IIAR JOURNALS

Innovating Knowledge. Advancing Research.

IIAR Journal of Economics, Finance, and Development Studies

Vol 1, No 1, 2026

Artificial Intelligence in Public Financial Management: A Theoretical Framework of Revenue Mobilisation, Expenditure Control and Fiscal Accountability in Developing Economies

NWOGBOJI Emmanuel Usulor[1], IGWE Emmanuel Sunday[2], OGAH Sunday Ogbo[3], BAMIGBOLA Paul Idowu[4], ATOLAGBE Umar Kawu[5]

[1] International Institute of AI and Research; [2,3,4,5] Veristzon Global Research Lab

ABSTRACT

Developing economies collect a small share of national income in tax and lose a large share of what they collect to leakage, and artificial intelligence is now being offered to tax authorities, treasuries and auditors as the remedy for both. The literature on tax administration indicates that information technology raises revenue only when the information it produces is acted upon, yet the newer literature on algorithmic government has rarely been read alongside it, so that claims about what AI can do for public finance rest on evidence drawn from the wrong institutions. This paper develops a framework for assessing what AI can and cannot do for public financial management in low-capacity states. It is a theoretical synthesis built from twenty-eight peer-reviewed journal articles published between 2023 and 2026, each verified against its Crossref publisher record and screened for retractions and expressions of concern, drawn from public economics, tax administration, procurement, state capacity and public administration. The framework distinguishes three functions that any enforcement technology performs: detecting non-compliance and irregular expenditure, deciding whom to audit and which payments to hold, and deterring taxpayers and officials through the perceived probability of consequence. AI improves detection substantially, improves decision only where the institution that acts on its outputs is independent of those it targets, and improves deterrence only where the state is seen to act on what it detects. Five propositions follow, concerning the conditions under which AI raises revenue rather than reallocating the burden, the tendency of algorithmic enforcement to fall hardest on the visible formal sector while the informal and the politically connected remain untouched, and the accountability requirements for systems that decide whom the state pursues. Nigeria's domestic resource mobilisation challenge anchors the analysis. The paper recommends that governments sequence investment in detection behind, or at least alongside, investment in the independence and visibility of enforcement, and that algorithmic selection be subject to published audit of its distributional incidence.

Keywords: public financial management; tax administration; artificial intelligence; revenue mobilisation; expenditure control

Artificial Intelligence in Public Financial Management: A Theoretical Framework of Revenue Mobilisation, Expenditure Control and Fiscal Accountability in Developing Economies

1. INTRODUCTION

The promise made on behalf of artificial intelligence in public finance is a promise about information. A tax authority that could see every transaction, a treasury that could see every payment and an auditor who could see every contract would, on this view, collect what is owed and spend only what is authorised. Vendors of risk-scoring engines, anomaly detectors and contract-screening tools make exactly this argument, and ministers under pressure to raise low tax-to-GDP ratios find it attractive. The argument is not foolish: a central strand of the economics of tax enforcement, as we summarise it, treats the authority's independent information about the taxpayer as a key determinant of compliance, and leakage persists where nobody is positioned to observe it. If machines can generate observation at scale, the case for them appears to write itself.

This paper argues that the case is incomplete in a way that matters for developing economies in particular. What the recent empirical literature on tax administration suggests, on our reading, is that information is necessary for enforcement but is not enforcement. Electronic filing and invoicing, digital payments and third-party data have raised revenue in some settings and not in others, and reviews of this evidence point to institutional constraints, among them resistance by taxpayers and collectors, a lack of institutional mainstreaming and the deployment and political incentives of officials (Okunogbe & Santoro, 2023; Okunogbe & Tourek, 2024). On our reading, the difference lies in what happened after the information arrived: whether anybody acted on it, whether the taxpayers believed anybody would, and whether the officials who were supposed to act were themselves free of the pressures that produced non-compliance in the first place. These are questions about institutions, not about data. An AI system that improves the quality of information without altering the institution that receives it changes the first link in a chain whose later links are the ones that were broken.

The paper's contribution is to turn this observation into a framework that separates the work of any enforcement technology into three functions: detection, identifying the taxpayer who under-reports, the payment that duplicates another, the contract whose price is out of line; decision, choosing which of the cases detected to pursue and with what intensity; and deterrence, the change in behaviour among taxpayers and officials who are not themselves pursued but who revise their estimate of the probability that they would be. AI acts on each differently. Its effect on detection is large and conditional on the quality and coverage of the state's data rather than on its institutions, because it is an improvement in the statistical processing of information the state already holds. Its effect on decision is conditional on the independence of the institution making the decision, because a system that scores risk does not pursue anyone; a person does, and that person can be instructed, bribed or frightened. Its effect on deterrence is the most conditional of all, because deterrence depends on what taxpayers and officials believe the state will do, and that belief is formed by observing what it has done. The aim is not to argue against AI in public finance but to specify the conditions under which it does what is claimed for it, so that investment is sequenced rather than made on faith. Nigeria, whose revenue problem is severe and which has committed publicly to digital fiscal systems across tax, treasury and procurement, anchors the analysis.

