

CENTRAL ASIAN JOURNAL OF THEORETICAL AND APPLIED SCIENCE

https://cajotas.centralasianstudies.org/index.php/CAJOTAS

Volume: 07 Issue: 04 | October 2026 ISSN: 2660-5317

Article

Artificial Intelligence Evolution, Emerging Risks and Safety Challenges: Implications For National Security and Society

Muhammad Auwal Sule, Ph.D[1]

1. Makieze International, Abuja, Nigeria
* Correspondence: albalagu1@gmail.com

Abstract: Artificial intelligence (AI) has evolved from rule-based computational systems into increasingly capable generative, multimodal and agentic systems able to perform complex cognitive tasks across language, science, coding, education, commerce, security and public administration. This rapid evolution presents substantial opportunities for productivity, innovation and national development, but it simultaneously creates new risks involving cybercrime, misinformation, deepfakes, privacy violations, discrimination, labour-market disruption, autonomous decision-making and potential loss of human control. This study examines the evolution of AI, emerging risks and safety challenges, with particular attention to implications for national security and society. Using a qualitative documentary research design, the study analyses secondary evidence published principally between 2021 and 2026 from the International AI Safety Report, Stanford AI Index, UNESCO, OECD, NIST, United Nations, World Bank and relevant Nigerian institutions. The analysis finds that AI capability is advancing faster than many existing governance, evaluation and institutional mechanisms. Globally, organisational AI adoption reached 88% in 2025, while generative AI reached approximately 53% population adoption within three years. Documented AI incidents increased from 233 in 2024 to 362 in 2025. Nigeria presents both significant opportunity and vulnerability: its National Artificial Intelligence Strategy seeks to position AI as a driver of productivity and human development, while high workplace adoption, digital expansion and emerging AI-generated disinformation create new security and governance challenges. The study argues that AI safety should be treated as a national security and societal resilience issue rather than solely a technology-sector concern. It recommends a whole-of-government AI safety governance architecture, mandatory risk assessment for high-impact AI applications, strengthened national capacity for AI-enabled cyber and information-security threats, and broad-based AI literacy and responsible-use programmes.

Keywords: Artificial intelligence; AI safety; emerging risks; national security; generative AI; AI governance

Citation: Sule, Ph.D M. A. Artificial Intelligence Evolution, Emerging Risks and Safety Challenges: Implications For National Security and Society. Central Asian Journal of Theoretical and Applied Science 2026, 7(4), 167-180.

Received: 10th Jun 2026
Revised: 11th Jul 2026
Accepted: 20th Aug 2026
Published: 18th Sep 2026

Copyright: © 2026 by the authors. Submitted for open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/)

1. Introduction

Artificial intelligence has moved from a specialised field of computer science into a general-purpose technology increasingly embedded in economic, social, governmental and security systems. Earlier generations of AI were largely associated with rule-based systems, expert systems and narrowly defined machine-learning applications. Contemporary AI has expanded into generative AI, multimodal systems and increasingly agentic applications capable of processing text, images, audio, video and code, while performing sequences of tasks with limited human intervention [1]. The International AI Safety Report observes that leading general-purpose AI systems can now converse in

numerous languages, generate realistic images and short videos, write computer code and solve advanced mathematics and scientific problems. AI agents are also becoming more capable of carrying out multi-step tasks, particularly in software engineering. The speed of this evolution is particularly significant. The 2026 Stanford AI Index reports that industry produced more than 90% of notable frontier AI models in 2025, while global organisational adoption reached 88%. Generative AI reached approximately 53% population adoption within three years, a rate faster than the historical diffusion of the personal computer or the internet. The same report indicates that documented AI incidents increased from 233 in 2024 to 362 in 2025. These developments demonstrate a widening gap between technological capability and the capacity of institutions to evaluate, regulate and safely deploy increasingly powerful systems [2].

The global AI economy is also expanding rapidly. Corporate AI investment reached approximately US$581.7 billion in 2025, while private AI investment increased by 127.5%. Generative AI accounted for a substantial proportion of this expansion. At the same time, AI infrastructure requires increasing amounts of computing power, energy, water and specialised semiconductor capacity. Stanford HAI reports that global AI compute capacity grew approximately 3.3 times per year since 2022 and reached 17.1 million H100-equivalents. The environmental and infrastructure dimensions of AI therefore increasingly form part of the broader safety discussion [3]. The evolution of AI nevertheless presents a fundamental policy paradox. The same capabilities that generate economic and social benefits can also be used maliciously. AI can improve medical research, education, agricultural planning, financial services, scientific discovery and public administration, while similar technologies can generate fraudulent content, automate cyberattacks, manipulate public opinion or amplify discrimination. The International AI Safety Report categorises major risks into misuse, malfunction and systemic risks. Misuse includes fraud, manipulation, cybercrime and potentially harmful biological and chemical applications; malfunction includes unreliable or unexpected system behaviour; while systemic risks arise from widespread deployment and include labour-market, societal and human-autonomy concerns. The emergence of generative AI has intensified the information-security dimension of these risks. AI systems can generate realistic text, images, audio and video at low cost and high speed. The International AI Safety Report notes that harmful incidents involving AI-generated content have increased substantially since 2021 [4]. It further reports that participants in one study mistook AI-generated text for human-written text 77% of the time, while another study found that listeners mistook AI-generated voices for real speakers 80% of the time. Such capabilities create vulnerabilities for fraud, identity deception, blackmail, reputational attacks, political manipulation and social engineering.

