© MAR 2026 | IRE Journals | Volume 9 Issue 9 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV9I9-1715244

AI for Tax and Customs Anomaly Detection: A PRISMA-Guided Systematic Review and Proposed Hybrid AI Framework for Trade and Revenue Administration in Emerging Digital Economies

OBAID UR REHMAN QURESHI
Kafaat business solutions

Abstract- Tax and customs administrations in emerging digital economies face simultaneous pressure to accelerate trade clearance, protect revenue, and maintain procedural fairness under rapidly digitising commercial activity. Traditional rule-based targeting, while operationally important, is increasingly outpaced by cross-border e-commerce growth, valuation manipulation, synthetic invoicing, shell entity proliferation, and adaptive fraud behaviour. This review synthesises peer-reviewed literature and high-value institutional publications from 2020 to early 2026 on artificial intelligence (AI), machine learning (ML), graph analytics, anomaly detection, explainable AI (XAI), and human-in-the-loop (HITL) risk management applied to tax and customs declarations. Following PRISMA 2020 reporting principles, 38 principal sources were selected from 518 initially identified records across Scopus, Web of Science, IEEE Xplore, ACM Digital Library, SSRN, Google Scholar, and institutional repositories of the OECD, WCO, APEC, CIAT, UNDP, and the International Growth Centre. The synthesis reveals three structural findings. First, high-performing anomaly detection systems consistently employ hybrid architectures that integrate supervised fraud scoring, unsupervised or semi-supervised anomaly detection, graph-based relational modelling, and active learning — rather than any single model family. Second, performance claims are highly sensitive to label scarcity, inspection bias, concept drift, and operational constraints including audit budgets, explainability mandates, and workflow integration requirements. Third, governance, legal explainability, human oversight, and continuous model monitoring are necessary design requirements — not post-implementation considerations — for public-sector deployment. Based on this evidence, the paper proposes a seven-layer AI-driven anomaly detection framework specifically designed for revenue administrations in emerging digital economies. The framework integrates data identity and interoperability foundations, risk-feature engineering, hybrid model stacking, explainable risk presentation, human-in-the-loop case management, continuous learning and drift control, and institutional governance and recourse mechanisms. The framework is directly applicable to ZATCA (Zakat, Tax and Customs Authority) in Saudi Arabia, GCC customs administrations, and comparable digital revenue administrations across emerging economies. The review concludes that technically credible AI-enabled anomaly detection is now achievable, but that scaled and sustainable deployment depends less on model novelty than on data quality, interoperable digital infrastructure, inspection feedback loops, legal explainability standards, and governance maturity.

Index Terms- Customs Fraud Detection; Tax Anomaly Detection; AI In Government; PRISMA Systematic Review; Emerging Digital Economies; Explainable AI; Graph Neural Networks; VAT Fraud; ZATCA; Vision 2030; Hybrid Detection Architecture; Human-In-The-Loop; Concept Drift; Active Learning.

I. INTRODUCTION

Digital transformation is fundamentally reshaping the operating model of tax and customs administrations worldwide. Where compliance monitoring once depended on paper declarations, manual audit selection, and periodic batch reviews, it now operates across real-time data streams, e-invoicing networks, electronic trade documents, and interconnected entity registries. This structural shift has elevated anomaly detection — the capacity to identify irregular, suspicious, or non-compliant behaviour within large and heterogeneous administrative datasets — from a specialist analytics function to a core strategic capability for protecting public revenue [OECD-2020, OECD-2025a, OECD-2025b].

The operational challenge this creates is substantial. Customs risk engines must distinguish legitimate trade from undervaluation, tariff misclassification, concealment, and the rapidly expanding surface area of cross-border e-commerce parcel flows. Tax administrations must identify anomalies across VAT declarations, e-invoice networks, corporate registries, payroll records, and third-party data — while simultaneously preserving due process, service quality, and legal taxpayer rights. The central difficulty is not simply whether AI can classify previously observed fraud. It is whether administrations can surface unknown or evolving non-compliance in environments defined by sparse confirmed labels, severe class imbalance, fragmented legacy data systems, and adversarially adaptive fraud behaviour.

Customs declarations are particularly exposed to this challenge because only a small fraction of shipments can be physically inspected, and critical risk attributes are relational — embedded in networks of importers, intermediaries, routing jurisdictions, and declaration histories — rather than contained within single transaction records [Kim-2020, Singh-2023]. This structural reality has driven measurable research momentum toward dual-task learning, active learning, graph neural networks (GNNs), and revenue-aware ranking models as more operationally appropriate alternatives to standard binary classifiers.

Tax administration presents a structurally parallel problem. Studies on VAT and income-tax compliance consistently show that supervised classification alone is insufficient when observed fraud cases are non-representative of the true population of evasion — a condition that is endemic to administratively-generated labels. Anomaly detection methods, relational network approaches, hybrid learning architectures, and interpretable case-ranking systems are accordingly gaining prominence across the tax-compliance research literature [Vanhoeyveld-2020, Murorunkwere-2022, Zheng-2024, Alexopoulos-2025, Belahouaoui-2025]. Institutional evidence confirms that AI is already deployed across a substantial share of tax administrations for fraud detection and risk assessment, but that implementation quality varies significantly with legal frameworks, data architecture, and organisational maturity [OECD-2025a, OECD-2025d].

The convergence of technical capability and institutional demand is especially consequential in the Gulf Cooperation Council (GCC) region. Saudi Arabia's Zakat, Tax and Customs Authority (ZATCA) has deployed mandatory e-invoicing through the Fatoorah platform in progressive phases, established the FASAH single-window customs system, and operates within Vision 2030's strategic mandate to diversify non-oil revenue, contract the shadow economy, and construct world-class digital public infrastructure [OECD-2025b]. These developments collectively generate both the data foundations and the governance obligation for AI-driven anomaly detection at national scale — yet the academic literature has not produced a thorough, PRISMA-structured synthesis that bridges the technical AI research with the specific institutional requirements of GCC and comparable emerging-economy revenue authorities.

Despite substantial growth in relevant publication volume, the literature remains structurally fragmented. Technical papers frequently optimise for predictive performance metrics without engaging institutional constraints, while policy reports describe digitalisation pathways without specifying how anomaly detection systems should be architecturally designed, operationally validated, or democratically governed. A systematic review integrating these strands is therefore both timely and necessary.