2. LITERATURE REVIEW

2.1 CONCEPTUAL REVIEW

Public financial management is the set of institutions and processes through which a government raises revenue, executes spending and accounts for both. In developing economies the revenue side is dominated by domestic resource mobilisation, the share of national income the state collects through taxation rather than aid, borrowing or resource rents, and the expenditure side by leakage, the gap between what is appropriated and what reaches its purpose. Fiscal accountability refers to the institutions through which the executive answers for both: audit, legislative scrutiny, publication and the political consequences that follow.

© IIARJournal. All rights reserved.

Artificial intelligence, for the purposes of this paper, means the use of statistical learning methods to produce predictions or classifications from administrative data at a scale and speed that human analysts cannot match: risk-scoring of taxpayers, matching of returns against third-party information, anomaly detection in invoices, payroll and payment streams, price-benchmarking in procurement and risk-based selection of entities for audit. The literature on AI in government has recently attempted to impose some order on these applications; Straub et al. (2023) propose a unified framework for AI in government built on three concepts, operational fitness, epistemic alignment and normative divergence, each tied to emerging technical measurement standards. Those concepts map onto the present framework in a useful way: detection is a question of operational fitness, whether the system processes the state's information well; decision is a question of epistemic alignment, whether a risk score means what the official acting on it takes it to mean; and the accountability problem taken up in Proposition 5 is one of normative divergence, whether the values embedded in the model are those the public would endorse.

The concepts of detection, decision and deterrence that organise this paper are usually treated separately, by the tax-compliance, public-administration and deterrence-oriented enforcement literatures respectively. Bringing them together makes visible that an improvement in one does not automatically propagate to the others, which is the paper's central point.

2.2 THEORETICAL REVIEW

Three theoretical traditions inform the framework. The first is the economics of tax enforcement under incomplete information, which in its modern form, as we summarise it, holds that compliance depends less on audit rates and penalties than on the authority's access to information the taxpayer cannot manipulate. The most recent contributions complicate this picture usefully. Almunia et al. (2024) examine mismatches between buyers' and sellers' reports of the same transactions in Uganda's value-added tax and find that the two sides disagree in most cases; most firms misreport to their own advantage, but about a quarter misreport in ways that raise their own liability, a pattern that points to confusion and poor record-keeping alongside evasion and, on our reading, implies that third-party information identifies evaders less cleanly in low-capacity settings than the theory assumes. Mascagni et al. (2023) reach a complementary conclusion for Rwanda, documenting reporting inconsistencies in the value-added tax data, effective burdens that differ across firms despite a proportional rate, and frequent breaks in the VAT chain, so that the tax as applied differs substantially from its theoretical functioning. If the state's information is noisier than the theory assumes, the gain from processing it faster is bounded by that noise.

The second tradition is the political economy of state capacity, where the question is why some states can enforce at all. Bergeron et al. (2024) show, through randomised variation in property tax rates in the Democratic Republic of Congo, that the status quo rate lies above the revenue-maximising rate because compliance falls as rates rise, and that the revenue-maximising rate itself rises with enforcement; low enforcement capacity sets a ceiling on what the rate can collect. Weigel and Kabue Ngindu (2023) trace a pathway out of this low-capacity trap in the same setting: a systematic door-to-door collection campaign raised compliance from near zero, and citizens who observed taxes being collected in a regular, non-arbitrary way updated their view of the government's procedural performance in a manner consistent with a virtuous cycle between perceived performance and fiscal capacity. Bah (2024) documents at the cross-country level for sub-Saharan Africa that institutional quality is positively associated with revenue across tax types. The tradition's implication for this paper is that enforcement technology is a complement to capacity, not a substitute for it, and that where capacity is lowest the technology has the least to work with.

The third tradition is the study of discretion and algorithmic decision in public administration. The classical concern was that discretion permits favouritism; the newer concern is that its removal permits a different kind of error. Szucs (2024) shows for Hungarian procurement that a high-discretion procedure raised prices, selected less productive contractors and benefited firms connected to the governing party, while Decarolis et al. (2025) show for Italy that discretionary procedures are associated with corruption only where competition is also limited, and that officials who were themselves under investigation chose those procedures more often. Best et al. (2023) find, for Russian procurement, that

© IIARJournal. All rights reserved.

close to two-fifths of the variation in prices paid for narrowly defined items is attributable to the individual bureaucrats and organisations doing the buying. On the algorithmic side, Alon-Barkat and Busuioc (2023) find in three experiments with Dutch citizens and civil servants no general tendency to over-rely on algorithmic advice and, in one study with citizens, a stronger propensity to adopt advice when it matched prevailing group stereotypes, a result not replicated with civil servants; de Boer and Raaphorst (2023) find among Dutch inspectors that automation shifts enforcement style, toward both more legalistic and more accommodating behaviour, without perceived discretion being the channel; and Grimmelikhuijsen (2023) finds that explaining an algorithm's reasoning raises the perceived trustworthiness of its decisions more than merely making the algorithm accessible does. A scoring model, in short, enters an organisation populated by people who will use it in ways its designers did not anticipate.

