

AI-Driven Anti-Money Laundering Systems and Compliance Costs in Malaysian Fintech Smes: the Roles of Operational Efficiency and Toe Readiness

Edmund Mok Kar Meng^{1*}, Norizah Mohamad²

^{1,2}University Malaysia of Computer Science and Engineering (UNIMY), Malaysia

***Corresponding author**

Edmund Mok Kar Meng

University Malaysia of Computer Science and Engineering (UNIMY), Malaysia

Email ID:- edmund_1221@outlook.com

ABSTRACT

Anti-money laundering (AML) compliance has become more technologically-intensive and costly, creating a difficult trade-off for small and medium-sized financial technology (FinTech) firms. This conceptual paper develops a framework explaining how AI-powered AML systems could reduce regulatory compliance costs for Malaysian FinTech SMEs and how these cost effects depend on operational and organisational conditions. Informed by AI-AML research, the Technology-Organisation-Environment (TOE) framework, and the Resource-Based View (RBV), the article contends that artificial intelligence will not automatically reduce compliance expenditure. In contrast, operational efficiency can be enhanced through machine learning, natural language processing, and graph-based monitoring, which can help reduce unnecessary alert review, speed up investigation and reporting, and shift compliance capacity from low-risk to higher-risk cases. Thus, operational efficiency is identified as the primary mediating mechanism between AI-AML adoption and compliance costs. Technological preparedness, organisational preparedness, and environment-related pressures are theorised as boundary conditions that strengthen/weaken the value realised from adoption. Due to tightening AML obligations and Bank Negara Malaysia's simultaneous advancement of policy agenda for responsible AI adoption in financial services, Malaysia is a timely setting. In the article, a parsimonious framework and testable propositions have been provided for empirical validation through PLS-SEM and explanatory interviews with Malaysian FinTech SMEs.

Keywords: artificial intelligence; anti-money laundering; FinTech SMEs; operational efficiency; compliance costs; TOE framework; Malaysia

INTRODUCTION

There is a growing structural tension altering AML compliance. Firms must be alert to increasingly sophisticated methods of money laundering, conduct ongoing customer monitoring throughout the relationship, apply risk-sensitive due diligence, and file high-quality suspicious transaction reports. Compliance functions are under pressure to rein in spending, manage increasing transaction volumes, and avoid slowing legitimate customers with unnecessary friction. Small and Medium Enterprises (SMEs) in Financial Technology (FinTech) often find themselves in such a conundrum. Their digital investment may be higher, but they do not have the large incumbent banks' budget, technology, compliance headcount or historical data assets.

Artificial intelligence offered as a solution to the issue is not feasible. Machine learning can be applied to AML operations in various ways, such as ranking or classifying transaction-monitoring alerts, using natural language processing for adverse media and sanctions screening, applying graph methods to reveal links across accounts and entities, and providing automated decision support to speed up customer risk assessment. The tasks help alleviate a conventional weakness of rule-based techniques: a large share of alerts still needs to be examined by a human, not because they represent suspicious behaviour, but because they do not. False positives call for analysts' investigation. They create queues, consume time and raise follow-up compliance costs. AI most certainly has a credible mechanism for lowering costs, although it is not automatic.

Live banking applications provide evidence of the difference. Halford et al. (2025) created a machine-learning scoring layer for a transaction-monitoring system. By the end of the observation period, 19% of the alerts were automatically hibernated (deemed low risk), and 18% were automatically escalated (deemed high risk). Furthermore, the time taken to investigate and file SARs dropped 61%. As noted by De Oliveira and Leal (2024), the use of the Random Forest approach significantly reduces false positives and is also linked to operational cost efficiency. Even if that is the case, it does not mean the evidence suggests that adopting AI will reduce the net compliance cost for a resource-constrained SME. Purchasing technology or validating models will be expensive if done with the organisation's needs in mind. The essential inquiry of this research is not whether AI technology improves Anti-Money Laundering (AML) detection, but under what organisational conditions these improvements lead to lower overall compliance costs.

Malaysia is in an especially adept situation in this matter. The AML/CFT/CPF framework in the country has been continuously strengthened, which includes the 2024 AML/CFT/CPF and Targeted Financial-s sanctions pol-icy

documents and the most recent one, amendments to the Anti-Money Laundering Act A1761, which came into force on March 1, 2026 (Bank Negara Malaysia [BNM], 2024, 2026). According to the FATF/APG mutual evaluation release in December 2025, Malaysia had materially improved its defences against illicit finance, while still facing some effectiveness challenges, especially in converting money-laundering investigations into prosecutions and convictions (Financial Action Task Force & Asia/Pacific Group on Money Laundering [FATF/APG], 2025). Simultaneously, in August 2025, BNM issued its Discussion Paper on Artificial Intelligence in the Malaysian Financial Sector. This statement indicates the presence of a regulatory agenda on AI, which does not oppose but expects responsible adoption and governance (BNM, 2025).

The compliance stakes are higher than the gains. In 2023, Bank Negara Malaysia (BNM) imposed an RM600,000 administrative monetary penalty on TNG Digital for sanctions-screening failures. In 2024, the Securities Commission Malaysia (SC) imposed a total penalty of RM550,000 on Tokenise Technology for four breaches of the anti-money laundering and counter-terrorism financing (AML/CFT) rules, including failures in the risk-based approach and ongoing due diligence (BNM, 2023; SC, 2024). None of these instances illustrates that AI would have offset the breaches. Nevertheless, they do demonstrate that tech-enabled financial firms work under enforceable compliance obligations, and that the quality of monitoring, customer screening and risk recalibration has financial consequences

Recent publications in CINEFORUM also reflect a wider interest in AI adoption, organisational outcomes and digital finance. For instance, Sharma and S (2026) employ PLS-SEM to model AI adoption through mediating and moderating mechanisms, while Joshi, Shushri, Shalini, Shree, Gajanand and Trivedi (2026) link AI-assisted practices in banking with employee efficiency and organisational performance. The current paper brings the interdisciplinary conversation into RegTech to examine a different outcome: the costs of complying with mandatory AML obligations in resource-constrained FinTech firms.

