



Tax360: A Hybrid Behavioral–Technological Analytics Framework for Enhancing SME Tax Compliance in Emerging Digital Economies

Dr. Stanley Mwangi Chege and Serena Waithera Rolloh

Department of Information and Computer Science, School of Science, Catholic University of
Eastern Africa, Nairobi, Kenya

Abstract: Small and Medium Enterprises (SMEs) drive productivity, innovation, and employment in emerging economies, yet they also account for a large share of persistent tax under-compliance. Digital transformations such as Kenya's iTax filing platform and the Electronic Tax Invoice Management System (eTIMS) have expanded transaction capture and improved the audit data footprint, but the benefits are constrained when platforms function as separate silos without automated, near-real-time reconciliation. In siloed environments, non-compliance behaviors such as revenue under-declaration, ghost invoicing, and strategic Nil-filing can persist, undermining domestic resource mobilization and eroding perceptions of fairness among compliant taxpayers (IMF, 2023; Slemrod, 2019). This study introduces Tax360, a hybrid behavioral–technological analytics framework that integrates (a) unsupervised anomaly detection using Isolation Forest (Liu et al., 2008), (b) transaction-network indicators to capture coordinated VAT risk propagation (Alexopoulos et al., 2021; Wang et al., 2023), (c) explainable AI (XAI) using SHAP-based local explanations for accountability (Li et al., 2023; Kass et al., 2023), and (d) automation features aimed at reducing taxpayer and auditor burden (Braithwaite, 2020; Ribeiro & Lopes, 2022). Following a Design Science Research (DSR) methodology (Hevner et al., 2004), a prototype was implemented and evaluated against a high-fidelity synthetic dataset of 10,000 SMEs and 200,000 VAT invoice records engineered to reflect realistic sector distributions and injected fraud patterns (De Souza & Ribeiro, 2021). Results show that a composite risk model combining Isolation Forest scores with reconciliation and network metrics achieved strong performance (Precision = 0.88; Recall = 0.82; AUC-ROC = 0.90) relative to common unsupervised baselines. By pairing high-precision risk scoring with transparent rationales and an auditable workflow, Tax360 aims to strengthen both deterrence and institutional legitimacy, offering a scalable governance model for modern tax administrations in emerging digital economies.

Keywords: RegTech; anomaly detection; explainable AI; VAT fraud detection; SMEs; digital government; tax compliance.

1. Introduction

1.1 Background: The SME Compliance Paradox

Domestic resource mobilization is essential for fiscal sustainability, public service delivery, and long-term development in emerging economies (IMF, 2023). SMEs frequently represent the majority of registered businesses and are central to job creation, local innovation, and supply-chain resilience. Yet, in many contexts SMEs contribute less tax revenue than expected relative to their economic role, creating a persistent compliance paradox. This discrepancy is shaped by structural factors: informality, limited bookkeeping capacity, heavy reliance on cash or semi-formal payment channels, and the opportunity to trade outside documented supply chains. Administrative complexity can amplify the gap because SMEs face higher relative compliance costs than large firms; they may lack professional accountants, integrated ERP systems, and stable digital skills within their staff (Braithwaite, 2020; Loo, 2006). As a result, enforcement strategies that rely on periodic audits or blanket compliance drives can produce mixed outcomes: they may recover some revenue but also increase resentment among compliant SMEs who experience audits as harassment rather than a fair, evidence-based process (Muehlbacher & Kirchler, 2021).

To address compliance challenges, tax administrations have digitized registration, filing, payment, and invoice capture. Kenya's iTax system provides electronic filing and taxpayer account management, while eTIMS captures invoice-level

data at the point of sale and supports transaction visibility for VAT administration (Kenya Revenue Authority, 2024). Global evidence suggests e-invoicing can increase data quality, reduce manual errors, and support business productivity when integrated into operational workflows (Australian Taxation Office, 2023). However, digitization does not automatically translate into intelligence: when invoice capture and return filing exist as parallel repositories, the administration gains data without the ability to reconcile, triage, and act on it at scale. This “silo effect” can be operationally costly because auditors must manually compare transaction streams, leaving enforcement reactive and selective based on limited human capacity. It is also strategically risky, because sophisticated actors can exploit timing and system boundaries—such as issuing invoices while under-declaring returns—especially when reconciliation is infrequent or occurs only after long delays.

VAT environments are particularly vulnerable to relational fraud patterns. Ghost invoicing can inflate input VAT claims, while missing trader strategies may involve networks of entities that distribute risk across counterparties, making single-entity audits insufficient (Alexopoulos et al., 2021; Chen et al., 2022). In such settings, the fairness of enforcement becomes inseparable from detection quality: if risk selection is noisy, compliant SMEs are audited and trust declines; if risk selection is accurate and explainable, the administration can improve deterrence while protecting legitimacy (Frey & Kirchler, 2022). Tax360 is motivated by this dual requirement: improve detection precision and build a governance-ready workflow that can be justified, audited, and communicated.

1.2 Problem Statement: Intelligence in the Age of Big Data

The problem addressed in this study is the lack of integrated, actionable intelligence within digitally transformed tax administrations. Although invoice capture systems and filing systems generate large volumes of data, the absence of automated reconciliation, anomaly detection, and decision-support narratives enables high-risk behaviors to persist undetected for long periods. This leads to revenue leakage and undermines domestic resource mobilization (IMF, 2023). It also creates a legitimacy gap: honest SMEs experience enforcement as arbitrary when audit selection appears inconsistent or when auditors cannot clearly explain why a case was targeted (Muehlbacher & Kirchler, 2021). From a governance perspective, the legitimacy gap is compounded by the increasing use of algorithmic tools in public administration, which raises accountability obligations around transparency, bias, and contestability (Cath, 2018; Crawford, 2021).

1.3 Research Objectives and Questions

The general objective is to design and evaluate an AI-enabled compliance analytics engine that improves SME tax monitoring while fostering voluntary compliance through transparency and burden reduction.

Specific objectives are to: (O1) engineer a synthetic SME dataset mirroring realistic compliance behaviors and fraud patterns; (O2) develop a hybrid anomaly detection model integrating unsupervised machine learning with fiscal rules and network metrics; (O3) implement XAI interfaces that provide auditors with transparent rationales; and (O4) evaluate performance and governance relevance against established baselines (Hevner et al., 2004; Khan & Ahmad, 2021).

Research questions are: RQ1: How effectively does a hybrid anomaly model detect high-risk SME reporting behaviors in a low-label environment? RQ2: Does explainability improve transparency and perceived fairness for auditor-facing decisions? RQ3: Can process automation reduce compliance burden for SMEs while improving administrative efficiency for auditors?

1.4 Contributions

This study makes four contributions. First, it proposes a hybrid detection architecture that is practical for emerging economies where confirmed fraud labels are limited and may reflect enforcement bias (Khan & Ahmad, 2021). Second, it integrates transaction-network indicators to capture coordinated VAT risk propagation, addressing the limitation of single-entity outlier detection in networked fraud scenarios (Alexopoulos et al., 2021; Wang et al., 2023). Third, it positions explainability as a governance requirement by designing SHAP-based local explanations into auditor workflow and logging, thereby supporting procedural justice, auditability, and defensibility in enforcement decisions (Li et al., 2023; Kass et al., 2023). Fourth, it links analytics to burden reduction through automation concepts that can reduce friction for SMEs and reduce manual reconciliation workload for auditors (Braithwaite, 2020; Ribeiro & Lopes, 2022).

