

Fiscal Digitalization and Tax Audit Performance: A Systematic Review of Technologies, Outcomes and Institutional Conditions

Soukaina BENMESSAOUD, (Ph.D. Candidate)

*Laboratory of Research in Innovation, Responsibility and Sustainable Development
 (INREDD)*

*Faculty of Legal, Economic and Social Sciences, Cadi Ayyad University
 Marrakech, Morocco*

Khalil MOKHLIS, (Full Professor)

*Laboratory of Research in Innovation, Responsibility and Sustainable Development
 (INREDD)*

*Faculty of Legal, Economic and Social Sciences, Cadi Ayyad University
 Marrakech, Morocco*

Correspondence address :	laboratory of Research in Innovation, Responsibility and Sustainable Development (INREDD) Faculty of Legal, Economic and Social Sciences, Cadi Ayyad University Marrakech, Morocco
Disclosure Statement :	The authors declare that they have not received any financial support that could have influenced the objectivity of this study. They take full responsibility for any potential plagiarism, the use of artificial intelligence in the writing process, as well as for the results presented in this article.
Conflict of Interest :	The authors report no conflicts of interest.
Cite this article:	BENMESSAOUD, S., & MOKHLIS, K. (2026). Fiscal Digitalization and Tax Audit Performance: A Systematic Review of Technologies, Outcomes and Institutional Conditions. <i>International Journal of Accounting, Finance, Auditing, Management and Economics</i> , 7(10), 133–151. https://doi.org/10.5281/zenodo.21384499
License	This is an open access article under the CC BY-NC-ND license

Received: 06/06/2026

Accepted: 18/08/2026

International Journal of Accounting, Finance, Auditing, Management and Economics - IJAFAME

ISSN: 2658-8455

Volume 7, Issue 10 (2026)

Fiscal Digitalization and Tax Audit Performance: A Systematic Review of Technologies, Outcomes and Institutional Conditions

Abstract:

Fiscal digitalization is changing tax auditing through electronic filing and invoicing, third-party data integration, and analytical tools that support risk-based case selection. Existing research remains fragmented across technology adoption studies, compliance evaluations, and fraud-detection models. This fragmentation makes it difficult to determine which digital tools improve which dimensions of audit performance and under what institutional conditions. This article addresses that gap through a systematic literature review structured in accordance with PRISMA 2020. The review covers 2010-2025 and uses Scopus, Web of Science, ScienceDirect, and supplementary Google Scholar searches. Of the 250 records identified, 35 studies met the eligibility criteria and were subjected to standardized data extraction, quality appraisal, and comparative thematic synthesis. The most recurrent findings concern shorter processing times, stronger risk-based targeting, improved transaction traceability, and lower compliance or administrative costs. Positive effects on audit yield, revenue, and voluntary compliance are also reported, but the evidence is more context dependent and does not justify unconditional generalization. Performance gains rely on data quality, interoperability, auditors' digital skills, the legal validity of electronic evidence, cybersecurity safeguards, and accountable algorithmic governance. The article makes three contributions. Theoretically, it develops an integrated framework connecting digital technologies, operational mechanisms, institutional conditions, and multidimensional audit performance. Methodologically, it introduces a transparent coding and appraisal structure that separates established evidence, convergent tendencies, and prospective claims. Managerially, it identifies an implementation sequence in which data governance, human capability, legal authorization, and security controls precede large-scale predictive deployment. The review therefore reframes fiscal digitalization not as a self-executing technological solution, but as a socio-technical transformation whose value depends on organizational and regulatory readiness.

Keywords: Fiscal digitalization, tax audit, tax administration, artificial intelligence, systematic review.

JEL Classification: H20, H26, M42, O33.

Paper type: Theoretical research, systematic literature review.

1. Introduction

Tax administrations are becoming data-intensive public organizations. Electronic filing, electronic invoicing, digital payment records, and third-party reporting have enlarged the volume, frequency, and granularity of information available for compliance management. At the same time, advances in data mining, machine learning, cloud infrastructure, and application programming interfaces have made it technically possible to compare transactions in near real time and to rank taxpayers according to risk. These developments have shifted fiscal digitalization from a service-delivery agenda toward a central component of audit strategy and state capacity (Kochanova et al., 2020; OECD, 2023).

The operational promise is substantial. Transaction digitization can reduce transcription errors and accelerate basic verification. Cross-matching can reveal inconsistencies between declarations, invoices, customs records, banking information, and other authorized sources. Predictive systems can prioritize cases that display patterns associated with non-compliance, while prefilled returns can reduce inadvertent errors and compliance costs. Empirical studies consequently report improvements in filing efficiency, reporting behavior, and case selection in settings as diverse as Peru, Tajikistan, Ethiopia, Taiwan, and several European tax systems (Bellon et al., 2022; Mascagni et al., 2021; Okunogbe & Pouliquen, 2022; Vanhoeyveld et al., 2020).

These gains are neither automatic nor uniform. A digital system may merely reproduce a fragmented process in electronic form. Incomplete identifiers, incompatible databases, weak connectivity, and poor data governance can increase false positives or exclude taxpayers whose activities remain partially informal. Likewise, an accurate predictive score does not by itself establish a legally sustainable audit finding. The result must remain explainable, contestable, and supported by admissible evidence. Digitalization therefore changes the performance frontier of tax auditing, but it also changes the conditions under which performance should be assessed.

The literature reflects this complexity, yet it is organized in separate streams. A first stream focuses on electronic services, including e-filing, e-invoicing, and prefilled returns, and generally examines adoption, compliance costs, or reporting behavior (Hesami et al., 2024; Qi & Che Azmi, 2021). A second stream studies data mining and machine-learning models for fraud detection and audit case selection, often emphasizing predictive accuracy (Baghdasaryan et al., 2022; Rahimikia et al., 2017; Savić et al., 2022). A third stream examines institutional capacity, corruption exposure, and government performance at an aggregate level (Kochanova et al., 2020; Okunogbe & Pouliquen, 2022; Uyar et al., 2021). Each stream captures an important part of the transformation, but their outcome variables, units of analysis, and standards of evidence differ.

This fragmentation creates three unresolved tensions. First, technical accuracy is frequently treated as equivalent to audit effectiveness, although a model can classify risk accurately without reducing audit cycle time or producing enforceable assessments. Second, studies of e-services often infer administrative performance from adoption or satisfaction measures, even when audit outcomes are not observed directly. Third, strong results from data-rich jurisdictions are sometimes generalized to administrations with different legal mandates, digital infrastructure, and taxpayer coverage. A systematic synthesis is therefore needed to connect technologies to mechanisms, mechanisms to performance dimensions, and performance effects to institutional conditions.

The scientific gap addressed in this article is the absence of a consolidated and methodologically explicit account of how fiscal digitalization affects the operational performance of tax auditing across multiple technologies and institutional contexts. Prior reviews have clarified the role of e-invoicing and prefilled returns, and conceptual work has mapped digital tax administration more broadly (Bassey et al., 2022; Hesami et al., 2024). However, the evidence on audit speed,

targeting quality, yield, compliance, traceability, and governance risks has not been integrated within a single analytical structure that distinguishes direct evidence from contextual inference. The main research question is therefore: How, and under what institutional conditions, does fiscal digitalization affect tax audit performance? Three sub-questions organize the analysis. The first asks which digital technologies and operational mechanisms are most frequently associated with audit improvement. The second asks which dimensions of performance are supported by the strongest and most recurrent evidence. The third asks which organizational, legal, ethical, and security conditions moderate these effects or generate new risks.

