**JOURNAL OF LEADERSHIP AND DEVELOPMENT**
Volume 2, Number 3 (September, 2026)
ISSN: 1595-9457 (online); 3093-0618 (print)
**Website:** https://jppssuniuyo.com/jld **Email:** jppssuniuyo@gmail.com
**Received**: August 25, 2026 **Accepted**: September 26, 2026 **Published**: September 30, 2026
**Citation**: Sodiq, Okunola O. (2026). "Large Language Models and Emerging Cybersecurity Challenges: Implications for Digital Security Systems in Nigeria." *Journal of Leadership and Development*, 2 (3): 181-194.

**Article** **Open Access**

# LARGE LANGUAGE MODELS AND EMERGING CYBERSECURITY CHALLENGES: IMPLICATIONS FOR DIGITAL SECURITY SYSTEMS IN NIGERIA

**Okunola Omobolaji Sodiq**
Department of Computer Science, Veritas University, Abuja, Nigeria

**Abstract**
Large Language Models (LLMs) are increasingly being incorporated into digital systems and cybersecurity workflows for tasks such as threat-intelligence analysis, security-log interpretation, incident documentation, vulnerability assessment and policy compliance. Their adoption, however, also creates new attack surfaces because LLMs can be manipulated through malicious instructions, exposed to sensitive information, connected to untrusted data, or granted excessive access to external tools. This paper examines the emerging cybersecurity challenges associated with LLM-enabled systems and their implications for digital security in Nigeria. It adopts a qualitative integrative literature review, Nigerian researches, official policy documents, cybersecurity guidance and AI risk-management frameworks. The analysis is organised around four themes: LLM capabilities for cybersecurity, LLM-specific and system-level threats, defensive applications, and socio-technical governance requirements. The review shows that security implications of LLMs extend beyond model behaviour to the data, retrieval systems, applications, interfaces, users and organisational processes surrounding the model. In the Nigerian context, these risks interact with infrastructure constraints, cybersecurity skills gaps, institutional capacity, data-protection obligations and dependence on external technology providers. The paper develops a Nigerian LLM Cybersecurity Resilience Framework comprising five mutually reinforcing layers: technical security, data governance, human oversight, organisational governance and national regulatory governance, with continuous evaluation as a cross-cutting principle. The framework contributes a context-sensitive structure for assessing and governing LLM deployment in Nigerian digital-security environments.

***Copyright** © 2026 By* JLD. **Publishers:** Omega Books
This is an open access article which permits unrestricted use provided the work is properly cited.

**Keywords:** Artificial Intelligence, Cybersecurity, Cyber Resilience, Digital Security, Generative Artificial Intelligence (GAI), Large Language Models (LLM), Nigeria.

## Introduction

Generative Artificial Intelligence (GenAI) has altered the way digital systems process information, generate content and support human decision-making. Large Language Models (LLMs) are particularly significant because they can process and generate natural language at scale and can be integrated into applications that retrieve information, analyse documents, assist with programming and interact with external tools. In cybersecurity, these capabilities create opportunities for improving information-intensive activities such as threat intelligence, security analysis, vulnerability assessment and incident response (Ferrag *et al*., 2025; Jaffal *et al*., 2025; Karras *et al*., 2025). The cybersecurity significance of LLMs is nevertheless broader than their usefulness as defensive tools. Once an LLM is incorporated into an operational system, the model becomes part of a larger computational and organisational environment. Inputs may originate from users, documents, websites, retrieval systems or other applications; outputs may subsequently be passed to software tools or used by security personnel to make decisions. Consequently, vulnerabilities can arise not only from the model but also from the interfaces, data, permissions, applications and processes surrounding it. ENISA (2023) distinguishes between using AI for cybersecurity and securing AI systems themselves, a distinction that is particularly useful for understanding LLM-enabled security environments.

The Open Worldwide Application Security Project (OWASP) Top 10 for Large Language Model Applications identifies major application-level risks including prompt injection, sensitive information disclosure, supply-chain vulnerabilities, data and model poisoning, improper output handling, excessive agency, system-prompt leakage, vector and embedding weaknesses, misinformation and unbounded consumption (OWASP Foundation, 2025). Recent systematic reviews similarly show that LLM security must be considered across the wider software and model lifecycle rather than only at the model layer (Jaffal *et al*., 2025; Moia *et al*., 2026). These developments are important for Nigeria because digital financial services, telecommunications, public digital services and other technology-dependent activities increasingly rely on interconnected information systems. Nigeria's National Cybersecurity Policy and Strategy (NNCPS) identifies cybersecurity as an important enabler of national security and economic development, while the National Artificial Intelligence Strategy identifies AI as a strategic area of technological and economic development (Federal Republic of Nigeria, 2021; Federal Republic of Nigeria, 2024). The Nigeria Data Protection Act 2023 also establishes a legal framework for personal-data processing and security. LLM deployment therefore intersects with existing cybersecurity, data-protection and AI-governance responsibilities.