1.1 Research Objectives

This review is structured around five original research objectives:

- Map the main AI and anomaly-detection approaches applied to tax and customs declarations in literature published from 2020 to early 2026, and synthesise their comparative performance characteristics.
- Compare how recent studies address the core operational constraints of revenue administrations, specifically class imbalance, label scarcity, inspection budget limitations, and concept drift.

- Identify the most policy-relevant design requirements for deploying AI-based anomaly detection in emerging digital economies, including interoperability standards, explainability mandates, governance structures, and human oversight protocols.
- Synthesise evidence on the relative strengths and limitations of supervised, unsupervised, semi-supervised, graph-based, and hybrid detection architectures across customs and tax application domains.
- Propose a practical, seven-layer implementation framework for AI-driven anomaly detection in trade and customs declarations that aligns technical performance requirements with administrative legitimacy, legal accountability, and Vision 2030 digital governance objectives.

1.2 Novelty and Contribution

This review makes three distinct contributions to the literature. First, it provides the first PRISMA 2020-compliant systematic synthesis that simultaneously addresses customs and tax anomaly detection, hybrid AI architectures, and emerging-economy institutional constraints within a single review. Second, it proposes a seven-layer implementation framework that explicitly incorporates GCC regulatory context — including ZATCA Fatoorah, FASAH single-window integration, and Vision 2030 digital governance requirements — rather than treating GCC as an afterthought to Western regulatory settings. Third, it introduces three assurance-oriented performance metrics — detection yield efficiency, concept drift resilience, and explainability compliance rate — specifically designed for public-sector AI evaluation contexts where offline accuracy metrics are insufficient.

II. REVIEW METHODOLOGY

2.1 Design Review and PRISMA 2020 Compliance

This paper adopts a PRISMA 2020-guided structured review design with thematic synthesis, rather than a statistical meta-analysis, because the underlying literature spans heterogeneous data types, outcome metrics, and administrative use cases that preclude meaningful quantitative aggregation [Page-2021a, Page-2021b]. The reporting structure follows PRISMA 2020 principles for search transparency, inclusion criteria specification, and synthesis structure. Database searches were executed between January 2020 and March 2026, drawing on six principal academic databases: Scopus, Web of Science, IEEE Xplore, the ACM Digital Library, SSRN, and Google Scholar. Institutional publications from multilateral and intergovernmental bodies — including the OECD, WCO, APEC, CIAT, UNDP, and the International Growth Centre — were incorporated where they provided implementation guidance, governance analysis, or digital infrastructure evidence directly relevant to revenue administration anomaly detection.

2.2 Search Strategy

The search strategy combined Boolean operators across three thematic clusters: (i) domain terms: "customs fraud," "trade declarations," "tax anomaly detection," "VAT fraud," "customs risk management," "tax compliance risk," "revenue administration AI"; (ii) methodological terms: "machine learning," "anomaly detection," "graph neural network," "active learning," "explainable AI," "semi-supervised learning," "concept drift," "hybrid detection"; and (iii) institutional terms: "digital tax administration," "e-invoicing," "single window," "ZATCA," "GCC customs," "emerging digital economy," "Tax Administration 3.0." Searches were conducted in English. Non-English sources were excluded to maintain review consistency. Citation tracking was applied to all included studies to identify additional relevant sources not captured in database searches.

2.3 Inclusion and Exclusion Criteria

Inclusion required each source to satisfy all four of the following criteria:

- Domain relevance: the source discusses tax, customs, trade declarations, or public-sector anomaly or risk detection as its primary or substantial secondary focus.
- Substantive contribution: the publication provides empirical evidence, methodological innovation, systematic synthesis, or institutional implementation guidance with specific technical or governance content.

- Publication window: the publication date falls between January 2020 and March 2026.
- Thematic alignment: the source contributes to one or more of the five review themes — detection architecture, data infrastructure, human oversight, explainability, or governance.

Exclusion criteria eliminated: purely generic fraud-detection studies with no tax, customs, or public-sector relevance; pre-2020 sources except foundational methodological references; conference abstracts and opinion editorials without empirical or methodological substance; and vendor white papers or commercial case studies without independently verifiable claims.

2.4 PRISMA 2020 Article Selection

Table 1 presents the full PRISMA 2020-compliant article selection summary. The final synthesis corpus of 38 principal sources — comprising 21 peer-reviewed journal articles and conference papers and 17 institutional reports — was selected from 518 database records and supplementary sources initially identified.

PRISMA 2020 Stage	Process Description	N (Records)
Records identified via database searches	Scopus, Web of Science, IEEE Xplore, ACM Digital Library, SSRN, Google Scholar	487
Records identified via supplementary methods	Institutional sources: OECD, WCO, APEC, CIAT, UNDP, IGC; citation tracking	31
Records after duplicate removal	Automated deduplication followed by manual verification	391
Records screened (title and abstract)	Inclusion/exclusion criteria applied against all five review themes	391
Records excluded at screening stage	Off-topic; no tax/customs/public-sector relevance; pre-2020; opinion pieces without empirical or methodological content	309
Full-text articles assessed for eligibility	Full-text content assessment against all five inclusion criteria	82
Full-text articles excluded (with reasons)	Insufficient technical detail (28); no direct tax/customs relevance (14); institutional report without anomaly-detection specificity (2)	44
Studies included in qualitative synthesis	Final corpus: 21 peer-reviewed studies + 17 institutional reports	38

Table 1. PRISMA 2020 article selection summary. Database and institutional searches conducted March 2026.

2.5 Data Extraction and Synthesis Protocol

Data were extracted across seven standardised dimensions: (a) domain and administrative context; (b) AI method family and specific techniques; (c) dataset characteristics and label availability; (d) evaluation metrics and reported performance; (e) explainability and traceability mechanisms; (f) human oversight and governance provisions; and (g) applicability to emerging digital economy contexts. Given significant methodological heterogeneity across included studies — encompassing different datasets, fraud definitions, evaluation protocols, and administrative environments — direct quantitative meta-analysis was not feasible. A structured thematic synthesis was therefore applied, consistent with established practice for systematic reviews in applied AI and public administration [Page-2021a, Nassif-2021].

III. RESULTS: MAJOR PATTERNS IN LITERATURE

Three structural patterns recur strongly across the reviewed literature. First, anomaly detection for tax and customs is becoming more hybrid: older rule-centric approaches are being complemented not replaced by supervised classification, semi-supervised learning, graph analytics, and anomaly ranking methods. Second, recent work increasingly treats inspection capacity and audit yield as explicit optimisation constraints rather than background assumptions. Third, governance issues explainability, contestability, data quality, and model monitoring are now integral to system design, not external afterthoughts [Kim-2023, OECD-2025d, Kuzniacki-2022, Ruvalcaba-2025].

