

Extending the Technology-Organization-Environment Framework with Adversarial Robustness Preparedness: A Conceptual Model of AI-Driven Cybersecurity Adoption in Nigerian Deposit Money Banks

Sreenath P. U.¹; Dr. A. Senthil Kumar²

¹PhD Research Scholar, Department of Computer Science School of Post Graduate Studies (SPGS) Skyline University Nigeria

²Dean & Associate Professor, Department of Computer Science School of Science and Information Technology (SSIT) Skyline University Nigeria

Publication Date: 2026/10/26

Abstract: Deposit Money Banks (DMBs) are embracing AI-powered intrusion and fraud detection, but there is no theory of adoption that asks if the AI is vulnerable to deception. Adversarial machine learning demonstrates how 'tricky' a small, purposely made change to the input can cause a detector to mistake malicious activity for benign activity. The TOE framework is thus insufficiently specified for the process of adopting AI cybersecurity in high-risk industries, the argument of this paper is that this is the case. By combining three literatures, the paper spans the research domains of TOE adoption studies (2022-2026), adversarial machine learning in intrusion detection, and recent financial and AI regulations, such as the CBN (Central Bank of Nigeria) 2024 Risk Based Cybersecurity Framework and Article 15 of the EU AI Act. Building on this synthesis it builds an extended TOE model with a new construct, Adversarial Robustness Preparedness (ARP): an organisation's ability to anticipate, test, defend against and govern attacks to its AI model. The model proposes eight propositions connecting technological, organisational and environmental factors and ARP with the adoption of AI for cybersecurity in Nigerian DMBs. It points to ARP as a direct cause of adoption, a partial mediator of organisational and regulatory influences, and a moderator of the impact of perceived AI capability. A measurement and validation roadmap based on structural equation modelling is provided. In addition to a testable theoretical extension, the paper provides a vocabulary for bank boards and regulators and a research agenda for new and developing financial markets.

Keywords: AI-Driven Cybersecurity; Adversarial Machine Learning; Adversarial Robustness Preparedness; Technology Adoption; TOE Framework; Intrusion Detection; Deposit Money Banks; Nigeria.

How to Cite: Sreenath P. U.; Dr. A. Senthil Kumar (2026) Extending the Technology-Organization-Environment Framework with Adversarial Robustness Preparedness: A Conceptual Model of AI-Driven Cybersecurity Adoption in Nigerian Deposit Money Banks. *International Journal of Innovative Science and Research Technology*, 11(10), 1795-1805. <https://doi.org/10.38124/ijisrt/26oct1113>

I. INTRODUCTION

Cyber risk is considered a financial-stability risk, as well as an operational risk. According to the International Monetary Fund (IMF, 2024), the probability of catastrophic losses due to cyber events has risen, and the financial sector is very vulnerable due to confidence impacts, disruptions of critical services and spillovers. The threat reports indicate that the same is true: The European Union Agency for Cybersecurity (ENISA, 2024) lists threats against availability, ransomware and threats against data as the top

threats for 2024. According to Nigeria Inter-Bank Settlement System (NIBSS) 2023, financial institutions in Nigeria lost about ₦17.67 billion to fraud (Agbo, 2024).

Regulators have responded. The Risk Based Cybersecurity Framework issued by the Central Bank of Nigeria (CBN) on 31 May 2024 and effective from 1 July 2024, mandates that the board of the banks be responsible for monitoring the framework, conduct a risk assessment annually, provide a 24-hour reporting window for banks to report any incident on the framework, and approve the

deployment of emerging technology like Artificial Intelligence and machine learning (AI and ML) by banks (CBN, 2024). On the other hand, Nigeria has also put in place the Nigeria Data Protection Act, 2023 and the Cybercrimes (Prohibition, Prevention, etc.) (Amendment) Act, 2024, and published a draft National Artificial Intelligence Strategy in August 2024 (Federal Ministry of Communications, Innovation and Digital Economy, 2024).

In response to this threat, many banks are increasingly using AI, with deep-learning intrusion detection systems (IDS), anomaly detection, and AI-based fraud detection (Achuthan et al., 2024; Kaur et al., 2023) being the three responses that are being offered. However, there's also an attack surface that AI has created for itself. There are three types of attacks: evasion attacks, where inputs are perturbed during inference to make the traffic appear benign; poisoning attacks, where the malicious traffic is seen during training to corrupt the model or training records; and privacy attacks, where the malicious traffic is used during training to extract models or training records (Vassilev et al., 2025). According to the surveys of ML-based IDS, these attacks are feasible for the network detectors and many of the proposed defences are ineffective against adaptive attacks (Alotaibi & Rassam, 2023; Ennaji et al., 2024; Khaleel et al., 2024).

Adoption research has not kept pace. The studies based on the TOE framework account for the variables of relative advantage, compatibility, organisational readiness, management support, government support, and competitive pressure which explain the adoption of AI (Badghish & Soomro, 2024; Hassan, 2024; Pinto et al., 2024). According to Nigerian banking studies, some of the factors that influence AI-related fraud detection are cost and compliance, as well as employee proficiency (Alaba et al., 2025). In any of these models, the question does not even arise of whether a bank is ready for an attack on its AI. It is possible for a bank to be very good on all of the TOE factors and yet implement the wrong detector, one that an astute adversary can circumvent.

➤ *Research Objective*

The paper presents a conceptual model extending TOE to incorporate Adversarial Robustness Preparedness (ARP) to explain the adoption of AI in the context of cybersecurity in Nigerian DMBs. It addresses three questions:

- Why is the classical TOE framework insufficient for AI-driven cybersecurity adoption in banking?
- What is adversarial robustness preparedness, and how does it relate to the established TOE contexts?
- What testable propositions and measurement approach follow from the extended model?

➤ *Approach*

Following Jaakkola's (2020) typology of conceptual articles, this is a theory adaptation paper. It builds on a framework (TOE) and adapts its scope with the second theory (adversarial machine learning) and the regulatory context of Nigerian banking.

➤ *Contributions*

The paper makes four contributions:

- A new construct: a definition and dimensions of ARP grounded in the NIST adversarial ML taxonomy (Vassilev et al., 2025).
- An extended model: a TOE extension with eight propositions on direct, mediating and moderating roles.
- A validation roadmap: indicative measurement items and an SEM-based analysis plan.
- Implications: guidance for bank boards, CISOs and the CBN.