2.3 EMPIRICAL REVIEW

The empirical literature on technology in tax administration has matured considerably since 2023, and its findings are more sobering than the vendor literature suggests. Okunogbe and Santoro (2023) review the evidence on technology in tax administration across developing economies and, alongside its potential to identify the tax base and monitor compliance, identify the limitations that recur: inadequate infrastructure and connectivity, resistance or non-adoption by taxpayers and by tax collectors, a lack of institutional mainstreaming of the new tools, and an unsupportive regulatory environment. Okunogbe and Tourek (2024) add the political dimension, emphasising that governments choose whether to invest in tax capacity and expand collection according to political incentives, and that the performance of technology depends on the deployment and incentives of the officials who use it. Read together, the two reviews imply what this paper takes as its starting point: technology that makes more taxpayers visible does not by itself oblige anybody to pursue them.

Studies of specific technologies confirm the pattern. Das et al. (2023) find that India's demonetisation pushed firms toward electronic payments and that firms which used them more reported more sales to the tax authority, accounting for roughly half of the rise in reported sales observed during the episode. Brockmeyer and Sáenz Somarriba (2025), by contrast, find that Uruguay's consumer rebates for card payments triggered an immediate and large increase in card transactions while leaving firms' tax compliance unaffected, which they attribute to endogenous adoption of card machines and to the small share of card sales in total reported sales. Traceability, these two studies together suggest, raises compliance only where it covers a large enough part of a firm's sales to change what it can safely report. Kotsogiannis et al. (2024) use Rwandan corporate tax records to show that the effect of an audit on future reporting depends on its scope: comprehensive audits had a significant positive effect on the corporate income and tax liability reported in the three years after the audit, while narrow-scope (desk-based and issue-oriented) audits had a negative net effect, a pattern we read as consistent with taxpayers learning what the authority does and does not examine. Bérgolo et al. (2023) show that letters providing information about audits changed compliance among Uruguayan firms in a way inconsistent with the canonical evasion model, the audit functioning, in their phrase, as a scarecrow.

Research on who bears enforcement has become a distinct strand. Elzayn et al. (2025) document that, despite race-blind selection, Black taxpayers in the United States were audited at 2.9 to 4.7 times the rate of others, and trace the disparity largely to a policy decision to prioritise over-claimed refundable credits over other forms of non-compliance; an algorithm targeting total under-reported tax would reduce the disparity and detect more non-compliance. Battaglini et al. (2025) show with Italian data that replacing the least promising tenth of planned audits with cases chosen by a machine-learning model would raise detected evasion by up to two-fifths, the model being trained, necessarily, on cases humans originally selected. Lediga et al. (2025) find that South Africa's forced registration of firms produced weak subsequent compliance, with few of the newly registered paying any tax, although the scale of the exercise made its aggregate revenue non-negligible. Moore (2023) argues that the drive to register ever more small taxpayers in sub-Saharan Africa creates an unduly favourable impression of policy and managerial effort, and that the 'informal sector' narrative diverts attention from the failure to tax wealthier individuals and larger firms, and Gallien and van den Boogaard (2023) locate the

© IIARJournal. All rights reserved.

disappointing results of formalisation policy in conceptual fallacies about what informality is.

On the expenditure side, Muralidharan et al. (2025) show that stricter biometric identification in India's largest social protection programme reduced corruption but cost between one and a half and two million legitimate beneficiaries their access at some point, so that, on our reading, the saving was partly a reallocation of loss from the state to the poor; they attribute much of the harm to how the transition was managed rather than to the technology itself. On the attitudinal side, Belmonte et al. (2025) find across thirty sub-Saharan African countries that aid projects delivering public goods without involving the state lower tax morale, which they read as a signal to citizens that the state cannot provide, and Santoro and Mascagni (2023) find that short videos emphasising deterrence and equity improved Rwandan taxpayers' perceptions, with the deterrence message alone raising the tax actually remitted. Both confirm that willingness to pay responds to what citizens observe the state doing and receiving. The lesson that runs through these literatures is that technology produces revenue and reduces leakage only through the institutions that receive its outputs.

3. MATERIALS AND METHODS

This is a theoretical paper whose research design is a conceptual synthesis: a framework is constructed from existing theory and tested for coherence against recent empirical findings rather than against new data. The unit of analysis is the enforcement function, the set of activities through which a public financial management system converts information about taxpayers, payments and contracts into revenue collected and leakage prevented. The functions of detection, decision and deterrence were separated by asking, for each documented technology intervention in the literature, at which point in the chain from information to consequence its effect arose or failed; the conditions under which AI affects each function were identified by comparing interventions that succeeded with those that did not, treating the institutional features that distinguished them as candidate moderators; and the propositions state those moderators as conditional claims that future empirical work could falsify.