The review also reveals a productive divergence between customs and tax analytics. Customs studies concentrate on transaction-level declarations with downstream physical inspection, making active learning and revenue-aware ranking particularly valuable. Tax studies more frequently exploit broader administrative networks — invoices, entities, counterparties, filing histories where graph and network methods are especially effective. Yet the two domains are converging around a common set of technical requirements: weak-label learning, adaptive monitoring, explainability, and workflow integration. Table 2 presents a comparative taxonomy of AI method families across both domains.

AI Method Family	Representative Techniques	Primary Use Case	Key Limitation
Supervised classification	Gradient boosting, XGBoost, neural networks, random forest	Known fraud patterns with labelled inspection outcomes	High precision on known patterns; degrades under class imbalance and label bias; requires representative training data
Unsupervised anomaly detection	Isolation Forest, autoencoders, one-class SVM, LOF	Novel or unknown non-compliance; cases without inspection labels	Detects emerging schemes; no labelled fraud required; higher false-positive burden
Semi-supervised learning	Graph-based propagation, self-training, pseudo-labelling	Mixed-label environments; administrative datasets with sparse confirmed cases	Exploits both labelled and unlabelled data; reduces label dependency
Graph-based detection	Graph Neural Networks (GNN), network centrality, link prediction	Relational fraud (VAT carousel, shell networks, importer rings)	Captures relational structure invisible to transactional models; requires entity linkage infrastructure
Active learning	Uncertainty sampling, query-by-committee, exploration-exploitation	Adaptive targeting where inspection feedback drives model improvement	Improves relevance under concept drift; balances exploration vs. exploitation
Hybrid architectures	Stacked ensembles, multi-objective optimisation, dual-task models	Comprehensive revenue protection combining known and novel risk	Most evidence-supported approach; operationally complex; requires orchestration layer

Table 2. Comparative taxonomy of AI method families for tax and customs anomaly detection. Source: author synthesis from included studies.

3.1 Customs Anomaly Detection: From Transaction Scoring to Adaptive Targeting

The modern customs literature originates from the recognition that customs fraud is not only a classification task but a selective inspection problem under severe resource constraints. DATE, proposed by Kim et al. [Kim-2020], was influential because it framed customs targeting as a dual task: identifying illicit declarations while simultaneously ranking them by expected revenue impact. This shifted the objective from accuracy alone to risk-adjusted revenue prioritisation. Subsequent work extended this operational framing: active learning for human-in-the-loop customs inspection demonstrated that exclusive exploitation of known suspicious patterns degrades performance over time by narrowing the labelled sample and reinforcing confirmation bias — some exploration is necessary to learn new fraud patterns under concept drift [Kim-2023].

More recent customs research advances toward relationship-aware models. GraphFC treats customs detection as a semi-supervised graph problem and demonstrates that label scarcity can be partially overcome by exploiting relational structure among declarations, importers, products, and routing intermediaries [Singh-2023]. In practice, this reflects real administrative systems in which risk is embedded in chains of counterparties, repeated valuation behaviour, and recurrent routing patterns. The APEC report on data analytics in customs risk management similarly identifies graph-based anomaly detection as a priority operational direction for detecting abnormal importer networks [APEC-2025].

Input data diversity in customs analytics is also growing. Dangsawang and Nuchitprasitchai [Dangsawang-2024] demonstrate that customs-relevant fraud signals extend beyond structured declarations into unstructured online content, using social-media analysis to detect unauthorised commercial activity linked to tax and customs leakage. For GCC administrations, this multimodal approach is particularly relevant given the rapid growth of e-commerce platforms operating across formal and informal channels. Alwanin et al. [Alwanin-2025] — working in a GCC customs context — demonstrate that joint classification-regression gradient boosting approaches produce superior operational prioritisation compared to binary classification alone, because customs officers require not only a suspicion score but a rational basis for allocating scarce inspections to highest-revenue-risk cases.

Institutional literature reinforces these technical developments. The WCO-WTO report on disruptive technologies identifies anomaly detection, intelligent audits, risk-based targeting, and HS classification support as priority customs use cases for AI [WCO-WTO-2022]. The WCO e-commerce compendium and the World Customs Journal review by Vijayakumar [Vijayakumar-2025] further document that customs administrations increasingly conceptualise analytics as part of a broader risk-management architecture spanning pre-arrival information, parcel flows, document digitalisation, supply-chain trust, and post-clearance audit. The emerging literature on customs risk management by Karklina-Admine et al. [Karklina-2024] provides a systematic perspective on institutional challenges that technical solutions must navigate. For ZATCA and GCC customs administrations, FASAH single-window data and Fatoorah e-invoice networks provide precisely the interoperable data infrastructure that these advanced approaches require.

3.2 Tax Anomaly Detection: VAT, Income Tax, and Compliance Risk Analytics

In the tax literature, one of the clearest findings is that anomaly detection becomes most valuable when labelled fraud cases are scarce, inspection bias is strong, and illicit behaviour is networked. Vanhoeyveld et al. [Vanhoeyveld-2020] provide a foundational example in VAT fraud, demonstrating that scalable anomaly detection outperforms purely supervised approaches when training labels are sparse and non-representative — a recurring issue in administrative datasets where observed fraud cases are conditioned on prior audit selection, meaning label distributions reflect earlier models and institutional choices rather than the true population of evasion.

Recent VAT work deepens the network orientation. Alexopoulos et al. [Alexopoulos-2025] demonstrate that VAT fraud is detected more effectively when transactional structures are analysed as networks, with network-derived features materially improving AUC over non-network baselines. This is especially relevant to carousel fraud, shell-company rotation, and organised non-compliance patterns that are invisible from single-return anomalies. For emerging digital economies with e-invoicing infrastructure — such as Saudi Arabia's Phase 2 Fatoorah integration requiring real-time e-invoice submission to ZATCA — this finding implies that invoice-network graphs can become strategic fraud-detection assets if the data architecture is designed to support network reconstruction.

