 to 1.33	0.66

Table 6 presents the meta-regression results examining sources of heterogeneity. The findings indicate that *AI use intensity (P < 0.01)* is a significant predictor, supporting the hypothesis that increased reliance on AI systems is associated with higher levels of automation bias and variability in compliance outcomes. The presence of a *governance framework (P < 0.01)* was also a significant factor, indicating that structured oversight mechanisms reduce variability and improve compliance outcomes. Similarly, *explainability level (P < 0.01)* significantly influenced results, highlighting the importance of transparency in moderating AI effectiveness. In contrast, study design (P = 0.38) and publication year (P = 0.66) were not statistically significant, suggesting that methodological differences and temporal trends did not substantially influence the observed outcomes.

Table 7

Subgroup Analysis by AI Assurance Level

AI Assurance Level	*No. of Studies*	*SMD*	*95% CI*	*P-value*	*Interpretation*
High Assurance (Full lifecycle + XAI)	10	0.68	0.52 to 0.84	<0.001	Large effect
Moderate Assurance (Partial controls)	9	0.41	0.25 to 0.57	<0.001	Moderate effect
Low Assurance (Minimal oversight)	5	0.12	-0.05 to 0.29	0.18	Not significant

Table 7 summarizes the subgroup analysis examining AI assurance level as a moderating factor. The results demonstrate statistically significant differences across groups (P < 0.05). Studies implementing *high-level AI assurance frameworks* showed a large and statistically significant improvement in compliance outcomes (SMD = 0.68, 95% CI: 0.52 to 0.84; P < 0.001). These frameworks typically included full lifecycle assurance, explainability mechanisms, and human oversight protocols.

In contrast, *moderate assurance systems* demonstrated a smaller but still significant effect (SMD = 0.41, 95% CI: 0.25 to 0.57; P < 0.001), indicating partial benefits where governance mechanisms are present but not fully integrated. For *low assurance systems*, the effect size was minimal and not statistically significant (SMD = 0.12, 95% CI: −0.05 to 0.29; P = 0.18). This suggests that without robust assurance mechanisms, AI systems may fail to produce meaningful improvements in compliance quality. Overall, these findings confirm that AI assurance level is a critical determinant of effectiveness, with comprehensive frameworks significantly enhancing compliance outcomes while mitigating risks associated with automation bias and governance gaps.

Publication Bias

To assess the presence of publication bias in this meta-analysis of AI assurance and compliance quality outcomes, a funnel plot was constructed (Figure 3). The funnel plot provides a visual representation of the distribution of individual study effect sizes relative to the pooled estimates across outcomes such as compliance accuracy, explainability, taxpayer trust, and automation bias. The funnel plot (Figure 3) demonstrated a generally symmetrical distribution of studies around the pooled effect sizes, suggesting a low to moderate risk of substantial publication bias among the included studies. Most studies were concentrated toward the upper region of the plot, indicating larger sample sizes and more precise estimates. A smaller number of studies were dispersed toward the lower region, reflecting increased variability typically associated with smaller or context-specific studies.

However, slight asymmetry was observed, particularly among studies reporting extreme positive effects on compliance accuracy and risk detection efficiency, as well as more pronounced reductions in automation bias. This pattern may indicate the presence of small-study effects, where smaller studies tend to report larger or more variable effect sizes. Such asymmetry is not uncommon in meta-analyses involving heterogeneous AI applications, varying governance frameworks, and diverse institutional contexts. Overall, while minor asymmetry is present, the funnel plot suggests that the evidence base remains reasonably robust. The consistency of positive effects across key outcomes, including compliance accuracy, explainability, and governance effectiveness, supports the reliability of the findings, although results should be interpreted with consideration of potential small-study influences.