The cyber-security implications are equally important. AI systems can assist users in identifying software vulnerabilities, analysing malware and generating code. The 2026 International AI Safety Report indicates that AI systems have increasingly been used in real-world cyber operations and that malicious actors have used AI to automate significant parts of cyberattacks. However, the report also stresses that fully autonomous end-to-end cyberattacks have not yet been demonstrated and that the ultimate balance between AI-enabled offensive and defensive cyber capabilities remains uncertain [5].

European experience demonstrates the importance of institutional preparedness. The European Union has adopted a risk-based approach to AI governance through the EU Artificial Intelligence Act, which distinguishes AI applications according to levels of risk and imposes stronger requirements on higher-risk systems. The European model reflects a recognition that AI cannot be governed adequately through broad ethical principles alone. Governance requires mechanisms for risk classification, transparency, human oversight, accountability and compliance. The United Kingdom provides another significant example. The UK's AI safety architecture has increasingly focused on evaluation of frontier models, risk monitoring and scientific assessment. The International AI Safety Report 2026 itself emerged from an international process involving more than 30

countries and international organisations, including the United Kingdom, European Union, United Nations and OECD. More than 100 experts contributed to the report, reflecting the increasingly international character of AI safety governance [6].

The European experience also demonstrates that AI safety cannot be separated from cybersecurity, privacy and human rights. The National Institute of Standards and Technology (NIST) AI Risk Management Framework, although developed in the United States, provides an influential example of lifecycle-based risk management. It emphasises that trustworthy AI should be valid and reliable, safe, secure and resilient, accountable and transparent, explainable and interpretable, privacy-enhanced and fair. NIST's Generative AI Profile further provides practical approaches for identifying and managing risks across the AI lifecycle.

Across Africa, AI adoption is occurring within a markedly different institutional and infrastructural environment. The continent has substantial youth populations, rapidly expanding digital connectivity and significant demand for technology-enabled development. At the same time, many African countries face gaps in digital infrastructure, computing capacity, specialised skills, cybersecurity and regulatory enforcement. The International AI Safety Report (2026) observes that AI adoption remains highly uneven globally, with adoption rates across much of Africa, Asia and Latin America likely below 10%, although individual countries and population groups can display substantially higher usage [7]. Africa's information environment creates a particularly important context for AI safety. The Africa Center for Strategic Studies reported that documented disinformation campaigns on the continent increased nearly fourfold between 2022 and 2024, reaching 189 documented campaigns across at least 39 African countries. Conflict-affected countries experienced substantially higher levels of disinformation. Nigeria has also been identified among countries where digital disinformation has increasingly intersected with political and electoral processes [8].

South Africa provides an important African illustration of the opportunities and challenges surrounding AI development. The country has comparatively advanced digital infrastructure, an established research ecosystem and growing interest in AI applications across financial services, public administration, health and industry. It has also emerged as one of the African countries developing advanced AI skills. Stanford HAI identifies South Africa among countries where AI engineering skills are growing rapidly. This illustrates the potential for African countries to participate actively in AI development rather than merely consuming foreign systems.

Kenya provides a second African comparison through its growing digital economy and technology ecosystem. The country's experience demonstrates the importance of digital public infrastructure, innovation and technology entrepreneurship, but also the need for effective data governance and responsible AI deployment [9]. Kenya was among the African countries represented on the Expert Advisory Panel for the International AI Safety Report 2026, alongside Nigeria and Rwanda, demonstrating increasing African participation in global AI safety discussions.

Nigeria presents a particularly important case because it combines a large population, a substantial technology ecosystem, rapidly expanding digital participation and a growing policy commitment to AI. The National Centre for Artificial Intelligence and Robotics (NCAIR) identifies AI, robotics and Fourth Industrial Revolution technologies as important components of Nigeria's digital transformation. NCAIR's stated mission includes promoting research, development and adoption of AI and robotics for economic growth, improved quality of life and global competitiveness. Nigeria formally developed its National Artificial Intelligence Strategy in 2025. The strategy identifies AI as an instrument for sustainable development, innovation, productivity and human well-being and seeks to strengthen Nigeria's position in AI adoption and development [10]. The strategy builds on initiatives including NCAIR and other government programmes and

recognises the need to address infrastructural and ecosystem constraints associated with AI development in an emerging economy

Nigeria's AI opportunity is strengthened by growing digital participation. The Stanford AI Index reports that in 2025 more than 80% of employees surveyed in Nigeria, India, China, the United Arab Emirates, Egypt and Saudi Arabia reported using AI at work either regularly or semi-regularly. Nigeria's high level of workplace AI use is significant because it suggests that AI is already entering professional and organisational environments rather than remaining a future technology [11].