Income-tax studies point in the same direction. Mururunkwere et al. [Murorunkwere-2022, Murorunkwere-2023] demonstrate that neural networks can effectively detect income-tax fraud, highlighting the importance of careful feature selection, class-handling strategies, and domain-specific feature engineering over generic data-mining approaches. Alsadhan [Alsadhan-2023] proposes a multi-module ML architecture for tax fraud detection, suggesting that staged or stacked architectures capture complementary patterns better than monolithic models. The wide-ranging survey by Zheng et al. [Zheng-2024] positions these developments within a broader taxonomy, documenting a clear shift from descriptive and rule-based methods toward machine learning, knowledge graphs, and integrated digital audit tools — while also noting that institutional deployment consistently lags behind methodological experimentation. Belahouaoui and Alm [Belahouaoui-2025] synthesise AI-based tax fraud detection under an Adaptive AI Tax Oversight framing, confirming that implementation involves simultaneous technical, organisational, ethical, and financial dimensions.

3.3 Hybrid Architectures: The Dominant Evidence-Supported Approach

Across both customs and tax domains, the most credible and operationally effective studies do not advocate a single universal algorithm. Instead, they consistently support hybrid architectures. Supervised learning remains valuable for recurring and labelled fraud patterns; unsupervised and semi-supervised anomaly detection is necessary for novel or unlabelled behaviour; graph models add detection power where entities and transactions form relational networks; and active learning continuously improves model relevance when manual verification generates the high-value labels that matter most for operational performance [Kim-2023, Singh-2023, Vanhoeyveld-2020, Alexopoulos-2025].

This pattern is consistent with the wider anomaly-detection literature. Nassif et al. [Nassif-2021] demonstrate in their systematic review that ML-based anomaly detection spans supervised, semi-supervised, unsupervised, and hybrid approaches, each with context-specific trade-offs. Li et al. [Li-2023] argue that anomaly detection systems in high-stakes settings require explainability by design rather than as a post-hoc addition. For public revenue administration, the implication is decisive: the most effective anomaly detector is not the one with the strongest offline accuracy metrics — it is the one that supports operational review, withstands audit defensibility requirements, and enables systematic model correction. Table 3 presents representative studies illustrating the range of approaches and their key contributions to this synthesis.

Study	Dataset / Context	Method	Key Performance Finding	Contribution to This Review
Kim et al. (2020) [DATE model]	WCO-linked customs declarations	Dual-task supervised learning + attention	Revenue-ranked fraud prioritisation; AUC > 0.90	Framed customs targeting as joint fraud detection + revenue-at-risk estimation; foundational architecture
Kim et al. (2023) [Active learning]	Customs inspection feedback data	Active learning with exploration–exploitation balance	Improved long-run precision under concept drift	Demonstrated that pure exploitation narrows feedback loop; exploration needed for model currency
Singh et al. (2023) [GraphFC]	Customs declaration networks	Semi-supervised graph neural network	Improved F1 under label scarcity vs. supervised baselines	Graph structure compensates for sparse confirmed fraud labels in customs administrations
Vanhoeyveld et al. (2020)	Belgian VAT declarations	Scalable unsupervised anomaly detection	Outperforms supervised baselines when labels sparse	Foundational evidence that anomaly detection is necessary when label distribution reflects audit bias
Alexopoulos et al. (2025)	Greek VAT transaction networks	Network-based fraud detection with graph features	Material AUC improvement over non-network baselines	Network features capture carousel fraud patterns invisible to single-return classification
Alwanin et al. (2025/2026)	Customs declarations (GCC context)	Gradient boosting for joint classification + revenue estimation	Operational prioritisation superior to binary classification alone	Revenue-at-risk estimation directly addresses customs resource allocation problem
Belahouaoui & Alm (2025)	Multi-jurisdiction survey	Systematic review of AI tax fraud detection	Confirms hybrid + governance requirement	Frames implementation as technical, organisational, ethical, and financial challenge

Malashin et al. (2025)	Tax audit dataset	Multi-objective optimisation for audit selection	Reduced unnecessary audits while maintaining fraud yield	Positions audit minimisation as co-objective alongside detection maximisation

Table 3. Representative 2020–2026 studies informing AI-driven tax and customs anomaly detection. Source: author synthesis from included studies.

3.4 Label Scarcity, Concept Drift, and Inspection Bias

Three technical obstacles dominate the literature and define the operational environment for AI deployment in revenue administration: label scarcity, concept drift, and inspection bias. Label scarcity arises because only a small subset of declarations or tax returns is ever manually verified, leaving the vast majority of the data without confirmed outcome labels. Inspection bias arises because verified cases are not randomly selected — they reflect earlier targeting heuristics, risk profiling decisions, and policy priorities, meaning the label distribution is itself a product of prior institutional choices rather than a representative sample of actual non-compliance. Concept drift arises because the distribution of trade and tax behaviour changes over time — sometimes abruptly — as fraud actors adapt to detection systems or as commerce shifts across products, platforms, and jurisdictions.

Customs research has addressed these obstacles most directly. Kim et al. [Kim-2023] demonstrate that exploration in active learning improves long-run performance relative to pure exploitation, precisely because exclusive focus on the most suspicious known items narrows the feedback loop and erodes model currency. GraphFC addresses weak labels by exploiting relational graph structure [Singh-2023]. Broader reviews of concept drift by Arora et al. [Arora-2024] and Hovakimyan and Bravo [Hovakimyan-2024] confirm that non-stationary data environments require explicit detection and adaptation strategies, particularly where class imbalance and evolving fraud patterns are pronounced. In practical revenue administration settings, this implies that annual or ad hoc model retraining is insufficient — administrations require routine monitoring of feature drift, score drift, and realised inspection yield.

3.5 Explainability, Human Oversight, and Legal Legitimacy

The reviewed literature strongly indicates that explainability is not optional in tax and customs anomaly detection. Unlike commercial applications, public-sector anomaly scores can trigger inspections, audits, payment holds, or enforcement actions that materially affect businesses and citizens. OECD reporting explicitly states that black-box AI in tax administration can undermine transparency, contestability, and taxpayer rights if officials cannot adequately explain why a case was selected or how a decision was reached [OECD-2025d]. CIAT [CIAT-2021] and legal scholarship make the same point from a rights perspective: if administrations deploy AI, they must preserve the capacity to explain contributing factors, enable challenge of errors, and document decision pathways [Kuzniacki-2022]. This requirement is directly relevant to GCC administrations under both ZATCA's internal administrative governance framework and Saudi Arabia's broader public AI governance obligations under Vision 2030 digital government policy.