This article develops a parsimonious conceptual model with three key elements. The initial adoption of AI will reduce regulatory compliance costs. In the second place, operational efficiency is defined as the means by which most of that effect happens. After that, the Technology-Organisation-Environment (TOE) framework helps identify boundary conditions that determine if adoption produces value. The Resource-Based View (RBV) is retained as a supplemental lens in explaining why an AI tool becomes economically valuable when complementary organisational capabilities are deployed properly. The narrower model, hence, is an exercise that is deliberately distinct from reworking an entire doctoral thesis into article form. Instead, it concentrates only on the mechanism most directly linked to cost and on the contextual factors most likely to explain heterogeneous outcomes across SMEs.

Malaysian AML Regulation and the FinTech SME Context

Spanning policy documents, legislation, supervisory expectations and enforcement, Malaysia's AML architecture for the financial sector takes a multidimensional approach. BNM was designated as the authority under the terms of the Anti-Money Laundering, Anti-Terrorism Financing, Anti-Restricted Activity Financing and Proceeds of Unlawful Activities Act 2001 (Act 613). As per the First Schedule, Reporting Institutions must perform the designated preventive obligations. The latest amendment, Act A1761, was gazetted on 14 May 2025 and commenced on 1 March 2026 (BNM, 2026). According to BNM (2024), financial institutions are expected to manage their own money laundering, terrorism financing, and proliferation financing risk exposure (AML/CFT/CPF), and targeted financial sanctions risk, on a risk-based approach, in line with the revised policy document issued on 5 February 2024, effective 6 February 2024.

Fintech businesses are concerned about these requirements because digital business models can easily handle higher transaction volumes without requiring scale compliance headcount. Payment systems, e-money issuers, remittance firms, digital lenders, identity-verification services, and related financial intermediaries may process many onboarding events and transactions despite being relatively small entities. Thus, the economic burden of compliance is partly a matter of scale. Controls often incorporate fixed or semi-fixed elements. Even if the firm's revenue base is much smaller than that of an incumbent bank, a licensed monitoring system, compliance personnel, independent reviews, legal advice, model validation, and a governance process must be in place.

In the context of the proposed research, a FinTech SME refers to a Malaysian financial technology firm engaged in regulated activities that meets the Malaysian SME national criteria for the service and other sectors. The SME Corporation Malaysia defines SMEs in these sectors as any firm whose sales turnover does not exceed RM20 million or which has not more than 75 full-time employees on an OR basis. The usefulness of this definition lies in its operationalisation of the model's defining resource constraint. The research problem involves organisational capacity rather than technological novelty; a small firm may be technically sophisticated but still face binding constraints in compliance staffing, specialist expertise, data engineering, and governance.

The regulators in Malaysia do not implement policies that promote innovation or hinder it. According to Bank Negara Malaysia (BNM) discussion paper (2025), a development and regulatory approach to financial sector AI adoption is outlined. The broader framework of global Monetary Policy still requires effective governance and risk management. Two kinds of pressures arise. Businesses are urged to leverage technology to improve service

and risk management; the use of artificial intelligence does not displace responsibility for results. This is an important distinction in this context: alerting may be automated, but reporting institutions remain responsible for their risk-based controls, customer due diligence, sanctions screening, and suspicious-transaction reporting.

Thus, it is a capital allocation problem from a managerial standpoint. An AI-enabled system is acquired and governed to enable counterparty investigation for intended or unintended risk. An SME FinTech must assess whether the savings from reduced manual review, faster investigations, fewer repeated tasks, and better targeting of risky counterparties justify this cost. The same technology can generate markedly different net benefits depending on factors such as data quality, system compatibility, staff expertise, management support, regulatory certainty, and vendor access. The inconsistencies in these findings pose challenges for research on AI adoption and firms' compliance costs.

AI-Driven AML Systems

Applications that harness machine learning, deep learning, natural language processing, graph analytics, or related computer techniques to improve one or more AML functions are AI-driven AML systems. These functions comprise monitoring and prioritisation of transaction alerts, customer identification and profiling, KYC/CDD/EDD workflows, sanctions and adverse media screening, network analysis, and support for suspicious transaction reporting. The key attribute is not that the system works unassisted. Instead, it employs statistical or computational learning to enhance the prioritisation, interpretation, or processing of information, exceeding static threshold rules.

The literature identifies many different application families. In supervised learning, historical outcomes with labels are used to predict the probability that a new case exhibits previously confirmed abnormal activity. Unsupervised methodologies look for anomalous behaviour in the absence of labelled examples. Methods that use graphs examine connections between customers, accounts, counterparties, devices, or transactions. The scope of Natural Language Processing (NLP) Monitoring extends beyond structured transaction data to include text, such as adverse media and case narratives, for information extraction. Han et al. (2020) view these advances as part of steps towards next-generation AML systems designed to reduce investigator workload and false alerts, while Jensen and Iosifidis (2023) arrange statistical and machine-learning methods according to client risk profiles and suspicious-behaviour alerts.

The mapping system confirms that the area has grown exponentially but remains uneven. According to a review done by Khan et al. (2024), multiple AI-based approaches have been proposed to restrict money laundering. In this regard, the authors observed considerable research activity on machine learning and related methods. Meanwhile, the literature has an emphasis on the technical model and its detection performance. The topics of organisational implementation and the economics of compliance SME-scale ROI are less developed. The problem is that a model may be technically accurate yet fail to create organisational value if it is expensive to integrate, hard to explain, poorly governed or dependent on data that the adopting firm cannot reliably generate.

