SPECIAL EDITION
ISSUE 01

DIGISPACE AFRICA

AFRITECH

Synthetic Truths, Real Harm



Digispace Research Fellows intellectual contributions and research that advances conversations on digital rights, emerging technologies, and internet governance in Africa.

AFRICA'S

DIGITAL FUTURE

18 SEPTEMBER, 2026

© DIGISPACEAFRICA.ORG

Published in August 2026

Report produced by Digispace Africa

This publication may be reproduced for non-commercial use in any form provided due credit is given to the publishers, and the work is presented without any distortion.

Copyright © 2026 Digispace Africa

Email: info@digispaceafrica.org

www.digispaceafrica.org

Credits

Illustrations by: Emmanuel G Boateng

Editorial

Babara Zipporah Ayiku
Theorose E. Dzineku
Grace Dorcas Annan

Layout and design by: Grace Dorcas & Co. Studios

Creative Commons Attribution 4.0 International (CC BY 4.0)

TABLE OF CONTENT

1 **What Is the Cost of Getting AI Wrong?**
Human Development Audit of Ghana's Artificial Intelligence Strategy.

2 **Synthetic Truths, Real Harm**
Artificial Intelligence, Information Literacy, And the Growing Misinformation Crisis in Africa

3 **Towards an AI Legal Practice Framework:**
What Kenya Must Regulate Now

4 **Rethinking Account Security in Nigeria:**
Behavioral Vulnerabilities, Shared-Device Practices, and Platform Trust & Safety Design.

5 **A Strategic Framework for Zimbabwe's meaningful participation in Internet Governance Forums**

6 **Exploring the Protection of Community Data Rights in Digital Medical Research in Kenya**
Digital Surveillance

7 No Human, No Criminal Liability?
Agentic AI and the Accountability Gap in Nigeria's Child Online Protection Framework.

Exploring the Protection of Community Data Rights in Digital Medical Research in Kenya

Beyond nominal inclusion: A Strategic Framework for Zimbabwe's meaningful participation in Internet Governance Forums

Generative AI misuse: How Generative AI Affects Women in Morocco

Linguistic Sovereignty and Digital Transparency: Using GenAI to Democratize Freedom of Information in Nigeria

12

Generative AI Misuse: Policy Brief for Gender-Sensitive Safeguards in Morocco

AI Governance in Morocco

13

Governing The Algorithm: Assessing Ghana's Legal And Institutional Readiness For Algorithmic Decision-Making

14

Outing as a digital weapon:

Why global social media platforms fail to safeguard criminalised LGBTIQ+ identities in Ethiopia and what a Contextual Harm Policy should include.

15

AI as a Tool for Electoral Integrity: Governance Frameworks for AI-Assisted Disinformation Monitoring during Nigeria's Elections

16

Developing Adaptive Regulatory Frameworks for Artificial Intelligence in Africa: Balancing Innovation, Accountability, and Human Rights in Emerging Digital Economies.

17

From Protests to Policy: Architecting Digital Infrastructure for Inclusive Youth Governance in Ghana

Acknowledgment

We sincerely thank the fellows of the Digispace Research Fellowship for their dedication, intellectual contributions, and commitment to producing research that advances conversations on digital rights, emerging technologies, and internet governance in Africa.

We are grateful to our mentors, advisors, and reviewers for their guidance, feedback, and support throughout the fellowship. We extend special appreciation to the Fellowship Manager Barbara Zipporah Ayiku, whose coordination, dedication, and behind-the-scenes work were instrumental in making the fellowship and this compilation possible.

Finally, we acknowledge the Digispace Africa team for their continued commitment to creating spaces that support emerging researchers and amplify critical research from Africa.