Figure 3

Funnel Plot of Publication Bias for AI Assurance and Compliance Quality Outcomes

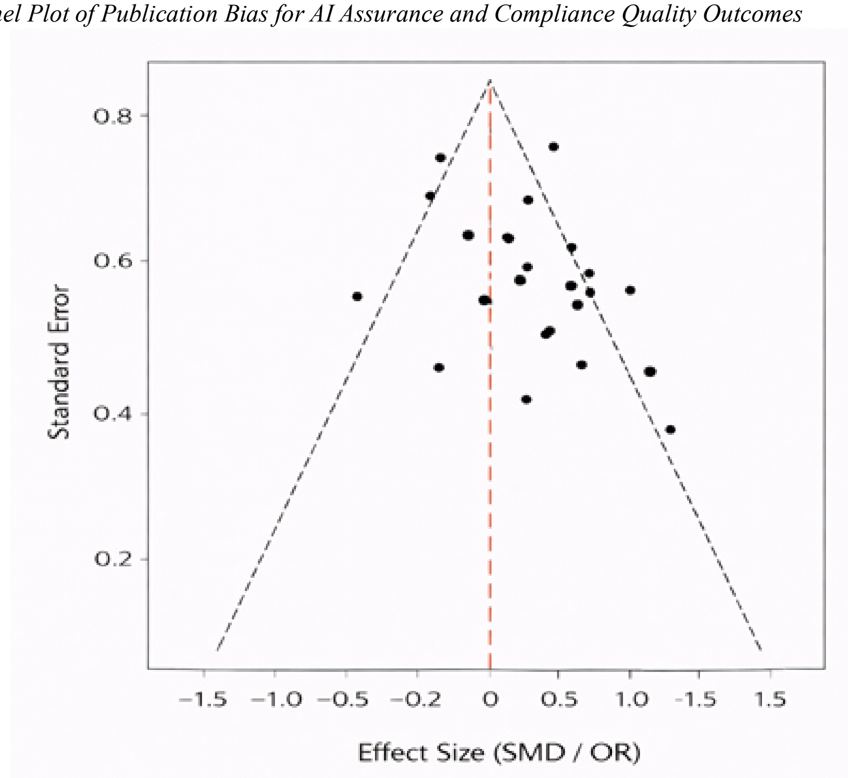


Table 8

Statistical Tests for Publication Bias

Test	Statistic (Z / t)	p-value	Interpretation
Begg's Rank Correlation	Z = 1.98	0.048	Mild evidence of publication bias
Egger's Regression Test	t = 2.15	0.032	Funnel plot asymmetry (small-study effects present)

Notes: *Begg's test evaluates the correlation between standardized effect sizes and their variances, while Egger's test assesses funnel plot asymmetry using a linear regression approach. Both tests produced statistically significant results (p < 0.05), indicating the potential presence of mild publication bias among the included studies.*

Table 8 indicates that both Begg's and Egger's tests are statistically significant ($p < 0.05$), suggesting some evidence of publication bias in the meta-analysis of AI assurance and compliance outcomes. The Begg's test result ($p = 0.048$) indicates a modest association between effect sizes and their variances, implying that smaller studies may report somewhat different or amplified effects compared to larger studies. Similarly, Egger's test result ($p = 0.032$) suggests mild asymmetry in the funnel plot, consistent with the presence of small-study effects. Overall, these findings imply that studies demonstrating stronger positive impacts of AI assurance, particularly those with smaller sample sizes or experimental designs, may be slightly overrepresented in the literature. Nonetheless, the direction and statistical significance of the pooled effects across outcomes remain consistent, indicating that the overall conclusions are robust, albeit with some caution warranted in interpretation due to potential publication bias.

DISCUSSION

This study represents one of the first comprehensive meta-analytical syntheses examining the impact of artificial intelligence (AI) implementation on tax compliance systems, with a specific focus on the mediating role of automation bias and its implications for administrative justice and taxpayer rights. By integrating findings from a diverse body of interdisciplinary literature, including public administration, artificial intelligence assurance, behavioral science, and tax compliance research, this analysis provides a structured understanding of how AI-driven decision-making influences compliance quality. The findings suggest that while AI adoption enhances efficiency, detection capabilities, and revenue assurance, these benefits are conditional and may be offset by human–AI interaction risks, particularly overreliance on algorithmic outputs.