The evidence base comprises twenty-eight peer-reviewed journal articles with volume years between 2023 and 2026. Each was identified through bibliographic search of the Crossref database and citation tracing from the core tax-administration and public-administration literatures, verified against its Crossref publisher record for digital object identifier, authorship, journal, volume, issue and pagination, and screened through the Retraction Watch data integrated in Crossref; no article carrying a retraction or expression of concern was retained. Working papers, policy reports, conference proceedings and book chapters were excluded, at a cost: some of the most directly relevant evidence on AI in revenue agencies exists only in institutional reports, so the paper's empirical anchor is the broader evidence on technology and enforcement. Attributed findings were drawn from each source's published abstract or record and, where it was openly available, its full text; where the paper extends a finding beyond what the source reports, the extension is identified as the authors' own reading, and where studies disagreed the disagreement was treated as information about moderating conditions rather than resolved by selection.

Nigeria is used as an anchoring case rather than as an object of empirical analysis, chosen because it combines a low tax-to-GDP ratio, a large informal economy, a federal division of fiscal authority, publicly announced investment in digital tax and treasury systems, and recurring public debate about the independence of enforcement from political influence. Two kinds of statement about Nigeria should be distinguished. Empirical findings about revenue, compliance and leakage are drawn only from the comparative literature that includes it, and no country statistic is reported that cannot be sourced to a verified reference. Descriptive statements about the existence and stated purpose of Nigerian fiscal systems, such as the treasury single account or the biometric payroll, are matters of public record reported for illustration, and no conclusion about their performance is drawn from them.

Generative artificial intelligence tools were used to assist in drafting, in organising the literature and in managing references; every reference was independently retrieved and verified by the authors against its publisher record, and the authors bear full responsibility for the content, including the interpretation of the sources. The principal limitation of the design is that a framework built from the behaviour of earlier technologies may not fully

© IIARJournal. All rights reserved.

capture the behaviour of systems that learn, so the propositions are offered as hypotheses for testing.

4. RESULTS AND DISCUSSION

4.1 RESULTS

The framework begins from an observation that is obvious once stated and routinely ignored in practice: nothing the state detects produces revenue or prevents leakage until somebody decides to act on it, and nothing done to one taxpayer or official changes the behaviour of others until they believe it would be done to them. Enforcement is a chain of three functions, and the value of an improvement at any link is bounded by the links that follow.

Detection is the function to which AI contributes most directly, and the one whose condition lies in the data rather than in the institution, because the methods are improvements in the statistical processing of information the state already holds. The evidence that better processing identifies more non-compliance is strong; Battaglini et al. (2025) show that replacing the least promising tenth of planned audits with cases chosen by a machine-learning model would raise detected evasion substantially over the Italian authority's existing selection. The condition on detection is the quality and coverage of the underlying information. Where a substantial minority of firms misreport to their own disadvantage, as Almunia et al. (2024) find in Uganda, part of what the model learns to separate is confusion from competence rather than honesty from evasion, and where much economic activity yields little usable information, as in the economies Moore (2023) describes, in which many registered small taxpayers pay little or nothing, it cannot detect what the record does not contain. Detection improves substantially, but over a domain bounded by the reach of the state's information rather than the sophistication of its algorithms.

Decision is the function at which the institution enters. A risk score ranks cases; it does not open an audit, hold a payment or cancel a contract. Those are acts of an official who has a superior, a budget, a career, sometimes a relative in the taxpayer's firm and occasionally a threat on the telephone. Their quality depends on the integrity and competitiveness of the environment in which discretion is exercised, which a scoring model does not change; Szucs (2024) shows discretion used to favour connected firms, and Decarolis et al. (2025) show that it attracts corruption where competition is limited and that officials under investigation for corruption choose it more often. What the model changes is the distribution of discretion. It can narrow it, by making deviation from the ranked list visible and therefore costly; that is the framework's own mechanism, and the finding of de Boer and Raaphorst (2023) that automation shifts how inspectors enforce without working through their perceived discretion is a caution that the shift need not run in the direction designers expect. It can also disguise discretion, by lending the authority of a machine to a selection whose incidence was fixed when its objective was chosen; Elzayn et al. (2025) show that a policy decision to prioritise over-claimed refundable credits produced racial disparities in audit rates despite race-blind selection, disparities the agency, which does not observe race, could not readily see. Whether AI improves decision therefore depends on whether the institution that acts on its outputs is independent of those it targets. Where the revenue agency's leadership can be removed by the executive whose allies are the largest under-payers, or where the treasury officer who would hold a payment reports to the minister whose contractor is being paid, the model's ranking will be consulted selectively, and the decision function will improve only for cases that nobody powerful cares about. The experimental finding of Alon-Barkat and Busuioc (2023) that advice was adopted more readily when it matched prevailing stereotypes, observed in one citizen sample and not replicated with civil servants, is at most a laboratory analogue of that selectivity.