Two recent empirical studies illustrate the operational value proposition. According to research by Reite et al. (2025), the utilisation of machine-learning client-risk classification has the potential to enhance the precision of AML targeting and reduce unwarranted false-positive investigations. This research used bank data on SMEs. Evidence presented by Halford et al. (2025) is particularly strong, as it examines a production deployment rather than an experimental model. When implemented as a filter on top of our client's existing rule-based transaction-monitoring system, our machine-learning 'bolt-on' layer flags low-risk alerts and escalates high-risk alerts to compliance officers. The architecture in this case is significant for SMEs because it suggests that AI adoption need not be predicated on a wholesale replacement of existing compliance infrastructure. Rather, incremental augmentation may be more economically and organisationally feasible.

AI, however, gives rise to control requirements. The data used during training and inference is incomplete and unrepresentative. As a reality, model performance may drift as customer behaviour changes. Compliance officers, auditors and regulators often find systems highly complex to explain. There is a strong incentive to use conservative thresholds, as false negatives tend to outnumber false positives. According to the FATF (2021), new technology can create both governance challenges and opportunities. For SMEs, implementing the model is not an issue of passing the test. It's whether the organisation can do Data Governance, model monitoring, human oversight and document accountability at an acceptable cost.

This evidence suggests a nuanced understanding of the AI-AML adoption process. The level of AI use should depend on how extensively it is incorporated into substantive compliance tasks, not just a software licence. An organisation has technically adopted the technology but has not realised its operational affordances by taking a hand-based approach to nearly every alert, despite having acquired an AI platform. Conversely, a firm that slowly automates low-risk triaging, prioritises investigations, and embeds onboarding and transaction-monitoring risk signals adopts deeper systems. Hence, AI-AML adoption is conceptualised as an organisational implementation construct, and its economic value depends on the extent to which the AI-AML alters compliance work through this construct.

Regulatory Compliance Costs and Operational Efficiency

The resources used to comply with mandatory AML/CFT rules impose a compliance cost. The definition includes any expenditure on direct items such as technology subscription, compliance staff salaries, external audit and legal fees, training, and reporting infrastructure. The management time wasted on indirect spending slows things down, creates friction with customers, and incurs a cost in staff capacity. The OECD (2014) cost framework for regulatory compliance is useful in that it conceptualises compliance as a resource-use problem rather than confining costs to a single accounting line. In FinTech SMEs, this broader point could have considerable economic significance: lower investigator hours could be meaningful in and of themselves, even if software spending increases, as long as total resource demand decreases or capacity increases, with no significant increase in staff. Cost is conceptually distinct from operational efficiency. Efficiency refers to the relationship between compliance outputs and the resources which are needed to produce them. In effective AML operations, alert review, customer risk assessment, sanctions screening, case investigation, and suspicious transaction reporting are conducted promptly and accurately. Analyst hours, management time, processing time, technology resources, and rework are inputs. An AI system enhances operational efficiency by enabling the company to conduct a certain level of compliance work with fewer inputs, or, alternatively, to perform a much higher level of compliance work with the same inputs, without reducing the quality of control.

This differentiation shows that operational efficiency should be modelled as a mediator. An external auditor's fee, an analyst's salary, or a regulatory filing cost is not directly reduced by AI. The quantity and quality of the work needed have changed. Reducing low-value investigations by fewer false positives. Analyst alerts for analysis should be a priority. During case preparation, automated information retrieval reduces time. Quicker screening can lessen onboarding bottlenecks. Should these changes lead to a drop in staff-hours per case, prevent unnecessary growth in headcount, reduce rework, or free up additional capacity for other compliance activities, the gain in efficiency will, in turn, lower total cost or impede its growth.

A clear example comes from Halford et al. (2025). The system they implemented significantly reduced the number of alerts requiring traditional manual treatment. Further, it has drastically cut down the investigation-to-report time. Reduced processing burden and fast case flow were the immediate result, not a reduced licence fee. De Oliveira and Leal (2024) link a lower false-positive rate to lower operational costs. According to Han et al. (2020), using AI-enabled AML will reduce investigators' workload, thereby following the same causal sequence. Hence, the conceptual chain runs as follows: AI capability changes the quality of decisions and workflows; improvements in workflows change resource utilisation; improvements in resource utilisation affect the cost of compliance.

The mediation of operational efficiency presents an opportunity for adoption failure. An SME may invest more in AI without yielding significant efficiency gains. There are various reasons why your company may not get on board with a machine learning solution. Perhaps data prep is still largely manual, legacy systems can't communicate with the new tool, or staff don't trust model output on compliance policy. Perhaps your compliance policy requires the same manual review to happen twice. For example, if you're only dealing with low volumes, the fixed implementation cost doesn't make sense. When this happens, the technology's direct cost will increase the total compliance cost. Accordingly, the mediation framework is more informative than a mere adoption-cost correlation, as it specifies the value created in the compliance process.

Developing a conceptual separation between efficiency and cost further aids measurement. The effectiveness of operations can be measured either using perceived or objective changes in alert-triage time, investigation time, KYC turnaround, false-positive burden, cases handled per person, compliance, employee workload capacity, or at constant resourcing. The money will be invested in technology spending, personnel costs, external assurance, training and opportunity costs. The separation limits contamination of constructs and leaves room for future empirical assessments to check whether firms reporting 'efficiency' improvements also report lower or more slowly growing compliance costs. Table 1 describes the definitions and conceptual roles of all the constructs.

Table 1. Construct Definitions and Conceptual Roles

Construct	Role	Operational meaning
AI-driven AML adoption	Independent variable	Extent to which AI/ML, NLP, graph analytics or related techniques are embedded in transaction monitoring, KYC/CDD, screening, risk profiling or reporting workflows.
Operational efficiency	Mediator	Improvement in compliance outputs relative to time, analyst-hours, and technology inputs, reflected in lower false-positive workload, faster case cycles, and greater capacity.
Regulatory compliance costs	Dependent variable	Direct and indirect resources used to meet AML/CFT obligations, including technology, personnel, external assurance, training, management time and process friction.