This does not require that every model be intrinsically simple. It requires that anomaly-detection systems be embedded within explainable workflows. Case selection can rely on complex ensembles or graph methods, but investigators must receive interpretable risk reasons: valuation deviation from peer benchmarks, invoice-network irregularity, improbable commodity-country pairings, entity-link concentration patterns, or mismatch with prior filing history. Explainability also supports model debugging — enabling analysts to distinguish true fraud signals from spurious correlations or data integration artefacts. The explainability requirement further strengthens the operational case for human-in-the-loop designs: human review is not merely a legal safeguard but a learning mechanism, converting inspection outcomes into the high-value labels that maintain model currency [Kim-2023].

3.6 Digital Public Infrastructure and the Emerging-Economy Context

Emerging digital economies face a double asymmetry in AI deployment for revenue administration. They often have strong incentives to digitalise quickly — because compliance gaps are large and staff resources constrained — yet may lack the interoperable data systems, governance capacity, standardised legal identifiers, and institutional safeguards required for high-quality, sustainable AI deployment. The OECD Tax Administration 3.0 literature consistently argues that digital identity systems, e-invoicing mandates, interoperable data exchange mechanisms, and embedded tax logic are foundational prerequisites for modern tax administration, not peripheral enhancements [OECD-2020, OECD-2022a, OECD-2022b, OECD-2025c].

Saudi Arabia's ZATCA represents a leading case study in this regard. The Fatoorah e-invoicing platform — Phase 1 (invoice generation) implemented from December 2021 and Phase 2 (real-time API integration with ZATCA) deployed in rolling phases from January 2023 — provides precisely the structured, standardised, and continuously updated transactional data that enables effective VAT network anomaly detection [ZATCA-2023, OECD-2022b]. The FASAH single-window customs platform similarly provides interoperable pre-arrival and clearance data that supports the multi-source risk feature engineering central to advanced customs targeting. The broader Vision 2030 digital government agenda — including the Saudi National Data Management Office (NDMO) data governance framework and the National Programme for Combating Commercial Concealment — creates the institutional architecture within which AI-driven anomaly detection systems must operate. These GCC developments position ZATCA and comparable authorities as practical implementation contexts for the framework proposed in Section 4.

For other emerging digital economies, the central message from the literature is one of sequencing. Mittra [Mittra-2026] and UNDP [UNDP-2025] both argue that AI can substantially strengthen public revenue services through real-time analytics and pattern recognition, but only when states can manage digital records, organisational workflows, and fairness risks systematically. Ruvalcaba-Gómez and García-Benitez [Ruvalcaba-2025] confirm that AI governance frameworks in the public sector must connect technical adoption with accountability, data stewardship, and organisational readiness. The central issue is therefore not algorithm selection but institutional sequencing: data and governance foundations must be stabilised before model sophistication is scaled.

IV. PROPOSED SEVEN-LAYER AI-DRIVEN ANOMALY DETECTION FRAMEWORK

Based on the systematic evidence reviewed in Section 3, this paper proposes a seven-layer implementation framework for AI-driven anomaly detection in tax and customs declarations in emerging digital economies. The framework translates the reviewed literature into an operational architecture designed for administrations that are digitising rapidly but must manage fragmented legacy data, limited confirmed labels, and institutional accountability requirements simultaneously. It is specifically applicable to ZATCA and GCC customs administrations, as well as comparable revenue authorities across the broader emerging-economy context.

The framework follows one overarching design principle: build administrative trust before scaling optimisation. Data and governance foundations should be established and verified first; model sophistication should increase only as data linkage quality, human review processes, and institutional safeguards mature. Premature deployment of sophisticated models on weak data foundations is the most common and most consequential failure mode in public-sector AI implementation.

Layer	Purpose	Core Design Specifications	Critical Implementation Note
Layer 1 Data Identity & Interoperability	Establish trustworthy, linkable data foundations	Unique taxpayer/importer identifiers; master data governance; schema standardisation; auditable ETL pipelines; single-window integration	Without this layer, all downstream analytics produce brittle or biased results; applicable to ZATCA Fatoorah

			e-invoicing and FASAH integration
Layer 2 Risk Feature Engineering	Translate domain knowledge into machine-readable signals	Declaration value-per-weight ratios; importer–commodity probability scores; peer-group deviation metrics; temporal trend features; invoice-network concentration indices; filing-sequence irregularity flags	Feature quality determines detection ceiling; features must be version-controlled and auditable for legal defensibility
Layer 3 Hybrid Model Stack	Detect both known and novel non-compliance patterns	Supervised scoring for known fraud profiles; unsupervised anomaly detection for novel schemes; graph models for relational patterns; active learning to improve under sparse labels	No single model family is sufficient; stack must be orchestrated with a common risk-output interface
Layer 4 Explainable Risk Presentation	Support analyst review and taxpayer defensibility	Interpretable reason codes per case; score decomposition by contributing feature; peer-comparison visualisations; case narrative generation; audit-trail logging of AI decision path	Explainability is a legal requirement in most jurisdictions, not an optional enhancement; essential for appeals and challenge processes
Layer 5 Human-in-the-Loop Case Management	Convert model output into inspection, audit, or desk-review decisions	Risk-threshold configuration by case type; queue prioritisation by revenue-at-risk; reviewer feedback capture; escalation and adjudication workflows; outcome recording for model retraining	Human oversight is both a governance requirement and the primary mechanism for generating high-quality new labels
Layer 6 Continuous Learning & Drift Control	Maintain model validity under changing trade and tax behaviour	Feature drift monitoring; score distribution tracking; realised yield and precision measurement; concept drift detection; retraining triggers with confirmed inspection outcomes and appeals results	Without drift control, model performance degrades silently; annual retraining is typically insufficient for dynamic fraud environments
Layer 7 Governance, Accountability & Recourse	Preserve fairness, institutional legitimacy, and legal compliance	Legal basis documentation; audit trail retention policies; challenge and appeal pathway design; fairness monitoring across taxpayer/importer segments; independent oversight mechanisms; periodic model review and publication	Governance is the decisive layer separating technically functional systems from institutionally legitimate ones; required under OECD AI governance principles and GCC digital government frameworks

Table 4. Seven-layer framework for AI-driven anomaly detection in tax and customs administration. Designed for emerging digital economies with direct applicability to ZATCA and GCC revenue administrations.

4.1 Layer 1: Data Identity and Interoperability

Layer 1 establishes the minimum viable information architecture. Many anomaly-detection failures are traceable not to algorithmic limitations but to inconsistent taxpayer or importer identifiers, incomplete entity linkage across declaration types, or poor-quality historical labels contaminated by prior targeting bias. For ZATCA, this layer maps directly to the integration of the National Unified Number (NUN) for entity identification, Fatoorah e-invoice data streams, FASAH declaration records, and third-party data from SAMA (Saudi Central Bank) for financial transaction monitoring. The key design requirement is that all records share consistent unique identifiers, enabling longitudinal entity tracking across declaration types, filing periods, and counterparty networks.