Deterrence is the function at which the public enters. Taxpayers and officials who are not themselves audited or sanctioned change their behaviour because they revise their estimate of the probability that they would be, and that estimate is formed by observing what the state visibly does. Bérgolo et al. (2023) show that letters alone can change compliance, though not in the way the canonical evasion model predicts, and Kotsogiannis et al. (2024) show that subsequent reporting responds to the scope of what the authority actually examined, rising after comprehensive audits and falling, on net, after narrow-scope ones. AI contributes to deterrence only indirectly, through its effect on what the state is seen to do, and this is the most conditional of its effects. A state that detects more and pursues the same people as before produces no new deterrence. A state that detects more and pursues only the small and the

© IIARJournal. All rights reserved.

visible deters the small and the visible while demonstrating to everyone else where enforcement stops, so that a large or connected taxpayer who watches many small cases pursued and large ones left untouched revises the probability of consequence downward; visible selectivity is itself evidence of impunity. A state that detects more, pursues the large and the connected and is seen to do so produces deterrence that may exceed the direct revenue of the cases pursued. The closest measured analogue is the finding of Boning et al. (2025) that an audit's effect on the audited taxpayer's own later reporting yields at least three times the initial revenue and that returns are highest at the top of the income distribution; the framework extends that individual deterrence to the observers of enforcement, an extension the study itself does not estimate.

Table 1 summarises the framework. Five propositions follow from it.

Table. 1. The detecting–deciding–deterring framework for AI in public financial management

Function	What AI does	Binding condition	Where it fails	Nigeria illustration
Detecting	Ranks taxpayers, payments and contracts by estimated risk from administrative data	Quality and coverage of the state's information	Noisy records; activity outside the administrative record	Large informal economy limits the domain over which risk scoring operates
Deciding	Narrows or disguises official discretion over whom to pursue	Independence of the acting institution from those it targets	Selective adherence to rankings; objective functions that prioritise one form of non-compliance over others	Whether federal and state revenue and treasury agencies can act on large cases free of executive and political pressure is the condition the framework makes decisive
Deterring	Changes perceived certainty of consequence through what the state is seen to do	Visibility of enforcement against the large and connected	Enforcement concentrated on the small and visible; perceived impunity above	Tax morale conditioned by perceptions of elite impunity and of what revenue buys

Source: authors' synthesis of the cited literature.

Proposition 1. AI raises revenue and reduces leakage in proportion to the independence of the institution that acts on its outputs, so that in states where enforcement agencies are subordinate to the political executive the revenue gain from improved detection is concentrated in cases that carry no political cost.

Proposition 2. Where detection improves but decision does not, enforcement intensity shifts toward taxpayers and transactions that are visible in administrative data, and the effective burden of taxation and audit falls more heavily on the formal sector, on wage earners and on compliant firms relative to the informal and the connected.

Proposition 3. The deterrent effect of AI-enabled enforcement is a function of the visibility of enforcement against large and politically protected non-compliers, not of the volume of cases detected, so that systems that generate many small enforcement actions can reduce perceived certainty of consequence among the large.

Proposition 4. AI-enabled expenditure control reduces measured leakage partly by reallocating loss from the state to entitled beneficiaries and compliant suppliers, so that the net welfare effect depends on the error structure of the detection model and on the remedy available to those wrongly flagged.

Proposition 5. Algorithmic selection of whom the state pursues is legitimate only where its objective function, its distributional incidence and its error rates are published and subject to independent audit, because the discretion it embodies is otherwise exercised by its designers without the accountability that attached to the officials it replaces.

4.2 DISCUSSION

The propositions can be read against the Nigerian case, which displays each of the framework's conditions in a demanding form. Nigeria's tax-to-GDP ratio is low by international standards and its public revenue has historically depended heavily on oil. The comparative literature associates weak tax collection with institutional quality across sub-Saharan Africa (Bah, 2024) and with the political incentives governments face in expanding collection (Okunogbe & Tourek, 2024), although neither study examines Nigeria or oil dependence specifically. Its reform programme over the past decade has announced or implemented exactly the systems the framework examines: an online tax administration

© IIARJournal. All rights reserved.

platform at the federal tax authority, now the Nigeria Revenue Service (formerly the Federal Inland Revenue Service), a treasury single account, a biometric personnel and payroll system, electronic procurement and risk-based compliance management. The framework predicts that, to the extent these systems operate as described, they improve detection, and that any gain in revenue or reduction in leakage will be bounded by the decision and deterrence functions.