Technological readiness	Moderator (TOE)	Data quality, infrastructure compatibility, systems integration capability and technical expertise supporting AI implementation.
Organisational readiness	Moderator (TOE)	Management support, compliance expertise, budget, governance and change-management capacity.
Environmental pressure	Moderator (TOE)	Regulatory pressure, competitive pressure, vendor maturity and industry diffusion shape the economics and constraints of AI use.

THEORETICAL FOUNDATION

Technology-Organisation-Environment Framework

Tornatzky and Fleischer (1990) introduced the Technology-Organisation-Environment (TOE) framework to explain technology adoption by organisations. The framework proposed that three contextual domains determine an organisation's adoption of a technology: the characteristics and availability of the technology; the attributes and resources internal to the organisation; and the external environment in which the organisation operates. According to Baker (2012), the framework remains useful in organisational innovation settings inasmuch as it is sufficiently broad to encompass different technologies while also attending to concrete conditions affecting the adoption and realisation of value.

The technological context is particularly critical for AI-AML because it hinges on infrastructure and data. How well will the AI tool integrate into existing transaction monitoring, customer data and case management systems? Application from effort. If the quality or completeness of data is poor, models may produce unreliable risk signals. The capability to incorporate outputs into operational processes without generating manual reconciliation tasks. Knowledge of data science and technology can affect a company's ability to challenge vendor statements, modify models, and track changes. As a result, AI is being adopted for actual efficiency gains rather than complexity if organisations are technologically advanced.

This refers to the availability of resources, governance and managerial capacity of the organisations. Organisations require more than a budget for technology to adopt AI-AML. For the model to succeed, it is essential that senior management backs up process redesign. It is vital for the compliance staff to know the regulatory obligations and model constraints. Model oversight and escalation roles must be clearly defined and understood. In addition, the organisation must change established routines. Organisational readiness may be very important because, in their case, very few will determine these conditions in a small firm. Varma (2019) demonstrates the applicability of the technological, organisational and environmental (TOE) framework to the adoption of FinTech by small businesses. Rebekah and Afjal (2025) show that both technological and organisational capabilities are closely related to digital transformation and FinTech adoption in banking.

Regulatory pressure, competitive pressure, vendor ecosystems, and industry diffusion constitute the environmental context. In the context of AML, regulatory pressure does not just encourage adoption; it also shapes the extent of discretion a firm is willing to afford to AI. If a company faces stronger supervisory expectations, it will likely invest in monitoring capabilities more quickly but may also maintain more manual controls until the technology is explainable and auditable. Competitive forces often drive automation to reduce hiring friction and lower operating costs. In this model, environmental pressure is treated as a boundary condition rather than assumed to have a uniformly positive main effect.

In this article, the term 'TOE readiness' refers to the contextual conditions surrounding adoption, and this shorthand denotes three dimensions of readiness. However, while their meaning is closely related, the three developments themselves are not identical in meaning. The Internal Preparedness is Technological Readiness and Organisational Readiness. The dimension of the environment is better defined as external pressure and support. By preserving that distinction, we prevent the characterisation of regulation as an internal capability, while still allowing all three TOE domains to be modelled as moderators of the value realised from AI-AML adoption.

Resource-Based View as a Supporting Lens

RBV helps complement TOE by explaining that even with access to the same technology, not all firms achieve the same success. According to Barney (1991), a sustained advantage depends on the organisation's resources and its capability to deploy them. An AI platform is unlikely to be of strategic value in today's environment. Similar tools may be available to many businesses via cloud-based RegTech vendors. Value is created by the combination of technology and data, which is specific to the firm, compliance expertise, workflows, managerial routines, and the ability to learn.

Consequently, the RBV elucidates that organisational readiness is a condition that extends beyond pre-adoption. Management support, staff expertise, data governance, and process knowledge are complementary resources that enable the AI capability to translate into operational efficiency. The same logic applies to SMEs, which may buy sophisticated tools but have no slack resources to either redesign the process or track the model over time. TOE explains the conditions; RBV describes the mechanisms. For simplicity, TOE is the main theoretical framework, whereas the RBV assists in interpreting the organisational dimension and the heterogeneity of expected outcomes, respectively.

Conceptual Framework and Research Propositions

The proposed framework creates a pathway that balances operational efficiency and compliance costs while preserving a straight adoption-cost path. The direct pathway includes cost impacts that the operational-efficiency construct does not fully capture. Examples include vendor consolidation, reduced external service needs, and lower implementation costs. The central theoretical mechanism is a mediated path whereby AI changes compliance work, which in turn changes resource use, which then changes cost. The relationship between nominal adoption and net benefit is moderated by TOE conditions, such that the same nominal adoption can yield different net cost outcomes under varying technological, organisational, and environmental conditions.

Proposition 1 (P1): AI-driven AML adoption is negatively associated with regulatory compliance costs among Malaysian FinTech SMEs.

The first proposition reveals what the overall effect should be if it works. Intelligent and automated transaction monitoring and risk classification can reduce unnecessary manual review, automate the processing of repetitive information and improve targeting of compliance resources. Evidence from Halford et al. (2025) and Reite et al. (2025) suggests that machine learning could reduce manual workload through its production and efficiencies, respectively. The proposal is probabilistic rather than deterministic because implementation costs could outweigh savings for unprepared firms.

Proposition 2 (P2): AI-driven AML adoption is positively associated with operational efficiency.

The first stage of the mechanism is specified by this. When adoption occurs, they should reduce processing time, false-positive burden, rework, and staff-hours per investigation, and increase the number of cases that can be handled at a given resource level. Han and colleagues (2020), de Oliveira and Leal (2024), and Halford et al. (2025) provide converging evidence of improvements in workload and process efficiency across AI-enabled AML activities.

Proposition 3 (P3): Operational efficiency is negatively associated with regulatory compliance costs.