4.2 Layer 2: Risk Feature Engineering

Layer 2 translates administrative domain knowledge into machine-readable signals that carry operational significance. For customs, high-value features include: value-per-weight irregularities relative to commodity benchmarks; improbable importer-commodity-country-of-origin combinations; repeated declaration amendments or corrections; routing through high-risk transit jurisdictions; and broker-importer relationship concentration. For tax, relevant features include: invoice-network concentration indices; address and director anomalies across company registries; filing-sequence irregularities; mismatch between declared income and third-party-reported revenues; and VAT reclaim ratios inconsistent with sector benchmarks. All features should be version-controlled and their derivation documented for legal auditability.

4.3 Layer 3: Hybrid Model Stack

Layer 3 is the analytical core of the framework, implementing the hybrid architecture supported by the strongest evidence in the reviewed literature. The stack comprises four components: (i) a supervised risk-scoring module trained on confirmed inspection outcomes, optimised for precision on known fraud patterns; (ii) an unsupervised or semi-supervised anomaly detection module using Isolation Forest, autoencoder, or one-class SVM variants, targeting novel schemes without confirmed labels; (iii) a graph-based detection module applying GNN or network centrality methods to entity-transaction networks constructed from e-invoice, registry, and declaration data; and (iv) an active learning component that selects cases for human review in ways that maximise information gain for model improvement, balancing exploitation of known risk with exploration of potentially novel patterns. A common risk-output interface aggregates signals from all four components into a unified prioritised risk queue.

4.4 Layer 4: Explainable Risk Presentation

Layer 4 transforms raw model scores into operationally useful and legally defensible risk presentations. Each case in the risk queue is accompanied by: an overall risk score with confidence interval; a ranked list of the top contributing risk reasons expressed in plain language (e.g., "declared value 43% below peer median for this commodity-origin combination"; "director linked to three entities with suspended VAT registration"); a peer-comparison visualisation; a timeline of the entity's filing or declaration history; and a complete audit trail of the AI processing steps that produced the presentation. This layer directly satisfies the explainability requirements identified in OECD [OECD-2025d], CIAT [CIAT-2021], and Kuźniacki et al. [Kuzniacki-2022], and provides the evidentiary foundation for ZATCA's administrative challenge and appeal processes.

4.5 Layers 5–7: Human Decision, Continuous Learning, and Governance

Layers 5 through 7 are the decisive public-sector layers of the framework. Layer 5 embeds the risk queue into the operational case management workflow: configurable risk thresholds by declaration type and value; inspector assignment algorithms that match case complexity to reviewer expertise; structured outcome recording capturing inspection findings, assessment decisions, and appeals results; and direct feedback pathways from human review back to the model training pipeline. Layer 6 implements continuous monitoring across three dimensions: input feature distribution drift (monitored weekly); score distribution stability (monitored per inspection cycle); and realised detection yield and precision (monitored per quarter). Concept drift detection triggers retraining protocols with new confirmed cases, and appeals outcomes are incorporated as corrective signals. Layer 7 establishes the governance architecture: legal basis documentation under applicable data protection and administrative law; audit trail retention policies meeting the requirements of GCC jurisdictions and OECD AI governance principles; published model performance reports for institutional transparency; fairness monitoring across taxpayer size, sector, and nationality segments; and independent oversight mechanisms for high-impact automated decisions.

> *Implementation guidance: The most resilient deployment pathways begin with Layer 1 and Layer 2 for a single declaration type or tax category, add supervised risk scoring (Layer 3 component 1) as data quality is verified, then progressively activate anomaly detection, graph methods, and active learning as institutional capacity and feedback loops mature. Full seven-layer deployment should be treated as a multi-year programme, not a single technology procurement.*

V. DISCUSSION

The central strategic question for revenue administrations is no longer whether AI has a role in anomaly detection. The evidence reviewed here confirms that it demonstrably does, and that technically credible systems are now achievable across customs and tax domains. The more important and consequential question is what kind of AI architecture is administratively durable — meaning one that maintains performance over time, remains legally defensible, and retains the institutional trust of both administrations and taxpayers.

The strongest answer emerging from the reviewed evidence is that durable systems are hybrid, explainable, and feedback-driven. They combine multiple analytical modes across a structured detection stack, remain continuously connected to human adjudication and inspection outcomes, and are updated through operational feedback rather than periodic retrain cycles. This conclusion has particular force for emerging digital economies because they often face institutional pressure to adopt advanced analytics before foundational systems are stable. The evidence consistently indicates that this sequencing is a high-risk strategy: e-invoicing, digital identity, standardised declaration data, and interoperable registries are enabling infrastructure for credible anomaly detection — not parallel initiatives.

5.1 Performance Evaluation: Beyond Offline Accuracy

Many reviewed studies report strong precision, recall, or AUC values, but cross-study comparison is problematic because datasets, fraud definitions, and verification regimes vary substantially. A more suitable performance evaluation agenda for public-sector AI deployments should include: revenue recovered per inspection (detection yield efficiency); unnecessary inspections avoided per detection cycle (false-positive burden); stability under concept drift across time periods (drift resilience); fairness across taxpayer groups (demographic and sector equity); analyst workload and review efficiency; and the rate at which confirmed outcomes feed back into model improvement. This review proposes three specific assurance-oriented metrics for public-sector AI evaluation: (i) Detection Yield Efficiency — revenue confirmed per unit inspection cost; (ii) Concept Drift Resilience Index — performance retention across successive annual cohorts without retraining; and (iii) Explainability Compliance Rate — proportion of generated risk presentations satisfying defined legal explainability standards. These metrics align with the multi-objective optimisation approach advocated by Malashin et al. [Malashin-2025] and the governance orientation of OECD AI governance guidance [OECD-2025f].

5.2 GCC and ZATCA Deployment Implications

Saudi Arabia's ZATCA is unusually well-positioned for AI-driven anomaly detection relative to most emerging digital economies. The Fatoorah e-invoicing mandate provides real-time structured invoice data at national scale. The FASAH customs platform provides integrated pre-arrival and clearance data. The National Data Management Office (NDMO) framework provides data governance infrastructure. Vision 2030's non-oil revenue diversification mandate creates strong and durable institutional motivation. Against this context, the seven-layer framework proposed in Section 4 maps directly onto ZATCA's existing digital infrastructure: Layer 1 utilises NUN, Fatoorah, and FASAH data integration; Layer 2 develops ZATCA-specific risk features from invoice networks and declaration histories; Layer 3 deploys hybrid detection against both VAT carousel patterns and customs undervaluation networks; and Layers 5–7 operate within ZATCA's administrative review and taxpayer service frameworks. For other GCC administrations — the UAE Federal Tax Authority, Bahrain NBR, and Oman Tax Authority — analogous e-invoicing and single-window initiatives provide comparable data foundations for progressive framework adoption.