The rest of the paper is organised as follows. TOE and adversarial machine learning is reviewed in Section 2. Related work and gap are synthesized in Section 3. The Model and propositions are presented in section 4. The validation roadmap is provided in section 5. Please note that implications are covered in section 6 and any limitations, future research and conclusions are covered in section 7.

II. THEORETICAL BACKGROUND

➤ *The Technology–Organization–Environment Framework*

TOE explains an organisation's adoption of technological innovation through three contexts (Tornatzky & Fleischer, 1990):

- The technological context: the characteristics and availability of the technology;
- The organisational context: the firm's resources, structures, readiness and leadership;
- The environmental context: industry structure, competitors, regulators and partners.

The firm-level lens of TOE is still the primary focus for AI adoption. In a 2024 meta-analysis of 12 studies and 3,398 respondents, 7 of 8 TOE factors were found to be statistically significant in contributing to the adoption of AI, with the factors of vendor partnership and organisational readiness being the most significant (Pinto et al., 2024). Badghish and Soomro (2024) found relative advantage, compatibility, human capital, market demand and government support to be significant for AI adoption among Saudi SMEs. According to Hassan (2024), technological infrastructure, leadership support and clarity of regulations are pivotal factors in Financial Services in South Africa.

The advantage of TOE is also its disadvantage when it comes to AI. It aims to be deliberately generic in its three contexts, and the researcher should explain, for a specific technology, which factors are important. Recent reviews claim that AI is distinct from previous technologies and thus needs extensions. Within each context, there are proposals for a human factor (TOE-H; Dönmez et al., 2025) and AI-specific factors (Cioli et al., 2025). Wallace et al. (2020) have added the concepts of “cyber catalysts” and “practice standards” to the TOE in the context of cybersecurity. This demonstrates that TOE lacks constructs for security adoption decisions.

➤ *Adversarial Machine Learning and AI-Driven Intrusion Detection*

Adversarial machine learning is the field of studying how attackers can take advantage of ML systems and how they can be protected. There are three main categories of attacks on predictive AI, according to NIST taxonomy (Vassilev et al., 2025):

- Evasion: manipulating inputs at deployment;
 - Poisoning: manipulating training data or models;
 - Privacy: extracting information about the model or its data.
- *Generative AI Introduces Three Additional Types of Misuse Attacks*

Small perturbations were enough to fool deep networks as shown by the Gradient Based Evasion Attacks, for example, Fast Gradient Sign Method (FGSM) (Goodfellow et al., 2015) and Projected Gradient Descent (PGD) (Madry et al., 2018). Adversarial training on these examples (Madry et al., 2018) and defensive distillation (Papernot et al., 2016) were the reference defences.

The problem is different in the case of network IDS. Network and transaction data is tabular and constrained: features are dependent on each other, and some are unchangeable for an attacker if they are to be successful in their attack. Simonetto et al. (2022) demonstrate that realistic attacks and defences must adhere to these feature space limitations. Evaluations of ML-based IDS demonstrate that adversarial examples can render malicious traffic indistinguishable, and that there are defenses such as adversarial training and detection and input transformation (Alotaibi & Rassam, 2023; Khaleel et al., 2024). Many defenses have subsequently been defeated by adaptive attacks, said Ennaji et al. (2024). That makes continuous testing, rather than a one-off defence, the realistic goal.

These attack types are related to the lifecycle of a bank's AI-powered IDS, as shown in Figure 2 (Section 4).

➤ *From Technical Robustness To Organisational Preparedness*

Robustness is not only a property of a model. It is also reliant on the organisation's ability to monitor, test and govern its models.

Practice evidence indicates that there is a gap in this area. Grosse et al. (2023) conducted a survey of 139 industrial practitioners about the number of ML defences deployed and their perceptions of the threat, which revealed that the number of ML defences deployed is dependent on the perceived threat exposure, and the perceived threat exposure is influenced by pre-existing knowledge about ML security. One reason why security teams don't give academic adversarial ML much attention is that it has moved away from the cheap ways of real attackers, according to Apruzzese et al. (2023). Mink et al. (2023) examine security practitioners' perception of ML-based security tools and their explanations. In addition, Rjoub et al. (2023) show the importance of explainability to foster trust in AI security tools.

Robustness is becoming an obligation for the organisation through the means of regulation. European Parliament & Council of the European Union (2024), Article 15 of the EU AI Act, the systems under the scope of this Act shall be robust against attacks, including data poisoning and model evasion or adversarial examples. Cyber vulnerabilities (such as model poisoning) and model risk are featured among the AI vulnerabilities that should be monitored, according to the Financial Stability Board (FSB, 2024). Robustness and stability are listed by the Monetary Authority of Singapore (MAS, 2024) as some of the major considerations in financial institutions' model risk management of AI models.

This development calls for a construct between the technology and the organisation. This paper refers to it as adversarial robustness preparedness.

III. RELATED WORK AND RESEARCH GAP

Table 1 summarizes selected studies in the three streams which are brought together in this paper, and which are representative of the past five years (2022-2026). The trend is evident. Adoption studies are based on measures of readiness, regulation and environment, not of adversarial preparedness. Adversarial ML focuses on robustness of the model rather than its uptake.