Nigerian researches on AI and cybersecurity indicate that adoption is shaped by infrastructure, workforce capacity, institutional readiness and governance. Eke (2026), for example, reported evidence of AI-enabled cybersecurity adoption in Nigerian banking while identifying workforce readiness, access to appropriate tools and infrastructure as constraints. Bade, Babate and Bara (2026) identified AI-assisted threats and defensive applications within Nigeria's digital economy and emphasise governance, workforce and infrastructure challenges. Ndoh, Iwuh and Ugochukwu (2026) similarly argued from a socio-technical perspective that effective AI adoption requires technological capability to be integrated with institutional capacity, governance and human development. Despite this emerging literature, there remains a need for a context-sensitive analysis that can connect LLM-specific

cybersecurity threats with the infrastructural, organisational, human and regulatory conditions of Nigerian digital systems. Existing international reviews provide detailed taxonomies of applications, threats and mitigations, but they do not establish how those findings should be interpreted within Nigeria's institutional and technological environment. This paper addresses that gap by synthesising international LLM-cybersecurity research with Nigerian AI, cybersecurity and governance literature and translating the synthesis into a Nigerian LLM Cybersecurity Resilience Framework (LLM-CRF).

The paper is guided by three research questions:

i. **RQ1:** What emerging cybersecurity threats are associated with the deployment of LLMs in digital-security systems?
ii. **RQ2:** What defensive cybersecurity functions can LLMs support, and what limitations constrain their safe use?
iii. **RQ3:** What technical, organisational, human and regulatory safeguards are required for resilient LLM deployment in Nigeria?

The paper makes three contributions. First, it synthesises LLM capabilities and vulnerabilities from a systems-oriented cybersecurity perspective. Second, it relates those risks to Nigerian infrastructure, workforce, institutional and regulatory conditions. Third, it proposes a layered resilience framework that maps LLM-related threats to technical and socio-organisational controls.

**Conceptual Background**

Artificial intelligence refers broadly to computational systems designed to perform tasks associated with functions such as learning, pattern recognition, reasoning, prediction and decision support. Within contemporary AI, machine-learning systems learn patterns from data, while generative models produce new content. LLMs are generative models designed primarily for language processing and generation, although contemporary systems can also work with code, structured information and multimodal inputs. Their cybersecurity relevance follows partly from their ability to transform large quantities of technical and textual information into forms that human analysts can interpret. The practical distinction between an LLM and an LLM-enabled system is important. An LLM by itself generates outputs in response to inputs; an LLM-enabled application may additionally include retrieval systems, databases, APIs, plugins, identity systems, monitoring components and automated actions. Security risks can therefore emerge at multiple layers. Moia *et al*. (2026) emphasise this wider system perspective by reviewing threats across real-world LLM deployment and the software and model lifecycles.

On the other hand, cybersecurity concerns the protection and resilient operation of digital systems, networks, applications and data against threats that may compromise confidentiality, integrity or availability. Cains *et al*. (2022) show that cybersecurity is inherently contextual and socio-technical, with human factors and resilient system functionality forming important components of cyber-risk analysis. This understanding is appropriate for LLM-enabled environments because security cannot be reduced to the protection of model parameters. Digital security systems should therefore be understood as interconnected arrangements of software, hardware, data, communication infrastructure, human users, organisational procedures and governance mechanisms. An LLM inserted into such an arrangement inherits some existing security risks while introducing additional ones associated with natural-language interaction, probabilistic output, external knowledge retrieval and tool use. Two related but distinct questions arise. The first concerns AI for

cybersecurity: how LLMs can assist security personnel in detecting, analysing and responding to threats. The second concerns security of AI: how LLM-based systems themselves can be protected. ENISA (2023) identifies both as important research areas. The distinction prevents an organisation from assuming that an LLM is secure merely because it is being used for security work. A model may assist a security operation while simultaneously becoming an additional attack surface.

**Literature Review**

The literature on large language models (LLMs) and cybersecurity has expanded rapidly as these systems have moved from general-purpose language applications into security operations, threat intelligence, vulnerability analysis, incident response, and security-policy processes. At the same time, researchers have identified significant risks associated with their deployment, including prompt injection, sensitive information disclosure, data and model poisoning, supply-chain vulnerabilities, improper output handling, and overreliance on potentially inaccurate outputs. The literature therefore presents LLMs as both cybersecurity enablers and sources of new security vulnerabilities. This section reviews existing scholarship on the applications, vulnerabilities, and governance challenges associated with LLMs in cybersecurity, with particular attention to their implications for Nigeria's emerging digital-security environment. It first examines the defensive applications of LLMs, then considers the principal vulnerabilities of LLM-enabled systems. The review subsequently situates these issues within Nigeria's evolving artificial intelligence, data-protection, and cybersecurity policy landscape. It then identifies the gap in existing literature that provides the basis for the Nigerian LLM Cybersecurity Resilience Framework proposed in this study.

Recent literature documents a wide range of LLM applications in cybersecurity. Jaffal *et al*. (2025) reviewed applications across network security, software and system security, cloud security, threat intelligence, social engineering, critical infrastructure and Internet-of-Things security. Ferrag *et al*. (2025) similarly reviewed applications in intrusion detection, software engineering, cyber-threat intelligence, malware detection and phishing detection. Karras *et al*. (2025) synthesised research on threat detection, incident response, vulnerability assessment and security analytics. The literature suggests that the strongest immediate role of LLMs is often analytical augmentation rather than autonomous replacement of specialist security technologies. LLMs therefore summarise threat reports, classify textual information, explain code, organise incident information and help analysts navigate complex documentation. These functions may reduce information-processing burdens, but they remain dependent on the quality of input data and the reliability of model outputs. The second literature strand concerns vulnerabilities. OWASP (2025) identifies a broad set of LLM application risks, while Ferrag *et al*. (2025), Jaffal *et al*. (2025) and Moia *et al*. (2026) showed that threats arise across model, application, data and integration layers. Prompt injection is particularly important because malicious content can influence how an LLM interprets instructions, especially where the model can access external data or tools. Improper output handling creates another risk when generated content is trusted or passed directly to another application. Data and model poisoning, sensitive information disclosure and supply-chain vulnerabilities further demonstrate that LLM security is inseparable from data and infrastructure security. Excessive agency increases the significance of these risks because a manipulated or erroneous model output can have consequences beyond the production of incorrect text when the system has permission to perform actions.