Nigeria has also invested in developing an AI ecosystem through research, skills and infrastructure. The Nigeria Artificial Intelligence Research Scheme was established to support AI research and develop skills in areas such as data science, AI and cloud computing. UNCTAD notes that Nigeria's approach reflects a broader attempt to use AI policy to promote structural transformation and workforce development. The organisation also emphasises that AI policy requires a whole-of-government approach encompassing infrastructure, data, skills, regulation, education, industry and trade [12].

UNESCO's engagement with Nigeria further demonstrates the growing importance of AI governance. In May 2025, UNESCO, the Federal Ministry of Communications, Innovation and Digital Economy and the European Union conducted a strategic policy engagement on Nigeria's Artificial Intelligence Readiness Assessment Methodology. The assessment examined AI governance, infrastructure, education and research, the AI economy and the socio-cultural environment, with ethics and inclusion cutting across the assessment. The security dimension is becoming increasingly relevant. Nigeria already faces cybercrime, fraud, terrorism, violent extremism, misinformation, identity theft and other forms of technology-enabled crime. AI can lower barriers to some of these activities by making content production, social engineering and certain cyber operations easier to conduct. African security experts have warned that AI is already being employed by criminal networks, terrorist organisations and state actors to amplify cyber and information threats.

The electoral environment is particularly sensitive. African disinformation campaigns have increasingly targeted elections, while Nigeria's 2023 election demonstrated the capacity of digital platforms to influence political narratives. Ahead of Nigeria's 2027 electoral cycle, AI-generated misinformation, synthetic audio, manipulated video and fabricated political statements could increase the speed and credibility of false narratives. Reports in 2026 have already raised concerns about AI-generated disinformation and manipulated multimedia content in Nigeria's political environment. The central concern of this study is therefore not that AI is inherently dangerous. Rather, the concern is that AI capabilities are evolving faster than the institutions, skills, safeguards and public understanding required to manage them. AI safety must consequently be understood as a multidimensional issue encompassing technological reliability, cybersecurity, human rights, social stability, economic resilience, national security and responsible governance [13].

2.0 Statement of the Problem

Artificial intelligence is developing at a speed that challenges conventional approaches to technological regulation. AI systems are becoming more capable, more accessible and increasingly embedded in economic and governmental activities, while many institutions still lack adequate technical expertise, evaluation mechanisms and governance structures.

The first problem is the widening capability-governance gap. Stanford HAI reports rapid advances in AI capability and adoption alongside persistent weaknesses in responsible-AI measurement. Documented AI incidents increased to 362 in 2025, while transparency among leading AI companies declined. The average Foundation Model Transparency Index score reportedly fell from 58 in 2024 to 40 in 2025.

The second problem is the expansion of AI-enabled criminality. Generative AI can facilitate fraud, impersonation, social engineering, cybercrime and synthetic media production. The International AI Safety Report identifies increasing evidence of harmful AI-generated content and real-world use of AI in cyber operations.

The third problem is information integrity. Deepfakes and synthetic content can undermine citizens' ability to distinguish authentic information from manipulated material. In fragile or polarised environments, this can weaken public trust, intensify communal tensions and complicate security operations [14].

The fourth problem is the socioeconomic distribution of AI benefits and risks. AI can increase productivity and create new economic opportunities, but it may also displace particular categories of work, especially routine and entry-level tasks. Stanford HAI (2026) reports that employment among software developers aged 22–25 declined nearly 20% from 2024, while one-third of surveyed organisations expected AI-related workforce reductions in the following year.

The fifth problem concerns institutional preparedness in Nigeria. Although Nigeria has developed a National AI Strategy and several AI-related programmes, implementation requires coordination among technology institutions, security agencies, regulators, educational institutions, businesses and civil society. Without such coordination, AI policy may become fragmented between innovation promotion and risk management.

The research problem, therefore, is not whether Nigeria should adopt AI. AI adoption is already occurring. The more important question is whether Nigeria can capture AI's developmental benefits while preventing technological evolution from creating unmanaged security, economic and societal vulnerabilities [15].

3.0 Objectives of the Study

The study seeks to:

i. examine the evolution and expanding capabilities of artificial intelligence and their implications for contemporary society;

ii. analyse the major emerging risks and safety challenges associated with generative, multimodal and increasingly autonomous AI systems;

iii. assess the implications of AI-related risks for national security, social stability, economic resilience and human rights, with particular reference to Nigeria; and

iv. evaluate existing AI governance and safety responses and propose measures for responsible, secure and inclusive AI development in Nigeria.

4.0 Research Questions

i. How has artificial intelligence evolved, and what new capabilities distinguish contemporary AI systems?

ii. What major risks and safety challenges are emerging from generative, multimodal and increasingly autonomous AI?

iii. What implications do AI-related risks have for Nigeria's national security, social stability, economic resilience and human rights?

iv. How adequate are existing AI governance arrangements, and what measures should Nigeria adopt to strengthen responsible and safe AI development?