Table 1 Representative Related Studies and their Treatment of Adversarial Robustness

Study	Context	Lens / method	Key factors or findings	Adversarial robustness addressed?
Pinto et al. (2024)	AI adoption, multi-industry	TOE meta-analysis (12 studies, n = 3,398)	7 of 8 TOE factors significant; organisational readiness and vendor partnership critical	No
Badghish & Soomro (2024)	AI adoption, Saudi SMEs	TOE; PLS-SEM	Relative advantage, compatibility, human capital, market demand, government support	No
Hassan (2024)	AI adoption, South African financial services	TOE; logistic regression	Infrastructure, leadership support, regulatory clarity	No

Alaba et al. (2025)	AI fraud detection, Nigerian DMBs	Survey; ordered logit	Top management support, IT infrastructure, regulatory compliance, staff competency; cost inhibits	No
Owusu (2025)	AI adoption, Ghanaian banks (n = 331)	Resource-based view; PLS-SEM	Absorptive capacity, agility and capabilities drive adoption; AI improves operational excellence	No
Rahman et al. (2023)	AI in Malaysian banking services	Mixed methods; PLS-SEM	Bankers cite missing regulation, privacy/security and skills gaps as barriers	Security as a barrier only
Uren & Edwards (2023)	Organisational AI adoption	Qualitative, expert interviews	People, process and data readiness needed besides technology readiness	No
Wallace et al. (2020)	Cybersecurity adoption	Extended TOE (conceptual)	Adds cyber catalysts and practice standards	Not AI-specific
Grosse et al. (2023)	ML security in industry (n = 139)	Quantitative practitioner survey	Defences track perceived threat exposure; ML-security knowledge shapes threat perception	Yes — practice, not adoption
Alotaibi & Rassam (2023)	ML-based IDS	Survey	Adversarial attacks and defences for IDS; open gaps	Yes — technical only
Khaleel et al. (2024)	Network and cybersecurity ML	Systematic review (42 studies)	Defence strategies; need for reliable mitigation	Yes — technical only

➤ *Three Gaps Follow from Table 1.*

- Theoretical gap. There are no TOE-based adoption studies that explicitly feature the readiness of the adopting organisation with respect to an attack on the AI itself. From the main, security seems to be a perceived blockage (Rahman et al., 2023).
- Contextual gap. Evidence from African banking is growing (Alaba et al., 2025; Hassan, 2024; Ofuani et al., 2024; Owusu, 2025). It sees regulatory compliance as a motivator and is prior to or has nothing to do with the AI-approval requirement contained in the CBN's framework of 2024 (CBN, 2024).
- Integration gap. The adversarial ML literature is technical, and focuses on the defended model as the unit of analysis (Alotaibi & Rassam, 2023; Khaleel et al., 2024; Simonetto et al., 2022). The practitioner literature demonstrates that robustness is the result of organisational decisions (Apruzzese et al., 2023; Grosse et al., 2023) but this work has not yet been incorporated in the framework of adoption theory.

IV. CONCEPTUAL MODEL AND PROPOSITIONS

➤ *Model Overview*

The proposed model (figure 1) retains the three contexts of TOE, but introduces ARP as another, cross-cutting construct. Every context context is presented with the most relevant factor to AI-based cybersecurity in Nigerian DMBs:

- Technological context: perceived AI security capability (PASC).
- Organisational context: organisational readiness (OR).

The first category is environmental context, which is a regulatory compliance pressure plus environmental threat and competitive pressure (ETC).

In the model, ARP has three functions:

- A direct responsibility for the adoption; or
- A partial mediator through which readiness and regulation act;

A facilitator which influences the extent to which the perceived capability leads to adoption.

- AI-driven cybersecurity adoption (AICA): The level of a bank's AI-based detection and response integration into its security operations. The model is presented in Figure 1, and the definition of each construct is given in Table 2.

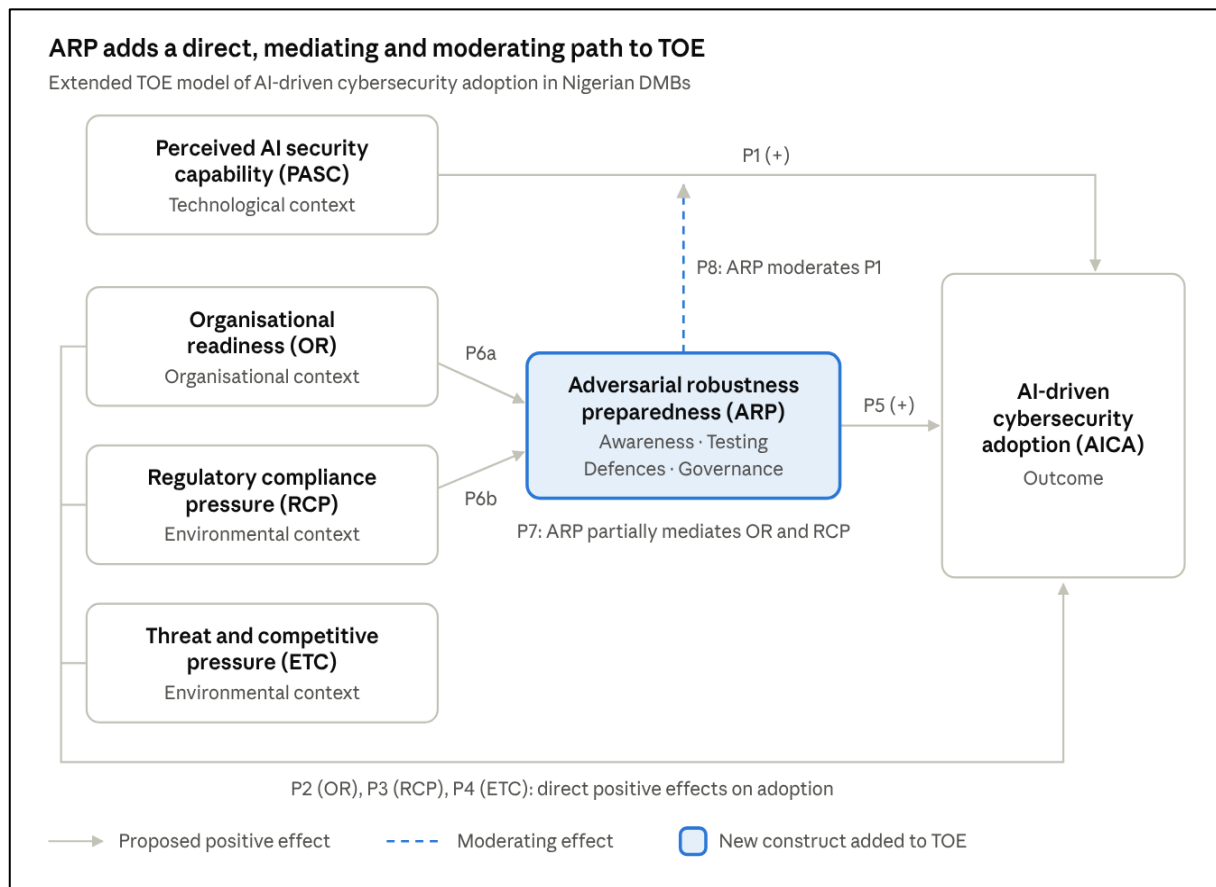


Fig 1 Proposed Extension of the TOE Framework with Adversarial Robustness Preparedness (ARP). Solid Arrows are Proposed Positive Effects (P1–P6); ARP Partially Mediates Organisational and Regulatory Effects (P7) and Moderates the Effect of Perceived Capability (P8).