When a FinTech SME can process its AML workload with fewer analyst-hours, shorter case cycle times, less duplication, and better prioritisation, the organisation should require fewer incremental resources for a given regulatory output. Lower direct spending, avoided future hiring, reduced external support or lower opportunity cost can all manifest as cost reductions. The idea, therefore, refers to the economic effect of effective compliance operations, not to the limited economising of cash flows in a single accounting period.

Proposition 4 (P4): Operational efficiency mediates the relationship between AI-driven AML adoption and regulatory compliance costs.

The proposal for mediation is the conceptual centre of this article. Acquiring AI does not exempt businesses from conducting a cost-benefit analysis. The greatest cost advantages should be where the mechanism appreciably alters compliance. According to this proposition, the future diagnosis would yield the measure associated with the proposition at hand. Thus, when we have a non-significant direct cost effect and a significant indirect effect, this indicates that AI creates value through process improvement. Further, the opposite would hold if we had a non-significant efficiency path. Thus, this will indicate widespread adoption, poor integration, or weak organisational capability.

Proposition 5a (P5a): Technological readiness strengthens the negative relationship between AI-driven AML adoption and regulatory compliance costs.

Elements such as compatibility, data quality, and integration capability reduce implementation friction for AI-AML. Organisations with higher technological readiness should dedicate less effort to manual data assembling and exception handling. A greater portion of AI capability should be allowed to shift to usable cost savings.

Proposition 5b (P5b): Organisational readiness strengthens the negative relationship between AI-driven AML adoption and regulatory compliance costs.

The degree to which the outputs of artificial intelligence (AI) are trusted and embedded in organisational decision-making processes is determined by management commitment, adequate budget, staff expertise, a culture of compliance, and change-management capacity. As a result, organisational preparedness should enhance cost advantages by reducing redundancy between automated and manual activities and enabling managers to redesign workflows to leverage higher-value human judgment.

Proposition 5c (P5c): Environmental pressure strengthens the relationship between AI-driven AML adoption and regulatory compliance cost outcomes, although the magnitude and direction of the net effect may depend on the balance between regulatory demands and enabling support.

The environmental options call for more nuance than the internal readiness options. Intensive regulatory pressure can increase the value of automation, as firms must meet greater control and reporting obligations. However, it can also necessitate further validation and human scrutiny. When competition heats up, companies seek to improve the onboarding experience. Barriers to implementation can be reduced as vendors mature and the industry diffuses. The anticipated empirical effect is therefore a stronger contextual sensitivity, tested in the Malaysian setting rather than assumed globally.

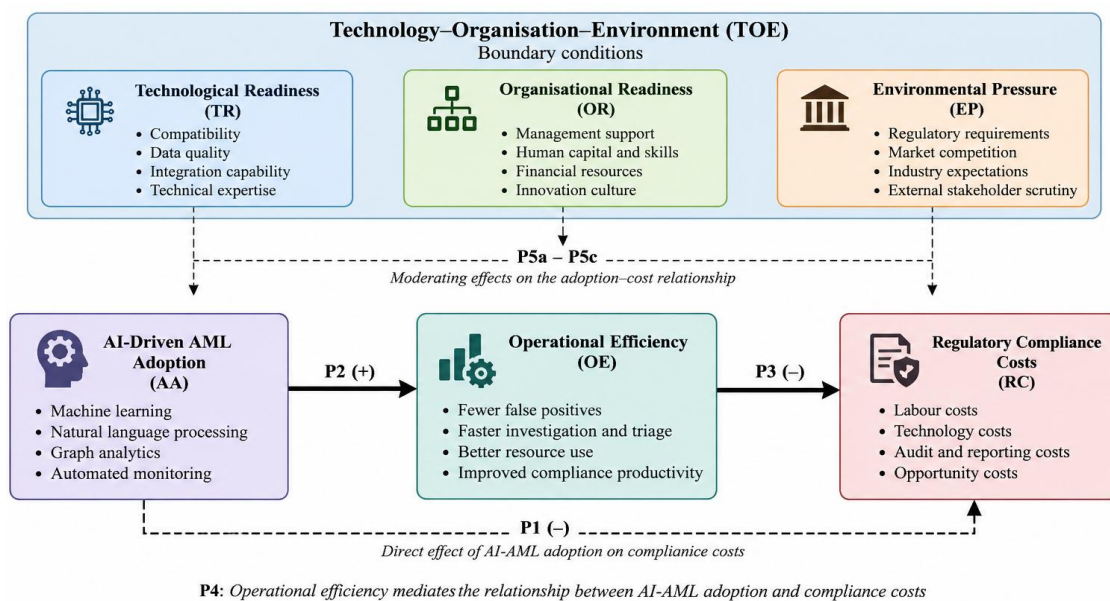


Figure 1. Conceptual Framework

Note. TOE = Technology-Organisation-Environment. The dashed P1 path represents the direct adoption-cost relationship; P2-P4 represent the mediation mechanism. P5a-P5c represent TOE boundary conditions.

Table 2. Research Propositions

Code	Proposition
P1	AI-driven AML adoption is negatively associated with regulatory compliance costs.
P2	AI-driven AML adoption is positively associated with operational efficiency.
P3	Operational efficiency is negatively associated with regulatory compliance costs.
P4	Operational efficiency mediates the relationship between AI-driven AML adoption and regulatory compliance costs.
P5a	Technological readiness strengthens the cost-reducing effect of AI-driven AML adoption.
P5b	Organisational readiness strengthens the cost-reducing effect of AI-driven AML adoption.
P5c	Environmental pressure conditions the AI-AML adoption-cost relationship; the net effect depends on the balance between regulatory demands and enabling support.

Proposed Empirical Validation

The conceptual model is not intended to be displayed as if validation has occurred, but is designed to be empirically tested. An appropriate design is a quantitative-dominant sequential explanatory design. The quantitative part can evaluate structural connections and interaction effects, while a follow-up interview can elaborate on why firms with seemingly similar adoption levels yield different operational and cost outcomes.