Nigerian literature provides evidence that AI adoption is influenced by conditions beyond technical capability. Eke (2026) reported that AI-enabled cybersecurity adoption in the banking sector is affected by workforce readiness, access to appropriate tools and infrastructure. Bade *et al*. (2026) identifed AI-assisted phishing, deepfake impersonation, automated malware deployment and intrusion-detection evasion alongside AI-based anomaly detection, intrusion prevention and incident response in Nigeria's digital economy. Ndoh *et al*. (2026) emphasised infrastructure, policy, institutional capacity and human capability as interconnected conditions of AI-security adoption. These studies are relevant to LLMs but do not by themselves provide an LLM-specific Nigerian resilience framework. The research gap therefore lies less in establishing that AI creates opportunities and risks and more in specifying how LLM-specific vulnerabilities should be integrated with Nigerian organisational and regulatory conditions. Existing international reviews are increasingly comprehensive, but their primary concern is the general LLM-cybersecurity landscape. Nigerian studies, by contrast, focus more broadly on AI, cybersecurity, digital transformation or security governance. The gap addressed here is the intersection of these literatures: the conditions under which LLM-enabled security systems can be deployed and governed within Nigeria. The proposed framework is intended to bridge this gap rather than compete with existing technical taxonomies.

**Theoretical Framework**

The primary theoretical lens is Sociotechnical Systems Theory (STS). Baxter and Sommerville (2011) argued that technical systems should be understood in relation to the organisational and social environments in which they are developed and used. The approach is appropriate because the security outcome of an LLM depends on the interaction among model capabilities, data, users, organisational procedures, infrastructure and governance. Applied to LLM cybersecurity, STS implies that a technically capable model may still generate insecure outcomes if it receives poorly governed data, has excessive permissions, operates without monitoring, or is used by personnel without adequate verification procedures. Conversely, a model with known limitations can provide useful support when embedded in a layered architecture with appropriate controls. The second analytical lens is the dual-use character of AI. LLM capabilities that assist defenders can also assist attackers. Natural-language generation can support security documentation and awareness training but can also facilitate convincing social-engineering content. Code assistance can support secure development but can also lower barriers to harmful coding activities. The dual-use perspective therefore prevents a one-sided assessment of LLMs as either beneficial or harmful. The two perspectives are complementary. STS explains how cybersecurity outcomes emerge from interactions among technical and social components, while the dual-use perspective explains why a single capability may produce different outcomes depending on its use and level of access. They both support a systems-oriented analysis in which LLM capability, threat exposure, organisational context and governance are treated as interdependent.

**Methodology**

The study adopts a qualitative integrative literature review. An integrative review is appropriate where evidence from different source types must be synthesised to develop a conceptual understanding rather than calculate a pooled statistical effect. The review therefore combines peer-reviewed research, Nigerian studies, official policy documents and authoritative cybersecurity and AI-risk frameworks. The study is deliberately not presented

as a PRISMA systematic review. It does not claim exhaustive retrieval or statistical synthesis. Instead, it uses a structured process of source identification, relevance screening, thematic organisation and conceptual synthesis. This distinction is important because recent systematic reviews such as Jaffal *et al*. (2025), Karras *et al*. (2025) and Moia *et al*. (2026) already provide systematic evidence bases; the present study's contribution is the contextual synthesis and Nigerian resilience framework. Priority was given to sources published from 2023 to 2026 for rapidly changing LLM-security issues, while earlier sources were retained for theoretical and conceptual foundations. Sources were considered relevant when they addressed one or more of the following: LLM cybersecurity applications; LLM vulnerabilities; AI-enabled cyber threats; cybersecurity governance; AI risk management; Nigerian cybersecurity; Nigerian AI governance; or data protection. The principal source categories were: (1) peer-reviewed LLM and cybersecurity research; (2) Nigerian studies on AI and cybersecurity; (3) official Nigerian cybersecurity, AI and data-protection documents; and (4) established international guidance from bodies such as NIST, OWASP and ENISA. Sources were excluded when they addressed AI only in general terms without a meaningful connection to cybersecurity, digital security, governance or the Nigerian context.

The literature was organised into four analytical themes: LLM capabilities for cybersecurity; LLM-enabled threats; defensive applications and limitations; and governance, resilience and Nigerian contextual conditions. The themes were then compared to identify relationships between threats and controls. Particular attention was given to whether a risk originated at the model layer or emerged through interaction with data, retrieval systems, applications, users or organisational procedures. The final analytical stage involved synthesising the findings into the Nigerian LLM Cybersecurity Resilience Framework. The framework is therefore a conceptual product of the literature synthesis, not an empirically validated operational model. Its proposed controls should consequently be tested in future studies involving Nigerian organisations and real-world LLM deployments. The use of authoritative standards and recent peer-reviewed reviews strengthens the conceptual basis of the study. However, an integrative review is subject to selection and interpretation bias and does not establish causal effects. In addition, Nigeria-specific empirical evidence concerning operational LLM deployment remains limited. The framework should therefore be understood as a research-informed conceptual framework requiring empirical validation.