Table 2 Construct Definitions and Theoretical Anchors

Construct	TOE context	Definition in this model	Anchors
Perceived AI security capability (PASC)	Technological	The extent to which decision-makers believe AI-based detection outperforms existing controls and fits existing systems (relative advantage and compatibility)	Badghish & Soomro, 2024; Kaur et al., 2023; Tornatzky & Fleischer, 1990
Organisational readiness (OR)	Organisational	The availability of infrastructure, skilled staff, data, budget and top-management support for AI security	Alaba et al., 2025; Pinto et al., 2024; Uren & Edwards, 2023
Regulatory compliance pressure (RCP)	Environmental	Pressure from binding rules and supervisory expectations to adopt and govern AI-based security	CBN, 2024; Hassan, 2024; Nigeria Data Protection Act, 2023
Environmental threat and competitive pressure (ETC)	Environmental	The perceived intensity of cyber threats and fraud, and the adoption of AI security by competing banks	Agbo, 2024; IMF, 2024; Pinto et al., 2024
Adversarial robustness preparedness (ARP)	New, cross-cutting	The organisation's capability to anticipate, test, defend against, monitor and govern attacks that target its AI models (evasion, poisoning, privacy)	European Parliament & Council of the European Union, 2024; Grosse et al., 2023; Vassilev et al., 2025
AI-driven cybersecurity adoption (AICA)	Outcome	The extent to which AI-based detection and response are integrated into the bank's security operations	Alaba et al., 2025; Owusu, 2025

➤ Defining Adversarial Robustness Preparedness

ARP is about an organisation's ability to forecast, test, protect, counter and manage hostile attacks against the security models they use for AI. It cannot be an attribute of any one model; it is an organisational capability. A bank

with a high ARP can have a model that it later finds sensitive and still identify, contain and fix the issue.

Based on the NIST taxonomy (Vassilev et al., 2025), the literature in IDS (Alotaibi & Rassam, 2023) and

practical experience (Apruzzese et al., 2023; Grosse et al., 2023), ARP is described by the four dimensions:

- Threat awareness and modelling: staff and leaders are aware of potential threats such as evasion, poisoning and privacy attacks and know which of these are realistic with the bank's data and attackers (Apruzzese et al., 2023; Grosse et al., 2023; Vassilev et al., 2025).
- Robustness testing: models are evaluated prior to deployment and periodically in response to adversarial examples, such as attacks that obey real feature constraints (Ennaji et al., 2024; Simonetto et al., 2022).
- Defensive engineering: defences like adversarial training (Madry et al., 2018), distillation (Papernot et al., 2016), input validation and ensembles are used, and data

pipelines are defended from poisoning (Alotaibi & Rassam, 2023; Khaleel et al., 2024).

- Governance and monitoring: accountability of the board and CISO for AI model risk, model inventories, monitoring for drift and attack, explainability for analysts and incident reporting in line with regulation (CBN, 2024; FSB, 2024; MAS, 2024; Rjoub et al., 2023).

ARP is hypothesised to have a reflective first-order relationship for parsimony reasons. It is discussed as a higher-order construct, with these four dimensions, in section 5.

Figure 2 shows where each attack class enters a bank's AI-driven IDS, and therefore where each ARP dimension acts.

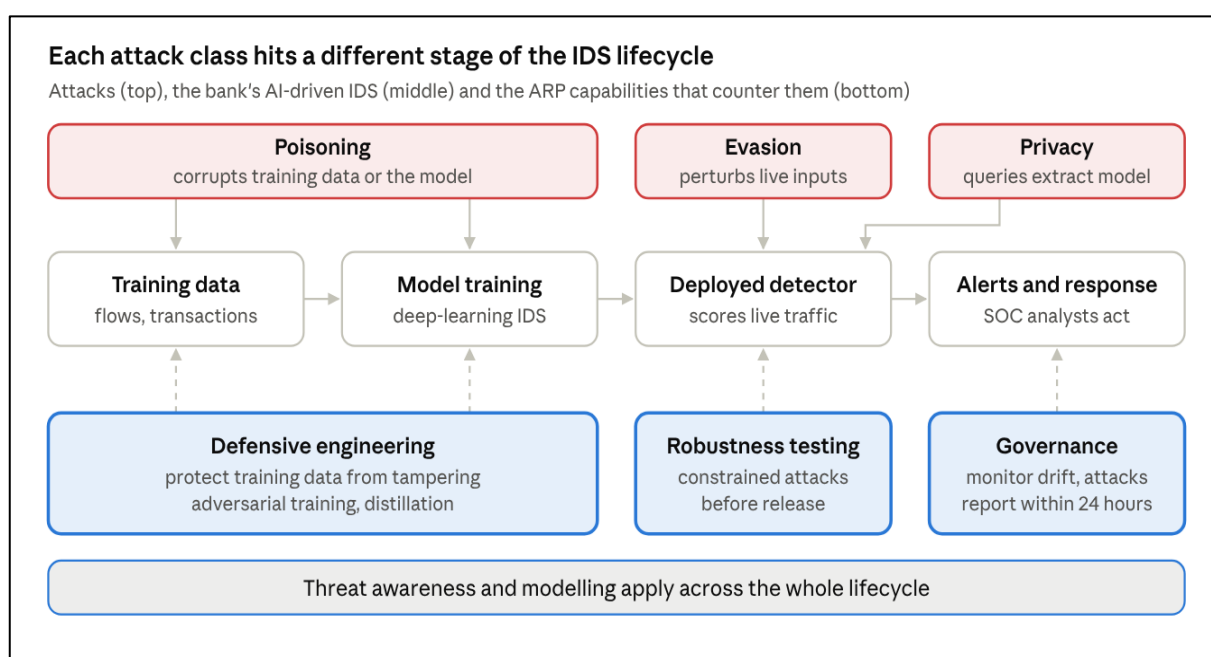


Fig 2 Adversarial Attack Classes, Following the NIST Taxonomy (Vassilev et al., 2025), Mapped to the Lifecycle of a Bank's AI-Driven IDS, with the ARP Capabilities that Counter Each One. Dashed Arrows Show where Each Capability Acts.