5.0 Conceptual Clarification

5.1 Artificial Intelligence Evolution

Artificial intelligence evolution refers to the progressive development of computational systems from rule-based and narrowly specialised applications towards machine-learning, deep-learning, generative, multimodal and increasingly agentic systems. The evolution is not merely an increase in computational speed. It involves

expanding capabilities in language, reasoning, perception, content generation, coding, decision assistance and autonomous task execution.

Contemporary general-purpose AI differs from earlier specialised systems because it can perform multiple categories of tasks through a common underlying model. The International AI Safety Report notes that leading systems can perform language translation, image and video generation, coding and advanced scientific reasoning.

5.2 AI Risk

AI risk refers to the possibility that an AI system, its developer, user or broader deployment environment may produce harmful outcomes. Risks may arise through deliberate misuse, technical malfunction or systemic effects from widespread adoption. The International AI Safety Report distinguishes these three categories and emphasises that some risks are already observable while others remain uncertain.

5.3 AI Safety

AI safety refers to the technical, institutional, legal and social measures designed to prevent AI systems from causing unacceptable harm and to ensure that their behaviour remains reliable, secure and subject to appropriate human control. UNESCO identifies safety and security, human oversight, transparency, accountability, privacy, fairness and sustainability as central principles for responsible AI.

5.4 AI Governance

AI governance refers to the institutions, laws, standards, policies, procedures and accountability mechanisms through which AI development and deployment are directed and monitored. It encompasses regulation, risk assessment, data governance, procurement, testing, auditing, human oversight, incident reporting and enforcement.

5.5 AI-Enabled National Security Risk

AI-enabled national security risk refers to threats arising when AI capabilities are used to undermine the security of individuals, institutions, critical infrastructure, democratic processes or the state. Such threats may involve cyberattacks, disinformation, identity manipulation, automated fraud, surveillance abuse, attacks against critical infrastructure or AI-assisted activities by violent non-state actors.

The study is anchored primarily in Risk Society Theory, complemented by Human Security Theory and the Technology-Enabled Threat Perspective.

Risk Society Theory, associated particularly with Beck, argues that modern societies increasingly generate risks through the very technologies and systems designed to promote development. Technological advancement therefore produces not only benefits but also new categories of uncertainty whose consequences can transcend national borders and institutional boundaries. AI fits this framework because the technology generates productivity, knowledge and innovation while simultaneously creating risks that may be difficult to predict or contain.

The theory is particularly relevant because AI risks are not necessarily proportional to the intentions of individual users. A system designed for legitimate purposes may generate harmful outputs through unexpected behaviour, biased data or misuse. The International AI Safety Report similarly distinguishes risks from deliberate misuse, malfunctions and systemic deployment.

Human Security Theory provides a second analytical lens. Human security shifts attention from the state alone towards the safety and dignity of individuals and communities. AI can affect economic security through labour displacement, personal security through fraud and identity theft, political security through manipulation, health security through unreliable information, and community security through misinformation and social polarisation.

The Technology-Enabled Threat Perspective complements the two theories by recognising that emerging technologies can alter the capabilities of both legitimate institutions and malicious actors. AI can strengthen security agencies through intelligence analysis, anomaly detection and predictive tools, while simultaneously strengthening cybercriminals, fraudsters and hostile influence actors.

Together, these perspectives generate the central proposition of the study: Where AI capabilities expand faster than institutional capacity, safety standards, digital literacy and accountability mechanisms, technological benefits may coexist with increasingly complex security and societal risks.

7.0 Methodology

The study adopts a qualitative documentary research design. It relies primarily on secondary evidence published between 2021 and 2026, supplemented where necessary by foundational theoretical and policy sources. The principal sources include the International AI Safety Report, Stanford AI Index, UNESCO, OECD, NIST, United Nations, World Bank, UNCTAD, African Center for Strategic Studies, Nigeria's National Artificial Intelligence Strategy, NCAIR and relevant Nigerian government documents. The documentary approach is appropriate because AI safety is a rapidly evolving field in which important evidence is generated through technical assessments, institutional reports, incident databases, policy frameworks and international governance documents. Data were analysed thematically under four dimensions: AI evolution and capability; emerging risks; national security and societal implications; and governance and safety responses. Evidence was subjected to comparative interpretation across global, European, African and Nigerian contexts.

The study does not assume that every potential AI risk will materialise. Distinctions are therefore maintained between documented harms, emerging risks supported by evidence, and future risks characterised by substantial uncertainty. This approach is consistent with the International AI Safety Report's emphasis on distinguishing present evidence from uncertain future trajectories.

8.0 Evolution of Artificial Intelligence and the New Capability Environment

The evolution of AI can broadly be understood through five stages. The first stage involved rule-based and symbolic AI, in which systems relied heavily on predefined rules and expert knowledge. These systems were useful in controlled environments but had limited ability to learn from large datasets.