Consider first the distributional proposition. The most visible taxpayers in Nigeria, on our reading, are the formal firms that file electronically, the employees whose tax is withheld at source and the banks through which traceable payments flow. A risk-scoring model will find its highest-confidence discrepancies among them because they are the ones about whom it has data, and an agency under pressure to show results will act on them, while the informal trader whose activity never enters the record and the connected firm whose case is quietly closed are both outside the model's effective reach, for different reasons. The effect is the pattern Elzayn et al. (2025) describe in a very different setting: an objective function that prioritises one form of non-compliance concentrates enforcement on the people who exhibit it, with a distributional incidence that, on our reading, nobody explicitly chose. The alternative strategy of forcing the informal into the register does not resolve this. Lediga et al. (2025) find that few of the firms registered by force in South Africa paid any tax afterwards, although the aggregate revenue gain was non-negligible, and Moore (2023) argues that the 'informal sector' narrative surrounding registration diverts attention from the larger taxpayers who go untaxed. Bachas et al. (2024) add that, because the share of spending in informal outlets falls steeply with income, consumption taxes are more progressive than assumed, which, on our reading, implies that formalising small vendors could erode some of that progressivity. The lesson for Nigeria is not that enforcement against the formal sector is wrong but that a strategy whose incidence is set by the structure of the data rather than by judgement will reproduce, in a new and apparently neutral form, the old pattern in which the compliant pay for the non-compliant.

Consider next the deterrence proposition. The attitudinal evidence from sub-Saharan Africa finds that willingness to pay tax responds to what citizens observe: aid projects that bypass the state depress tax morale (Belmonte et al., 2025), and a deterrence message raised both perceptions and the tax actually remitted in Rwanda (Santoro & Mascagni, 2023). Where elite impunity is widely perceived, the deterrent value of AI-enabled enforcement depends almost entirely on whether it is used against the elite, and a system that generates thousands of notices to small firms while the cases that would signal seriousness remain unresolved produces, on the framework's logic, a net reduction in perceived certainty of consequence among those whose compliance matters most. Boning et al. (2025) show that audits at the top of the income distribution return many times their cost once the audited taxpayer's later reporting is counted; if the observers of such audits respond as the audited do, which the framework proposes and the study does not test, the return is larger still. The implication is uncomfortable but clear: the revenue return to AI in Nigeria's tax administration is determined less by the quality of its models than by whether the Nigeria Revenue Service and its state counterparts are permitted to pursue, and are seen to pursue, the cases their models identify at the top.

The expenditure side raises the fourth proposition. The evidence from India's biometric identification reforms (Muralidharan et al., 2025) is the clearest demonstration available that a detection technology can reduce leakage and harm the entitled simultaneously, and their finding that the harm turned on how the transition was managed is the point the framework generalises. Nigeria's biometric payroll system was introduced with the stated aim of removing ghost workers, and its treasury single account with that of consolidating government accounts and curbing leakages, and the framework expects detection gains of that kind wherever such systems work as designed. It draws attention, though, to the person wrongly flagged: the genuine employee whose salary is suspended, the supplier whose invoice is held on an anomaly score. If the remedy available to them is slow, costly or dependent on the office that flagged them, the system has redistributed loss toward those least able to contest it, and the headline saving overstates the welfare gain. The procurement literature makes a parallel point: Best et al. (2023) attribute close to two-fifths of the variation in prices paid to the individual bureaucrats and organisations doing the buying, Szucs (2024) shows

© IIARJournal. All rights reserved.

discretion being used to favour connected firms, and Decarolis et al. (2025) show that officials under investigation for corruption choose discretionary procedures more often, which suggests that an anomaly detector in an environment of weak integrity will flag the clumsy and miss the sophisticated, who are usually the expensive.

The fifth proposition concerns accountability. An audit-selection or payment-screening model embodies a judgement about whom the state should pursue, formerly made by officials answerable in principle to superiors, auditors, courts and legislatures and now made by whoever chose the objective and training data; Elzayn et al. (2025) show that this choice can have large distributional consequences, arising in their case under race-blind selection and therefore not directly visible to the agency applying it. Grimmelikhuijsen (2023) shows that explaining an algorithm's reasoning raises the perceived trustworthiness of its decisions more than merely making it accessible, though the effects varied across decision contexts, and Straub et al. (2023) tie each dimension of their framework, including the normative divergence between a system's embedded values and the public's, to measurable standards, which is the kind of apparatus a published audit would require. The framework's requirement is modest in principle and demanding in practice: that the objective function, distributional incidence and error rates of any model that decides whom the state pursues be published and independently audited. In Nigeria this would require the audit institutions whose own capacity the technology is meant to augment to acquire the competence to audit the technology, a sequencing problem the literature has barely addressed.