**Emerging Cybersecurity Challenges of LLMs**

The growing use of LLMs in digital systems also introduces new cybersecurity risks that can affect the confidentiality, integrity and reliability of information systems. These risks arise from both the capabilities of LLMs and the ways in which they are integrated with data, applications and external tools. Some of the major cybersecurity challenges associated with LLM deployment, including prompt injection, information disclosure, data poisoning, supply-chain vulnerabilities, hallucination, excessive agency and social engineering.

i. **Prompt Injection:** Prompt injection occurs when untrusted content influences an LLM's interpretation of instructions. The risk becomes particularly consequential when the model retrieves external information or has access to software tools. OWASP (2025) treated prompt injection as a leading LLM application risk. Mitigation should therefore include separation of trusted instructions from untrusted content, least-privilege permissions, input and output validation, adversarial testing and human approval for high-risk actions.
ii. **Sensitive Information Disclosure:** LLMs may process personal, financial, commercial or institutional information supplied through prompts, documents, retrieval systems or

connected applications. The risk is therefore partly a data-flow problem. In Nigeria, this intersects directly with the Nigeria Data Protection Act 2023, which establishes requirements concerning lawful and accountable personal-data processing and security (Nigeria Data Protection Commission, 2023). Organisations should apply data minimisation, access control, retention limits and appropriate contractual and technical safeguards when external LLM services are used.

iii. **Data and Model Poisoning:** Poisoning occurs when malicious or misleading information is introduced into training, fine-tuning, retrieval or related data pipelines. In retrieval-augmented systems, the integrity and provenance of documents become security concerns because manipulated information may influence downstream outputs. Security controls should therefore cover the provenance, integrity and access permissions of retrieval corpora and model-development datasets (OWASP, 2025; Zou *et al*., 2025).

iv. **Supply-Chain Vulnerabilities:** LLM applications depend on models, APIs, datasets, libraries, embeddings, cloud services and other third-party components. A weakness in one component can affect the security of the complete system. Nigerian organisations should therefore assess provider security, data-processing arrangements, service continuity, contractual controls, dependency concentration and incident-notification obligations (Autio *et al*., 2024).

v. **Improper Output Handling:** LLM outputs should not automatically be treated as trusted commands or verified facts. Where outputs are passed to software systems, inadequate validation can convert a model error into an application-level vulnerability. Output validation, structured interfaces, permission boundaries and human approval should therefore be applied before high-impact actions are executed (OWASP, 2025).

vi. **Hallucination and False Security Judgements:** LLM hallucination refers to plausible but inaccurate generated content. In cybersecurity, inaccurate classification, vulnerability explanations or incident recommendations can produce false assurance or inappropriate responses. The central risk is therefore not merely factual error but the possibility that fluent output will be mistaken for reliable evidence. Security workflows should require verification against logs, threat intelligence, system telemetry and established procedures (Ji *et al*., 2023).

vii. **Phishing and Social Engineering:** LLMs can generate fluent and context-sensitive language. The same capability may assist security-awareness activities while also increasing the quality and scalability of deceptive communications. The defensive response should therefore combine technical email and identity controls with awareness, detection and verification mechanisms rather than relying on the linguistic quality of generated content as evidence of legitimacy (Heiding *et al*., 2024).

viii. **Misinformation and Overreliance:** Overreliance occurs when users defer excessively to model outputs despite uncertainty or lack of evidence. This is a socio-technical risk because the model's limitation becomes consequential through human interpretation. Appropriate use policies, training, verification requirements and accountability structures are therefore necessary (Passi *et al*., 2024; Passi & Vorvoreanu, 2022).

**Defensive Capabilities of LLMs**

Large Language Models (LLMs) are not only sources of new cybersecurity risks but can also provide useful capabilities for strengthening digital security. Their ability to analyse large amounts of information, identify patterns and generate useful responses can support different areas of cybersecurity operations, including threat intelligence, security log analysis,

incident response, vulnerability analysis, policy compliance and Security Operations Centre activities. In threat intelligence, LLMs can assist analysts in summarising threat reports, extracting indicators, comparing descriptions and organising intelligence drawn from multiple textual sources. Recent research identifies this kind of multilingual, cross-source threat-intelligence extraction as an emerging application of LLMs in cybersecurity practice (Özbay *et al*., 2026). Their principal value in this area lies in information processing and analyst augmentation rather than in replacing specialised detection systems. A related capability concerns security log and event analysis. Because security operations generate large volumes of textual and semi-structured records, LLMs can help interpret event data and summarise relationships across logs that would otherwise be time-consuming for analysts to review manually (Kheddar, 2025). This value is nonetheless conditional: because automated summaries can omit or misinterpret evidence, analysts should retain access to the underlying logs and telemetry rather than relying on LLM-generated summaries alone.

LLMs can also support incident response by assisting with incident classification, documentation, identification of affected systems and coordination of response activities.