2. Literature review

2.1 Compliance Theory: Deterrence, Legitimacy, and Burden

Economic deterrence theory explains compliance as a rational choice under uncertainty: taxpayers weigh expected gains from evasion against expected costs shaped by audit probability and penalty severity (Allingham & Sandmo, 1972). Empirical and policy work continues to validate deterrence mechanisms, especially when audit targeting becomes more effective and visible (Slemrod, 2019). However, deterrence is incomplete as a standalone explanation because taxpayers are also social actors embedded in institutions; they respond to norms, trust, and perceived legitimacy of public authority. Legitimacy theory emphasizes that voluntary compliance is strengthened when tax administrations are perceived as fair, transparent, and respectful, and when decisions are procedurally just (Frey & Kirchler, 2022; Muehlbacher & Kirchler, 2021). In practice, legitimacy is influenced by how enforcement is experienced: frequent arbitrary audits can reduce tax morale even if penalties are strong. Burden reduction theory adds that compliance cost matters: simplifying processes and reducing friction can move SMEs from avoidance to cooperation by lowering the non-monetary cost of being compliant (Braithwaite, 2020; Loo, 2006).

2.2 Digital Government, RegTech, and Tax Administration 3.0

Digital government scholarship highlights an evolution from digitizing forms and records to delivering intelligent, data-driven public services that can adapt to complex realities (Medaglia & Zhu, 2017; Wirtz & Müller, 2019). In this evolution, public value depends not only on technology capability but also on governance, trust, and institutional design. Tax360 is a part of a broader innovation-centric transformation of public administration (Maxamadumarovich et al., 2012). Studies emphasize persistent hurdles: data interoperability, skills gaps, procurement constraints, legacy system fragmentation, and unclear accountability when algorithms influence public decisions (Janssen et al., 2020; Sun & Medaglia, 2019). These hurdles are especially salient in emerging economies, where infrastructure and capacity constraints can slow adoption and increase risks of uneven implementation.

RegTech extends these insights by focusing on technology-enabled compliance and regulation, and it is increasingly applied to public revenue functions. RegTech in emerging markets must account for contextual constraints, including limited formal records and heterogeneous SME ecosystems (Al-Omari et al., 2023). In tax, the OECD's Tax Administration 3.0 vision emphasizes standardized data availability and exchange, allowing compliance to be embedded in business processes rather than performed after-the-fact through manual returns (OECD, 2021, 2023). Tax Administration 3.0 further implies a shift in enforcement: administrations move from sampling-based audits to continuous risk monitoring supported by real-time data streams. Where integration is weak, the administration may still depend on batch audits, limiting deterrence and increasing friction.

2.3 AI Governance and Algorithmic Accountability in Enforcement

Public-sector AI adoption raises governance and accountability questions because state algorithms can alter resource allocation, enforcement intensity, and citizen experience. AI governance research emphasizes transparency, contestability, and human oversight as safeguards against opaque decisions and unintended harms (Cath, 2018). Crawford (2021) argues that AI systems can obscure underlying political and economic choices while reinforcing power asymmetries, making accountability mechanisms essential. Within tax administration, these concerns are amplified because audit selection and enforcement actions have direct financial and reputational consequences for businesses. Therefore, any AI-enabled audit selection system should be designed as a decision-support tool with auditable rationale rather than as an autonomous judge.

Explainable AI methods provide technical mechanisms to support accountability. SHAP is widely used to generate local explanations by attributing feature contributions to individual predictions or scores, which can be mapped to evidence and reviewed by auditors (Li et al., 2023). However, explainability is only useful if integrated into workflow and governance processes. Kass et al. (2023) emphasize the role of explainability in digital taxation as a bridge between analytics and legitimate decision-making, especially when decisions must be defended internally and externally. Tax360 treats explainability as a first-class design requirement and integrates it with logging and role-based access controls consistent with AI management system thinking (ISO, 2022).

2.4 VAT and Tax Fraud Detection Approaches

Tax anomaly detection approaches can be broadly classified as supervised, semi-supervised, and unsupervised. Supervised models require reliable labels, which are frequently scarce in tax contexts and can embed enforcement bias (Khan & Ahmad, 2021). Unsupervised methods are therefore attractive, especially for continuous monitoring, because they can detect unusual patterns without confirmed labels. Isolation Forest is a prominent unsupervised approach because it isolates anomalies through random partitioning and scales efficiently to large datasets (Liu et al., 2008). Baseline methods such as LOF and One-Class SVM are also commonly used, although they can be sensitive to parameterization and local density assumptions.

For VAT fraud, network analysis is particularly important because many schemes are relational rather than purely individual. Graph-based approaches model invoice and counterparty relationships to identify suspicious communities, central nodes, and risk propagation dynamics (Chen et al., 2022; Wang et al., 2023). Alexopoulos et al. (2021) demonstrate that combining network signals with machine learning can improve VAT fraud detection, especially when fraud behaviors are coordinated. Tax360 builds on this foundation by explicitly integrating network metrics into a composite risk score alongside reconciliation features and machine-learning anomaly scores.

2.5 Kenyan Context and Adoption Considerations

Kenya’s tax digitization has made substantial progress, but operational effectiveness depends on system integration and adoption. The Kenya Revenue Authority has provided guidance to support eTIMS adoption, indicating the strategic importance of invoice capture for VAT compliance (Kenya Revenue Authority, 2024). Evaluation work on eTIMS adoption suggests that outcomes vary by sector and by the maturity of taxpayer systems, implying a need for supportive interventions and well-designed analytics to avoid penalizing capability constraints (Kibet, 2022). SME compliance behavior in Kenya is influenced by deterrence, capacity, and legitimacy considerations, including perceptions of fairness and value-for-tax (Kariuki, 2023). Therefore, a compliance analytics framework must not only detect anomalies but also support fair treatment and guidance to reduce unintentional non-compliance.

2.6 Research Gap

Three gaps motivate Tax360. First, digitized tax platforms often remain fragmented, preventing automated reconciliation and near-real-time risk triage (OECD, 2023). Second, many tax administrations operate in low-label environments that require scalable anomaly detection approaches with safeguards against false positives and enforcement bias (Khan & Ahmad, 2021). Third, AI-enabled enforcement requires governance-by-design: explainability, traceability, and human oversight must be embedded to protect legitimacy and comply with public-sector accountability expectations (Cath, 2018; ISO, 2022). Tax360 addresses these gaps by combining hybrid analytics with explainable decision support and burden reduction concepts.