The target population must consist of Malaysian FinTech SMEs engaged in activities with significant AML/CFT obligations, including, but not limited to, the offering of e-money and digital wallet services, digital lending, e-KYC or identity-verification services, remittance and cross-border payments, and relevant FinTech intermediation. The ideal respondent from the organisation is the person with direct responsibility for technology-enabled compliance decisions or for AML compliance. This is the Money Laundering Reporting Office or the compliance officer. The next preferred respondent, when necessary, will be the chief executive or the chief technology officer. This approach to responding to the conceptual model corresponds to the organisational level of measurement.

The research design of this article is based on aims for some 180 organisational responses, with a minimum sample size determined by a prior power analysis. If we work across FinTech sub-segments, it will reduce the risk that the model reflects only the largest or most accessible category of firms. Some eligible companies may not have sufficient compliance staff, so that recruitment could be organised through direct contact with these organisations, industry associations and professional networks, with due regard to the confidentiality of respondents.

The extent to which AI-AML is substantively used across relevant compliance functions should be measured as opposed to an adopter/non-adopter binary. Indicators of operational efficiency should include reductions in alert-review time, investigation time, KYC turnaround time, staff hours per case, false-positive workload, and the ability to manage higher volumes without a proportional increase in resources. The cost of regulatory compliance should include direct and indirect use of resources. Technological readiness, organisational readiness, and environmental pressure should each be operationalised separately so that the three TOE dimensions do not become conflated into a single undifferentiated index.

Partial least squares structural equation modelling (PLS-SEM) is suitable for the validation proposed as the model in this study contains mediation and interaction effects, emphasises prediction and theory development, and may involve a sample restriction due to the finite population of eligible FinTech SMEs. Hypothesis testing should be preceded by measurement model assessment of indicator reliability, internal consistency, convergent validity and discriminant validity. The evaluation of collinearity, path estimates, effect sizes, explanatory power, predictive relevance, mediation and interaction effects deserves consideration in line with PLS-SEM Guideline (Hair et al., 2019). Consistent with latent-variable interaction procedures, interaction terms can be used to test moderation (Henseler & Chin, 2010).

A subsequent qualitative segment consisting of about eight to ten semi-structured interviews could then examine implementation mechanisms that survey coefficients cannot detect. The topic can include data availability, vendor selection, model governance, how you will interact with the regulator, staff trust in AI, duplicated controls, cost allocation, circumstances in which AI failed to reduce work, etc. Do not use the qualitative evidence to create causal confirmations. Its role is to help explain the heterogeneity in quantitative patterns and to refine the interpretation of the TOE boundary conditions

Table 3. Proposed Empirical Validation Design

Element	Proposed approach
Research design	Quantitative-dominant sequential explanatory design (survey followed by explanatory interviews).
Unit of analysis	FinTech SME at organisational level.
Target respondents	MLRO/compliance officer; where unavailable, CEO or CTO with direct knowledge of AML technology and costs.
Proposed quantitative sample	Approximately 180 organisational respondents; final adequacy to be confirmed through power analysis and achieved response distribution.
Core analysis	PLS-SEM measurement and structural model assessment, mediation analysis, and latent-variable moderation/interactions.
Qualitative follow-up	Approximately 8-10 semi-structured interviews to explain implementation mechanisms and divergent outcomes.

DISCUSSION

The framework reframes the AI-AML debate's focus from technical capability to organisational value creation. Much of the technical literature raises the question of whether a machine-learning model can operationalise risk more accurately than a rules-based system. That question remains relevant, but it is not enough for FinTech SMEs. A small regulated firm must know whether gains in accuracy decrease the total resources it must devote to its obligations. Thus, operational efficiency is where model performance and economic outcome is brought together through the proposed model.

This categorisation results in many theoretical consequences. To begin with, it steers clear of technological determinism. The fact that a task will become automated does not make it less costly. Whether automation

influences workflow and resource utilisation determines cost benefits. In addition, the framework separates technology adoption from technology value. TOE has been employed extensively to explain why an organisation adopts an innovation; here, we use the same domains to explain why realised value varies during or after adoption. This extension aligns with the broader technology–value literature, in which post-adoption uses and complementary capabilities affect organisational outcomes.

In addition, unlike many broad studies of digital transformation, the framework assigns a more specific role for operational efficiency. Operational efficiency is considered not a generic performance advantage. KYC throughput influences the accuracy of the KYC completion date, which can be part of a transaction alert investigation review cycle. Because of this specificity, the mediator is measurable. In turn, it gives managers a way to diagnose if an AI project is generating value without waiting for aggregate cost reductions to be visible. Another merit of the model is that it analytically centralises SMEs' resource constraints. Although vendor-based regulatory technology and cloud delivery might reduce the initial investment in sophisticated artificial intelligence, organisations will still need internal capabilities. In demand, scarce expertise is still engaged in data preparation, policy interpretation, exception handling, model oversight and vendor governance. The resource-based perspective helps explain why two similar SMEs that acquire the same product can yield different results. The more advanced user will be able to integrate the technology into personal routines and achieve higher efficiency, while the less competent user will increasingly stack similar manual tasks alongside automatic controls in parallel.

The framework also sheds light on the environmental paradox confronting regulated innovators. Regulatory pressure can create an impetus to use AI to mitigate the scale-up challenge of manual controls. However, strict accountability may impede automation if firms cannot explain or justify model outputs. A regulator that promotes responsible AI adoption with outcome-based accountability can affect adoption and costs. Malaysia serves as an informational context, as BNM has mapped out an AI policy direction and is strengthening AML expectations. Future empirical studies can reveal whether this combination serves as enabling pressure, constraining pressure, or has divergent impacts that vary across different segments of FinTech.

Value realisation has another timing implication. AI projects may increase compliance costs, at least initially, as they involve integration and testing, adjustments to controls, staff training, and parallel running. The adoption trajectory may thus be obscured by cross-sectional comparisons between adopters and non-adopters. Because they are investing in capability, early adopters may appear more costly, while mature adopters may already have removed duplicate controls and stabilised their workflows. Those who develop predictive models for the future should include adoption and implementation maturity chronologies to avoid mistaking temporary transition costs for evidence that AI cannot relieve the compliance burden.