The second stage involved machine learning, where algorithms learned patterns from data rather than relying entirely on manually defined rules. This significantly expanded AI applications in areas such as financial prediction, recommendation systems, fraud detection and image recognition.

The third stage involved deep learning, which enabled substantial advances in speech recognition, computer vision and natural-language processing. Large neural networks trained on extensive datasets produced capabilities that were difficult to achieve through conventional programming.

The fourth stage is generative AI. Large language models and related systems moved AI from classification and prediction towards content creation. AI can now produce text, images, audio, video and software code. The rapid diffusion of generative AI has transformed public access to sophisticated computational capabilities.

The fifth emerging stage is agentic and multimodal AI. These systems can combine multiple capabilities and undertake sequences of tasks. The International AI Safety Report (2026) notes that AI agents are increasingly capable in areas such as software engineering, while the duration of software tasks they can complete with high success has been increasing rapidly.

This evolution changes the security environment because capability is no longer restricted to large institutions. Many advanced AI tools are accessible through ordinary digital devices. This democratisation of capability creates a dual-use environment: the same tool can empower a student, entrepreneur, doctor or public servant while also assisting a fraudster, propagandist or cybercriminal.

The capability environment is therefore characterised by three important features: speed, accessibility and scalability. AI can generate content much faster than humans; many tools are inexpensive or free; and one malicious actor can potentially produce content for thousands or millions of targets.

Result and Discussion

9.0 Emerging Risks and Safety Challenges

9.1 AI-Generated Disinformation and Deepfakes

Synthetic media represents one of the most immediate societal risks. AI can generate realistic images, audio and video that imitate public figures or ordinary individuals. Deepfake technology can therefore be used for fraud, political manipulation, blackmail, reputational attacks and communal provocation. The International AI Safety Report (2026) reports that AI-generated content can already be highly convincing, with substantial difficulties in distinguishing synthetic material from authentic content.

For Nigeria, the electoral implications are particularly significant. False statements attributed to political candidates, security agencies or community leaders could spread rapidly during periods of political tension. The danger lies not only in convincing people of false information but also in creating uncertainty about genuine information. When citizens can no longer determine whether a recording is authentic, trust itself becomes vulnerable.

9.2 Cybersecurity and AI-Assisted Crime

AI is increasingly relevant to cybersecurity because it can support both attack and defence. It can assist legitimate defenders in detecting anomalies, analysing malware and identifying vulnerabilities [16]. The same capabilities can be misused to assist cybercriminals.

The International AI Safety Report documents increasing evidence of AI use in real-world cyber operations. It notes that AI can identify software vulnerabilities and generate code associated with cyberattacks, although fully autonomous end-to-end attacks have not yet been demonstrated.

For Nigeria, this creates particular risks for banking, telecommunications, government databases, electricity systems, healthcare infrastructure and digital public services. AI-enabled attacks could potentially increase the speed and scale of attacks while reducing the level of technical expertise required.

9.3 Fraud, Identity Theft and Social Engineering

Generative AI can make fraudulent communication more convincing. Voice cloning can imitate relatives, executives, political leaders or public officials. AI-generated text can make phishing messages grammatically sophisticated and personalised. The International AI Safety Report identifies fraud, scams, blackmail, extortion and identity-related harms among the criminal uses of AI-generated content. This creates a significant societal challenge because conventional public-awareness messages often assume that citizens can recognise fraudulent communication through spelling mistakes, poor grammar or obvious inconsistencies. AI reduces the reliability of these indicators [17].

9.4 Bias, Discrimination and Exclusion

AI systems learn from data. Where training data contain historical inequalities or insufficient representation, AI systems can reproduce or amplify those patterns. UNESCO emphasises fairness, non-discrimination, diversity and inclusion as central principles of AI

governance. It also stresses that AI systems should respect human rights and dignity. This is especially relevant to Nigeria because imported AI systems may not adequately represent Nigerian languages, cultures and social contexts. The International AI Safety Report notes that AI performance can decline significantly in less represented cultural and linguistic environments. AI deployment in recruitment, education, policing, financial services and public administration could therefore produce unfair outcomes if systems are not tested against local populations.

9.5 Privacy and Data Protection

AI systems depend heavily on data. The greater the volume and sensitivity of data processed, the greater the consequences of poor data governance. UNESCO identifies privacy and data protection as core AI ethics principles. NIST similarly includes privacy enhancement among the characteristics of trustworthy AI. The Nigerian challenge is to ensure that AI development does not undermine data-protection rights. Government agencies and private companies increasingly possess large datasets containing financial, health, identity, educational and behavioural information. AI can generate substantial value from such data, but improper collection, processing or disclosure can produce serious individual and institutional harm.

9.6 Reliability and Hallucination

AI systems can generate false information while presenting it confidently. This creates risks when users treat AI-generated outputs as authoritative. The 2026 AI Index found substantial variation in hallucination rates across leading models, demonstrating that reliability remains an unresolved issue. The International AI Safety Report similarly describes false information, flawed code and misleading advice as examples of AI malfunction. This problem becomes particularly serious when AI is used in medicine, law enforcement, military decision-making, emergency management or public policy without appropriate human verification.