3. Conceptual Model and Hypotheses

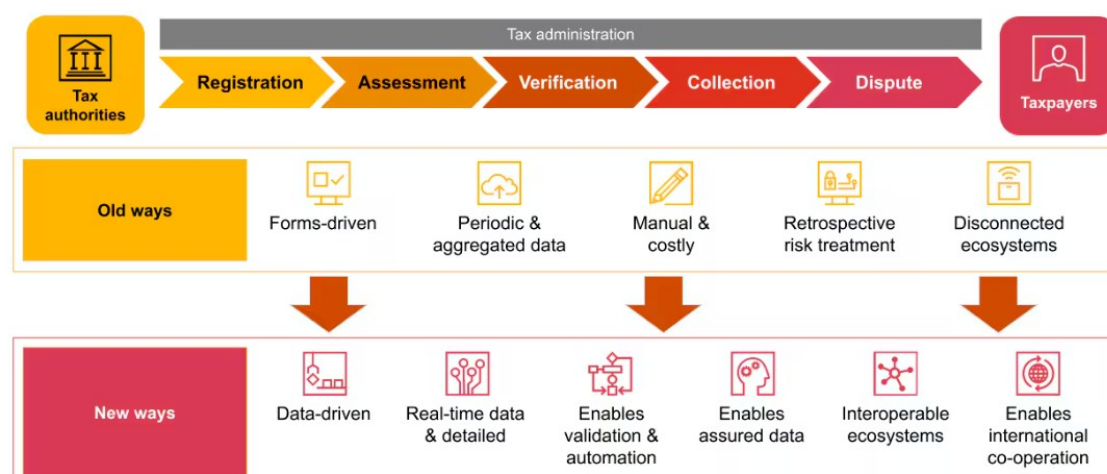
3.1 Model Overview

Tax360 conceptualizes compliance outcomes as shaped by three interacting mechanisms. The deterrence pathway posits that improved detection precision increases perceived audit probability and the expected cost of evasion (Allingham & Sandmo, 1972; Slemrod, 2019). The legitimacy pathway posits that explainable, evidence-based enforcement improves perceptions of fairness and institutional legitimacy, strengthening voluntary compliance (Frey & Kirchler, 2022). The burden pathway posits that automation and guidance reduce compliance friction, directly improving compliance intent and moderating the impact of fairness on compliance by making compliance practically achievable (Braithwaite, 2020; Ribeiro & Lopes, 2022). This integrated view aligns with the idea that enforcement power and trust can be mutually reinforcing under conditions of procedural justice (Kirchler et al., 2008).

3.2 Hypotheses

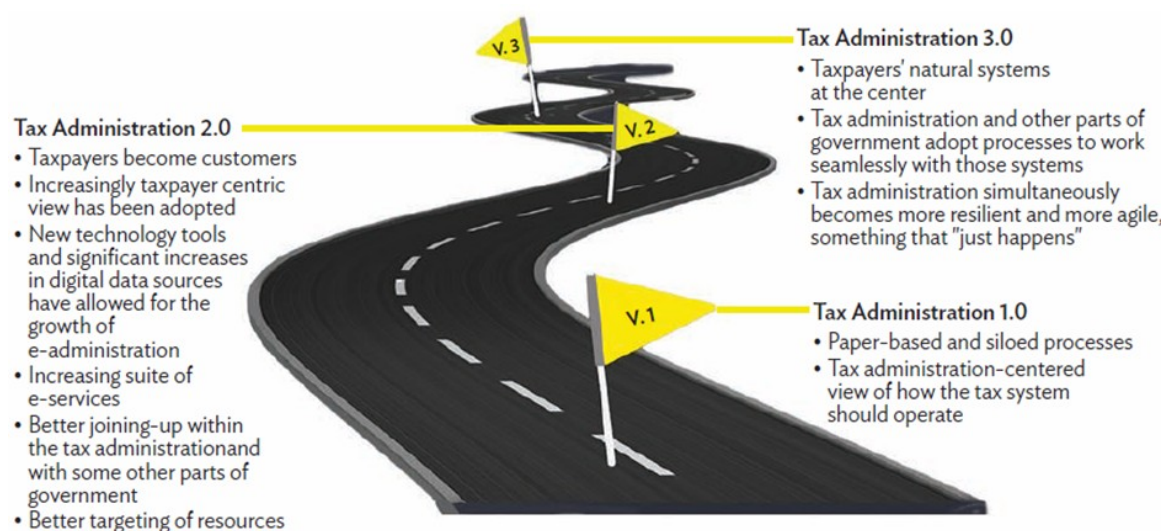
- H1: Higher perceived detection risk from anomaly modeling positively influences voluntary compliance intent.
- H2: Explainable audit flags enhance perceptions of fairness in tax governance.
- H3: Fairness positively influences voluntary compliance intent.
- H4: Reduced compliance burden increases voluntary compliance intent.
- H5: Perceived burden reduction moderates the relationship between fairness and compliance intent.
- H6: The combined effect of detection risk, fairness, and burden reduction yields higher compliance outcomes than any individual pathway alone.

Figures and Tables



Source: Tax Administration 3.0: The Digital Transformation of Tax Administration, OECD, 2020

Figure 1: Tax Administration 3.0



Source: OECD, 2020. Forum on Tax Administration.

Figure 2: Roadmap to effectiveness

Figure 3: Tax360 Framework

The above Tax360 Framework visualizes an integrated pipeline that transforms fragmented data into actionable, accountable tax intelligence. The following is a short description of its core components:

Data Integration Layer: The process begins by reconciling two parallel information silos: iTax (self-declarations) and eTIMS (real-time invoices).

Tax360 Analytics Engine: This hybrid core utilizes three distinct analytical methods:

Unsupervised Anomaly Detection: Uses Isolation Forest to identify outliers without needing historical labels.

Transaction-Network Indicators: Maps relationships between entities to detect coordinated fraud, such as "missing trader" schemes.

Explainable AI (SHAP-based XAI): Breaks down complex risk scores into understandable drivers for human review.

Governance & Auditor Dashboard: The output provides auditors with a Composite Risk Score and a SHAP Waterfall Plot that lists



specific drivers—such as high invoice variance or strategic filing lags—ensuring "human-in-the-loop" accountability.

Behavioral Outcomes: By moving from batch audits to precise, explainable targeting, the framework seeks a synergy between deterrence power and institutional trust, ultimately fostering higher voluntary compliance.

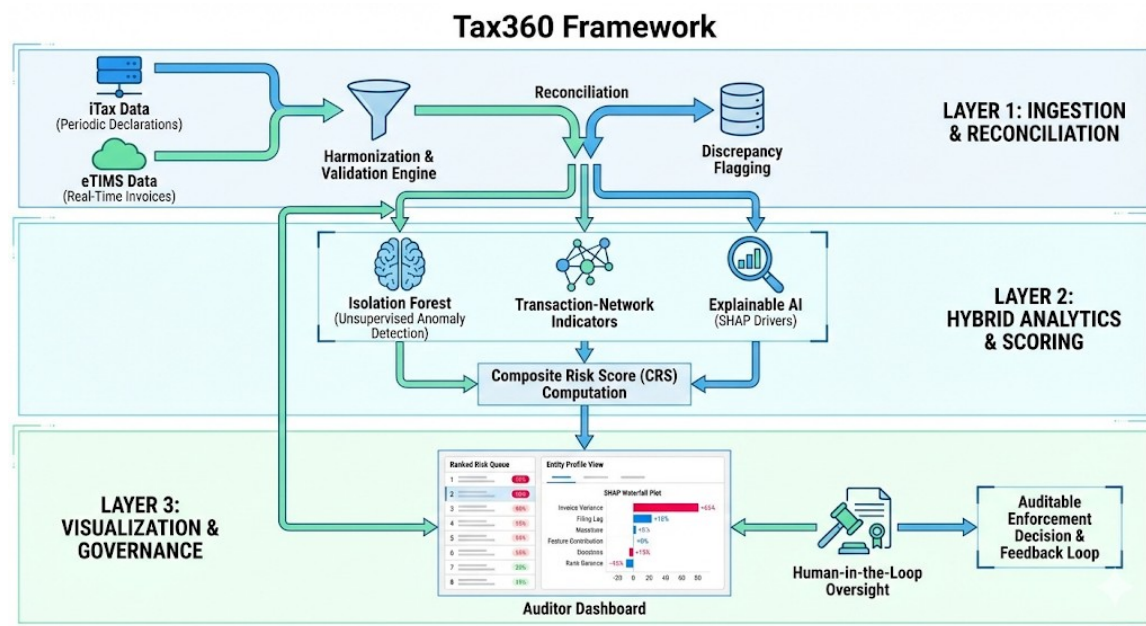


Figure 4: Tax360 Data Flows

The **Tax360 data flow** above represents an integrated pipeline that transforms raw, siloed tax data into actionable compliance intelligence through the following stages:

- **Data Ingestion and Harmonization:** The framework ingest parallel data streams from **iTax** (periodic summary-level declarations) and **eTIMS** (near-real-time invoice transactions). This layer harmonizes schemas and validates records to prepare them for reconciliation.
- **Reconciliation and Feature Engineering:** The system performs automated reconciliation to identify discrepancies—the "silo effect"—between point-of-sale invoices and self-declared returns. It computes critical features such as **invoice variance**, **filing lag**, and **network exposure** .
- **Hybrid Analytics and Risk Scoring:** Reconciled data is processed by a hybrid engine combining an **unsupervised anomaly detector** (Isolation Forest), a **deterministic fiscal rule layer**, and a **network risk component** . These inputs are fused into a **Composite Risk Score (CRS)** used to rank SME profiles .

Explainable Visualization and Workflow: The final intelligence is delivered to an **auditor dashboard**. This interface provides ranked queues and **SHAP-based local explanations**, allowing auditors to view the specific drivers of a risk flag (e.g., high invoice-return gaps) and choose appropriate interventions, such as discrepancy notifications or full audits

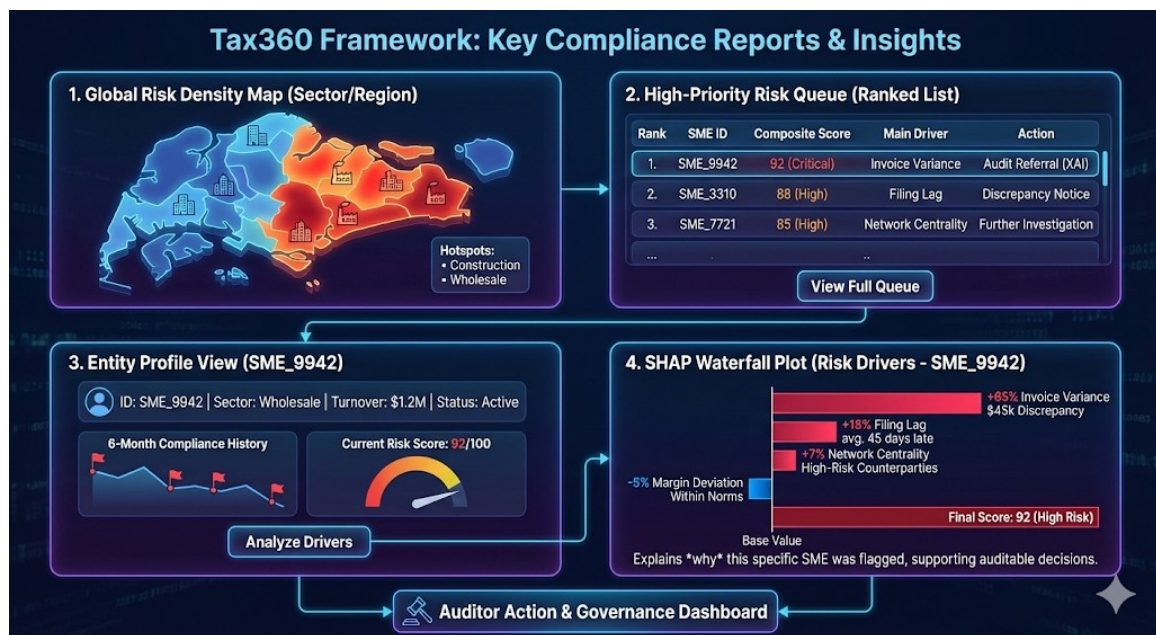


Figure 5: Tax360 reports and insights

The **Tax360** framework above generates several key reports and dashboard views to provide actionable insights for tax auditors and administrators:

1. Governance and Auditor Dashboard Reports

The primary interface for insight generation is the **Auditor Dashboard**, which includes the following components:

- **Global Risk Map:** Visualizes SME risk density across different economic sectors and geographic regions.
- **Entity Profile View:** Provides a comprehensive look at a specific SME, displaying its historical compliance indicators alongside its current **Composite Risk Score (CRS)**.
- **Ranked Risk Queues:** A prioritized list of SMEs ranked by their risk scores, allowing administrators to manage audit workloads efficiently.
- **Evidence Summaries:** Direct links to underlying transaction records (invoices and declarations) that support the identified anomalies.

2. Analytical and Explainability Insights

These reports break down the "why" behind a risk flag to ensure algorithmic accountability:

- **SHAP Waterfall Plots:** Generates local feature attributions that explain exactly how specific behaviors (e.g., invoice variance or filing lags) contributed to an entity's risk score.
- **Ranked Feature Contributions:** A list for each flagged SME showing which features had the most significant impact on its risk profile.
- **Network Centrality Reports:** Identifies relational risk by highlighting entities that serve as central nodes in suspicious transaction clusters or circular flows.

3. Operational and Intervention Outputs

The framework translates analytical insights into specific intervention reports:

- **Discrepancy Notifications:** Automated reports for mid-risk cases that detail variances and invite voluntary correction from the taxpayer.
- **Audit Referral Briefs:** Detailed dossiers for high-risk cases that include the risk score, evidence links, network context, and SHAP explanations for formal audit proceedings.

- **Audit Trails:** A comprehensive log of all human interactions with AI-generated flags, supporting internal governance and external defensibility of enforcement decisions.

4. Research Methodology - Materials and Methods

4.1 Research Design: Design Science Research (DSR)

The study follows the Design Science Research paradigm, which emphasizes constructing and evaluating artifacts that solve organizational problems and contribute design knowledge (Hevner et al., 2004). DSR is suitable for Tax360 because the objective is to create an actionable compliance analytics framework and evaluate its feasibility and performance rather than to test a single behavioral theory in isolation. The study progressed through iterative phases: problem diagnosis (identifying silo-driven reconciliation gaps and governance risks), design (specifying architecture, risk model, and explainability requirements), implementation (building the prototype services and dashboard), demonstration (running the system on synthetic data), and evaluation (benchmarking performance and assessing interpretability outputs). The artifact is reported with sufficient detail to allow replication and extension in future empirical deployments.

4.2 Data Engineering and Synthetic Dataset Construction

Access to real taxpayer datasets is often restricted due to legal, ethical, and political risks, and using such data in research can expose SMEs to harm. Synthetic data provides a privacy-preserving alternative when engineered carefully to reflect plausible distributions and relationships (De Souza & Ribeiro, 2021). Tax360 constructed a synthetic dataset comprising 10,000 SME profiles and 200,000 VAT invoice records. SME profiles included sector, turnover band, filing frequency, and historical compliance proxies (e.g., filing timeliness and frequency of Nil declarations). Invoice records included transaction timestamps, VAT values, and counterparty identifiers to support network analytics. Fraud patterns were injected using rules designed to mirror common schemes: under-declaration (invoices exceed returns), ghost invoicing (input VAT spikes without plausible output), and missing trader structures (clusters with unusual counterparty centrality and circular flows). Because injected fraud points are known, evaluation can compute precision, recall, and AUC-ROC with clear ground truth labels for anomaly cases.