The model also helps distinguish between digitisation and intelligent automation. Although the firm may have digitised onboarding, case management and reporting, it is not utilising machine learning for risk prioritisation. On the other hand, the AI model may be deployed only for a limited transaction-monitoring function, while the surrounding workflow is fragmented and manual. There are various organisational conditions. According to a study by Sharma and S (2026), “adoption outcomes depend on mechanisms and contextual factors, and not just exposure to technology”. When it comes to AML, the essential point is not whether AI exists but whether it is properly integrated into governed processes to change the amount, speed, and allocation of compliance work.

Practical and Policy Implications

According to the framework for FinTech SME managers, the decision to invest in AI-AML should be based on the economics of workflow rather than the vendor's feature list. A credible business case must indicate what tasks will change, how many analyst-hours are expected to be avoided, which manual controls can be retired, what new governance work will be introduced and when the net resource impact will be visible. If an organisation is unable to show a reduction in case-handling time or an increased capacity for the work, then a pilot that only reports model accuracy is incomplete.

Order of completion also matters in execution. Companies with poor data quality or siloed systems will generally see far more value from investment in their data pipeline and case-management integration than in advanced machine learning. Organisations that do not have sufficient internal capabilities should consider training and governance design before starting automation of extension. From a TOE perspective, readiness should not be a stable characteristic but rather a condition that can be deliberately developed. The sequence in place aligns with it.

Doesn't mean what it places don't get; simply, distance may be a lot smarter; proximity may be involved. The product must minimise integration costs and explain the risk score. In addition, the product should enable configurable human-review thresholds, maintain an auditable decision trail, and maximise the firm's ability to scale usage, except where large fixed infrastructure costs are necessary. It is imperative to measure vendor metrics

that reflect operational results, such as hours saved in reviewing and cases worked, rather than detection accuracy alone. Transparent evidence on implementation effort is especially needed in light of hidden integration and governance costs that can make cloud solutions more costly than they appear.

The framework demonstrates to the regulators the importance of proportionality and clarity. Small reporting institutions need to be educated on why AI is permitted and how supervisory expectations apply to them. Models must be controlled. Human oversight, validation, third-party risk, and explainability must be aspects. With clear expectations, uncertainty costs can be reduced without lowering compliance standards. Coordinated interaction with RegTech providers and regulatory sandboxes can help demarcate responsible automation from irresponsible delegation of compliance decisions.

The framework should not be interpreted as a case for substituting compliance specialists. According to the best evidence in favour of augmentation, automated prioritisation, information retrieval and pattern detection can reduce low-value manual work, leaving humans in charge to make the final call in case of complexity or high risk. Human-in-the-loop architecture is widely considered useful in AML due to the regulatory and social ramifications of false negatives, as well as the contextual interpretation of model output.

LIMITATIONS AND FUTURE RESEARCH

The current paper is conceptual and not based on primary data. Thus, empirical testing of its claims is necessary to draw any causal or policy conclusions. Apart from all this, the framework is also narrow. The emphasis is on the cost of regulatory compliance and operational efficiency. The quality of suspicious transaction reports, customer experience, fraud losses, reputational risk, and long-term innovation capability, which may matter to firms and regulators, are not included. After testing the core cost systems, the testing results can be fed into the subsequent models.

The Malaysian setting may restrict direct generalisation. The Regs, RegTech market, law enforcement, license fees, job expensiveness, and regulatory framework differ across regimes. Comparative studies across ASEAN economies could clarify whether TOE conditions operate similarly within regulatory infrastructure or vendor ecosystems. Longitudinal investigations might also be highly useful, as AI typically incurs an initial cost before efficiencies emerge. For this reason, a cross-sectional design might undercount delayed benefits and overcount those from recent adopters.

It will also require carefully designed measures. Most organisations never disclose how much they actually spend on compliance, model functioning, and enforcement weaknesses. Using perceptual measures may address barriers to disclosure, but it also introduces common-method bias and reporting error. In future reports, whenever possible, surveys should be cross-checked with and supported by real indicators. These include, but are not limited to, alert volume, average time on case, compliance headcounts, technology spending bands, and vendor reports on efficiency. The actual efficiency, or the simple migration of work to another function, can likely be assessed through qualitative interviews.

Ultimately, breakthroughs in generative AI, large language models, privacy-protecting learning and graph methods may shift the capabilities and costs of RegTech. The governance risks of predictive models for risk scoring will differ from generative systems for narrative summarisation or analyst support. Researchers should draw this distinction in future research. The framework can accommodate these changes, but construction measures should be revised as industry practice evolves.

CONCLUSION

A promising AI-facilitated AML mechanism, while potentially decreased, is conditional upon the employment of the human element in the system's decision-making process. The most compelling argument is not that AI directly reduces compliance costs. According to a survey, AI can facilitate tasks such as enhanced investigation, compliance, and task prioritisation. As a result, those operational savings could reduce or contain the resources needed to comply.

The proposed framework, therefore, places operational efficiency at the centre of the adoption-cost relationship and uses TOE conditions to explain why identical technologies may produce different outcomes across firms. Technological readiness determines whether data and systems can support the tool; organisational readiness determines whether the tool can be integrated into governed workflows; and environmental pressure shapes both the urgency and constraints of adoption. RBV provides a supporting explanation for why complementary organisational capabilities are necessary to turn a widely available technology into firm-level value.

For Malaysian FinTech SMEs, the model offers a testable way to move beyond the general claim that AI improves compliance. Future empirical validation can determine whether adoption genuinely lowers regulatory compliance costs, whether operational efficiency carries that effect, and which TOE conditions make the difference between a useful RegTech investment and an additional layer of compliance expense.