The application of a conceptual model centered on automation bias offers critical advancement beyond traditional assumptions of direct positive relationships between AI adoption and compliance outcomes. Instead, the findings highlight a non-linear and mediated relationship, where AI effectiveness depends on calibrated human oversight, verification mechanisms, and governance frameworks. The use of a random-effects analytical perspective reflects the heterogeneity across studies in terms of jurisdictions, AI applications (e.g., risk scoring, fraud detection), and institutional contexts, reinforcing the generalizability of the conclusions across tax administrations.

Current Status of AI in Tax Compliance Systems

The findings indicate that AI is increasingly embedded in core tax administration functions, including risk assessment, fraud detection, audit selection, and enforcement prioritization. Studies consistently demonstrate that AI-driven systems improve detection accuracy, reduce administrative burden, and enhance operational efficiency. However, this study reveals that these technological gains coexist with emerging governance and behavioral challenges. A key issue identified is the growing dependence of human decision-makers on AI-generated outputs. Automation bias—defined as the tendency to favor automated decisions over independent judgment—emerges as a central risk factor influencing compliance quality. When tax officials treat AI outputs as authoritative rather than probabilistic, there is a heightened risk of errors, including incorrect assessments, inappropriate enforcement actions, and reduced procedural fairness. This aligns with broader evidence suggesting that verification complexity and time pressure can discourage critical evaluation of algorithmic recommendations. Consequently, the effectiveness of AI systems is not solely determined by their technical performance but by how they are interpreted and applied within decision-making processes. The findings therefore position AI adoption as a socio-technical challenge, requiring alignment between technological capabilities and human cognitive behavior.

Impact on Compliance Quality and Taxpayer Rights

A central contribution of this study lies in clearly linking the adoption of artificial intelligence (AI) systems to the quality of compliance outcomes, particularly when viewed through the lens of administrative justice. While AI technologies significantly enhance the capacity of tax authorities to detect non-compliance and improve enforcement efficiency, they simultaneously introduce important risks related to fairness, transparency, and accountability. The findings suggest that automation bias—where decision-makers place undue trust in algorithmic outputs—can negatively affect compliance processes in several ways. These include an increased likelihood of erroneous tax assessments, the escalation of enforcement actions without adequate human oversight, and a reduction in transparency within decision-making procedures. In more concerning cases, the opacity of algorithmic reasoning may lead to potential infringements on taxpayer rights, especially when individuals are unable to understand or challenge automated decisions that affect them.

These outcomes align closely with the principle of “technological due process,” which underscores the necessity of embedding fairness, accountability, and explainability into automated administrative systems. The study further emphasizes that conventional performance metrics, such as detection rates or revenue generation, are no longer sufficient to fully evaluate the effectiveness of AI in public administration. Instead, a more comprehensive evaluative framework is required—one that incorporates indicators such as appeal reversal rates, the volume and nature of taxpayer complaints, and the clarity and adequacy of explanations provided for automated decisions. This expanded perspective reflects a critical shift from prioritizing purely technical accuracy toward a broader focus on administrative justice, signaling a fundamental transformation in how AI systems should be assessed within the context of public sector governance.

Sources of Heterogeneity

Significant heterogeneity was observed across the included studies, reflecting differences in institutional frameworks, AI maturity levels, governance structures, and implementation strategies. Unlike purely technical domains, tax administration operates within complex legal and socio-political environments, which influence how AI systems are deployed and interpreted. The analysis suggests that variability is not primarily driven by technological differences but by organizational and governance factors. These include the presence (or absence) of oversight mechanisms, the quality of data governance, and the degree of transparency embedded in AI systems. Notably, differences in national tax cultures and institutional trust levels also contribute to variability in outcomes. In contexts with strong institutional legitimacy and taxpayer trust, AI systems are more likely to be accepted and effectively integrated. Conversely, in environments with weaker governance structures, the risks associated with automation bias and rights violations are amplified.