Research examining LLMs specifically within cybersecurity incident-management workflows found that they can meaningfully support analysts in interpreting incidents, while cautioning that this support functions best as an aid to human judgement rather than a substitute for it (Jones *et al*., 2025). Consistent with this caution, high-impact containment or system-modification decisions should remain subject to controlled approval and verification rather than being executed automatically. Beyond incident handling, LLMs can assist with vulnerability analysis and secure coding by explaining code and helping to identify possible vulnerabilities. This capability should be treated as a starting point rather than a conclusion: because generated code and vulnerability assessments may themselves contain errors, their suggestions should be tested using established software-security methods before being relied upon. A similar logic applies to compliance and policy analysis, where LLMs can assist with mapping organisational controls to complex cybersecurity and regulatory requirements. Cadet *et al*. (2024) illustrate this potential in the context of policy compliance and risk mitigation, showing how LLMs can help organisations interpret regulatory obligations more efficiently. Even so, regulatory interpretation generated in this way should be verified against authoritative legal and policy texts rather than treated as a final determination. Within Security Operations Centres, LLMs are increasingly explored as components of Retrieval-Augmented Generation (RAG) and agentic architectures intended to support analyst workflows (Alqahtani & gKumar, 2026). The evidence nonetheless points toward a layered rather than autonomous architecture: specialised detection technologies identify events, the LLM assists with interpretation, human analysts must verify that interpretation, as controlled mechanisms execute only approved responses. This layered architecture is important precisely because it limits the consequences of hallucination, prompt manipulation and excessive agency, each of which becomes more dangerous when an LLM is allowed to act without human verification.

**Implications for Digital Security Systems in Nigeria**

The growing use of Large Language Models (LLMs) has important implications for digital security systems in Nigeria. While LLMs can improve threat detection, incident response and other cybersecurity activities, their use may also introduce new risks to government systems, financial institutions, telecommunications networks and digital platforms. These implications are particularly important in Nigeria because of existing infrastructure limitations,

cybersecurity skills gaps and data governance concerns that recent Nigerian research on AI and security has identified as socio-technical constraints on responsible adoption (Ndoh *et al*., 2026). Government systems can potentially use LLMs for document processing, information retrieval, citizen-service support and security analysis. Because such systems may process sensitive personal and administrative information, deployment should be accompanied by strict access control, data governance, auditability and human oversight. Nigeria's national cybersecurity strategy and AI strategy provide broader policy contexts within which such deployment should be considered (Federal Republic of Nigeria, 2021; National Information Technology Development Agency [NITDA], 2024). Digital financial systems present a related but distinct case: they are attractive settings for LLM-assisted security because of the volume of transaction, fraud and incident information they generate, yet unrestricted model access could create significant confidentiality and integrity risks. Evidence from Nigerian banking research indicates that workforce readiness and infrastructure remain relevant constraints on AI-enabled cybersecurity (Eke, 2026). Telecommunications and digital platforms raise similar concerns at greater scale: they generate extensive operational and security information, and while LLMs may assist with analysis and incident management, the scale of these systems increases the consequences of weak permissions, data leakage or automated errors.

These sector-specific risks point to a broader workforce implication. LLM deployment changes the skill profile required of cybersecurity personnel, who increasingly need AI literacy, understanding of model limitations, output verification, secure prompt and context management, AI-security assessment, data governance and regulatory awareness. Nigerian capacity-building programmes should therefore treat AI security as a component of cybersecurity professional development rather than a separate specialism. This human-capacity dimension is closely tied to infrastructure: LLM deployment may require reliable connectivity, cloud or local computing resources, monitoring infrastructure and skilled personnel, and Nigerian organisations should assess whether they can securely operate the complete LLM application rather than equating access to an external model with cybersecurity capability. Where organisations depend on external LLM providers, this assessment must also extend to questions of data sovereignty, including where data are processed, how long they are retained, what access providers retain, how service continuity is assured, and how contractual and architectural changes on the provider's side are managed. These concerns do not make external services inherently insecure, but they introduce third-party and data-governance risks that should be assessed before deployment rather than after. Nigeria already has important components of a governance environment relevant to LLM deployment, including the National Cybersecurity Policy and Strategy, the Nigeria Data Protection Act 2023 and the National Artificial Intelligence Strategy (Federal Republic of Nigeria, 2021; Nigeria Data Protection Commission, 2023; National Information Technology Development Agency [NITDA], 2024). The regulatory challenge is to translate these broad principles into operational requirements for LLM-enabled systems, covering risk assessment, data governance, model evaluation, incident reporting, human oversight, third-party risk management, security testing and accountability.

**Discussion**

The review indicates that LLMs can expand the information-processing capacity of cybersecurity teams, with their contribution strongest where large volumes of textual information must be summarised, interpreted or organised — a strength consistent with what recent reviews describe as LLMs' emerging role as semantic reasoning and decision-

support components in security workflows (Kasri et al., 2025). This can be particularly valuable where organisations face shortages of specialised personnel, as is often the case in the Nigerian context discussed earlier in this paper. However, augmentation should not be confused with autonomous reliability: security decisions remain dependent on evidence, context and verification, not on the fluency of a model's output alone. At the same time, the integration of an LLM into an operational system adds new security-relevant interfaces. The relevant security boundary extends beyond the model itself to include prompts, retrieved information, model outputs, APIs, databases, plugins and tools, which explains why securing the model alone is insufficient. The transition toward retrieval-augmented and agentic systems makes this systems perspective increasingly important, since autonomy and tool integration introduce new points at which manipulation or error can propagate through a security pipeline (Moia et al., 2026; Jabbar et al., 2025). These two findings converge on a single socio-technical conclusion: capability does not automatically produce security. A technically advanced model can generate insecure outcomes if it is connected to excessive privileges, poorly governed data or inadequate monitoring, while an imperfect model can still be useful when its role is bounded, its outputs are verified and its permissions are limited. This is consistent with the view, drawn from sociotechnical systems theory, that the performance of a technical system cannot be separated from the human and organisational arrangements within which it operates (Baxter & Sommerville, 2011). Resilience, in other words, depends less on the sophistication of the model than on the interaction between technical safeguards and human and organisational controls.