The review has four objectives. It maps the technologies examined in the literature; consolidates the reported effects across a multidimensional performance framework; identifies contextual moderators and implementation risks; and evaluates the strength and transferability of the evidence. The study adopts a PRISMA 2020 reporting structure and a qualitative systematic synthesis rather than a meta-analysis because the included studies use heterogeneous designs, measures, and units of observation (Page et al., 2021; Snyder, 2019).

The contribution is deliberately bounded. The article does not claim that digitalization has a single average causal effect across tax administrations. Its theoretical contribution is an integrated technology-mechanism-condition-performance model. Its methodological contribution is a transparent coding and quality-appraisal protocol that enables results from econometric studies, experiments, case studies, and predictive-model evaluations to be compared without treating them as equivalent. Its managerial contribution is a sequencing logic for implementation: reliable digital transactions and identifiers first, interoperable data and skilled staff second, and predictive automation only after legal, ethical, and cybersecurity controls are operational.

The remainder of the article proceeds as follows. Section 2 develops the conceptual background and analytical framework. Section 3 explains the systematic review protocol. Section 4 presents the descriptive and thematic results. Section 5 discusses the answers to the research questions, theoretical implications, managerial implications, limits of transferability, and priorities for future research. Section 6 concludes.

2. Conceptual Background and Analytical Framework

2.1. From transaction digitization to predictive auditing

Fiscal digitalization should be understood as a layered transformation rather than a single technology. The first layer converts taxpayer interactions and transaction records into structured digital data. E-filing, e-invoicing, online payment, digital registration, and prefilled returns belong to this layer. Their immediate contribution is not prediction but data availability, standardization, and timeliness. Without stable identifiers and consistent transaction formats, later analytical layers cannot generate reliable audit signals (Chen et al., 2015; Cuylen et al., 2016).

The second layer integrates information across internal and external sources. Tax declarations can be reconciled with invoices, customs records, payroll information, withholding statements, property data, and other sources authorized by law. Integration changes the audit process because discrepancies can be identified before a conventional field audit begins. It also creates a more complete audit trail. The value of this layer depends on matching accuracy, common data definitions, update frequency, and clear responsibilities for correcting errors.

The third layer applies descriptive, anomaly-detection, network, or supervised-learning techniques to rank cases and identify patterns. Data-mining studies demonstrate that predictive models can outperform undifferentiated or purely rule-based selection when fraud is rare and audit resources are limited (Castellón González & Velásquez, 2013; Rahimikia et al., 2017; Wu et al., 2012). More recent studies use unsupervised or hybrid techniques because labeled fraud

data are incomplete and historically audited cases may be a biased sample of the taxpayer population (Savić et al., 2022; Vanhoeyveld et al., 2020).

The fourth layer concerns workflow automation and decision support. Risk scores, alerts, case-management systems, and automated reconciliation can reduce the time spent on repetitive checks. The auditor's role then shifts from collecting and transcribing information toward evaluating evidence, investigating complex relationships, and deciding whether an anomaly has a credible fiscal explanation. This transition is organizational as much as technical because it changes competencies, accountability, and the distribution of discretion.

Blockchain is often presented as a potential fifth layer because distributed ledgers can make records difficult to alter and can support automated transaction rules. Yet the reviewed evidence is much thinner than for e-invoicing or machine learning. Blockchain should therefore be treated as a prospective infrastructure option rather than an established determinant of audit performance. The distinction between deployed evidence and technological possibility is essential throughout the synthesis.

2.2. A multidimensional conception of audit performance

Audit performance is frequently reduced to the amount assessed or collected, but that measure alone is incomplete. A tax administration can increase short-term yield by concentrating on easily collectible cases while failing to address complex or systemic risks. Conversely, a digital intervention may reduce compliance costs and improve voluntary reporting even if immediate audit yield changes little. This review therefore distinguishes five performance dimensions.

The first dimension is operational efficiency. It includes audit cycle time, the cost of processing and verification, the number of cases handled per auditor, and the proportion of routine tasks automated. E-filing and e-invoicing often affect this dimension directly because information becomes machine-readable and available earlier. The second dimension is targeting effectiveness, defined as the capacity to direct scarce audit resources toward cases with a higher probability or expected value of non-compliance. Predictive analytics and network methods act primarily through this mechanism.

The third dimension is fiscal effectiveness, which includes additional assessments, collections, recovery, and the prevention of revenue leakage. This dimension is important but difficult to compare because studies use different horizons and institutional rules. The fourth dimension is compliance and deterrence. Digital third-party reporting, prepopulation, and visible cross-checking can reduce opportunities for misreporting or increase the perceived probability of detection (Gillitzer & Skov, 2018; Kleven et al., 2011). The fifth dimension is procedural quality, including traceability, data integrity, consistency, transparency, contestability, and the legal sustainability of audit decisions.

These dimensions can move in different directions. A highly automated system may be fast but produce many false positives. A predictive model may improve the hit rate while concentrating attention on historically audited groups. Prefilled returns may simplify filing but induce taxpayers to accept incorrect information by default. Performance assessment must consequently include both productive gains and control safeguards.

2.3. Institutional conditions and risk mechanisms

The relationship between digital technology and audit performance is moderated by institutional capacity. Data quality is the first condition because predictive accuracy is limited by missing, delayed, or inconsistent input. Interoperability is the second because isolated databases cannot support cross-checking. Human capability is the third because auditors must understand model outputs, detect implausible results, document reasoning, and translate technical signals into legally relevant evidence.

Legal and governance conditions are equally important. Tax administration requires a lawful basis for collecting and linking data, a clear mandate for automated selection, retention rules, and procedures that allow taxpayers to challenge factual errors. Algorithmic accountability requires documentation of the system's objective, variables, validation, version history, and human oversight. These requirements do not eliminate discretion; they make discretion traceable and reviewable (Wieringa, 2020).

Cybersecurity and privacy risks grow as information becomes centralized and reusable. A breach can expose highly sensitive financial data, disrupt filing or payment services, and undermine trust. Security therefore belongs to performance governance rather than being treated as a separate technical function. The relevant controls include access management, encryption, monitoring, backup, incident response, supplier oversight, and separation between development, validation, and production environments.

Organizational culture also affects implementation. Staff may resist systems that appear to replace professional judgment, while managers may overstate model certainty because a quantitative score seems objective. Successful transformation requires participation by auditors, tax lawyers, data specialists, and information-security teams. Training must address not only software use but also data interpretation, bias, evidence, and communication with taxpayers.

The analytical framework used in this review therefore links four elements: the digital technology deployed; the operational mechanism it activates; the performance dimension affected; and the institutional conditions that enable or constrain the effect. This structure prevents the analysis from attributing outcomes to technology alone.