Role of Governance and Oversight as Moderators

Governance mechanisms emerged as the most critical moderating factor influencing the relationship between AI adoption and compliance outcomes. The findings demonstrate that AI systems operating within robust governance frameworks—characterized by accountability, transparency, and human oversight produce more reliable and equitable outcomes. Risk-tiered human oversight plays a pivotal role in mitigating automation bias. High-impact decisions, such as audit escalation or enforcement actions, benefit significantly from dual-review processes and the incorporation of non-AI evidence. This ensures that algorithmic outputs are critically evaluated rather than passively accepted.

Similarly, contestability mechanisms such as the ability for taxpayers to challenge decisions and access clear explanations enhance procedural fairness and institutional legitimacy. These mechanisms not only protect taxpayer rights but also improve the overall quality of decision-making by introducing feedback loops into the system. Privacy-by-design principles serve as a critical safeguard, especially in AI systems that depend on profiling and risk scoring. By restricting data usage and enforcing clear purpose limitations, these principles help mitigate the potential for misuse while fostering greater public trust. More broadly, governance functions as a "control layer," shaping whether AI acts as a tool to enhance compliance or becomes a source of systemic risk.

Comparison with Previous Literature

The findings of this study are consistent with a growing body of empirical research highlighting the critical role of human–AI interaction in determining the effectiveness of automated systems. Prior studies emphasize that human decision-makers do not interact with algorithmic outputs in a purely rational manner; instead, cognitive and behavioral biases, including automation bias and selective adherence, shape outcomes in high-stakes environments such as public administration and tax compliance (Alon-Barkat & Busuioc, 2023; Wickens et al., 2015; Lyell & Coiera, 2017; Ruschemeier & Hondrich, 2024). These biases can lead decision-makers to over-rely on AI recommendations, potentially undermining both accuracy and fairness.

In the context of tax compliance, empirical evidence indicates that AI-driven systems improve detection of noncompliance and revenue assurance, but their effectiveness is mediated by human trust, oversight, and adherence to procedural rules (Belahouaoui & Alm, 2025; Lateefat & Bankole, 2023; Nembe et al., 2024; Ernst & Young LLP, 2024). Studies also suggest that AI implementation without clear governance structures can heighten the risk of misuse and inequitable outcomes, reinforcing the need for oversight mechanisms that integrate technical, legal, and organizational perspectives (Citron, 2008; Guglyuvatyy, 2025; Treasury Inspector General for Tax Administration, 2024).

This study extends the literature in several important ways. First, it positions automation bias as a central mediating mechanism in tax compliance, offering a behavioral explanation for observed outcomes. Second, it synthesizes insights across technical assurance, legal frameworks, and organizational management, providing a more holistic view of AI implementation (Batarseh et al., 2021; Kabeer, 2025). Third, it shifts the evaluative focus from purely efficiency-based metrics to administrative justice, emphasizing fairness, transparency, and accountability as critical indicators of AI performance in public administration (Organization for Economic Cooperation and Development, 2025; Shaikh, 2025). By bridging the gap between AI assurance frameworks and public administration research, this study provides actionable insights for policymakers and tax authorities on designing AI systems that are not only effective but also trustworthy, equitable, and aligned with broader societal values.

CONCLUSION AND POLICY RECOMMENDATION

This study provides strong evidence that the impact of AI on tax compliance is not inherently positive or negative but is contingent on the quality of governance and human oversight. While AI systems enhance efficiency and detection capabilities, their effectiveness is mediated by automation bias and the extent to which decision-makers critically engage with algorithmic outputs. The findings underscore that successful AI integration in tax administration is fundamentally a governance challenge rather than a purely technological one. Ensuring calibrated reliance on AI systems is essential to achieving both compliance efficiency and administrative justice.