9.7 Labour-Market Disruption

AI has significant implications for employment. It can complement workers, increase productivity and create new occupations, but it can also reduce demand for particular tasks. Stanford HAI (2026) reports that employment among software developers aged 22–25 declined nearly 20% from 2024, while one-third of organisations expected workforce reductions in the coming year.

Nigeria faces a distinctive challenge because its youthful population requires substantial employment creation. The World Bank notes that Africa's working-age population is projected to more than double by 2050 and that the continent needs approximately 28 million new jobs annually. AI could increase productivity and generate new opportunities, but inadequate skills development could also widen inequality.

9.8 Loss of Human Oversight

As AI systems become increasingly agentic, the question of human control becomes more important. An AI system that merely provides information creates a different risk profile from an AI system capable of initiating transactions, modifying databases, communicating with third parties or controlling physical systems. UNESCO therefore emphasises human oversight and determination, while NIST stresses lifecycle-based testing, evaluation and monitoring. The fundamental safety principle should be that high-impact decisions remain subject to identifiable human responsibility.

9.9 Biological and Chemical Misuse

The increasing scientific capability of AI creates another category of concern. The International AI Safety Report (2026) indicates that general-purpose AI can provide detailed information relevant to biological and chemical research and that some systems have demonstrated high performance on specialised scientific tasks. Several developers

strengthened safeguards in 2025 because pre-deployment testing could not rule out meaningful assistance to novices in biological weapons development.

This does not mean that AI has independently created such weapons. Rather, it demonstrates the dual-use character of advanced scientific AI and the need for appropriate safeguards around high-risk applications.

10.0 AI, National Security and Society in Nigeria

AI changes the national security environment in at least five ways. First, it alters the information environment. Security agencies increasingly operate in environments where false information can spread faster than official responses. AI-generated content can complicate crisis communication and create uncertainty during security incidents.

Second, AI affects cybersecurity. Government networks, financial systems, telecommunications and critical infrastructure can face more sophisticated attacks. At the same time, AI can strengthen defensive monitoring if appropriately deployed.

Third, AI influences counterterrorism and violent-extremism prevention. AI may assist authorities with information analysis, early warning and identification of emerging patterns. However, automated profiling can also produce false positives, discriminatory outcomes or privacy violations if human oversight is inadequate.

Fourth, AI affects electoral security. AI-generated political content could amplify polarisation, impersonation and false narratives. African experience demonstrates that information manipulation already represents a significant security concern, and AI can increase the speed and scale of content generation.

Fifth, AI affects economic security. A country that fails to develop AI capabilities may become dependent on foreign technologies, data infrastructures and models. Such dependence can create strategic vulnerabilities in critical sectors.

Nigeria's National AI Strategy recognises the importance of AI for productivity, innovation, sustainable development and human well-being. The strategic challenge is therefore to combine AI adoption with AI safety rather than treating the two as competing objectives.

11.0 Governance and Safety Responses

Effective AI governance requires a shift from reactive regulation towards lifecycle risk management. NIST's AI Risk Management Framework provides one useful model through the functions of governing, mapping, measuring and managing AI risks. Its Generative AI Profile extends these principles to risks specific to generative systems.

UNESCO provides a complementary human-rights framework. Its Recommendation on the Ethics of AI emphasises proportionality, safety and security, privacy, multi-stakeholder governance, accountability, transparency, human oversight, sustainability, awareness and fairness.

The OECD has also emphasised the importance of common AI-incident reporting. Its 2025 framework identifies 29 criteria intended to help jurisdictions identify high-risk systems and assess impacts on people and the environment.

Nigeria already possesses important institutional building blocks, including NCAIR, NITDA and its National AI Strategy. UNESCO's AI readiness engagement also provides an opportunity to assess governance, infrastructure, education, research, economic and socio-cultural dimensions of AI development.

The central governance challenge is coordination. AI does not belong to one ministry. It intersects with communications, cybersecurity, education, health, finance, defence, policing, justice, elections, labour, industry and foreign affairs. AI governance must therefore become a whole-of-government and whole-of-society process.

12.0 Findings of the Study

The study produces four principal findings. First, AI capability is evolving faster than conventional governance systems. Technical performance, adoption and investment are increasing rapidly, while responsible-AI evaluation and transparency remain comparatively weaker. Stanford HAI (2026) identifies a widening gap between AI capabilities and society's capacity to measure and manage them.

Second, AI risks are increasingly dual-use and interconnected. Cybersecurity, misinformation, fraud, privacy, labour-market and national-security risks cannot be treated as isolated categories. A single AI capability may generate benefits in one context and vulnerabilities in another.

Third, Nigeria is simultaneously an AI opportunity environment and a risk environment. High workplace AI adoption and the National AI Strategy demonstrate significant potential, while cybersecurity vulnerabilities, information manipulation and institutional capacity gaps create risks requiring preventive governance.