DECLARATIONS

Ethics statement: No primary human-participant data are reported in this conceptual article. Ethical approval will be required for the subsequent empirical phase in accordance with UNIMY requirements

REFERENCES

1. Baker, J. (2012). The technology-organization-environment framework. In Y. K. Dwivedi, M. R. Wade, & S. L. Schneberger (Eds.), *Information systems theory: Explaining and predicting our digital society* (Vol. 1, pp. 231-245). Springer. https://doi.org/10.1007/978-1-4419-6108-2_12
2. Bank Negara Malaysia. (2023, September 1). Bank Negara Malaysia imposes administrative monetary penalty on TNG Digital Sdn. Bhd. for non-compliances with the Financial Services Act 2013. <https://www.bnm.gov.my/-/ea-pn-07-2023>
3. Bank Negara Malaysia. (2024, February 5). Policy document on anti-money laundering, countering financing of terrorism, countering proliferation financing and targeted financial sanctions for financial institutions. <https://www.bnm.gov.my/-/pd-amlcftcpftfs-fi>
4. Bank Negara Malaysia. (2025, August 5). Discussion paper on artificial intelligence in the Malaysian financial sector. <https://www.bnm.gov.my/innovation>
5. Bank Negara Malaysia. (2026). The AMLA: Anti-money laundering / countering financing of terrorism (AML/CFT). <https://amlcft.bnm.gov.my/the-aml>
6. Barney, J. (1991). Firm resources and sustained competitive advantage. *Journal of Management*, 17(1), 99-120. <https://doi.org/10.1177/014920639101700108>
7. de Oliveira, J. R., & Leal, A. G. (2024). Enhancing anti-money laundering protocols: Employing machine learning to minimise false positives and improve operational cost efficiency. In *Proceedings of the 2024 8th International Conference on Advances in Artificial Intelligence (ICAAI '24)* (pp. 8-13). Association for Computing Machinery. <https://doi.org/10.1145/3704137.3704156>
8. Financial Action Task Force. (2021). Opportunities and challenges of new technologies for AML/CFT. <https://www.fatf-gafi.org/en/publications/Digitaltransformation/Opportunities-challenges-new-technologies-aml-cft.html>
9. Financial Action Task Force, & Asia/Pacific Group on Money Laundering. (2025). Malaysia's measures to counter money laundering, terrorist financing and proliferation financing: Mutual evaluation report. <https://www.fatf-gafi.org/en/publications/Mutualevaluations/mer-malaysia-2025.html>
10. Hair, J. F., Risher, J. J., Sarstedt, M., & Ringle, C. M. (2019). When to use and how to report the results of PLS-SEM. *European Business Review*, 31(1), 2-24. <https://doi.org/10.1108/EBR-11-2018-0203>
11. Halford, E., Gibson, I., Newfield, M., & Dhanwala, M. (2025). Developing a scoring model for managing money laundering transactions using machine learning. *Journal of Money Laundering Control*, 28(7), 30-49. <https://doi.org/10.1108/JMLC-09-2024-0152>
12. Han, J., Huang, Y., Liu, S., & Towey, K. (2020). Artificial intelligence for anti-money laundering: A review and extension. *Digital Finance*, 2(3-4), 211-239. <https://doi.org/10.1007/s42521-020-00023-1>
13. Henseler, J., & Chin, W. W. (2010). A comparison of approaches for the analysis of interaction effects between latent variables using partial least squares path modeling. *Structural Equation Modeling: A Multidisciplinary Journal*, 17(1), 82-109. <https://doi.org/10.1080/10705510903439003>
14. Jensen, R. I. T., & Iosifidis, A. (2023). Fighting money laundering with statistics and machine learning. *IEEE Access*, 11, 8889-8903. <https://doi.org/10.1109/ACCESS.2023.3239549>
15. Joshi Shushri Shalini Shree Gajanand, & Trivedi, A. (2026). Impact of artificial intelligence assisted HR practices on employee performance and profitability of public sector bank in India. *CINEFORUM*, 66(2), 923-937. <https://doi.org/10.5281/zenodo.20434156>
16. Khan, H. U., Malik, M. Z., & Nazir, S. (2024). Identifying the AI-based solutions proposed for restricting money laundering in financial sectors: Systematic mapping. *Applied Artificial Intelligence*, 38(1), 2344415. <https://doi.org/10.1080/08839514.2024.2344415>

17. OECD. (2014). OECD regulatory compliance cost assessment guidance. OECD Publishing. <https://doi.org/10.1787/9789264209657-en>
18. Rebekah, E., & Afjal, M. (2025). Exploring the mediating role of FinTech adoption and digital transformation in enhancing bank performance: A TOE framework study using a hybrid SEM-ANN approach. *SAGE Open*, 15(4). <https://doi.org/10.1177/21582440251367149>
19. Reite, E. J., Karlsen, J., & Westgaard, E. G. (2025). Improving client risk classification with machine learning to increase anti-money laundering detection efficiency. *Journal of Money Laundering Control*, 28(1), 93-107. <https://doi.org/10.1108/JMLC-03-2024-0040>
20. Securities Commission Malaysia. (2024). Administrative actions in 2024: Tokenize Technology (M) Sdn Bhd. <https://www.sc.com.my/regulation/enforcement/actions/administrative-actions/administrative-actions-in-2024>
21. Sharma, N., & S, M. (2026). AI affordance and its influence on adoption intention of salesperson. *CINEFORUM*, 66(1), 567-596. <https://revistadecineforum.com/index.php/cf/article/view/654>
22. SME Corporation Malaysia. (n.d.). SME definition. <https://smecorp.gov.my/index.php/en/policies/2020-02-11-08-01-24/sme-definition>
23. Tornatzky, L. G., & Fleischer, M. (1990). The processes of technological innovation. Lexington Books.
24. Varma, A. (2019). Fintech adoption choices of small businesses: A technology organization environment (TOE) framework study. *Accounting and Finance Research*, 8(2), 86-101. <https://doi.org/10.5430/afr.v8n2p86>