Table 1: Technology-mechanism-condition-performance framework

Technology	Primary mechanism	Critical conditions	Expected performance effect
E-filing and e-invoicing	Structured and timely transaction data	Stable identifiers, taxpayer access, standard formats	Efficiency, traceability, compliance
Prefilled returns and third-party reporting	Automatic verification and reduced information burden	Accurate third-party data, correction mechanisms	Compliance, accuracy, cost
Data integration and cross-matching	Detection of inconsistencies across sources	Interoperability, lawful data sharing, provenance	Targeting, traceability, procedural quality
Machine learning and anomaly detection	Risk ranking and pattern recognition	Representative data, validation, human review, monitoring	Targeting, hit rate, resource allocation
Workflow and case-management systems	Automation of routine checks and documentation	Process redesign, staff capability, access control	Cycle time, consistency, accountability
Blockchain and smart contracts	Immutable records and rule execution	Scalable infrastructure, legal recognition, operational evidence	Prospective traceability and reconciliation

Source: Authors' synthesis of the reviewed evidence.

3. Research Methodology

3.1. Review design and reporting standard

The study is a systematic literature review with qualitative evidence synthesis. It is not a bibliometric review because its principal objective is to compare substantive findings rather than citation structures, and it is not a meta-analysis because the outcome definitions and research designs are too heterogeneous to support a pooled effect size. The protocol follows the transparency principles of PRISMA 2020 and the management-research guidance developed by Tranfield et al. (2003), Snyder (2019), and Booth et al. (2016).

The unit of analysis is a scholarly study that examines a digital technology or digital administrative process in relation to tax administration, compliance management, audit selection, fraud detection, or a closely connected performance outcome. The review covers publications from January 2010 through December 2025.

3.2. Information sources and reproducible search strategy

The principal sources were Scopus and Web of Science because they provide multidisciplinary coverage of accounting, public administration, information systems, economics, and computer science. ScienceDirect was searched to increase coverage of tax, information-systems, and applied analytics journals. Google Scholar was used only as a supplementary source to identify recent or cross-disciplinary items and to verify backward and forward citations. The review did not treat search-engine ranking as a quality criterion.

The search combined three concept groups: tax administration; digital technology; and audit or performance outcomes. Database syntax was adapted while preserving the same conceptual structure. Searches were limited to titles, abstracts, and keywords where the platform allowed this option. The exact strings are reported in Table 2 so that the procedure can be reproduced and updated.

Table 2: Reproducible database search strategy

Database	Search equation	Filters and use
Scopus	TITLE-ABS-KEY(("tax administration" OR "tax authority" OR "fiscal administration") AND (digital* OR "electronic filing" OR e-filing OR "electronic invoicing" OR e-invoicing OR "big data" OR "machine learning" OR "artificial intelligence" OR blockchain) AND (audit* OR compliance OR fraud OR "risk selection" OR performance))	2010-2025; journal and reviewed conference literature; English
Web of Science	TS(("tax administration" OR "tax authority" OR "fiscal administration") AND (digital* OR e-filing OR e-invoicing OR "big data" OR "machine learning" OR "artificial intelligence" OR blockchain) AND (audit* OR compliance OR fraud OR performance))	2010-2025; article and proceedings paper; English
ScienceDirect	("tax administration" OR "tax audit") AND (digitalization OR e-invoicing OR e-filing OR "machine learning" OR "data mining") AND (performance OR compliance OR fraud)	2010-2025; research article and review article
Google Scholar	All phrase combinations above, plus backward and forward citation checking of included studies	Supplementary only; first 200 relevance-ranked results reviewed

Source: Authors' protocol, structured from Page et al. (2021), Snyder (2019), and Tranfield et al. (2003).

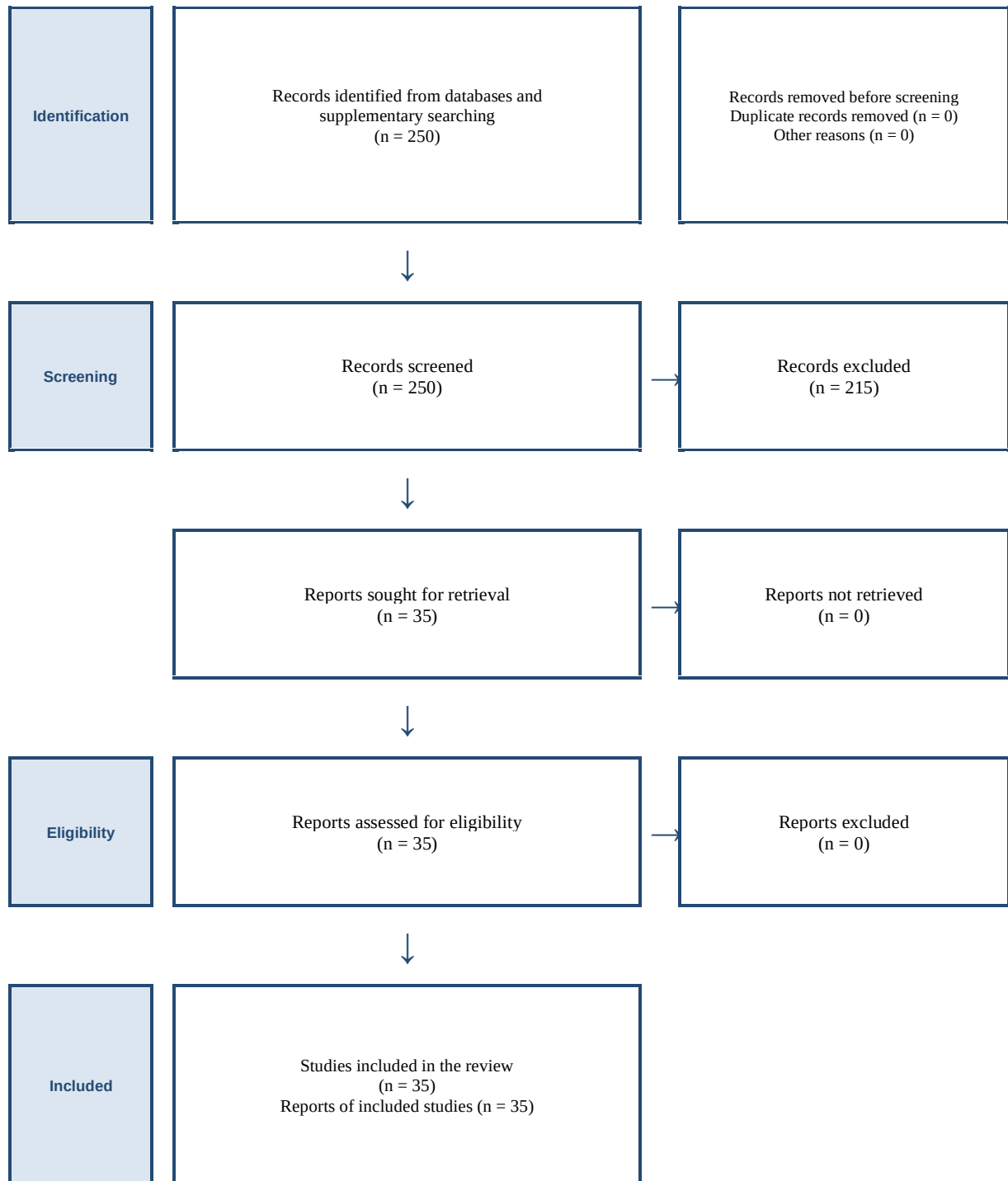
3.3. Eligibility criteria and study selection