➤ Propositions

- Technological context. The more decision-makers believe that AI is more effective at detecting intrusions and fraud than rule-based defences and that it can integrate with current systems, the more likely it is to be adopted. This is similar to the strong impact of relative advantage and compatibility found in the literature on TOE research (Badghish & Soomro, 2024; Pinto et al., 2024).
- ✓ P1. Perceived AI security capability is positively associated with AI-driven cybersecurity adoption in DMBs.
- Organisational context. Organisational context. Adoption of AI is always predicted by readiness in infrastructure, skills, data and leadership support (Pinto et al., 2024; Uren & Edwards, 2023). The Nigerian context of banks

suggests that the adoption of AI fraud detection is driven by the speed of IT infrastructure, the strength of top management support, and the competence of staff in Nigerian banks (Alaba et al., 2025).

- ✓ P2. Organisational readiness is positively associated with AI-driven cybersecurity adoption in DMBs.
- Environmental context. Regulation shapes AI adoption in finance (Alaba et al., 2025; Hassan, 2024). The CBN framework not only calls for investment in cybersecurity, but it also requires CBN clearance to deploy AI (CBN, 2024). Regulation is therefore expected to push banks towards governed adoption.
- ✓ P3. Regulatory compliance pressure is positively associated with AI-driven cybersecurity adoption in DMBs.

- ✓ Moreover, the growing intensity of the threat or the adoption by other players increases the perceived cost of not adopting.
- ✓ P4. Environmental threat and competitive pressure is positively associated with AI-driven cybersecurity adoption in DMBs.
- Adversarial robustness preparedness. A security tool that can itself be evaded is a liability. For banks unable to test or govern their models, the risk of models is unacceptable to the boards and supervisors (CBN, 2024; FSB, 2024). In high ARP areas, there are times when decision-makers can afford to use AI in critical security roles and continue it after incidents.
- ✓ P5. Adversarial robustness preparedness is positively associated with AI-driven cybersecurity adoption in DMBs.
- ARP as a mediator. Preparedness has to be built. Skilled staff, data pipelines and leadership attention (organisational readiness) are prerequisites for robustness testing and model governance (Grosse et al., 2023; Uren & Edwards, 2023). Adversarial attacks and accountable AI risk management are increasingly coming into regulatory focus, with resilience against adversarial attacks increasingly required, while

accountable AI risk management is required (CBN, 2024; European Parliament & Council of the European Union, 2024; MAS, 2024).

- ✓ P6a. Organisational readiness is positively associated with adversarial robustness preparedness.
- ✓ P6b. Regulatory compliance pressure is positively associated with adversarial robustness preparedness.
- ✓ P7. Adversarial robustness preparedness partially mediates the effects of organisational readiness and regulatory compliance pressure on AI-driven cybersecurity adoption.
- ARP as a moderator. ARP as a moderator. When the bank can prove it, the perceived capability is a reality. Where an IDS makes high accuracy claims, these are less persuasive if the bank is not able to demonstrate accuracy under adversarial inputs. The acceptance of ML-based security applications also relies on the practitioners' understanding of their behaviour (Mink et al., 2023; Rjoub et al., 2023).
- ✓ P8. Adversarial robustness preparedness positively moderates the relationship between perceived AI security capability and AI-driven cybersecurity adoption, so that the relationship is stronger when ARP is high. Table 3 summarises the propositions.

Table 3 Summary of Propositions

No.	Relationship	Expected direction	Role of ARP
P1	PASC → AICA	Positive	—
P2	OR → AICA	Positive	—
P3	RCP → AICA	Positive	—
P4	ETC → AICA	Positive	—
P5	ARP → AICA	Positive	Direct antecedent
P6a	OR → ARP	Positive	Outcome of readiness
P6b	RCP → ARP	Positive	Outcome of regulation
P7	OR, RCP → ARP → AICA	Partial mediation	Mediator
P8	ARP × PASC → AICA	Strengthens the effect	Moderator

V. ROADMAP FOR EMPIRICAL VALIDATION

The model is designed to be tested with a cross-sectional survey of IT, cybersecurity, risk and compliance staff in Nigerian DMBs. The unit of analysis is the bank, with key informants reporting on it. The data would be analysed using the structural equation modelling.

➤ Indicative Measurement Items

There are existing scales that can be modified for PASC, OR, RCP, ETC, and AICA from the TOE studies cited in Table 2. ARP is in need of new items. Table 4 presents a preliminary pool, using the four dimensions and grounded in the literature. All items would use a five- or seven-point agreement scale. They need to be refined by expert review and a pilot before they can be used.

Table 4 Proposed Initial Items for Adversarial Robustness Preparedness

Dimension	Item (respondent rates agreement for their bank)	Anchor
Threat awareness	ARP1: Our security staff understand how attackers can manipulate the inputs of our AI detection models.	Grosse et al., 2023; Vassilev et al., 2025
Threat awareness	ARP2: We assess which attacks on our AI models (evasion, data poisoning, model theft) are realistic for our bank.	Apruzzese et al., 2023; Vassilev et al., 2025
Robustness testing	ARP3: Our AI security models are tested against adversarial examples before deployment.	Alotaibi & Rassam, 2023; Ennaji et al., 2024
Robustness testing	ARP4: Adversarial tests use realistic changes that an attacker could actually make to network or transaction data.	Simonetto et al., 2022
Defensive	ARP5: We apply defences such as adversarial training or input validation to	Khaleel et al., 2024; Madry

Dimension	Item (respondent rates agreement for their bank)	Anchor
engineering	our AI security models.	et al., 2018
Defensive engineering	ARP6: Training data for our AI models is protected against tampering (poisoning).	Vassilev et al., 2025
Governance and monitoring	ARP7: A named executive is accountable for the risk of attacks on our AI models.	CBN, 2024; MAS, 2024
Governance and monitoring	ARP8: We monitor deployed AI models for signs of attack or performance drift and can report incidents within regulatory timelines.	CBN, 2024; FSB, 2024

➤ Analysis Plan

Table 5 sets out the validation stages and their usual decision criteria. Because P8 involves an interaction and ARP may be specified as a higher-order construct, PLS-

SEM is appropriate for an exploratory first test (Hair et al., 2022). Covariance-based SEM can then confirm the measurement model with global fit indices (Hu & Bentler, 1999).