Fourth, AI safety requires institutional coordination rather than technology regulation alone. International evidence from UNESCO, NIST, OECD and the International AI Safety Report demonstrates that effective AI safety involves risk assessment, monitoring, incident reporting, human oversight, transparency, skills and public awareness.

13.0 Implications for National Security and Sustainable Development

The security implications of AI extend beyond conventional military concerns.

13.1 Information Security

AI-generated misinformation can undermine public confidence in government institutions, security agencies, elections and public-health messages. The ability to manufacture convincing synthetic content creates a new challenge for strategic communication.

13.2 Cybersecurity

AI can increase the sophistication, speed and scale of cyber threats while simultaneously improving defensive capabilities. Nigeria therefore requires AI-enabled cyber defence alongside stronger human expertise.

13.3 Democratic Stability

Election periods are particularly vulnerable because political narratives spread rapidly and citizens make decisions under conditions of uncertainty. AI-generated impersonation and deepfakes could complicate electoral communication and increase tensions.

13.4 Economic Resilience

AI can increase productivity and create new industries, but it can also disrupt employment. Nigeria's policy challenge is to develop AI skills faster than AI displaces routine tasks.

13.5 Human Security

AI affects personal privacy, employment, access to services, reputation, financial security and political participation. A national AI policy that focuses only on economic growth would therefore be incomplete.

13.6 Strategic Autonomy

AI infrastructure, chips, cloud services, data and foundation models are concentrated internationally. Nigeria's long-term security will partly depend on its capacity to develop domestic expertise and strategic partnerships without creating excessive technological dependency.

14.0 Recommendations

Recommendation One

The Federal Government should establish a coordinated national AI safety and risk-governance architecture covering high-impact public and private AI applications.

Implementation Strategies

i. NITDA, in collaboration with the Federal Ministry of Communications, Innovation and Digital Economy and relevant regulators, should develop a national AI risk-classification and assessment framework requiring high-impact AI systems to undergo documented safety, privacy, bias, cybersecurity and human-oversight assessments before deployment.

ii. Office of the National Security Adviser (ONSA) to coordinate with NITDA, the National Centre for Artificial Intelligence and Robotics, the Nigeria Data Protection Commission and relevant security institutions to establish a national AI-risk monitoring mechanism for applications with significant implications for national security and public safety.

Recommendation Two

The Federal Government should strengthen Nigeria's capacity to detect and respond to AI-enabled cybercrime, deepfakes, synthetic misinformation and technology-enabled security threats.

Implementation Strategies

Office of the National Security Adviser in collaboration with the Nigeria Police Force and relevant cybersecurity institutions to establish an AI-enabled threat-monitoring and incident-reporting capability covering major cyber, synthetic-media and AI-enabled fraud risks.

Independent National Electoral Commission (INEC), in collaboration with security agencies, technology platforms, civil society and fact-checking organisations, should develop an election-period rapid-response mechanism for identifying and communicating verified information concerning AI-generated political content and impersonation. The Federal Government should expand national AI skills, literacy and responsible-use capacity to ensure that citizens and institutions can benefit from AI without becoming disproportionately exposed to its risks.

Federal Ministry of Education and NITDA, in collaboration with universities and technical institutions, should integrate AI literacy, digital verification, cybersecurity awareness, data protection and responsible AI use into relevant education and professional-development programmes. NCAIR, in collaboration with industry and development partners, should expand practical training in AI engineering, safety evaluation, cybersecurity and responsible AI development, with emphasis on Nigerian languages, local data and locally relevant applications. The Federal Government should institutionalise human-centred safeguards for AI applications used in national security, public administration and other high-impact decision environments.

MDAs to require documented human oversight, audit trails and periodic independent evaluation for AI systems used in policing, security analysis, public services, recruitment, justice, healthcare and other high-impact functions.

NITDA in collaboration with the Nigeria Data Protection Commission, and sector regulators to establish mandatory incident-reporting and post-deployment review requirements for high-impact AI systems to ensure that harmful or discriminatory outcomes are identified and corrected promptly.

15.0 Conclusion

Artificial intelligence has evolved from specialised computational systems into a general-purpose technological infrastructure capable of transforming economic production, scientific research, education, public administration and national security. Contemporary generative and agentic systems demonstrate capabilities that would have