Studies were included when they met five criteria: publication between 2010 and 2025; peer-reviewed journal article or academically reviewed conference paper; explicit focus on a public tax administration, taxpayer-facing fiscal system, or tax-compliance process; analysis of at least one digital technology or digital information process; and reporting of an audit, compliance, efficiency, targeting, revenue, traceability, governance, or implementation outcome. Empirical studies could use experiments, surveys, administrative data, econometrics, case studies, or predictive modeling. Systematic reviews were eligible when they directly synthesized e-invoicing, prefilled returns, or digital tax administration and were used to triangulate rather than duplicate primary-study claims.

Items were excluded when they concerned private-sector financial auditing without a tax-administration link, discussed digital taxation only as a tax-base problem, contained no assessable method or evidence, were editorial or promotional pieces, duplicated another record, or could not be retrieved in full text. English-language publications were included. This restriction is reported as a limitation because it may underrepresent national experiences published in French, Spanish, Portuguese, Chinese, or other languages.

The search identified 250 records. After title and abstract screening, 215 records were excluded, and 35 reports were sought and retrieved for full-text assessment. All 35 reports met the eligibility criteria and were included in the review. The dominant reasons for exclusion at the screening stage were the absence of a public tax-administration focus or an audit or performance outcome, a focus restricted to private-sector auditing, and non-scholarly status.

Figure 1: PRISMA 2020 flow diagram for study selection



Source: Authors' elaboration, adapted from Page et al. (2021).

3.4. Quality appraisal

Each included study was assessed through a five-criterion appraisal matrix: clarity of the research question; appropriateness and transparency of the design; adequacy of the data or evidence; coherence between analysis and conclusions; and explicit treatment of limitations or validity. Each criterion was scored as not demonstrated, partially demonstrated, or clearly demonstrated. Studies with strong designs and transparent evidence were rated high; studies with useful but more limited designs were rated moderate. No study was excluded solely because of appraisal because the review also aimed to map technologies and implementation

contexts. However, moderate-quality studies were not used as the sole basis for strong causal claims.

The appraisal produced 27 high-quality and 8 moderate-quality studies. This distribution is reported at study level in Appendix A. The procedure reduces the risk of presenting conceptual statements, predictive accuracy, and causal administrative effects as equivalent forms of evidence.

Table 3: Quality-appraisal criteria and decision rules

Criterion	Operational test	Scoring
Research question	Clear object, context, and outcome	0 = not demonstrated; 1 = partial; 2 = clear
Design transparency	Method, sample or data source, and analytical procedure are described	0 = not demonstrated; 1 = partial; 2 = clear
Evidence adequacy	Data and measures are appropriate for the stated claim	0 = not demonstrated; 1 = partial; 2 = clear
Analytical coherence	Results support the conclusions without material overstatement	0 = not demonstrated; 1 = partial; 2 = clear
Validity and limitations	Biases, uncertainty, or limits are recognized	0 = not demonstrated; 1 = partial; 2 = clear

Source: Authors' appraisal matrix adapted to heterogeneous systematic-review evidence.

3.5. Data extraction and synthesis

A standardized matrix captured bibliographic details, year, jurisdiction, technology, research design, data source, principal outcome, performance dimension, institutional moderators, risks, and quality rating. The first analytical cycle used descriptive coding to map the evidence. The second used mechanism coding to identify how a technology was expected to affect performance, for example through data standardization, cross-matching, prepopulation, risk ranking, anomaly detection, workflow automation, or third-party reporting.

The third cycle grouped findings into operational efficiency, targeting effectiveness, fiscal yield, compliance and deterrence, procedural quality, moderators, and risks. Counts indicate how many included studies addressed or supported a theme, not a pooled effect size. A finding was classified as established evidence when it was supported by multiple high-quality empirical studies using direct outcomes; as a convergent tendency when studies pointed in the same direction but used heterogeneous or indirect measures; and as prospective when the claim mainly reflected conceptual reasoning or early-stage technology.

Comparisons were made across technology type, research design, jurisdictional setting, and quality level. Contradictory or conditional findings were retained rather than averaged away. This approach is appropriate for a heterogeneous body of evidence and supports a more cautious interpretation of transferability.

4. Results

4.1. Descriptive profile of the evidence base

The final corpus contains 35 studies published between 2010 and 2025. Publication activity accelerated after 2020, when digital service expansion and advanced analytics became central to tax-administration reform. The corpus includes 9 studies from 2010-2014, 9 from 2015-2019, and 17 from 2020-2025. This distribution supports the view that the field has moved from early adoption and process-standardization questions toward performance evaluation, machine learning, and governance.

The studies are methodologically diverse. Experiments and quasi-experiments evaluate reporting behavior, compliance, or administrative effects. Surveys and structural models examine adoption and process efficiency. Administrative-data and econometric studies measure

compliance, revenue, or government capacity. Machine-learning studies evaluate classification, anomaly detection, and case selection. Case and design studies explain system architecture, interoperability, or implementation maturity. This diversity broadens the evidence base but limits direct statistical aggregation.

Geographic coverage is also uneven. European and Asian tax systems are prominent, while African and Latin American settings are represented by fewer studies despite the importance of digital leapfrogging and informality. Cross-country studies provide broader coverage but often use aggregate proxies for digital government rather than direct audit measures. The evidence is therefore strongest for transaction digitization, prepopulation, and model-based case selection in administrations with relatively structured data.

Table 4: Descriptive profile of the 35 included studies

Dimension	Category	Number of studies
Publication period	2010-2014	9
Publication period	2015-2019	9
Publication period	2020-2025	17
Quality appraisal	High	27
Quality appraisal	Moderate	8
Technology group	Transaction digitization and prepopulation	15
Technology group	Advanced analytics and machine learning	7
Technology group	Broader digital administration and information systems	13

Source: Authors' coding of the final review corpus.

4.2. Technologies, mechanisms, and maturity levels

Transaction digitization and prepopulation form the largest technology group in the corpus. E-filing, e-invoicing, and prefilled returns convert taxpayer information into standardized records and reduce the need for manual transcription. Their most direct mechanisms are timelier data, automatic validation, cross-checking, and reduced filing burden. The evidence is mature because these tools have been implemented in multiple jurisdictions and evaluated through several designs (Bellon et al., 2022; Chen et al., 2015; Doxey et al., 2021; Hesami et al., 2024). Advanced analytics and machine learning form a smaller but rapidly expanding group. Data-mining classifiers, anomaly-detection systems, hybrid intelligent models, network data, and variational autoencoders are used to prioritize cases or detect suspicious relationships. These studies commonly report better discrimination than simple random or rule-only selection, especially when non-compliance is rare. However, predictive performance depends on labeled data, historical audit selection, feature quality, validation design, and changes in taxpayer behavior (Baghdasaryan et al., 2022; Kleanthous & Chatzis, 2020; Vanhoeyveld et al., 2020). Broader e-government and tax-administration studies examine how digital capability affects compliance costs, service quality, corruption exposure, or institutional capacity. Their contribution is to show that audit performance is embedded in wider administrative systems. Electronic filing may reduce direct contact and discretionary opportunities, but only when users can access the system and when procedures are redesigned rather than merely transferred online (Kochanova et al., 2020; Okunogbe & Pouliquen, 2022).