One objection holds that the framework underestimates the capacity of detection to change the other two functions over time, since a flood of accurate detections creates pressure on institutions to act. There is something in this; Weigel and Kabue Ngindu (2023) show that systematic, visible collection can initiate a compliance dynamic from a low base. The framework does not deny that detection can be a lever on institutions, only that it is not automatically one. The lack of institutional mainstreaming that Okunogbe and Santoro (2023) list among the recurring limitations of tax technology is the reminder that information can accumulate without being acted upon, and the Ugandan and Uruguayan studies add two cautions: part of what arrives is confusion rather than evasion (Almunia et al., 2024), and making transactions traceable did not by itself change what firms reported in Uruguay, where card sales were a small share of total sales (Brockmeyer & Sáenz Somarriba, 2025). A second objection, that the conditions are so demanding as to make the technology irrelevant, misreads the argument: the framework is a guide to where investment has high returns and in what order.

Future research should test the propositions directly: the distributional proposition with administrative data on the incidence of algorithmically selected audits by sector and firm size, the deterrence proposition by comparing compliance responses to publicised enforcement against large and small taxpayers, and the expenditure proposition by measuring the exclusion error of screening systems and the cost of remedy. The accountability proposition cannot be tested until deployed models' objective functions are disclosed, which is itself a finding.

5. CONCLUSION AND RECOMMENDATION

This paper has argued that artificial intelligence improves the detection function of public financial management substantially, the decision function conditionally on the independence of the institution that acts, and the deterrence function only where the state is seen to act on what it detects, and it has stated five propositions through which those claims can be tested.

The recommendations are addressed to governments considering or expanding investment in AI for public finance, with Nigeria in view. Investment in detection should be sequenced behind, or at least alongside, investment in the independence of the enforcement agencies that will act on it, because the return to detection is bounded by that independence. Enforcement strategy should be set by deliberate judgement about where revenue and deterrence lie rather than left to the structure of the data, and the distributional incidence of algorithmically selected audits and holds should be monitored and published. Expenditure-control systems should treat the remedy for the wrongly flagged as a first-order design feature, with time limits and independent review. The objective functions, incidence and error rates of any model that decides whom the state pursues should be disclosed to and audited by an institution independent

© IIARJournal. All rights reserved.

of the one deploying it. And governments should resist letting the volume of small enforcement actions substitute for the visible resolution of large ones, because the former can reduce rather than raise the certainty of consequence that drives compliance.

6. CONTRIBUTION TO EXISTING KNOWLEDGE

The paper makes three contributions. The detecting–deciding–deterring framework separates functions that the tax-compliance, public-administration and deterrence literatures treat in isolation and shows that an improvement at one link does not propagate automatically to the next. The five propositions translate it into conditional claims about revenue, leakage, distribution and accountability that can be tested with administrative data from agencies that have adopted algorithmic selection. The application to Nigeria shows that the framework's conditions describe the institutional features of a large developing economy in which the technology is already being deployed, and identifies the sequencing of investment in detection, independence and visibility as the central policy problem.

7. REFERENCES

Almunia, M., Hjort, J., Knebelmann, J., & Tian, L. (2024). Strategic or confused firms? Evidence from "missing" transactions in Uganda. *Review of Economics and Statistics, 106*(1), 256–265. https://doi.org/10.1162/rest_a_01180

Alon-Barkat, S., & Busuioc, M. (2023). Human–AI interactions in public sector decision making: "Automation bias" and "selective adherence" to algorithmic advice. *Journal of Public Administration Research and Theory, 33*(1), 153–169. https://doi.org/10.1093/jopart/muac007

Bachas, P., Gadenne, L., & Jensen, A. (2024). Informality, consumption taxes, and redistribution. *Review of Economic Studies, 91*(5), 2604–2634. https://doi.org/10.1093/restud/rdad095

Bah, M. (2024). Tax revenue mobilization and institutional quality in sub-Saharan Africa: An empirical investigation. *African Development Review, 36*(2), 201–221. https://doi.org/10.1111/1467-8268.12752

Battaglini, M., Guiso, L., Lacava, C., Miller, D. L., & Patacchini, E. (2025). Refining public policies with machine learning: The case of tax auditing. *Journal of Econometrics, 249*, 105847. https://doi.org/10.1016/j.jeconom.2024.105847

Belmonte, A., Bove, V., & Di Salvatore, J. (2025). Bad signals? Foreign aid and tax morale across sub-Saharan Africa. *World Development, 194*, 107055. https://doi.org/10.1016/j.worlddev.2025.107055

Bergeron, A., Tourek, G., & Weigel, J. L. (2024). The state capacity ceiling on tax rates: Evidence from randomized tax abatements in the DRC. *Econometrica, 92*(4), 1163–1193. https://doi.org/10.3982/ECTA19959

Bérgolo, M., Ceni, R., Cruces, G., Giaccobasso, M., & Perez-Truglia, R. (2023). Tax audits as scarecrows: Evidence from a large-scale field experiment. *American Economic Journal: Economic Policy, 15*(1), 110–153. https://doi.org/10.1257/pol.20200321

Best, M. C., Hjort, J., & Szakonyi, D. (2023). Individuals and organizations as sources of state effectiveness. *American Economic Review, 113*(8), 2121–2167. https://doi.org/10.1257/aer.20191598