4.3 Feature Engineering and Risk Indicators

Feature engineering aimed to create indicators that are simultaneously predictive and interpretable. Invoice variance measures the difference between invoice totals and declared sales, serving as the primary reconciliation signal. Filing lag captures timing anomalies and can reflect strategic delay behaviors. Input–output distortion reflects VAT claim anomalies relative to expected output VAT. Margin deviation flags implausible profitability trends compared to sector baselines. Network exposure and centrality capture relational risk, reflecting the likelihood that an entity is part of a coordinated structure (Alexopoulos et al., 2021; Wang et al., 2023). The feature set was designed to map to audit narratives: each feature can be traced to evidence in invoices or returns and described in business terms.

4.4 Hybrid Detection Model

The model combines three components: a machine-learning anomaly score (Isolation Forest), a deterministic fiscal rule layer, and a network risk component. Isolation Forest identifies observations that are easier to isolate through random partitioning, which corresponds to outlier-like behavior in the feature space (Liu et al., 2008). The fiscal rule layer encodes domain logic such as repeated Nil filing with high invoice volume or extreme invoice–return gaps, ensuring known regulatory violations are surfaced even if they are not extreme outliers. The network component computes normalized risk indicators from transaction graphs, capturing coordinated behaviors that may appear normal at the single-entity level but suspicious in relational structure (Chen et al., 2022). These components are combined into a Composite Risk Score (CRS), with weights tuned through grid search to balance enforcement objectives. The CRS is used to rank cases for potential audit or intervention.

$$\text{CRS} = w1 \cdot V_var + w2 \cdot I_dev + w3 \cdot ML_anom + w4 \cdot N_risk$$

4.5 Explainability, Workflow, and Human Oversight

To support accountability, Tax360 integrates SHAP explanations into the auditor dashboard. For each flagged SME, the system displays a risk score, a ranked list of feature contributions, and evidence links to underlying records. This enables auditors to validate and document decisions, supporting human-in-the-loop enforcement and reducing automation bias (Cath, 2018). Explainability is treated as a governance mechanism, not only a technical interpretation output: the dashboard logs auditor actions and rationales, supporting audit trails and internal review. Such mechanisms align with public-sector AI governance guidance and AI management system thinking where accountability, transparency, and monitoring are built into the system lifecycle (ISO, 2022; Misuraca & van Noordt, 2020).

4.6 Evaluation Design

Technical evaluation compared Tax360 with unsupervised baselines commonly used for anomaly detection: LOF and One-Class SVM. Metrics reported include precision, recall, F1-score, and AUC-ROC. The evaluation emphasized precision because false positives have outsized legitimacy and burden costs in SME contexts, and because high false positive rates can undermine the adoption of analytics by auditors who lose trust in the system (Muehlbacher & Kirchler, 2021). The evaluation also considered governance-oriented criteria: whether explanations were coherent, traceable to evidence, and usable in an audit workflow. Because the dataset is synthetic, the evaluation is a feasibility demonstration rather than a definitive claim of real-world effectiveness; future field validation is required.

4.7 Baseline Configuration, Tuning, and Practical Deployment Constraints

Baseline configuration matters in anomaly detection because parameter choices can significantly alter the false-positive rate. For LOF, we explored neighborhood sizes representing small, medium, and large local neighborhoods (k values tuned across a grid) and selected the configuration that maximized F1-score while constraining precision to a policy-relevant minimum. For One-Class SVM, we tuned kernel type and ν (an upper bound on training errors and a lower bound on support vectors), again prioritizing precision because false positives are costly in SME contexts. These tuning practices mirror operational realities: tax administrations rarely accept black-box tuning that produces high alert volumes without clear gains, because auditors have limited capacity and because high false-positive rates undermine institutional legitimacy and auditor trust in the tool (Muehlbacher & Kirchler, 2021; Janssen et al., 2020).

In addition to statistical tuning, Tax360 considered practical deployment constraints. First, data pipelines must cope with incomplete or delayed records. SMEs may upload invoices late, misclassify transactions, or experience connectivity issues. Second, sector heterogeneity implies that a single global threshold can be unfair. For example, seasonal businesses may exhibit high variance during peak periods without evasion intent. Third, audit capacity constraints require that scores be convertible into a manageable queue size. Tax360 therefore supports configurable risk thresholds per sector and turnover band, and it exposes ranking rather than binary decisions to allow administrators to manage workload while preserving transparency.

Finally, because public-sector analytics systems are subject to scrutiny, reproducibility and traceability were treated as design requirements. Every score is generated from versioned model parameters, logged feature values, and time-stamped ingestion records. This practice supports accountability and internal governance review and aligns with broader AI oversight expectations and management-system thinking (ISO, 2022; Misuraca & van Noordt, 2020).

4.8 Behavioral Evaluation Plan for Field Pilots

The current study evaluates technical feasibility using synthetic data. In an operational pilot, a complementary behavioral evaluation would be necessary to test the conceptual model's legitimacy and burden pathways. A practical design would combine (a) a stepped-wedge rollout across regions or sectors, (b) survey measures capturing perceived fairness, trust, and burden, and (c) administrative outcomes such as filing timeliness, variance reductions, and voluntary corrections. Such designs help isolate whether explainable discrepancy notifications produce behavior change beyond enforcement intensity. Intervention studies show engagement with curated content shapes behavioral response (Abueva et al., 2025). They also help detect unintended harms, such as SMEs disengaging from formal systems due to perceived surveillance or confusion, which would be counterproductive. This pilot logic is consistent with public-sector AI adoption research emphasizing the need to evaluate not only accuracy but also public value, legitimacy, and organizational fit (Sun & Medaglia, 2019; Wirtz & Müller, 2019).

5. System Architecture and Implementation

5.1 Architecture Overview

Tax360 is designed to function as an integration and analytics layer across iTax and eTIMS streams, consistent with the direction of Tax Administration 3.0 where standardized data exchange enables continuous compliance monitoring (OECD, 2021, 2023). The ingestion layer harmonizes schemas, validates records, and computes features at appropriate temporal granularities (e.g., invoice-level features and filing-period aggregates). The analytics layer executes the hybrid scoring engine, supports batch scoring and near-real-time updates, and maintains model monitoring statistics for drift and stability. The visualization layer provides an auditor-facing dashboard with ranked risk queues, case drill-down, SHAP explanations, and a decision audit trail. This layered design enables separation of concerns and supports secure access control boundaries between data ingestion, scoring logic, and presentation.

5.2 Implementation and Security Controls

The prototype was implemented using Python (FastAPI) for analytics services, PostgreSQL for data management, and React for the user interface. Security controls include RBAC, transport encryption, and logging. In production deployments, additional controls would include strict data minimization, configurable retention policies, segregation of duties, and periodic access reviews, reflecting data protection and governance requirements. Tax360’s governance checklist aligns with Kenya’s Data Protection Act obligations and general public-sector AI principles by emphasizing transparency, privacy, and human oversight. Because enforcement decisions can be contested, auditable logs and reproducible scoring are treated as core requirements rather than optional features.