Blockchain and smart contracts appear frequently in conceptual discussions but are not supported by the same volume of direct tax-audit evidence. The review therefore does not rank blockchain alongside e-invoicing or machine learning as an established performance tool. Its plausible contributions to immutability and automated reconciliation remain research propositions that require evaluation in operational tax environments.

4.3. Effects on multidimensional audit performance

Operational efficiency is the most recurrent outcome. Twenty-eight of the 35 studies report or examine shorter processing, lower compliance burden, improved workflow, or more efficient allocation of administrative effort. The evidence is strongest where digital records replace

repeated manual entry and where prepopulation or electronic invoices allow automatic validation. These effects should be interpreted as a convergent tendency rather than a universal cost reduction because implementation expenditure, maintenance, cybersecurity, and support costs are not consistently measured.

Targeting and detection are addressed in 10 studies. Predictive and anomaly-detection models generally improve the identification of high-risk cases relative to undifferentiated selection or basic rules. The strongest evidence comes from evaluations that use out-of-sample validation, administrative audit outcomes, or direct comparison with existing selection processes (Battaglini et al., 2025; Rahimikia et al., 2017; Savić et al., 2022). Nevertheless, model accuracy is not identical to realized audit yield. Cases must still be investigated, evidence must be documented, and assessments must survive administrative or judicial review.

Compliance and deterrence are addressed in 17 studies. Third-party reporting, prefilled forms, electronic invoicing, and electronic services can reduce unintentional errors, narrow opportunities for underreporting, or change the perceived probability of detection. Field and quasi-experimental evidence from Denmark, Peru, Tajikistan, Ethiopia, and other settings supports positive effects, but behavioral responses differ. Prefilled information can improve accuracy when it is reliable, yet it can also create passive acceptance of errors or strategic adjustment in non-prefilled fields (Doxey et al., 2021; van Dijk et al., 2020).

Fiscal yield is directly coded in only 3 studies, which is much less than the number discussing efficiency or compliance. Positive revenue or assessment effects are reported, but outcome definitions and time horizons vary. This imbalance is important because the claim that digitalization automatically raises revenue is stronger than the direct evidence permits. The most defensible conclusion is that digitalization improves information and targeting conditions that can support yield when collection capacity, legal follow-up, and taxpayer coverage are adequate.

Traceability and procedural quality appear in 9 studies. Electronic invoices, third-party reports, integrated identifiers, and system logs can strengthen the audit trail and reduce some forms of manual error. However, procedural quality also requires correction mechanisms, access controls, documentation, and the ability to explain why a taxpayer was selected. Digital traces are useful evidence only when their provenance and integrity are established.

Table 5: Consolidated effects of fiscal digitalization on audit performance

Performance dimension	Studies	Evidence assessment	Reported contribution	Principal qualification
Operational efficiency	28	Convergent tendency	Shorter processing, reduced manual entry, improved workflow and lower filing burden	Implementation and recurring costs are not consistently measured
Risk targeting and detection	10	Established for model discrimination; conditional for realized audit yield	Higher-risk case ranking, anomaly detection, network-based signals	Historical selection bias, false positives, drift, weak labels
Compliance and deterrence	17	Established in several settings; context dependent	Improved reporting, fewer errors, stronger third-party verification	Behavioral adaptation and passive acceptance of incorrect prefilled data
Fiscal yield	3	Limited direct evidence	Additional assessment, collection, reduced leakage	Different horizons and collection rules; causal attribution difficult
Traceability and procedural quality	9	Convergent tendency	Audit trail, data integrity, reproducibility of checks	Requires provenance, correction rights, legal validity and access controls

Source: Authors' qualitative synthesis. Counts indicate thematic coverage and are not pooled effect sizes.

4.4. Moderators, implementation constraints, and risks

Institutional moderators are explicitly addressed in 13 studies. Data quality and interoperability are the most consistent conditions. Missing identifiers, inconsistent classifications, delayed reporting, and fragmented databases weaken cross-matching and model reliability.

Infrastructure and user access remain important in developing-country settings, where electronic systems may coexist with informal activity and uneven connectivity.

Human capability is a second moderator. Auditors require data literacy, but the objective is not to convert every auditor into a data scientist. They must understand the purpose and limits of a model, identify implausible outputs, request supporting evidence, and document the decision pathway. Data specialists, tax lawyers, cybersecurity teams, and operational auditors therefore need a shared governance process. Training that focuses only on software interfaces is insufficient.

Legal frameworks determine whether digitally generated information can support enforceable action. Relevant issues include the authority to obtain and link data, recognition of electronic records, retention periods, taxpayer access and correction rights, and the degree of permissible automation in case selection. When these rules are unclear, digital systems may accelerate detection but create disputes about evidence, proportionality, or procedural fairness.

Risk-related findings or cautions appear in 9 studies. The principal risks are cybersecurity incidents, privacy violations, false positives, opaque selection, automation bias, and unequal coverage of taxpayers. Historical audit data can reproduce earlier enforcement priorities, while groups with sparse digital records may be incorrectly classified or overlooked. Model monitoring must therefore include error rates, subgroup performance, drift, and the outcomes of human review.

The evidence also reveals a sequencing problem. Administrations sometimes seek advanced analytics before establishing reliable transactional data, common identifiers, and case-management integration. The reviewed studies suggest the reverse sequence: digitize and standardize transactions; establish data governance and legal authority; develop staff capability; validate decision-support models; and only then scale predictive automation.

5. Discussion

5.1. Integrated answer to the research questions

The main research question asked how and under what institutional conditions fiscal digitalization affects tax audit performance. The synthesis shows that digitalization operates through distinct mechanisms rather than a single effect. Transaction digitization improves availability, timeliness, and standardization. Data integration enables cross-checking. Analytics improves prioritization. Workflow systems translate risk signals into operational action. These mechanisms affect different performance dimensions and should not be collapsed into a general claim of efficiency.

The first sub-question concerned the technologies with the strongest utility. The most mature evidence supports e-filing, e-invoicing, prefilled returns, third-party reporting, and data-driven case selection. Transaction technologies provide the data foundation; predictive tools add value only after that foundation is reliable. Machine learning is particularly useful for ranking scarce audit resources and detecting complex patterns, but its practical value depends on investigation and legal follow-up. Blockchain remains prospective because direct evaluations of tax-audit outcomes are limited.