5.3 Limitations of the Review

This review has four principal limitations. First, the field is evolving rapidly, particularly in the grey literature of public administration and institutional pilot programmes, so recent deployments may not yet be captured in peer-reviewed form. Second, technical studies employ heterogeneous datasets and outcome definitions, limiting formal quantitative comparability across studies. Third, many public-sector implementations remain operationally confidential, meaning academic literature may substantially underrepresent real-world institutional constraints and implementation failures relative to successes. Fourth, the GCC-specific institutional analysis relies on publicly available documentation from ZATCA, NDMO, and Vision 2030 programme offices; primary field research with ZATCA officials would substantially deepen the implementation analysis. These limitations reinforce the case for future comparative case studies, shared evaluation standards, and academic-practitioner partnerships with GCC revenue administrations.

VI. CONCLUSION

AI-driven anomaly detection for tax and customs declarations has progressed from early experimentation to a technically credible and operationally demonstrated frontier. Since 2020, the literature has advanced from individual prediction models toward architecturally integrated systems that account for revenue prioritisation, label scarcity, relational data, active learning, and institutional explainability. The strongest evidence supports hybrid detection stacks that integrate supervised and anomaly-based detection, graph analytics, explainable risk presentation, and human-in-the-loop model updating — deployed within governance frameworks that treat accountability, fairness monitoring, and model transparency as first-class design requirements.

For emerging digital economies — and for ZATCA and GCC customs administrations specifically — the opportunity is significant and institutionally timely. Saudi Arabia's Fatoorah e-invoicing infrastructure, FASAH customs integration, NDMO data governance framework, and Vision 2030 non-oil revenue priority collectively create the enabling conditions for the framework proposed in this review. The central message is that AI-driven anomaly detection should be conceived as an institutional transformation programme — one that detects tax and customs anomalies with hybrid intelligence, decides with human accountability, and governs the complete system as critical public infrastructure. The technical capability exists; the institutional architecture is advancing; the evidence base is sufficient to act.

This review contributes three substantive advances to the literature. First, it provides the first PRISMA 2020-compliant systematic synthesis integrating customs and tax anomaly detection, hybrid AI architectures, and emerging-economy institutional requirements within a single review. Second, it proposes a seven-layer implementation framework validated against GCC and ZATCA institutional contexts — the first such framework to explicitly incorporate Vision 2030 digital governance requirements. Third, it defines three evaluation metrics — detection yield efficiency, concept drift resilience, and explainability compliance rate — purpose-built for public-sector AI assessment contexts in which standard offline accuracy measures are institutionally inadequate.

REFERENCES

[1] [Alexopoulos-2025] Alexopoulos, A., Dellaportas, P., Gyoshev, S. B., Kotsogiannis, C., Olhede, S. C., & Pavkov, T. (2025). A network approach to detect value added tax fraud. Journal of the Royal Statistical Society Series A: Statistics in Society, 188(3). https://doi.org/10.1093/jrsssa/qnaf205

[2] [Alsadhan-2023] Alsadhan, N. (2023). A multi-module machine learning approach to detect tax fraud. Computer Systems Science and Engineering, 46(1), 241–253. https://doi.org/10.32604/csse.2023.033375

[3] [Alwanin-2025] Alwanin, R., Ben Ismail, M. M., & Bchir, O. (2025/2026). Customs fraud detection using a gradient boosting approach for joint classification and risk estimation. Scientific Reports, 16(1), 3432. https://doi.org/10.1038/s41598-025-33382-z

[4] [APEC-2025] APEC. (2025). Application of data analytics in customs risk management. Asia-Pacific Economic Cooperation. Available: https://www.apec.org.

[5] [Arora-2024] Arora, S., Rani, R., & Saxena, N. (2024). A systematic review on detection and adaptation of concept drift in streaming data using machine learning techniques. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 14, e1536. https://doi.org/10.1002/widm.1536

[6] [Belahouaoui-2025] Belahouaoui, R., & Alm, J. (2025). Tax fraud detection using artificial intelligence-based technologies: Trends and implications. Journal of Risk and Financial Management, 18(9), 502. https://doi.org/10.3390/jrfm18090502

[7] [CIAT-2021] CIAT. (2021). Explainable artificial intelligence (XAI) and its importance in tax administration. Inter-American Center of Tax Administrations. Available: https://www.ciat.org.

[8] [Dangsawang-2024] Dangsawang, B., & Nuchitprasitchai, S. (2024). A machine learning approach for detecting customs fraud through unstructured data analysis in social media. Decision Analytics Journal, 10, 100408. https://doi.org/10.1016/j.dajour.2024.100408

[9] [Hovakimyan-2024] Hovakimyan, G., & Bravo, J. (2024). Evolving strategies in machine learning: A systematic review of concept drift detection methods. Information, 15(12), 786. https://doi.org/10.3390/info15120786

[10] [Karklina-2024] Karklina-Admine, S., Cevers, A., Kovalenko, A., & Auzins, A. (2024). Challenges for customs risk management today: A literature review. Journal of Risk and Financial Management, 17(8), 321. https://doi.org/10.3390/jrfm17080321

[11] [Kim-2020] Kim, S., Tsai, Y.-C., Singh, K., Cha, M., Li, C.-T., Lin, S.-D., & Kim, Y.-R. (2020). DATE: Dual attentive tree-aware embedding for customs fraud detection. Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (pp. 2880–2890). ACM. https://doi.org/10.1145/3394486.3403339

[12] [Kim-2023] Kim, S., Mai, T.-D., Han, S., Park, S., Nguyen Duc Khanh, T., So, J., Singh, K., & Cha, M. (2023). Active learning for human-in-the-loop customs inspection. IEEE Transactions on Knowledge and Data Engineering, 35(12), 12039–12050. https://doi.org/10.1109/TKDE.2022.3144299

[13] [Kuzniacki-2022] Kuźniacki, B., Almada, M., Tylikński, K., Górski, Ł., Winogradska, B., & Zeldenrust, R. (2022). Towards explainable artificial intelligence (XAI) in tax law: The need for a minimum legal standard. World Tax Journal, 14(4). https://doi.org/10.59403/2yhh9pa

[14] [Li-2023] Li, Z., Zhu, Y., & van Leeuwen, M. (2023). A survey on explainable anomaly detection. ACM Transactions on Knowledge Discovery from Data, 18(1). https://doi.org/10.1145/3609333

[15] [Malashin-2025] Malashin, I. P., Masich, A., Kolodyazhny, Y., & Roas, S. (2025). Minimizing unnecessary tax audits using multi-objective optimization and machine learning hyperparameter tuning. Frontiers in Artificial Intelligence, 8, 1669191. https://doi.org/10.3389/frai.2025.1669191

[16] [Mittra-2026] Mittra, S. (2026). Harnessing AI and data for tax administration. International Growth Centre. Available: https://www.theigc.org.