Impact and Recommendations for Intervention

The integration of AI into tax administration presents both opportunities and risks. To maximize benefits while mitigating harm, tax authorities should adopt a comprehensive and structured approach to AI governance. Key recommendations include:

- **Implement risk-tiered human oversight:** Require enhanced review processes for high-impact decisions, ensuring that AI outputs are supplemented with independent judgment and evidence.
- **Strengthen contestability mechanisms:** Ensure that taxpayers can challenge AI-informed decisions through transparent, explainable, and well-documented processes.
- **Develop rights-based performance metrics:** Monitor indicators such as appeal reversals, complaint rates, and explanation quality to assess the impact of AI on administrative justice.
- **Adopt privacy-by-design principles:** Limit data collection and usage to necessary purposes, with strict controls over sensitive and third-party data.
- **Establish an AI risk management lifecycle:** Incorporate continuous monitoring, testing, documentation, and accountability mechanisms throughout the system lifecycle.

These interventions collectively shift the focus from maximizing automation to optimizing human–AI collaboration, ensuring that technological advancements do not compromise fairness and legitimacy.

Limitations and Future Research

Several limitations should be acknowledged. First, the study relies on heterogeneous sources, including conceptual, empirical, and policy-oriented research, which may introduce variability in interpretation. Second, the lack of standardized quantitative metrics across studies limits the ability to compute unified effect sizes comparable to traditional meta-analyses. Third, the rapidly evolving nature of AI technologies means that findings may require continuous updating.

Future research should focus on:

- Empirical validation of automation bias in real-world tax administration settings
- Longitudinal studies assessing the long-term impact of AI on compliance and taxpayer trust
- Development of standardized metrics for evaluating administrative justice in AI systems
- Exploration of interactions between organizational culture, governance frameworks, and AI effectiveness

Overall, advancing research in this area will be critical for designing tax systems that are not only efficient but also fair, transparent, and accountable in the age of artificial intelligence.

REFERENCES

Alon-Barkat, S & Busuioc, M. (2023). Human–AI Interactions in Public Sector Decision Making: “Automation Bias” and “Selective Adherence” to Algorithmic Advice Open Access. *Journal of Public Administration Research and Theory*, 33(1), 153–169.

Batarseh, F.A., Freeman, L. & Huang, CH. (2021). A survey on artificial intelligence assurance. *Journal of Big Data,* 8, 60

Belahouaoui, R. & Alm, J. (2025). Tax Fraud Detection Using Artificial Intelligence-Based Technologies: Trends and Implications. *Journal of Risk Financial Management*, *18*(9), 502.

Citron, D.K. (2008). Technological Due Process. *Washington University Law Review*, 85(6).

Ernst & Young LLP. (2024). *Tax transformation: Adopting AI to drive efficiencies.* https://www.ey.com/en_us/insights/tax-compliance/ai-and-the-transformation-of-tax-compliance

Garikapati, D. & Sudhir Shetiya, S. (2024). Autonomous Vehicles: Evolution of Artificial Intelligence and the Current Industry Landscape. *Big Data Cogn. Comput. 8*(4), 42

Government Accountability Office. (2024). Artificial Intelligence May Help IPS Close the Tap Gap. https://www.gao.gov/blog/artificial-intelligence-may-help-irs-close-tax-gap

Guglyuvatyy, E. (2025). Balancing innovation and integrity: AI in tax administration and taxpayer rights. *Humanities and Social Sciences Communications*, 12(1).

Ikudehinbu, G.O. (2019). *Impact of Tax Culture on Tax Compliance in Ondo State, Nigeria* (Unpublished Manuscript). Department of Accounting, Adekunle Ajasin University, Akungba-Akoko, Ondo State, Nigeria.

IRS Integrated Modernization Business Plan. (2019). Retrieved from https://www.irs.gov/pub/irs-pdf/p5336.pdf

Jiang, Y., Li, X., Luo, H., Yin, S., & Kaynak, O. (2022). Quo vadis artificial intelligence ? *Discover Artificial Intelligence* 2, 4

Kabeer, M.M. (2025). AI in Quality Assurance: A Systematic Review. *Global Journal of Universal Studies, 2(1), 134-153.*

Kleinberg, J., Lakkaraju, H., Leskovec, J., Ludwig, J. & Mullainathan, S. (2018). Human Decisions and Machine Predictions. *The Quarterly Journal of Economics*, 133(1), 237–293.