The second sub-question concerned performance dimensions. Operational efficiency and risk targeting receive the most consistent support. Compliance and traceability also improve in many settings, although behavioral and institutional effects vary. Direct evidence on audit yield and long-term net cost is more limited. The review therefore supports a conditional performance hypothesis: digitalization is likely to improve audit performance when it reduces information frictions and improves resource allocation, but the size and durability of the gain depend on implementation quality and institutional capacity.

The third sub-question concerned moderators and risks. Data quality, interoperability, staff capability, legal authority, cybersecurity, and algorithmic accountability are not peripheral controls. They are productive conditions. Weakness in any of them can convert technical sophistication into operational error, legal vulnerability, or loss of trust.

5.2. Theoretical implications

The results refine general digital-transformation theory by specifying the mechanisms that connect technology to public-sector performance. Digital transformation is not defined by the acquisition of new tools but by changes in value creation, organizational structures, and decision processes (Mergel et al., 2019; Vial, 2019). In tax auditing, value is created when digital records alter information availability, when analytics changes case prioritization, and when workflows reorganize the relationship between machine output and professional judgment.

The synthesis also qualifies technology-acceptance explanations. Perceived usefulness and ease of use matter for taxpayers and staff, particularly in e-invoicing and e-filing adoption. Yet acceptance is insufficient to explain audit performance. A widely used system can still produce poor evidence if identifiers are unreliable or data are not interoperable. Conversely, a technically accurate system can fail if auditors do not trust, understand, or lawfully apply its outputs. The proposed framework therefore complements adoption models with data, legal, and accountability conditions.

A further implication concerns algorithmic governance. Risk models do not remove discretion; they relocate it into data definitions, target variables, thresholds, training samples, and workflow rules. Accountability must therefore cover system design and institutional use, not only the final auditor decision. This finding supports process-based approaches that document who defined the model objective, what evidence was used, how performance was validated, and how affected taxpayers can correct errors (Wieringa, 2020).

5.3. Managerial and policy implications

For tax administrators, the principal implication is that digital audit investment should be governed as a portfolio of capabilities. The first priority is reliable digital data: stable taxpayer identifiers, standardized invoices, metadata, data-quality controls, and correction procedures. The second is integration: secure interfaces, common definitions, and case-management systems that allow information to be traced from source to decision. The third is people: multidisciplinary teams and continuous training for auditors, analysts, lawyers, and security specialists.

Predictive models should be introduced through controlled pilots. Validation should compare the model with the existing selection process using out-of-sample data and operational outcomes, not accuracy alone. Relevant indicators include hit rate, additional assessment, collection, cycle time, false-positive rate, appeal outcomes, subgroup effects, and net cost. Human review should be documented, particularly when staff override or confirm a model recommendation.

Legal and ethical governance should be developed before scaling. The administration should publish an intelligible description of the purpose and safeguards of automated selection without revealing details that would facilitate evasion. Internal documentation should be more detailed and include variables, prohibited data, validation results, version history, access rights, and incident procedures. Taxpayer-facing processes should permit correction of source data and meaningful review of adverse decisions.

Cybersecurity should be integrated into program evaluation. A system that improves audit selection but exposes financial records or disrupts filing cannot be considered high performing.

Procurement and supplier contracts should address data location, subcontractors, patching, audit rights, model updates, business continuity, and exit arrangements.

5.4. Transferability, limitations, and priorities for future research

The findings are not equally transferable across administrations. Systems with broad taxpayer registration, mature e-invoicing, reliable third-party data, and strong appeal procedures can exploit advanced analytics more effectively than administrations where transactions remain informal or identifiers are fragmented. Context matters at four levels: digital coverage, administrative capacity, legal authority, and taxpayer behavior. Policy transfer should therefore focus on mechanisms and preconditions rather than copying a specific model.

This review has several limitations. First, the English-language restriction and reliance on major databases may omit national studies and technical evaluations. Second, the included studies are heterogeneous in design, outcome, and quality, which precludes meta-analysis and limits comparison of effect size. Third, publication bias may favor successful digital projects and high-performing models. Fourth, quality appraisal involved interpretive judgment.

Future research should be prioritized in four areas. The first priority is longitudinal evaluation of net performance, including implementation and maintenance costs, realized collections, appeals, and model drift. The second is causal evidence from low- and middle-income administrations where digital coverage and informality differ from OECD settings. The third is algorithmic accountability, including explainability, bias, human oversight, and taxpayer remedies. The fourth is comparative evaluation of implementation sequences, particularly whether investment in data quality and staff capacity produces larger returns than early investment in sophisticated models.

These priorities are feasible when tax administrations create secure research partnerships, document system changes, and retain evaluation datasets. Future studies should predefine outcomes and distinguish technical validation from operational impact.

6. Conclusion

This systematic review consolidates evidence from 35 studies on the relationship between fiscal digitalization and tax audit performance. The strongest and most recurrent findings concern faster processing, improved data availability, more focused risk selection, and stronger transaction traceability. Compliance effects are also positive in several settings, while direct evidence on long-term net cost and revenue yield is more limited and context dependent.

The scientific contribution is an integrated framework that connects technology, operational mechanism, institutional condition, and performance dimension. It explains why e-invoicing, prefilled returns, data integration, and machine learning cannot be assessed through a single indicator. The methodological contribution is a structured synthesis that separates high-quality direct evidence from convergent but heterogeneous tendencies and from prospective claims. The managerial contribution is a sequencing principle: data quality, interoperability, skills, legal authority, and cybersecurity should precede large-scale predictive automation.

For tax administrations, the central implication is that digitalization should be evaluated as a socio-technical reform. Software acquisition is only one component. Performance depends on the reliability of the data, the competence of the people using the system, the enforceability of digital evidence, and the capacity to explain and review automated recommendations. Governance safeguards are therefore part of productive capacity rather than constraints added after deployment.

Further research should concentrate on longitudinal outcomes, developing-country settings, net implementation costs, appeal and error data, and accountable use of artificial intelligence. International exchange can support common principles for data quality, security,

documentation, and evaluation, but technical standardization should remain adaptable to national legal and administrative contexts.