been considered highly advanced only a few years ago. Global organisational adoption reached 88% in 2025, while generative AI achieved approximately 53% population adoption within three years. At the same time, documented AI incidents increased to 362 in 2025, demonstrating that technological expansion is accompanied by increasingly visible risks. The evidence shows that AI risk should not be understood only as a distant or hypothetical concern. Fraud, synthetic media, cyber operations, misinformation, unreliable outputs, privacy violations and labour-market disruption are already observable. Other risks, including loss of control and advanced biological or chemical misuse, remain less certain but potentially more consequential. The appropriate policy response is therefore neither technological pessimism nor unrestricted technological optimism. Nigeria occupies a strategically important position in this emerging environment. Its National Artificial Intelligence Strategy, NCAIR, AI research initiatives and expanding digital economy provide a foundation for AI-enabled development. Nigeria also demonstrates significant AI usage in the workplace, indicating that the technology is already influencing professional and economic activity. However, the country's national-security environment creates particular vulnerabilities. Cybercrime, misinformation, electoral tensions, identity fraud and institutional capacity constraints could magnify the consequences of poorly governed AI. The emergence of increasingly realistic synthetic content is particularly significant because the credibility of information is itself becoming a security asset. The central argument of this study is therefore that AI safety should be integrated into Nigeria's national security and development architecture rather than treated as a narrow technology-sector issue. AI governance should cover the entire lifecycle of systems—from design, data and procurement to deployment, monitoring, incident reporting and retirement. The international evidence provides useful principles. UNESCO emphasises human dignity, safety, privacy, accountability, transparency, human oversight, sustainability and fairness. NIST provides a practical risk-management approach based on governing, mapping, measuring and managing risks. The OECD emphasises systematic incident reporting, while the International AI Safety Report demonstrates the importance of continuous evaluation as AI capabilities evolve. Nigeria should consequently pursue an approach that combines innovation with precaution, technological development with human oversight, and AI adoption with institutional preparedness. The objective should not be to slow technological progress unnecessarily but to ensure that progress remains aligned with national interests, human security and public safety. Ultimately, the strategic question is not whether artificial intelligence will transform Nigerian society. It already is. The more important question is whether Nigeria will shape that transformation through anticipatory governance, domestic capability and responsible innovation, or allow technological risks to develop faster than its institutions can manage them. A whole-of-government and whole-of-society approach to AI safety provides the stronger pathway towards ensuring that artificial intelligence becomes an instrument of national resilience, productivity and human development rather than a multiplier of existing security and social vulnerabilities.

REFERENCES

[1] Africa Center for Strategic Studies, Mapping a Surge of Disinformation in Africa. Washington, DC, USA: Africa Center for Strategic Studies, 2024.

[2] Africa Center for Strategic Studies, Artificial Intelligence Strategy in the Security Domain: AI-Amplified Security Threats. Washington, DC, USA: Africa Center for Strategic Studies, 2025.

[3] C. Autio, R. Schwartz, J. Dunietz, S. Jain, M. Stanley, E. Tabassi, P. Hall, and K. Roberts, Artificial Intelligence Risk Management Framework: Generative Artificial Intelligence Profile, NIST AI 600-1. Gaithersburg, MD, USA: National Institute of Standards and Technology, 2024, doi: 10.6028/NIST.AI.600-1.

[4] U. Beck, Risk Society: Towards a New Modernity. London, U.K.: Sage, 1992.

[5] Y. Bengio, S. Clare, C. Prunkl, M. Murray, M. Andriushchenko, B. Bucknall, R. Bommasani, S. Casper, T. Davidson, R. Douglas, D. Duvenaud, P. Fox, U. Gohar, R. Hadshar, A. Ho, T. Hu, C. Jones, S. Kapoor, A. Kasirzadeh, et al., International AI Safety Report 2026. AI Safety Institute, 2026.

[6] International AI Safety Report, 2026 Report: Extended Summary for Policymakers. AI Safety Institute, 2026.

[7] National Information Technology Development Agency, National Artificial Intelligence Strategy. Abuja, Nigeria: Federal Government of Nigeria, 2025.

[8] National Centre for Artificial Intelligence and Robotics, National Centre for Artificial Intelligence and Robotics. Abuja, Nigeria: NITDA, 2026.

[9] National Institute of Standards and Technology, Artificial Intelligence Risk Management Framework (AI RMF 1.0), NIST AI 100-1. Gaithersburg, MD, USA: U.S. Department of Commerce, 2023, doi: 10.6028/NIST.AI.100-1.

[10] OECD, Towards a Common Reporting Framework for AI Incidents, OECD Artificial Intelligence Papers, no. 34. Paris, France: OECD Publishing, 2025, doi: 10.1787/f326d4ac-en.

[11] OECD, Trends in AI Incidents and Hazards Reported by the Media, OECD Artificial Intelligence Papers, no. 53. Paris, France: OECD Publishing, 2026, doi: 10.1787/4f5ff43c-en.

[12] Stanford Institute for Human-Centered Artificial Intelligence, The 2026 AI Index Report. Stanford, CA, USA: Stanford University, 2026.

[13] Stanford Institute for Human-Centered Artificial Intelligence, Responsible AI. Stanford, CA, USA: Stanford University, 2026.

[14] United Nations in Nigeria, UNESCO Hosts Policy Engagement to Strengthen Nigeria's AI Readiness Assessment Report. Abuja, Nigeria: United Nations, 2025.

[15] United Nations Conference on Trade and Development, Technology and Innovation Report 2025: Inclusive Artificial Intelligence for Development. Geneva, Switzerland: United Nations, 2025.

[16] UNESCO, Recommendation on the Ethics of Artificial Intelligence. Paris, France: United Nations Educational, Scientific and Cultural Organization, 2021.

[17] World Bank, People, Skills and Digital Innovation for Jobs in Africa. Washington, DC, USA: World Bank, 2026.