**Nigerian LLM Cybersecurity Resilience Framework**

The Nigerian LLM Cybersecurity Resilience Framework translates the preceding analysis into a layered model. Its central proposition is that resilient LLM deployment requires simultaneous attention to the technical system, information flows, human decision-making, organisational governance and national regulation, with continuous evaluation cutting across all five layers.

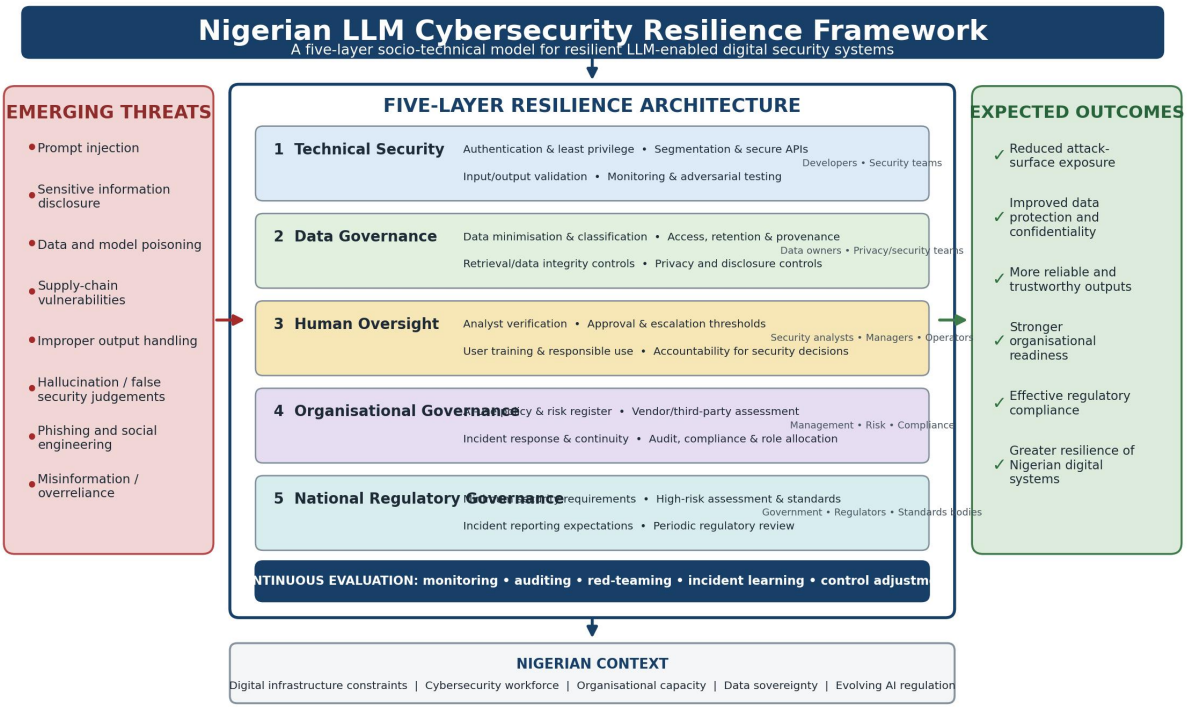



**Figure 1. Nigerian LLM Cybersecurity Resilience Framework**

**Source:** *Author's synthesis, drawing on OWASP (2025), Autio et al. (2024), Baxter and Sommerville (2011), Federal Republic of Nigeria (2021, 2024) and Nigeria Data Protection Commission (2023).*

| Layer | Primary risks addressed | Core controls | Primary responsibility |
|---|---|---|---|
| 1. Technical security | Prompt injection; insecure outputs; excessive agency; API and integration vulnerabilities | Authentication; least privilege; segmentation; input/output validation; monitoring; adversarial testing; secure APIs | Developers, architects, security teams |
| 2. Data governance | Sensitive-data disclosure; poisoning; retrieval compromise; provenance failure | Data minimisation; classification; access control; retention; provenance; retrieval protection; privacy controls | Data owners, privacy officers, security teams |
| 3. Human oversight | Hallucination; false security judgements; automation bias | Analyst review; approval thresholds; verification; escalation; training; accountability | Security analysts, managers, system operators |
| 4. Organisational governance | Unsafe use; unmanaged third-party risk; weak incident response | AI-use policy; risk register; audits; vendor assessment; incident plans; role allocation | Boards, management, risk and compliance units |
| 5. National regulatory governance | Regulatory gaps; inconsistent standards; systemic risk | Guidelines; high-risk assessment; minimum controls; incident reporting; provider standards; periodic review | Government, regulators, standards bodies |
| Cross-cutting: continuous evaluation | Changing threats and model capabilities | Testing; monitoring; red-teaming; incident learning; control adjustment; re-evaluation | All stakeholders |

**Figure 2. Nigerian LLM Cybersecurity Resilience Framework.**

**Source:** *Author's synthesis, drawing on OWASP (2025), Autio et al. (2024), Federal Republic of Nigeria (2021, 2024) and Nigeria Data Protection Commission (2023).*

Technical security forms the first line of defence. Controls should be proportionate to the model's function and level of access: a document-summarisation system should not be treated in the same manner as an LLM connected to a banking platform, government database or security operations environment. Least privilege is especially important here, because it limits the consequences of manipulated or erroneous outputs before they can propagate further into a system. This technical layer must be matched by a second layer of data governance, protecting information entering, passing through and leaving the LLM-enabled system through data minimisation, classification, retention controls, provenance management and retrieval-database security. This layer should be aligned with the Nigeria Data Protection Act 2023 and applicable sectoral requirements (Nigeria Data Protection Commission, 2023).