References

- (1). Alm, J., Cherry, T., Jones, M., & McKee, M. (2010). Taxpayer information assistance services and tax compliance behavior. *Journal of Economic Psychology*, 31(4), 577-586.
- (2). Baghdasaryan, V., Davtyan, H., Sarikyan, A., & Navasardyan, Z. (2022). Improving tax audit efficiency using machine learning: The role of taxpayer's network data in fraud detection. *Applied Artificial Intelligence*, 36(1), 2012002. <https://doi.org/10.1080/08839514.2021.2012002>
- (3). Bassey, E., Mulligan, E., & Ojo, A. (2022). A conceptual framework for digital tax administration: A systematic review. *Government Information Quarterly*, 39(4), 101754. <https://doi.org/10.1016/j.giq.2022.101754>
- (4). Battaglini, M., Guiso, L., Lacava, C., Miller, D. L., & Patacchini, E. (2025). Refining public policies with machine learning: The case of tax auditing. *Journal of Econometrics*, 249, 105847. <https://doi.org/10.1016/j.jeconom.2024.105847>
- (5). Bellon, M., Dabla-Norris, E., Khalid, S., & Lima, F. (2022). Digitalization to improve tax compliance: Evidence from VAT e-invoicing in Peru. *Journal of Public Economics*, 210, 104661. <https://doi.org/10.1016/j.jpubeco.2022.104661>
- (6). Booth, A., Sutton, A., & Papaioannou, D. (2016). *Systematic approaches to a successful literature review* (2nd ed.). Sage.
- (7). Castellón González, P., & Velásquez, J. D. (2013). Characterization and detection of taxpayers with false invoices using data mining techniques. *Expert Systems with Applications*, 40(5), 1427-1436. <https://doi.org/10.1016/j.eswa.2012.08.051>
- (8). Chen, S.-C., Wu, C.-C., & Miao, S. (2015). Constructing an integrated e-invoice system: The Taiwan experience. *Transforming Government: People, Process and Policy*, 9(3), 370-383.
- (9). Cuylen, A., Kosch, L., & Breitner, M. H. (2016). Development of a maturity model for electronic invoice processes. *Electronic Markets*, 26(2), 115-127.
- (10). Dabla-Norris, E., Misch, F., Cleary, D., & Khwaja, M. (2020). The quality of tax administration and firm performance: Evidence from developing countries. *International Tax and Public Finance*, 27(3), 514-551.
- (11). de Clercq, B. (2019). The uberisation of e-filing in South Africa. *eJournal of Tax Research*, 16(3), 440-473.
- (12). Doxey, M. M., Lawson, J. G., & Stinson, S. R. (2021). The effects of prefilled tax returns on taxpayer compliance. *Journal of the American Taxation Association*, 43(2), 63-85.
- (13). Evans, C., & Tran-Nam, B. (2010). Managing tax system complexity: Building bridges through prefilled tax returns. *Australian Tax Forum*, 25(2), 245-274.
- (14). Fochmann, M., Hechtner, F., Kölle, T., & Overesch, M. (2021). Combating overreporting of deductions in tax returns: Prefilling and restricting the deductibility of expenditures. *Journal of Business Economics*, 91(7), 935-964.
- (15). Fochmann, M., Müller, N., & Overesch, M. (2021). Less cheating? The effects of prefilled forms on compliance behavior. *Journal of Economic Psychology*, 83, 102365.
- (16). Fonseca, M. A., & Grimshaw, S. B. (2017). Do behavioral nudges in prepopulated tax forms affect compliance? Experimental evidence with real taxpayers. *Journal of Public Policy & Marketing*, 36(2), 213-226.

- (17). Gillitzer, C., & Skov, P. E. (2018). The use of third-party information reporting for tax deductions: Evidence and implications from charitable deductions in Denmark. *Oxford Economic Papers*, 70(3), 892-916.
- (18). Hernández-Ortega, B., & Jiménez-Martínez, J. (2013). Performance of e-invoicing in Spanish firms. *Information Systems and e-Business Management*, 11(3), 457-480.
- (19). Hernández-Ortega, B. (2012). Key factors for adoption and subsequent use of e-invoicing. *Academia Revista Latinoamericana de Administración*, 50, 15-30.
- (20). Hesami, S., Jenkins, H., & Jenkins, G. P. (2024). Digital transformation of tax administration and compliance: A systematic literature review on e-invoicing and prefilled returns. *Digital Government: Research and Practice*, 5(3), Article 18. <https://doi.org/10.1145/3643687>
- (21). Kleanthous, C., & Chatzis, S. (2020). Gated mixture variational autoencoders for value added tax audit case selection. *Knowledge-Based Systems*, 188, 105048. <https://doi.org/10.1016/j.knosys.2019.105048>
- (22). Kleven, H. J., Knudsen, M. B., Kreiner, C. T., Pedersen, S., & Saez, E. (2011). Unwilling or unable to cheat? Evidence from a tax audit experiment in Denmark. *Econometrica*, 79(3), 651-692.
- (23). Kochanova, A., Hasnain, Z., & Larson, B. (2020). Does e-government improve government capacity? Evidence from tax compliance costs, tax revenue, and public procurement competitiveness. *The World Bank Economic Review*, 34(1), 101-120. <https://doi.org/10.1093/wber/lhx024>
- (24). Lian, J.-W. (2015). Critical factors for cloud based e-invoice service adoption in Taiwan: An empirical study. *International Journal of Information Management*, 35(1), 98-109.
- (25). Mascagni, G., Mengistu, A. T., & Woldeyes, F. B. (2021). Can ICTs increase tax compliance? Evidence on taxpayer responses to technological innovation in Ethiopia. *Journal of Economic Behavior & Organization*, 189, 172-193. <https://doi.org/10.1016/j.jebo.2021.06.007>
- (26). Mergel, I., Edelmann, N., & Haug, N. (2019). Defining digital transformation: Results from expert interviews. *Government Information Quarterly*, 36(4), 101385. <https://doi.org/10.1016/j.giq.2019.06.002>
- (27). OECD. (2023). Tax administration 2023: Comparative information on OECD and other advanced and emerging economies. OECD Publishing. <https://doi.org/10.1787/900b6382-en>
- (28). Okunogbe, O., & Pouliquen, V. (2022). Technology, taxation, and corruption: Evidence from the introduction of electronic tax filing. *American Economic Journal: Economic Policy*, 14(1), 341-372. <https://doi.org/10.1257/pol.20200123>
- (29). Okunogbe, O., & Santoro, F. (2023). The promise and limitations of information technology for tax mobilization. *The World Bank Research Observer*, 38(2), 295-324. <https://doi.org/10.1093/wbro/lkac008>
- (30). Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., ... Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ*, 372, n71. <https://doi.org/10.1136/bmj.n71>
- (31). Qi, Y., & Che Azmi, A. (2021). Factors affecting electronic invoice adoption and tax compliance process efficiency. *Transforming Government: People, Process and Policy*, 15(1), 150-168. <https://doi.org/10.1108/TG-04-2020-0070>