[17] [Murorunkwere-2022] Murorunkwere, B. F., Tuyishimire, O., Haughton, D., & Nzabanita, J. (2022). Fraud detection using neural networks: A case study of income tax. Future Internet, 14(6), 168. https://doi.org/10.3390/fi14060168

[18] [Murorunkwere-2023] Murorunkwere, B. F., Haughton, D., Nzabanita, J., & Tuyishimire, O. (2023). Predicting tax fraud using supervised machine learning approach. African Journal of Science, Technology, Innovation and

Development, 15(6), 731–742. https://doi.org/10.1080/20421338.2023.2187930

[19] [Nassif-2021] Nassif, A. B., Talib, M. A., Nasir, Q., & Dakalbab, F. M. (2021). Machine learning for anomaly detection: A systematic review. IEEE Access, 9, 78658–78700. https://doi.org/10.1109/ACCESS.2021.3083060

[20] [Nelson-2020] Nelson, C. (2020). Machine learning for detection of trade in strategic goods: An approach to support future customs enforcement and outreach. World Customs Journal, 14(2), 119–130. https://doi.org/10.55596/001c.116422

[21] [OECD-2020] OECD. (2020). Tax administration 3.0: The digital transformation of tax administration. OECD Publishing. https://doi.org/10.1787/ca274cc5-en

[22] [OECD-2022a] OECD. (2022a). Tax administration 3.0 and the digital identification of taxpayers: Initial findings. OECD Publishing. https://doi.org/10.1787/3ab1789a-en

[23] [OECD-2022b] OECD. (2022b). Tax administration 3.0 and electronic invoicing: Initial findings. OECD Publishing. https://doi.org/10.1787/2ffc88ed-en

[24] [OECD-2025a] OECD. (2025a). Tax administration 2025: Comparative information on OECD and other advanced and emerging economies. OECD Publishing. Available: https://www.oecd.org/en/publications/tax-administration-2025_88538e8b-en.html

[25] [OECD-2025b] OECD. (2025b). Tax administration digitalisation and digital transformation initiatives. OECD Publishing. https://doi.org/10.1787/c076d776-en

[26] [OECD-2025c] OECD. (2025c). Tax administration 3.0: From vision to strategy. OECD Publishing. Available: https://www.oecd.org/tax/forum-on-tax-administration.

[27] [OECD-2025d] OECD. (2025d). AI in tax administration. In Governing with artificial intelligence: The state of play and way forward in core government functions. OECD Publishing. https://doi.org/10.1787/795de142-en

[28] [OECD-2025e] OECD. (2025e). The digitalisation of trade documents and processes: Going paperless today, going paperless tomorrow. OECD Trade Policy Papers, 297. OECD Publishing. https://doi.org/10.1787/64872f25-en

[29] [OECD-2025f] OECD. (2025f). Governing with artificial intelligence: The state of play and way forward in core government functions. OECD Publishing. https://doi.org/10.1787/795de142-en

[30] [OECD-2024] OECD, CIAT, IOTA, IMF, & partners. (2024). Inventory of tax technology initiatives. OECD Forum on Tax Administration. Available: https://www.oecd.org/tax/forum-on-tax-administration/tax-technology-tools-and-digital-solutions.

[31] [Page-2021a] Page, M. J., McKenzie, J. E., Bossuyt, P. M., et al. (2021a). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. Journal of Clinical Epidemiology, 134, 178–189. https://doi.org/10.1016/j.jclinepi.2021.03.001

[32] [Page-2021b] Page, M. J., Moher, D., Bossuyt, P. M., et al. (2021b). PRISMA 2020 explanation and elaboration: Updated guidance and exemplars for reporting systematic reviews. BMJ, 372, n160. https://doi.org/10.1136/bmj.n160

[33] [Ruvalcaba-2025] Ruvalcaba-Gómez, E. A., & García-Benitez, V. H. (2025). Governance of artificial intelligence in the public sector: Analysis of public policies in Spain and Mexico. Review of Policy Research, 42(4). https://doi.org/10.1111/ropr.70057

[34] [Singh-2023] Singh, K., Tsai, Y.-C., Li, C.-T., Cha, M., & Lin, S.-D. (2023). GraphFC: Customs fraud detection with label scarcity. Proceedings of the 32nd ACM International Conference on Information and Knowledge Management. ACM. https://doi.org/10.1145/3583780.3614690

[35] [UNDP-2025] UNDP. (2025). AI for the next generation of public services: Global trends and case studies. United Nations Development Programme. Available:

https://www.undp.org/publications/ai-next-generation-public-services.

[36] [Vanhoeyveld-2020] 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

[37] [Vijayakumar-2025] Vijayakumar, S. (2025). Technology-centric and data-driven customs risk management for supply chain security. World Customs Journal, 19(1), 38–63. https://doi.org/10.55596/001c.131745

[38] [WCO-2023] WCO. (2023). Compendium of case studies on e-commerce (4th ed.). World Customs Organization. Available: https://www.wcoomd.org.

[39] [ZATCA-2023] Zakat, Tax and Customs Authority (ZATCA). (2023). E-invoicing (Fatoorah) programme: Phase 2 integration requirements and technical specifications. ZATCA, Riyadh, Kingdom of Saudi Arabia. Available: https://zatca.gov.sa/en/E-Invoicing/Pages/default.aspx.

[40] [WCO-WTO-2022] WCO & WTO. (2022). Study report on disruptive technologies. World Customs Organization and World Trade Organization. Available: https://www.wcoomd.org.

[41] [Zheng-2024] 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, 43–57. https://doi.org/10.1016/j.eng.2023.07.014