Table 5 Validation Stages and Decision Criteria

Stage	Purpose	Method	Criterion
1. Content validity	Check ARP items cover the construct	Expert panel (security, risk, compliance)	Item retained if experts agree it is relevant and clear
2. Pilot	Check wording and reliability	Small pilot sample; item analysis	Remove ambiguous or low-correlation items
3. Reliability	Internal consistency	Cronbach's α ; composite reliability	≥ 0.70 (Hair et al., 2022)
4. Convergent validity	Items reflect their construct	Outer loadings; AVE	Loadings ≥ 0.708 ; AVE ≥ 0.50 (Hair et al., 2022)
5. Discriminant validity	ARP distinct from OR and RCP	HTMT ratio	< 0.85 (conservative) or < 0.90 (Henseler et al., 2015)
6. Model fit (CB-SEM)	Global fit of the measurement model	CFI, TLI, RMSEA, SRMR	CFI/TLI ≥ 0.95 ; RMSEA ≤ 0.06 ; SRMR ≤ 0.08 (Hu & Bentler, 1999)
7. Structural model	Test P1–P6	Path coefficients, bootstrapping	Significant paths in the predicted direction (Hair et al., 2022)
8. Mediation and moderation	Test P7 and P8	Bootstrapped indirect effects; interaction term	Confidence interval of the indirect effect or interaction excludes zero (Hair et al., 2022)

The threat to the model is most likely discriminant validity between ARP and organisational readiness. Strong ARP can be reported by a bank that is well prepared in general. At stage 5, the question directly asks respondents if they distinguish between the two. The data are obtained from a single respondent, so common-method bias must be minimised in the processes and be statistically checked.

➤ Linking the Model to Technical Evidence

Survey measures of ARP can be triangulated with technical evidence. The accuracy, detection rate and false alarm rate of a bank's (or a research team's) IDS before and after adversarial attacks like FGSM and PGD (Goodfellow et al., 2015; Madry et al., 2018) can be measured when using an AI-powered IDS. This measured robustness gap can then be compared with the bank's ARP (self-reported). A small gap at the banks reporting high ARP would be a positive indicator of the construct's validity.

VI. DISCUSSION AND IMPLICATIONS

➤ Theoretical Implications

The model makes three theoretical moves.

- It extends TOE's scope. To drive adoption of a security technology that is a target, generic factors like relative advantage, readiness and regulation are not sufficient. It

builds upon the tradition of extension of Wallace et al. (2020) and recent TOE work for AI (Cioli et al., 2025; Dönmez et al., 2025), as well as introducing a construct based on adversarial ML.

- It treats robustness as an organisational capability. Technical studies view robustness as a property of models developed (Alotaibi & Rassam, 2023; Khaleel et al., 2024). Organisations' actions regarding robustness are evident from the practitioners and are shown to be dependent on threat perception and knowledge (Grosse et al., 2023). ARP provides a means to measure that ability to adoption research.
- It specifies mechanisms, not only predictors. The model proposes two mediating processes: mediation (P7) and moderation (P8) that account for the action of readiness and regulation and when perceived capability becomes adoption. This responds to the requests for going beyond the lists of TOE factors (Pinto et al., 2024).

➤ Practical Implications for Banks

- Ask "can it be fooled?" before deployment. The figures of vendor accuracy for benchmark datasets do not reflect the ability to withstand adaptive attackers (Apruzzese et al., 2023; Ennaji et al., 2024). Procurement and model validation should require adversarial test results using realistic, constrained attacks (Simonetto et al., 2022).

- Build ARP alongside adoption. Staff training on adversarial threats, a model inventory, and robustness testing prior to release, as well as robustness monitoring after release, are relatively low cost as compared to an evaded detector.
- Assign accountability. There is already a CBN framework which mandates board supervision and reporting on cybersecurity on a quarterly basis (CBN, 2024). If AI model risk (adversarial testing results) were included in those reports, boards would also be able to see ARP.

➤ *Implications for Regulators and Policy*

- Specify what AI approval requires. The CBN has said that it must approve the deployment of AI before it happens. (CBN, 2024). This could lead to explicit approval criteria, in line with Article 15 of the EU AI Act (European Parliament & Council of the European Union, 2024) and the MAS model-risk practice (MAS, 2024).
- Link national AI policy to sector rules. The National AI Strategy is currently being drafted, which includes an AI risk-management framework (Federal Ministry of Communications, Innovation and Digital Economy, 2024). It will harmonise the CBN cybersecurity rules and the Nigeria Data Protection Act, 2023, and minimise overlapping compliance obligations for banks.
- Treat AI vulnerabilities as a stability issue. The FSB identifies the cyber vulnerabilities as one of the financial-stability concerns in the context of AI (FSB, 2024). Increasingly, it is logical to monitor controls that depend on AI.

VII. LIMITATIONS, FUTURE RESEARCH AND CONCLUSION

➤ *Limitations*

- The model is conceptual. The propositions are taken from literature and regulation and have not been tested. Their direction and strength may differ in practice.
- The context is specific. The model is developed in the context of CBN 2024 for DMBs in Nigeria. The regulatory construct needs to be re-specified when transferred to other countries or to non-bank financial institutions.
- ARP is new. Its dimensions and items (Table 4) are proposals. They might require revision following expert review and pilot testing, and the relationship between ARP and general organisational readiness needs to be demonstrated empirically.
- Some sources are preprints or practitioner reports. These are identified in the reference list and should be revised as soon as peer-reviewed versions are available (Ennaji et al., 2024; Pinto et al., 2024).