- (32). Rahimikia, E., Mohammadi, S., Rahmani, T., & Ghazanfari, M. (2017). Detecting corporate tax evasion using a hybrid intelligent system: A case study of Iran. *International Journal of Accounting Information Systems*, 25, 1-17. <https://doi.org/10.1016/j.accinf.2016.12.002>
- (33). Salmony, M., & Harald, B. (2010). E-invoicing in Europe: Now and the future. *Journal of Payments Strategy & Systems*, 4(4), 371-380.
- (34). Savić, M., Atanasijević, J., Jakovetić, D., & Krejić, N. (2022). Tax evasion risk management using a hybrid unsupervised outlier detection method. *Expert Systems with Applications*, 193, 116409. <https://doi.org/10.1016/j.eswa.2021.116409>
- (35). Snyder, H. (2019). Literature review as a research methodology: An overview and guidelines. *Journal of Business Research*, 104, 333-339. <https://doi.org/10.1016/j.jbusres.2019.07.039>
- (36). Tranfield, D., Denyer, D., & Smart, P. (2003). Towards a methodology for developing evidence-informed management knowledge by means of systematic review. *British Journal of Management*, 14(3), 207-222. <https://doi.org/10.1111/1467-8551.00375>
- (37). Uyar, A., Nimer, K., Kuzey, C., Shahbaz, M., & Schneider, F. (2021). Can e-government initiatives alleviate tax evasion? The moderation effect of ICT. *Technological Forecasting and Social Change*, 166, 120597. <https://doi.org/10.1016/j.techfore.2021.120597>
- (38). van Dijk, W. W., Goslinga, S., Terwel, B. W., & van Dijk, E. (2020). How choice architecture can promote and undermine tax compliance: Testing the effects of prepopulated tax returns and accuracy confirmation. *Journal of Behavioral and Experimental Economics*, 87, 101574.
- (39). Vanhoeyveld, J., Martens, D., & Peeters, B. (2020). Value-added tax fraud detection with scalable anomaly detection techniques. *Applied Soft Computing*, 86, 105895. <https://doi.org/10.1016/j.asoc.2019.105895>
- (40). Vial, G. (2019). Understanding digital transformation: A review and a research agenda. *The Journal of Strategic Information Systems*, 28(2), 118-144. <https://doi.org/10.1016/j.jsis.2019.01.003>
- (41). Vieira, P. A., Pimenta, D. P., Cruz, A. F. da, & Souza, E. M. S. de. (2019). Effects of the electronic invoice program on the increase of state collection. *Revista de Administração Pública*, 53(2), 481-491.
- (42). Walpole, M. (2014). VAT compliance cost indicators. *Australian Tax Forum*, 29(3), 391-415.
- (43). Warren, N. (2016). E-filing and compliance risk: Evidence from Australian personal income tax deductions. *Australian Tax Forum*, 31(3), 577-602.
- (44). Wieringa, M. (2020). What to account for when accounting for algorithms: A systematic literature review on algorithmic accountabilities. *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, 1-18. <https://doi.org/10.1145/3351095.3372833>
- (45). Wu, R.-S., Ou, C. S., Lin, H.-Y., Chang, S.-I., & Yen, D. C. (2012). Using data mining technique to enhance tax evasion detection performance. *Expert Systems with Applications*, 39(10), 8769-8777. <https://doi.org/10.1016/j.eswa.2012.01.204>
- (46). Zheng, Q., Xu, Y., Liu, H., Shi, B., Wang, J., & Dong, B. (2024). A survey of tax risk detection using data mining techniques. *Engineering*, 34, 43-59. <https://doi.org/10.1016/j.eng.2023.07.014>

Appendix A: Characteristics of the 35 included studies

The matrix reports the study-level evidence used in the synthesis. The quality rating reflects the five-criterion appraisal described in Section 3.4.

Table A1: Study-level evidence matrix

No.	Study	Year	Context	Digital technology	Method	Primary outcome	Quality
1	Alm et al.	2010	United States	Digital taxpayer assistance	Field experiment	Compliance and service response	High
2	Evans and Tran-Nam	2010	Australia	Prefilled returns	Policy analysis	Compliance costs and simplification	Moderate
3	Salmony and Harald	2010	Europe	E-invoicing	Industry and policy analysis	Process standardization and cost	Moderate
4	Kleven et al.	2011	Denmark	Third-party reporting and audit data	Randomized audit experiment	Evasion, detection and deterrence	High
5	Wu et al.	2012	Taiwan	Data mining	Predictive classification	Tax evasion detection	High
6	Hernández-Ortega	2012	Spain	E-invoicing	Firm survey	Adoption and continued use	Moderate
7	Castellón González and Velásquez	2013	Chile	Data mining	Predictive classification	False-invoice detection	High
8	Hernández-Ortega and Jiménez-Martínez	2013	Spain	E-invoicing	Firm survey	Operational performance	Moderate
9	Walpole	2014	Cross-country	VAT digital reporting	Comparative analysis	Compliance costs	Moderate
10	Chen et al.	2015	Taiwan	Integrated e-invoice system	National case study	Interoperability and traceability	High
11	Lian	2015	Taiwan	Cloud e-invoicing	Adoption survey	Adoption determinants	Moderate
12	Cuylen et al.	2016	Germany	E-invoice process maturity	Design and validation study	Process maturity and control	High
13	Warren	2016	Australia	E-filing	Administrative-data analysis	Compliance risk and deductions	High
14	Fonseca and Grimshaw	2017	United Kingdom	Prefilled forms and nudges	Field experiment	Reporting accuracy	High
15	Rahimikia et al.	2017	Iran	Hybrid intelligent system	Predictive modeling	Corporate tax evasion detection	High
16	Gillitzer and Skov	2018	Denmark	Third-party information reporting	Quasi-experimental analysis	Deduction compliance	High
17	Vieira et al.	2019	Brazil	Electronic invoice program	Panel and policy evaluation	State tax collection	High
18	de Clercq	2019	South Africa	E-filing platform	Institutional case analysis	Service design and access	Moderate

19	Kochanova et al.	2020	Cross-country	E-government systems	Cross-country econometrics	Government capacity and compliance costs	High
20	Dabla-Norris et al.	2020	Developing countries	Tax administration quality and digital capability	Firm-level econometrics	Firm performance and administrative quality	High
21	Vanhoeeyveld et al.	2020	Belgium	Unsupervised anomaly detection	Machine-learning evaluation	VAT fraud targeting	High
22	Kleanthous and Chatzis	2020	European VAT context	Variational autoencoders	Machine-learning evaluation	Audit case selection	High
23	van Dijk et al.	2020	Netherlands	Prefilled returns	Behavioral experiment	Accuracy and compliance	High
24	Qi and Che Azmi	2021	Malaysia	E-invoicing	Survey and structural modeling	Adoption and process efficiency	Moderate
25	Uyar et al.	2021	Cross-country	E-government	Panel econometrics	Tax evasion	High
26	Fochmann, Hechtner et al.	2021	Germany	Prefilled returns and deduction rules	Laboratory experiment	Overreporting and compliance	High
27	Fochmann, Müller et al.	2021	Germany	Prefilled returns	Laboratory experiment	Compliance behavior	High
28	Doxey et al.	2021	United States	Prefilled returns	Behavioral experiment	Taxpayer compliance	High
29	Mascagni et al.	2021	Ethiopia	Electronic tax services	Administrative-data and survey analysis	Compliance response and access	High
30	Bellon et al.	2022	Peru	VAT e-invoicing	Quasi-experimental administrative-data analysis	VAT reporting and compliance	High
31	Okunogbe and Pouliquen	2022	Tajikistan	Electronic tax filing	Field experiment and administrative data	Compliance costs and corruption exposure	High
32	Baghdasaryan et al.	2022	Armenia	Machine learning and network data	Predictive modeling	Audit case selection and fraud detection	High
33	Savić et al.	2022	Serbia	Hybrid outlier detection	Machine-learning evaluation	Tax evasion risk management	High
34	Hesami et al.	2024	International	E-invoicing and prefilled returns	Systematic literature review	Compliance, costs and design conditions	High
35	Battaglini et al.	2025	Italy	Machine-learning policy refinement	Econometric and predictive evaluation	Tax audit targeting and policy design	High

Source: Authors' systematic review matrix.