5.3 Threat Modeling and Abuse-Resistance

Tax analytics platforms are attractive targets because they contain sensitive economic data and because adversaries may attempt to influence detection outcomes. Tax360 therefore anticipates multiple threat classes: (a) data exfiltration, (b) insider misuse, (c) model evasion by adversarial taxpayers, and (d) data poisoning through manipulated invoices. RBAC and audit logs mitigate insider threats by enforcing least privilege and enabling post-hoc accountability. Transport and storage encryption mitigate interception and exfiltration risks. To reduce evasion, the system supports feature-robust scoring that combines multiple signals (variance, lag, ratios, and network exposure), making it harder for adversaries to mask risk by manipulating a single indicator.

Model evasion is a realistic concern in any enforcement analytics system. If taxpayers learn the exact scoring logic, they may attempt to optimize behavior to avoid detection while continuing evasion. Tax360 addresses this by (1) separating public-facing guidance from full model internals, (2) using drift monitoring to detect shifts in feature distributions that could indicate adaptation, and (3) periodically re-tuning thresholds and rule parameters within a governed change process. Governance is central here: changes must be documented, reviewed, and tested to avoid inconsistent enforcement and to preserve administrative defensibility (Cath, 2018; ISO, 2022).

5.4 Operational Workflow: From Detection to Intervention

Tax360 is designed to support multiple intervention modes rather than a single ‘audit-or-nothing’ outcome. The default workflow is: (1) ingestion and reconciliation; (2) continuous scoring and queue ranking; (3) triage by auditors; (4) intervention selection; and (5) resolution and learning. For mid-risk cases, the intervention can be a discrepancy notification that invites voluntary correction, supported by evidence summaries and clear explanations of the variance. For high-risk cases, the intervention can be an audit referral with attached evidence, network context, and SHAP explanations. For repeated high-risk entities, escalation can include deeper network investigations that identify upstream and downstream counterparties.

This multi-mode workflow is important for SMEs because not all discrepancies reflect intent to evade. Some reflect recordkeeping errors, misinterpretation of VAT rules, or adoption friction with e-invoicing tools. By enabling graduated

interventions, Tax360 supports a cooperative compliance posture that can preserve tax morale while maintaining credible deterrence for deliberate evasion (Frey & Kirchler, 2022; Braithwaite, 2020).

6. Results

6.1 Detection Performance

On the synthetic evaluation dataset, the Tax360 composite model achieved Precision = 0.88, Recall = 0.82, and AUC-ROC = 0.90. These values indicate that most flagged cases are true anomaly cases (high precision) and that the model captures a substantial share of injected fraud patterns (high recall). Compared to baseline outlier detectors (LOF and One-Class SVM), the hybrid approach improved precision, supporting the goal of minimizing false positives. This is important because false positives have high legitimacy costs: auditing compliant SMEs can reduce tax morale and trigger resistance to future compliance initiatives.

6.2 Explainability and Dominant Drivers

SHAP analysis indicated that invoice variance dominated risk attribution (approximately 65% of contribution), followed by filing lag (18%), margin deviation (10%), and network centrality (7%). The pattern is consistent with the design logic: reconciliation between invoice capture and declared returns is the central compliance signal in a digitized VAT environment. Explainability outputs translated these signals into audit narratives, such as identifying periods where invoices exceed declared sales or where input VAT claims surge without plausible output. Such narratives support procedural justice because auditors can communicate the basis of scrutiny in concrete terms and SMEs can be guided toward remediation where discrepancies reflect capability rather than intent (Kass et al., 2023; Kibet, 2022).

6.3 Error Implications

Tax360’s precision implies lower false positive rates than typical unsupervised baselines, which is important for maintaining institutional legitimacy. False positives generate administrative overhead for auditors and impose compliance costs on businesses, potentially discouraging formalization and voluntary cooperation (Braithwaite, 2020). False negatives represent direct revenue leakage and deterrence failure. The precision–recall balance achieved by Tax360 indicates a pragmatic trade-off: it aims to protect legitimacy while maintaining sufficient deterrence through meaningful detection coverage. This balance supports the Slippery Slope perspective that enforcement power and trust must be jointly managed for sustainable compliance (Kirchler et al., 2008).

6.4 Ablation Perspective: Contribution of Each Component

To interpret why the composite model improves precision, an ablation perspective is useful. A machine-learning-only configuration (Isolation Forest without rules and without network metrics) tends to flag extreme outliers but can miss coordinated patterns that appear locally normal. A rules-only configuration captures known violations but is brittle and easier to game, and it can create high alert volumes when thresholds are conservative. A network-only configuration can highlight suspicious communities but may over-flag legitimate clusters (e.g., wholesalers with many counterparties). The composite approach reduces these weaknesses by triangulating. Variance and lag features anchor detection in reconciliation logic that is straightforward to explain. Isolation Forest captures multi-dimensional deviations that are difficult to enumerate as rules. Network metrics surface propagation risk and coordinated behaviors that would not appear anomalous at the single-entity level. This triangulation is also governance-supportive: when multiple independent signals point to risk, auditors can defend decisions more confidently, and taxpayers are more likely to perceive enforcement as evidence-based rather than arbitrary.

7. Discussion

7.1 Interpreting Performance in Low-Label Environments. Theoretical Contributions and Comparison with Prior Studies

The superior performance of the Tax360 framework provides significant theoretical and empirical contributions to the fields of digital government, RegTech, and computational tax enforcement, particularly within the context of emerging economies in the Global South. This study reinforces a growing body of evidence demonstrating the effectiveness of **unsupervised anomaly detection** in tax environments where labeled fraud data are scarce, biased, or strategically manipulated by taxpayers (Alexopoulos et al., 2021; Wang et al., 2023). By achieving an AUC of 0.90, Tax360 empirically validates **Isolation Forest** as a robust public-sector enforcement tool, especially when augmented with transaction-level and network-based metrics that capture how VAT fraud propagates across commercial relationships rather than remaining confined to firm-level irregularities (Alexopoulos et al., 2021; Wang et al., 2023).

A primary contribution of this research lies in the explicit contrast it draws between the **Tax360 “glass-box” approach** and the prevailing **“black-box” digital enforcement systems** currently deployed in several African jurisdictions. While digital reforms in countries such as Ethiopia and Rwanda have demonstrably increased reporting volumes through mandatory e-invoicing, their impact on actual fraud reduction has remained limited due to the absence of integrated, real-time analytical reconciliation (Mascagni & Mengistu, 2023). In such systems, audit selection is often driven by opaque, summary-level rules that lack granular transactional justification, thereby weakening taxpayer trust and procedural legitimacy. As Mascagni and Mengistu (2023) emphasize, **data capture alone does not translate into revenue gains unless accompanied by intelligent reconciliation and enforcement logic**. Tax360 directly addresses this “intelligence gap” by reconciling parallel digital streams—namely eTIMS transactional data and iTax declarations—into a single, auditable analytical framework.