➤ *Future Research*

- Validate ARP. Convelop and confirm an ARP scale using expert review and pilot testing and confirmatory analysis, including testing of the four-dimensional higher-order structure.
- Test the propositions. Research Nigerian DMBs and draw comparisons with the banks operating in other African markets such as Hassan (2024) and Owusu (2025).
- Triangulate with technical evidence. Compare self-reported ARP with measured robustness of deployed or prototype AI-driven IDS under FGSM, PGD and constrained attacks (Section 5.3).
- Study change over time. Support ARP and adoption on an ongoing basis as ARP becomes more developed and the national AI strategy evolves.
- Extend to generative AI. In addition, misuse attacks for generative AI (Vassilev et al., 2025) are added by NIST. Further research should determine if AI assistants in Security Operations Centres fall under the ARP.

➤ *Conclusion*

The technology of AI-based cybersecurity holds the promise of detecting fraud at Nigerian banks faster and with greater accuracy, but it also provides a new target for attackers: the model. Current research on TOE adoption identified factors that indicate whether the bank adopted TOE, but not how ready the bank is for the TOE to be attacked.

In this paper, adversarial robustness preparedness is defined as an organisational capability that has four dimensions: threat awareness, robustness testing, defensive engineering, and governance and monitoring. It incorporated ARP into an expanded TOE model comprising eight propositions, giving a measurement and validation process roadmap.

The central claim is simple. Sustainable uptake of AI security in a high-risk, regulated sector is ultimately about what the AI can do but whether or not the bank can demonstrate that it continues to do it when attacked.

DECLARATIONS

- *Funding:* [to be completed].
- *Conflicts of interest:* The author(s) declare no conflict of interest.
- *Data availability:* Not applicable; no new data were generated in this conceptual study.
- *Use of AI tools:* [to be completed in line with the target journal's policy].

REFERENCES

- [1]. Achuthan, K., Ramanathan, S., Srinivas, S., & Raman, R. (2024). Advancing cybersecurity and privacy with artificial intelligence: Current trends and future research directions. *Frontiers in Big Data*, 7, Article 1497535. <https://doi.org/10.3389/fdata.2024.1497535>
- [2]. Agbo, C. (2024, April 30). Financial institutions lost N17.6 billion to fraud in 2023 – Report [Reporting NIBSS, *Annual fraud landscape, January–December 2023*]. *21st Century Chronicle*. <https://21stcenturychronicle.com/financial-institutions-lost-n17-6-billion-to-fraud-in-2023-report/>
- [3]. Alaba, J. S., Ahmed, S. J., Farida, A. P., & Oluwatosin, O. V. (2025). Adoption of AI-driven fraud detection system in the Nigerian banking sector: An analysis of cost, compliance, and competency. *Financial Economics Letters*, 4(2), 40–53. <https://doi.org/10.58567/fel04020004>
- [4]. Alotaibi, A., & Rassam, M. A. (2023). Adversarial machine learning attacks against intrusion detection systems: A survey on strategies and defense. *Future Internet*, 15(2), Article 62. <https://doi.org/10.3390/fi15020062>
- [5]. Apruzzese, G., Anderson, H. S., Dambra, S., Freeman, D., Pierazzi, F., & Roundy, K. A. (2023). "Real attackers don't compute gradients": Bridging the gap between adversarial ML research and practice. In *2023 IEEE Conference on Secure and Trustworthy Machine Learning (SaTML)*. IEEE.
- [6]. Badghish, S., & Soomro, Y. A. (2024). Artificial intelligence adoption by SMEs to achieve sustainable business performance: Application of technology–organization–environment framework. *Sustainability*, 16(5), Article 1864. <https://doi.org/10.3390/su16051864>
- [7]. Central Bank of Nigeria. (2024). *Risk-based cybersecurity framework and guidelines for deposit money banks and payment service banks*. https://www.cbn.gov.ng/Out/2024/BSDBSD/CBN%20Risk-Based%20Cybersecurity%20Framework%20for%20DMBs%20and%20PSBs_2024.pdf
- [8]. Cioli, A., Marinelli, L., La Rocca, A., & Perna, A. (2025). AI adoption in B2B marketing: An analysis of technological, organizational, and environmental factors. *Micro & Macro Marketing*, (3), 491–527. <https://doi.org/10.1431/118666>
- [9]. Cybercrimes (Prohibition, Prevention, etc.) (Amendment) Act, 2024 (Nigeria). <https://digitalpolicyalert.org/change/12205>
- [10]. Dönmez, S., Tuncay Çelikel, A., Soykut Sarıca, Y. P., & Develi, E. İ. (2025). Understanding AI adoption at organizations: Literature review of TOE framework. *PressAcademia Procedia*, 21(1), 64–69. <https://doi.org/10.17261/Pressacademia.2025.1994>
- [11]. Ennaji, S., De Gaspari, F., Hitaj, D., Kbidı, A., & Mancini, L. V. (2024). *Adversarial challenges in network intrusion detection systems: Research insights and future prospects* (arXiv:2409.18736) [Preprint]. arXiv. <https://doi.org/10.48550/arXiv.2409.18736>
- [12]. European Parliament & Council of the European Union. (2024). Regulation (EU) 2024/1689 laying down harmonised rules on artificial intelligence (Artificial Intelligence Act). *Official Journal of the European Union*, L series. <https://ai-act-service-desk.ec.europa.eu/en/ai-act/article-15>
- [13]. European Union Agency for Cybersecurity. (2024). *ENISA threat landscape 2024*. <https://www.enisa.europa.eu/publications/enisa-threat-landscape-2024>
- [14]. Federal Ministry of Communications, Innovation and Digital Economy. (2024). *National artificial intelligence strategy* [Draft]. National Information Technology Development Agency & National Centre for Artificial Intelligence and Robotics.
- [15]. Financial Stability Board. (2024). *The financial stability implications of artificial intelligence*. <https://www.fsb.org/uploads/R14112024.pdf>
- [16]. Goodfellow, I. J., Shlens, J., & Szegedy, C. (2015). Explaining and harnessing adversarial examples. In *International Conference on Learning Representations (ICLR)* (2015). <https://arxiv.org/abs/1412.6572>
- [17]. Grosse, K., Bieringer, L., Besold, T. R., Biggio, B., & Krombholz, K. (2023). Machine learning security in industry: A quantitative survey. *IEEE Transactions on Information Forensics and Security*, 18, 1749–1762. <https://doi.org/10.1109/TIFS.2023.3251842>
- [18]. Hair, J. F., Hult, G. T. M., Ringle, C. M., & Sarstedt, M. (2022). *A primer on partial least squares structural equation modeling (PLS-SEM)* (3rd ed.). Sage.
- [19]. Hassan, A. S. (2024). Factors driving artificial intelligence adoption in South Africa's financial services sector. *Academic Journal of Interdisciplinary Studies*, 13(5). <https://doi.org/10.36941/ajis-2024-0173>
- [20]. Henseler, J., Ringle, C. M., & Sarstedt, M. (2015). A new criterion for assessing discriminant validity in variance-based structural equation modeling. *Journal of the Academy of Marketing Science*, 43(1), 115–135. <https://doi.org/10.1007/s11747-014-0403-8>
- [21]. Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, 6(1), 1–55. <https://doi.org/10.1080/10705519909540118>
- [22]. International Monetary Fund. (2024). Cyber risk: A growing concern for macrofinancial stability. In *Global financial stability report: The last mile: Financial vulnerabilities and risks* (Chapter 3). <https://www.imf.org/en/Publications/GFSR/Issues/2024/04/16/global-financial-stability-report-april-2024>
- [23]. Jaakkola, E. (2020). Designing conceptual articles: Four approaches. *AMS Review*, 10(1–2), 18–26. <https://doi.org/10.1007/s13162-020-00161-0>