Boning, W. C., Hendren, N., Sprung-Keyser, B., & Stuart, E. (2025). A welfare analysis of tax audits across the income distribution. *Quarterly Journal of Economics, 140*(1), 63–112. https://doi.org/10.1093/qje/qjae037

Brockmeyer, A., & Sáenz Somarriba, M. (2025). Electronic payment technology and tax compliance: Evidence from Uruguay's financial inclusion reform. *American Economic Journal: Economic Policy, 17*(1), 242–272. https://doi.org/10.1257/pol.20220434

Das, S., Gadenne, L., Nandi, T., & Warwick, R. (2023). Does going cashless make you tax-rich? Evidence from India's demonetization experiment. *Journal of Public Economics, 224*, 104907. https://doi.org/10.1016/j.jpubeco.2023.104907

© IIARJournal. All rights reserved.

de Boer, N., & Raaphorst, N. (2023). Automation and discretion: Explaining the effect of automation on how street-level bureaucrats enforce. *Public Management Review, 25*(1), 42–62. https://doi.org/10.1080/14719037.2021.1937684

Decarolis, F., Fisman, R., Pinotti, P., & Vannutelli, S. (2025). Rules, discretion, and corruption in procurement: Evidence from Italian government contracting. *Journal of Political Economy Microeconomics, 3*(2), 213–254. https://doi.org/10.1086/732654

Elzayn, H., Smith, E., Hertz, T., Guage, C., Ramesh, A., Fisher, R., Ho, D. E., & Goldin, J. (2025). Measuring and mitigating racial disparities in tax audits. *Quarterly Journal of Economics, 140*(1), 113–163. https://doi.org/10.1093/qje/qjae027

Gallien, M., & van den Boogaard, V. (2023). Formalization and its discontents: Conceptual fallacies and ways forward. *Development and Change, 54*(3), 490–513. https://doi.org/10.1111/dech.12768

Grimmelikhuijsen, S. (2023). Explaining why the computer says no: Algorithmic transparency affects the perceived trustworthiness of automated decision-making. *Public Administration Review, 83*(2), 241–262. https://doi.org/10.1111/puar.13483

Kotsogiannis, C., Salvadori, L., Karangwa, J., & Mukamana, T. (2024). Do tax audits have a dynamic impact? Evidence from corporate income tax administrative data. *Journal of Development Economics, 170*, 103292. https://doi.org/10.1016/j.jdeveco.2024.103292

Lediga, C., Riedel, N., & Strohmaier, K. (2025). What you do (not) get when expanding the net: Evidence from forced taxpayer registrations in South Africa. *Journal of Development Economics, 172*, 103388. https://doi.org/10.1016/j.jdeveco.2024.103388

Mascagni, G., Dom, R., Santoro, F., & Mukama, D. (2023). The VAT in practice: Equity, enforcement, and complexity. *International Tax and Public Finance, 30*(2), 525–563. https://doi.org/10.1007/s10797-022-09743-z

Moore, M. (2023). Tax obsessions: Taxpayer registration and the "informal sector" in sub-Saharan Africa. *Development Policy Review, 41*(1), e12649. https://doi.org/10.1111/dpr.12649

Muralidharan, K., Niehaus, P., & Sukhtankar, S. (2025). Identity verification standards in welfare programs: Experimental evidence from India. *Review of Economics and Statistics, 107*(2), 372–392. https://doi.org/10.1162/rest_a_01296

Okunogbe, O., & Santoro, F. (2023). The promise and limitations of information technology for tax mobilization. *The World Bank Research Observer, 38*(2), 295–324. https://doi.org/10.1093/wbro/lkac008

Okunogbe, O., & Tourek, G. (2024). How can lower-income countries collect more taxes? The role of technology, tax agents, and politics. *Journal of Economic Perspectives, 38*(1), 81–106. https://doi.org/10.1257/jep.38.1.81

Santoro, F., & Mascagni, G. (2023). Visual nudges: How deterrence and equity shape tax attitudes and behaviour in Rwanda. *Journal of Behavioral and Experimental Economics, 107*, 102118. https://doi.org/10.1016/j.socec.2023.102118

Straub, V. J., Morgan, D., Bright, J., & Margetts, H. (2023). Artificial intelligence in government: Concepts, standards, and a unified framework. *Government Information Quarterly, 40*(4), 101881. https://doi.org/10.1016/j.giq.2023.101881

Szucs, F. (2024). Discretion and favoritism in public procurement. *Journal of the European Economic Association, 22*(1), 117–160. https://doi.org/10.1093/jeea/jvad017

Weigel, J. L., & Kabue Ngindu, E. (2023). The taxman cometh: Pathways out of a low-capacity trap in the Democratic Republic of the Congo. *Economica, 90*(360), 1362–1396. https://doi.org/10.1111/ecca.12489

© IIARJournal. All rights reserved.