Furthermore, this study advances the **Slippery Slope Framework** proposed by Kirchler et al. (2008) by empirically demonstrating a **synergistic relationship between enforcement power (deterrence) and institutional trust (legitimacy)** within a digitally mediated tax environment. Traditional tax compliance literature often conceptualizes power and trust as competing forces, where increased enforcement intensity risks undermining voluntary compliance (Kirchler et al., 2008; Muehlbacher & Kirchler, 2021). In contrast, Tax360 operationalizes a model of **“precise power”**, where high-accuracy, explainable enforcement enhances rather than erodes trust. In emerging digital economies, arbitrary or batch-based audits of SMEs are frequently perceived as coercive harassment, which weakens tax morale and encourages informality (Muehlbacher & Kirchler, 2021). By leveraging SHAP-based Explainable AI (XAI), Tax360 provides localized feature attributions that clearly articulate why a specific SME was flagged, transforming enforcement from an opaque exercise of authority into a transparent, evidence-based interaction.

This research also contributes to the **digital governance and public-sector AI literature** by positioning explainability not merely as a technical interpretability feature, but as a **core governance control** necessary for institutional accountability and legitimacy. As tax administrations transition toward the **Tax Administration 3.0** paradigm—where compliance processes are increasingly embedded within taxpayer systems—the obligation to provide auditable rationales for automated decisions becomes a foundational requirement for democratic governance (OECD, 2023). Tax360 demonstrates that pairing high-precision machine-learning risk scoring with fairness-enhancing transparency and automation-driven burden reduction yields stronger compliance intent than any individual pathway alone. This integrated behavioral-technological architecture offers a scalable, ethical, and governance-aligned blueprint for modernizing revenue authorities in emerging digital economies.

7.2 Deterrence, Trust, and Procedural Justice

Tax360 is designed to influence compliance through both deterrence and legitimacy. Improved detection increases perceived audit probability, consistent with economic deterrence theory (Allingham & Sandmo, 1972). Explainability and reduced false positives support legitimacy by making enforcement appear fair and evidence-based, consistent with legitimacy theory (Frey & Kirchler, 2022). A cross-domain empirical studies support notion that perceived system trustworthiness shapes users’ response (Khudaykulova et al., 2026). In practical terms, explainability enables auditors to justify why a case was selected and to differentiate between intentional evasion and capability-driven errors. This is crucial in emerging contexts where SMEs may struggle with adoption of digital invoicing tools (Kibet, 2022). By enabling targeted assistance and early remediation, the system can convert enforcement encounters into learning opportunities, strengthening voluntary compliance rather than relying only on coercion.

7.3 Governance and Accountability

AI-enabled audit selection must be governed to prevent misuse and to maintain contestability. Tax360 embeds governance through human-in-the-loop workflow, auditable logs, and explanation outputs that can be reviewed internally

and defended externally (Cath, 2018). The design aligns with AI management system thinking where risk management, monitoring, and accountability are built into the system lifecycle (ISO, 2022). Crawford’s (2021) critique of AI power asymmetries highlights the importance of limiting function creep and ensuring proportionality; Tax360’s checklist therefore emphasizes purpose limitation, access controls, and transparency for audit selection decisions. These controls are essential if tax administrations adopt more advanced generative AI capabilities for taxpayer communication or case summarization, as such capabilities introduce new risks related to hallucination and misrepresentation (OECD, 2024).

7.4 Policy and Operational Implications

Tax360 supports a shift from batch enforcement to risk-based fairness. By prioritizing high-risk profiles, administrations can better allocate scarce audit capacity and reduce harassment of compliant SMEs. By integrating reconciliation and discrepancy alerts, administrations can intervene earlier, preventing under-declaration from accumulating across periods. By providing explainable rationales, administrations can strengthen accountability and taxpayer trust. The approach complements ongoing modernization initiatives and could be integrated into broader compliance strategies such as taxpayer education, pre-populated returns, and targeted nudges for high-risk behavior correction. Where e-invoicing is adopted primarily for data capture, adding integrated analytics may be the missing link that converts visibility into revenue impact, as suggested by mixed evidence on e-invoicing outcomes (Mascagni & Mengistu, 2023).

8. Policy Implications, Limitations, and Future Research

8.1 Policy Implications

For policymakers and tax administrators, Tax360 illustrates how to operationalize Tax Administration 3.0 principles in an emerging economy context. Integrated reconciliation analytics can strengthen domestic resource mobilization while supporting fairness and reducing friction for SMEs. In Kenya, such a framework aligns with KRA’s digitization strategy and may help translate eTIMS data capture into targeted, defensible enforcement (Kenya Revenue Authority, 2024). Policy design should explicitly account for capability gaps: when anomalies reflect adoption challenges rather than intent, remediation support and education may be more effective than punitive approaches (Kibet, 2022).

8.2 Limitations

The study has three limitations. First, synthetic data cannot fully capture real-world behavioral adaptation, strategic responses to enforcement, or macroeconomic shocks. Second, the evaluation is cross-sectional and does not measure longitudinal compliance changes after interventions. Third, network analytics are constrained by data availability; deeper detection of carousel fraud requires multi-layer and temporal graph modeling and potentially cross-border data exchange. These limitations indicate that operational field validation is needed before policy adoption, including controlled pilots and governance impact assessments.

8.3 Future Research

Future research should validate Tax360 on operational datasets under strong privacy safeguards. Methodologically, future work can extend the network layer using temporal graph models, integrate semi-supervised learning where verified cases exist, and develop drift monitoring that detects when fraudsters adapt to detection signals. Behaviorally, longitudinal studies should test whether explainable discrepancy notifications shift SME compliance intent, whether burden-reducing automation changes filing behavior, and whether precision improvements measurably increase institutional trust. Finally, as tax administrations explore generative AI for communication and case summarization, research must examine governance controls to prevent hallucinated explanations and to ensure consistency with public-sector accountability requirements (OECD, 2024).

9. Conclusion

Tax administrations in emerging economies face persistent compliance challenges despite digitization. Tax360 demonstrates that combining reconciliation metrics, unsupervised anomaly detection, network indicators, and explainability can improve risk selection performance while strengthening governance and legitimacy. By reducing false positives and providing auditable rationales, the framework supports trust-preserving enforcement. By incorporating automation concepts, it also addresses compliance friction that commonly constrains SMEs. Tax360 therefore offers a scalable pathway toward intelligent, fair, and accountable tax administration in emerging digital economies.

Appendices

Appendix A: Dataset Schema

Table A1: SME Profile Schema: SME_PIN, annual turnover, sector, filing frequency, compliance status.
Table A2: VAT Invoice Schema: invoice ID, SME_PIN, transaction date, VAT value, counterparty PIN.
Table A3: Risk Log Schema: log ID, SME_PIN, anomaly score, rule violation, timestamp.

Appendix B: SHAP Dashboard Prototype Description

The dashboard provides a global risk map showing SME risk density by sector and region, an entity profile view that displays historical compliance indicators alongside the Composite Risk Score, local SHAP waterfall plots that explain feature contributions to the score, and an audit trail that logs human interactions with AI flags. This workflow supports human-in-the-loop governance and maintains traceability of decisions for internal review and external defensibility.

Appendix C: Governance Compliance Checklist

Privacy (Kenya Data Protection Act): anonymization of PII before analytics ingestion.
Ethics (public-sector AI principles): algorithmic transparency and auditable rationales.
Security (ISO/IEC 42001): RBAC and least-privilege controls.
Transparency (procedural justice): SHAP explanations for automated flags.
Integrity (data sovereignty): processing within national digital boundaries.
Governance (AI management system): regular audits for bias and model drift.