- [24]. Kaur, R., Gabrijelčič, D., & Klobučar, T. (2023). Artificial intelligence for cybersecurity: Literature review and future research directions. *Information Fusion*, 97, Article 101804. <https://doi.org/10.1016/j.inffus.2023.101804>
- [25]. Khaleel, Y. L., Habeeb, M. A., Albahri, A. S., Al-Quraishi, T., Albahri, O. S., & Alamooodi, A. H. (2024). Network and cybersecurity applications of defense in adversarial attacks: A state-of-the-art using machine learning and deep learning methods. *Journal of Intelligent Systems*, 33(1). <https://doi.org/10.1515/jisys-2024-0153>
- [26]. Madry, A., Makelov, A., Schmidt, L., Tsipras, D., & Vladu, A. (2018). Towards deep learning models resistant to adversarial attacks. In *International Conference on Learning Representations (ICLR 2018)*. <https://arxiv.org/abs/1706.06083>
- [27]. Mink, J., Benkraouda, H., Yang, L., Ciptadi, A., Ahmadzadeh, A., Votipka, D., & Wang, G. (2023). Everybody's got ML, tell me what else you have: Practitioners' perception of ML-based security tools and explanations. In *2023 IEEE Symposium on Security and Privacy (SP)*. IEEE.
- [28]. Monetary Authority of Singapore. (2024). *Information paper on artificial intelligence model risk management*.
- [29]. Nigeria Data Protection Act, 2023 (Nigeria). <https://kpmg.com/ng/en/insights/2023/08/nigeria-data-protection-act-2023-review.html>
- [30]. Ofuani, A. B., Omoera, C. I., & Akagha, C. (2024). Artificial intelligence integration and performance of deposit money banks in Lagos Metropolis: A study of United Bank for Africa. *UNILAG Journal of Business*, 10(1), 54–81.
- [31]. Owusu, A. (2025). Achieving operational excellence through artificial intelligence: The case of Ghanaian banks. *International Journal of Information Management Data Insights*, 5(2), Article 100377. <https://doi.org/10.1016/j.jjime.2025.100377>
- [32]. Papernot, N., McDaniel, P., Wu, X., Jha, S., & Swami, A. (2016). Distillation as a defense to adversarial perturbations against deep neural networks. In *2016 IEEE Symposium on Security and Privacy (SP)* (pp. 582–597). IEEE. <https://doi.org/10.1109/SP.2016.41>
- [33]. Pinto, A. S., Abreu, A., Cota, M. P., & Paiva, J. (2024). *Exploring the drivers of AI adoption: A meta-analysis of technological, organizational and environmental (TOE) factors* [Preprint]. Research Square. <https://doi.org/10.21203/rs.3.rs-5634577/v1>
- [34]. Rahman, M., Ming, T. H., Baigh, T. A., & Sarker, M. (2023). Adoption of artificial intelligence in banking services: An empirical analysis. *International Journal of Emerging Markets*, 18(10), 4270–4300. <https://doi.org/10.1108/IJOEM-06-2020-0724>
- [35]. Rjoub, G., Bentahar, J., Abdel Wahab, O., Mizouni, R., Song, A., Cohen, R., Otrok, H., & Mourad, A. (2023). A survey on explainable artificial intelligence for cybersecurity. *IEEE Transactions on Network and Service Management*, 20(4), 5115–5140. <https://doi.org/10.1109/TNSM.2023.3282740>
- [36]. Simonetto, T., Dyrnishi, S., Ghamizi, S., Cordy, M., & Le Traon, Y. (2022). A unified framework for adversarial attack and defense in constrained feature space. In *Proceedings of the Thirty-First International Joint Conference on Artificial Intelligence (IJCAI-22)* (pp. 1313–1319). <https://doi.org/10.24963/ijcai.2022/183>
- [37]. Tornatzky, L. G., & Fleischer, M. (1990). *The processes of technological innovation*. Lexington Books.
- [38]. Uren, V., & Edwards, J. (2023). Technology readiness and the organizational journey towards AI adoption: An empirical study. *International Journal of Information Management*, 68, Article 102588. <https://doi.org/10.1016/j.ijinfomgt.2022.102588>
- [39]. Vassilev, A., Oprea, A., Fordyce, A., Anderson, H., Davies, X., & Hamin, M. (2025). *Adversarial machine learning: A taxonomy and terminology of attacks and mitigations* (NIST AI 100-2e2025). National Institute of Standards and Technology. <https://doi.org/10.6028/NIST.AI.100-2e2025>
- [40]. Wallace, S. A., Green, K. Y., Johnson, C. M., Cooper, J. T., & Gilstrap, C. M. (2020). An extended TOE framework for cybersecurity-adoption decisions. *Communications of the Association for Information Systems*, 47, 338–363. <https://doi.org/10.17705/1CAIS.04716>