Technical and data controls alone are insufficient without a third layer of human oversight, which should be risk-sensitive rather than universal: low-risk tasks may be

automated more extensively, whereas high-consequence actions should require verification and approval. The objective is to prevent a model output from becoming an unexamined security decision, a concern consistent with OWASP's recommendation that high-risk actions in LLM applications require human approval (OWASP Foundation, 2025). Beyond individual decisions, organisations need a fourth layer of organisational governance, establishing clear rules for selecting, deploying, monitoring and retiring LLM systems. This should include acceptable-use policies, risk registers, third-party assessments, incident-response arrangements, staff training and accountability mechanisms, so that oversight is not left to the judgement of individual analysts alone. At the national level, a fifth layer of regulatory governance should establish a baseline for secure LLM deployment without prescribing controls so rigidly that they become obsolete as technology changes. High-risk applications, critical systems, da ta protection, incident reporting, provider dependence and professional competence are areas in which regulatory guidance may be particularly relevant, building on the foundations already provided by Nigeria's National Cybersecurity Policy and Strategy and National Artificial Intelligence Strategy (Federal Republic of Nigeria, 2021; National Information Technology Development Agency [NITDA], 2024). Cutting across all five layers is a continuous evaluation cycle, testing, monitoring, red-teaming, incident learning, control adjustment and re-evaluation, which reflects the dynamic character of LLM capabilities and threat techniques. This is consistent with the literature's emphasis on lifecycle evaluation, which identifies fragmented evaluation practices and limited longitudinal testing as persistent weaknesses in current LLM-security research (Alqahtani & Kumar, 2026; Jabbar *et al*., 2025)

**Recommendations**

Based on the literature synthesis and the proposed framework, the following recommendations are made.

i. Organisations deploying LLMs for cybersecurity should conduct application-level threat modelling before granting access to sensitive data or external tools.
ii. High-risk LLM actions should operate under least-privilege permissions and appropriate human approval thresholds.
iii. Nigerian organisations should establish LLM-specific data-governance procedures covering input data, retrieval corpora, retention, provider access and data transfers.
iv. Cybersecurity teams should receive training in prompt injection, model limitations, output validation, AI supply-chain risk and AI-specific incident response.
v. Organisations should maintain an LLM inventory and risk register identifying models, providers, connected tools, datasets, permissions and responsible personnel.
vi. Government and relevant regulators should develop practical guidance for high-risk LLM deployments in critical digital systems while aligning new requirements with existing cybersecurity and data-protection frameworks.
vii. Third-party LLM providers should be assessed for security controls, data-processing practices, service continuity, incident notification and contractual accountability.
viii. Universities and professional institutions should expand interdisciplinary training connecting computer science, cybersecurity, data protection, AI governance and responsible AI.
ix. Future Nigerian research should empirically test the proposed framework in banking, telecommunications, government digital services and higher-education environments.

## Limitations and Future Research

This study is limited by its reliance on secondary literature. It does not measure the performance of a particular LLM, conduct penetration testing, or empirically validate the proposed framework in a Nigerian organisation. The Nigerian empirical literature on operational LLM deployment is also still developing. These limitations mean that the framework should be interpreted as a conceptual and research-informed model rather than a validated security standard. Future research should therefore operationalise the framework through case studies and field surveys of Nigerian organisations. Studies could develop measurable indicators for each framework layer and test relationships among LLM access, security incidents, human oversight, organisational maturity and regulatory compliance. Experimental research could also evaluate LLM-assisted threat detection, log analysis and incident response using Nigerian datasets. Such work would provide empirical evidence for refining the framework.

## Conclusion

LLMs are becoming relevant to cybersecurity because they can process and generate information at a scale useful for threat intelligence, security analysis, incident response, vulnerability assessment and compliance activities. Their adoption, however, also changes the cybersecurity boundary. Prompt injection, sensitive information disclosure, poisoning, insecure outputs, excessive agency, misinformation and supply-chain vulnerabilities demonstrate that the model cannot be secured independently of the wider system. For Nigeria, these risks are shaped by infrastructure, workforce capability, institutional readiness, data-protection requirements and dependence on external technology providers. The central argument of this paper is therefore that LLM cybersecurity should be treated as a socio-technical resilience problem rather than simply a model-performance problem. The proposed Nigerian LLM Cybersecurity Resilience Framework provides a structured response through five mutually reinforcing layers: technical security, data governance, human oversight, organisational governance and national regulatory governance, supported by continuous evaluation. Its contribution is to connect established LLM-security concerns with the particular institutional and technological conditions of Nigerian digital systems. The framework should now be subjected to empirical testing so that its controls, responsibilities and maturity indicators can be refined for practical adoption.