References

- Abueva, N., Wu, Y., Buzelo, A., Dzhanegizova, A., Karakushev, D., & Obrenovic, B. (2025). A study of the artistic techniques of the French writer Pascal Quignard in the context of developing emotional intelligence in high school students. *Frontiers in Psychology*, 16, 1617127.
- Alexopoulos, A., Karanikas, N., Lykourantzou, I., & Vergidis, K. (2021). A network and machine learning approach to detect VAT fraud. arXiv. <https://arxiv.org/abs/2109.08912>
- Allingham, M. G., & Sandmo, A. (1972). Income tax evasion: A theoretical analysis. *Journal of Public Economics*, 1(3–4), 323–338.
- Al-Omari, A., et al. (2023). RegTech in emerging economies: A review and research agenda. *Journal of Financial Regulation and Compliance*.
- Arner, D. W., et al. (2017). FinTech, RegTech, and the reconceptualization of financial regulation. *Northwestern Journal of International Law & Business*.
- Australian Taxation Office. (2023). E-invoicing for business productivity. <https://www.ato.gov.au/>
- Braithwaite, V. (2020). Hope, trust and taxes: Some public policy challenges. *Australian Journal of Social Issues*, 55(3), 303–320.
- Castro, P., Freitas, A., & Silva, M. (2022). Tax audit selection with explainable machine learning. *Government Information Quarterly*, 39(4), 101711.
- Cath, C. (2018). Governing artificial intelligence: Ethical, legal and technical opportunities and challenges. *Philosophical Transactions of the Royal Society A*.
- Chen, L., Zhang, Q., & Zhou, Y. (2022). Graph-based VAT fraud detection. *Information Systems Frontiers*, 24(6), 1543–1558.
- Crawford, K. (2021). *The atlas of AI: Power, politics, and the planetary costs of artificial intelligence*. Yale University Press.
- Cunningham, S., & Devereux, M. (2020). VAT digitalization and compliance behavior. *Fiscal Studies*, 41(3), 579–602.
- De Souza, M., & Ribeiro, A. (2021). Synthetic data for risk analytics: A privacy-preserving framework. *Journal of Big Data*, 8(1), 45.
- Dwivedi, Y. K., et al. (2021). Artificial intelligence (AI): Multidisciplinary perspectives on emerging challenges, opportunities, and agenda for research, practice and policy. *International Journal of Information Management*.
- European Commission. (2023). Report on VAT fraud trends in the EU. <https://ec.europa.eu/>
- Frey, S., & Kirchler, E. (2022). Explaining tax compliance through legitimacy. *Journal of Economic Behavior & Organization*, 202, 522–538.
- GAO. (2022). IRS modernization and AI-enabled audit analytics.
- Hevner, A. R., March, S. T., Park, J., & Ram, S. (2004). Design science in information systems research. *MIS Quarterly*, 28(1), 75–105.
- International Monetary Fund. (2023). Domestic resource mobilization report. <https://www.imf.org/>
- ISO. (2022). ISO/IEC 42001: Artificial intelligence management system standard. International Organization for Standardization.
- Janssen, M., et al. (2020). Data-driven decision-making in the public sector: Identifying hurdles and challenges. *Government Information Quarterly*.
- Kankanhalli, A., et al. (2019). AI and the public sector: Overcoming barriers to adoption. *Communications of the ACM*.
- Kariuki, M. (2023). SME tax compliance behavior in Kenya. *Journal of African Fiscal Policy*, 5(2), 45–60.

- Kass, A., Leemans, M., & Uten, S. (2023). AI governance for digital taxation: The role of explainability. *Journal of Information Policy*, 13, 221–240.
- Khan, R., & Ahmad, T. (2021). Machine learning for tax evasion detection: A systematic review. *ACM Computing Surveys*, 54(8), 1–30.
- Khudaykulova, M., He, Y., Obrenovic, B., Khudaykulov, A., & Abueva, N. (2026). The role of streamer trustworthiness and attractiveness in enhancing viewer engagement and mindfulness in live streaming. *Acta Psychologica*, 264, 106367.
- Kibet, S. (2022). Performance evaluation of eTIMS adoption in Kenya. *African Journal of Information Systems*, 14(1), 55–74.
- Kirchler, E. (2007). *The economic psychology of tax behaviour*. Cambridge University Press.
- Kirchler, E., et al. (2008). Enforced versus voluntary tax compliance: The “slippery slope” framework. *Journal of Economic Psychology*.
- Kenya Revenue Authority. (2024). eTIMS adoption guidance. <https://www.kra.go.ke/>
- Li, Q., Chen, J., & Sun, L. (2023). Explainable anomaly detection for tax administration. *Government Information Quarterly*, 40(4), 101–125.
- Liu, F. T., Ting, K. M., & Zhou, Z. (2008). Isolation forest. In 2008 IEEE International Conference on Data Mining (pp. 413–422). IEEE.
- Loo, E. C. (2006). *The influence of the self-assessment system on tax compliance* (Doctoral dissertation, University of Sydney).
- Mascagni, G., & Mengistu, A. T. (2023). The impact of e-invoicing on tax compliance: Evidence from Ethiopia. *Journal of Development Economics*, 164, 104073.
- Maxamadumarovich, U. A., Obrenovic, B., & Amonboyev, M. (2012). Understanding the innovation concept. *Journal on Innovation and Sustainability RISUS*, 3(3), 19-26.
- Medaglia, R., & Zhu, D. (2017). Public value creation through AI in the government sector. *Government Information Quarterly*.
- Misuraca, G., & van Noordt, C. (2020). *AI watch: AI in public services* (JRC Technical Report). Joint Research Centre.
- Muehlbacher, S., & Kirchler, E. (2021). Behavioral tax compliance. *Journal of Economic Psychology*, 87, 102423.
- OECD. (2021). *Implementing Tax Administration 3.0: Data availability and exchange*. OECD Publishing.
- OECD. (2023). *Tax Administration 3.0: The digital transformation of tax administration*. OECD Publishing.
- OECD. (2024). *Tax administration in the age of generative AI*. OECD Publishing.
- Park, J., & Lee, H. (2021). Fairness and trust in digital taxation. *Information Systems Frontiers*, 23(5), 1241–1256.
- Ribeiro, P., & Lopes, S. (2022). SME burden reduction through automation. *Electronic Commerce Research*, 22, 993–1012.
- Singh, A., & Patel, S. (2022). AI-based revenue assurance in public finance. *Digital Government: Research and Practice*, 3(1), 11–21.
- Slemrod, J. (2019). Tax compliance and enforcement. *Journal of Economic Literature*.
- Sun, T. Q., & Medaglia, R. (2019). Mapping the challenges of artificial intelligence in the public sector. *Government Information Quarterly*.
- Torgler, B. (2007). *Tax compliance and tax morale: A theoretical and empirical analysis*. Edward Elgar.
- Tornatzky, L. G., & Fleischer, M. (1990). *The processes of technological innovation*. Lexington Books.
- Wang, D., Yin, Q., & Yang, S. (2023). VAT fraud anomaly propagation analysis using graph neural networks. *Information Systems Frontiers*, 25(1), 139–153.
- Wirtz, B. W., & Müller, W. M. (2019). An integrated artificial intelligence framework for public management. *Public Management Review*.
- World Bank. (2022). *Technology for tax administration in the Global South*. World Bank Group.
- Zuiderwijk, A., Vhen, Y., & Salem, F. (2021). Implications of the use of artificial intelligence in public governance: A systematic literature review and a research agenda. *Government Information Quarterly*. <https://doi.org/10.1016/j.giq.2021.101577>