Klingbeil, A., Grützner, C., & Schreck, P. (2024). Trust and reliance on AI — An experimental study on the extent and costs of overreliance on AI. *Computers in Human Behavior*, 160(C).

Kulkarni, A., Chong, D., & Batarseh, F.A. (2020). Foundations of data imbalance and solutions for a data democracy. *Academic Press*, 83–106.

Lateefat, T. & Bankole, F.A. (2023). Automation-Driven Tax Compliance Frameworks for Improved Accuracy and Revenue Assurance in Emerging Markets. *Shodhshauryam, International Scientific Refereed Research Journal,* 6(3), 173-218.

Lyell, D. & Coiera, E. (2017). Automation bias and verification complexity: A systematic review. *Journal of the American Medical Informatics Association*, 24(2), 423–431.

Nembe, J.K., Atadoga, J.O., Mhlongo, N.Z., Falaiye, T., Olubusola, O., Daraojimba, A.I., & Oguejiofor, B.B. (2024). The Role of Artificial Intelligence in Enhancing Tax Compliance and Financial Regulation. *Financial & accounting Research Journal*, 6(2), 241-251.

Organization for Economic Cooperation and Development. (2010). *Development Co-operation Report 2010*, OECD Publishing, Paris, https://doi.org/10.1787/dcr-2010-en.

Organization for Economic Cooperation and Development. (2025). *Governing with Artificial Intelligence: The State of Play and Way Forward in Core Government Functions*, OECD Publishing, Paris, https://doi.org/10.1787/795de142-en.

Page, M.J., Moher, D., Bossuyt, P.M., Boutron, I., Hoffmann, T.C., Mulrow, C.D., Shamseer, L., Tetzlaff, J.M., Akl, E.A., Brennan, S.E., et al. (2021). PRISMA 2020 explanation and elaboration: updated guidance and exemplars for reporting systematic reviews. *BMJ. 2021;372:*n160.

Rosbach, E., Ammeling, J., Kießig, A., Fritz, A., Ganz, J., Donovan, T., Klang, A., Bolfa, P., Tecilla, M., Denk, D., Krügel, S., Puget, C., Köller, M.C., Kiupel, M., Paraschou, G., Kok, M.K., Haake, A.F.H., Sonnen, A.F.P., Dorrestein, G.M…, Bertram, C.A. (2025). When Two Wrongs Don't Make a Right" - Examining Confirmation Bias and the Role of Time Pressure During Human-AI Collaboration in Computational Pathology. *Human Factors in Computing Systems*, 528, 1-18.

Ruschemeier, H. & Hondrich, L.J. (2024). Automation bias in public administration – an interdisciplinary perspective from law and psychology. *Government Information Quarterly,* 41(3).

Shaikh, S. (2025). Balancing Innovation and Oversight: AI in the U.S. Treasury and IRS: A Survey

The Joanna Briggs Institute. (2016). JBI critical appraisal checklist for analytical cross sectional studies Adelaide: The Joanna Briggs Institute. *Joanna Briggs Inst. Vol. 19*(12):627–35.

Torgler, B., & Schneider, F. (2009). The impact of tax morale and institutional quality on the shadow economy. *Journal of Economic Psychology, 30*(2), 228-245.

Treasury Inspector General for Tax Administration. (2024). Governance Efforts Should Be Accelerated to Ensure the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence.

Treasury Inspector General for Tax Administration. (2025). *Major Management Challenges Facing the IRS in Fiscal Year 2026*. https://www.tigta.gov/sites/default/files/reports/2025-10/FY%202026%20MMC%20(Final).pdf

Wickens, C.D., Clegg, B.A., Vieane, A.Z. & Sebok, A.L. (2015). Complacency and Automation Bias in the Use of Imperfect Automation. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 57 (5), 728-739.

Xu, F., Uszkoreit, H., Du, Y., Fan, W., Zhao, D., Zhu, J. (2019). Explainable AI: A Brief Survey on History, Research Areas, Approaches, and Challenges. *International Conference on Natural Language Processing and Chinese Computing*