## References

Autio, C., Schwartz, R., Dunietz, J., Jain, S., Stanley, M., Tabassi, E., Hall, P., & Roberts, K. (2024). *Artificial intelligence risk management framework: Generative artificial intelligence profile* (NIST AI 600-1). National Institute of Standards and Technology. https://doi.org/10.6028/NIST.AI.600-1

Bade, A. M., Babate, A. I., & Bara, M. W. (2026). AI-driven cyber threats and defences in Nigeria's digital economy: A literature review. I*nternational Journal of Research and Innovation in Applied Science*, 11(2), 1505–1511. https://doi.org/10.51584/IJR IAS.2026.110200140

Baxter, G., & Sommerville, I. (2011). Socio-technical systems: From design methods to systems engineering. *Interacting with Computers*, 23(1), 4–17. https://doi.org/10.1016/j.intcom.201 0.07.003

Cains, M. G., Flora, L., Taber, D., King, Z., & Henshel, D. S. (2022). Defining cyber security and cyber security risk within a multidisciplinary context using expert elicitation. *Risk Analysis*, 42(8), 1643–1669. https://doi.org/10.1111/risa.13687

Cadet, E., Etim, E. D., Essien, I. A., Erigha, E. D., Babatunde, L. A., Ajayi, J. O., & Obuse, E. (2024). Large language models for cybersecurity policy compliance and risk mitigation. *International Journal of Scientific Research in Humanities and Social Sciences*, 1(2), 612–643. https://doi.org/10.32628 /IJSRSSH242560

Collins, C., Dennehy, D., Conboy, K., & Mikalef, P. (2021). Artificial intelligence in information systems research: A systematic literature review and research agenda. *International Journal of Information Management*, 60, 102383. https://doi.org/10.1016/ j.ijinfomgt. 2021.102383

ENISA. (2023). Artificial intelligence and cybersecurity research. European Union Agency for Cybersecurity. https://www.enisa.europa.eu/publications/artificial-intelligence-and-cybersecurity-research

Eke, R. N. (2026). Determinants of artificial intelligence adoption for cybersecurity in Nigeria's banking sector: Evidence from survey and thematic analysis. *African Journal of Management and Business Research*, 21(1), 245–265. https://doi.org/ 10.62154/ ajmbr.2025.021.01022

Federal Republic of Nigeria. (2021). National cybersecurity policy and strategy 2021. Office of the National Security Adviser/ngCERT.

Federal Republic of Nigeria. (2024). National artificial intelligence strategy 2024. National Centre for Artificial Intelligence and Robotics / NITDA.

Ferrag, M. A., Alwahedi, F., Battah, A., Cherif, B., Mechri, A., Tihanyi, N., Bisztray, T., & Debbah, M. (2025). Generative AI in cybersecurity: A comprehensive review of LLM applications and vulnerabilities. *Internet of Things and Cyber-Physical Systems*, 5, 1–46. https://doi.org /10.1016/j.iotcps.2025.01.001

Heiding, F., Schneier, B., Vishwanath, A., Bernstein, J., & Park, P. S. (2024). Devising and detecting phishing emails using large language models. IEEE Access, 12, 42131–42146. https://doi.org/10.1109/ACCESS.2024.3375882

Jaffal, N. O., Alkhanafseh, M., & Mohaisen, D. (2025). Large language models in cybersecurity: A survey of applications, vulnerabilities, and defense techniques. AI, 6(9), 216. https://doi.org /10.3390/ai6090216

Ji, Z., Lee, N., Frieske, R., Yu, T., Su, D., Xu, Y., Ishii, E., Bang, Y., Madotto, A., & Fung, P. (2023). Survey of hallucination in natural language generation. ACM Computing Surveys, 55(12), Article 248. https://doi.org/10.1145/3571730

Karras, A., Theodorakopoulos, L., Karras, C., Theodoropoulou, A., Kalliampakou, I., & Kalogeratos, G. (2025). LLMs for cybersecurity in the big data era: A comprehensive review of applications, challenges, and future directions. Information, 16(11), 957. https://doi.org/10.3390/ info16110957

Moia, V. H. G., Sanz, I. J., Rebello, G. A. F., de Meneses, R. D., Hitaj, B., & Lindqvist, U. (2026). LLM in the middle: A systematic review of threats and mitigations to real-world LLM-based systems. Computer Science Review, 61, 100916. https://doi.org/10.1016/j.cos rev.2026.100916

Ndoh, J. A., Iwuh, C. E., & Ugochukwu, S. C. (2026). Artificial intelligence and security in Nigeria: A socio-technical analysis of opportunities and constraints. African Journal of Social and Behavioural Sciences, 16(1).

Nigeria Data Protection Commission. (2023). Nigeria Data Protection Act, 2023. https://ndpc.gov.ng/ndp-act-2023/

OWASP Foundation. (2025). OWASP Top 10 for Large Language Model Applications 2025. https://genai.owasp.org/resource/owasp-top-10-for-llm-applications-2025/

Passi, S., Dhanorkar, S., & Vorvoreanu, M. (2024). Appropriate reliance on generative AI: Research synthesis (Technical Report MSR-TR-2024-7). Microsoft Corporation.

Passi, S., & Vorvoreanu, M. (2022). Overreliance on AI: Literature review (Technical Report MSR-TR-2022-12). Microsoft Corporation.

Zakari, M. (2024). Implication of artificial intelligence on national security for the Nigeria security agencies. Journal of Terrorism Studies, 6(1), Article 6. https://doi.org/10.7454/jts.v6i1.1075

Zou, W., Geng, R., Wang, B., & Jia, J. (2025). PoisonedRAG: Knowledge corruption attacks to retrieval-augmented generation of large language models. In 34th USENIX Security Symposium (USENIX Security 25). USENIX Association. https://www.usenix.org/conference/usenixsecurity25/ presentation/zou-poisonedrag