EXECUTIVE DIRECTOR	**THEOROSE ELIKPLIM DZINEKU**
FELLOWSHIP MANAGER	**BARBARA ZIPPORAH AYIKU**
EDITORORIAL	**EWURADWOA BANDELE**
GRAPHIC DESIGN	**GRACE DORCAS & CO. STUDIOS**
ILLUSTRATION	**EMMANUEL G BOATENG** Studio X GH
CONTRIBUTING WRITERS	**DYLAN HAROLD ABAKAH** **MABLE BARBARA AMURON** **BENEDICTA NAA ODARKOR ABLATEYE** **HASSAN HABIBAT TEMITAYO** **RICHMOND OSEI** **JOY JUNIAS ONG'ANYA** **GBEMISOLA ESHO** **HANAA DANY** **CHARITY NCUBE** **ISMAIL ABDRASHEED** **AYI HERMAS** **CLOTHILDE 'ANNABELLE' FRANÇOIS** **CHINAZA FRANCES AGBA** **AMINA MUSA** **LEYTI K (PSEUDONYM)**

DIGISPACE AFRICA

FROM THE FELLOWSHIP MANAGER



When we welcomed the inaugural cohort of the Digispace Africa Fellowship earlier this year, we knew we were launching more than a four-month research programme. We were taking the first steps towards building a community of African researchers, practitioners and advocates committed to shaping the future of digital policy, governance and rights across the continent.

Looking back now, I am incredibly proud of what this cohort has accomplished. Over the course of the fellowship, fifteen fellows from across Africa came together with different backgrounds, perspectives and lived experiences. Some approached questions of artificial intelligence through education and public policy. Others examined digital surveillance, online safety, governance, internet shutdowns, data protection or platform accountability. While their research topics differed, they shared a common commitment: producing work that is grounded in African realities and capable of informing better policy and practice.

One of the greatest privileges of serving as Fellowship Manager has been witnessing those ideas evolve. A concept note gradually became a research question. Conversations with mentors sharpened arguments. Expert sessions challenged assumptions. Drafts were rewritten, refined and strengthened. By the end of the programme, what emerged was not simply a collection of research papers, but a body of work that reflects the depth of thinking and potential that exists across our continent.

This fellowship has also reaffirmed something I have come to believe deeply: meaningful programmes are built through people. Behind every session was a speaker who generously shared their expertise. Behind every research paper was a mentor who invested their time and thoughtful feedback. Behind every deadline, discussion and milestone was a growing community of fellows who encouraged one another, exchanged ideas and remained committed to seeing the journey through together.

As this was our inaugural fellowship, we also spent considerable time building the foundations that will support future cohorts. Throughout the programme, we established structured mentorship processes, strengthened our research review systems, expanded our volunteer network, and laid the groundwork for an alumni community that we hope will continue long after this cohort graduates.

These are quiet achievements that may not always be visible, but they are essential to creating a fellowship that is sustainable, collaborative and capable of growing year after year.

Personally, this journey has reminded me that some of the most important work happens behind the scenes. It happens in planning meetings, in reviewing drafts late into the evening, in responding to questions, coordinating mentors across countries, solving unexpected challenges, and making sure that every fellow has the support they need to do their best work. While much of that work remains unseen, seeing the quality of the research produced by this cohort has made every step worthwhile.

To our fellows, thank you for trusting us with your ideas and allowing us to be part of your research journey. Your curiosity, resilience and commitment have been the driving force behind this programme.

To our mentors, speakers, Advisory Council members, volunteers and partners, thank you for sharing your expertise so generously. Your contributions have shaped not only individual research projects but also the culture of learning and collaboration that has defined this fellowship.

Finally, I would like to thank our Executive Director, Theorose Elikplim Dzineku, for her vision in creating this fellowship and for her unwavering belief that African voices must play a leading role in shaping conversations around digital policy and governance.

While this report marks the conclusion of the inaugural Digispace Africa Fellowship, I believe it also marks the beginning of something much larger. The conversations started here should not end with the final submission of a paper or the close of a symposium. They should continue through collaboration, research, advocacy and the growing community that now forms the Digispace Africa Alumni Network.

It has been an honour to walk this journey with our first cohort, and I look forward to seeing where their work and the fellowship itself go next.



Fellowship Manager

DIGISPACE AFRICA

Who Are We

Digispace Africa is a youth- and women-led nonprofit organization dedicated to advancing digital inclusion, digital rights, and responsible technology across Africa.

We work to equip young people, researchers, journalists, policymakers, and civil society actors with the knowledge and skills needed to navigate an increasingly digital world. Through research, capacity-building programs, policy engagement, and community-driven initiatives, Digispace Africa promotes safe, inclusive, and rights-respecting digital ecosystems.

Digispace Research Fellowship

The Digispace Research Fellowship is one of our flagship initiatives designed to strengthen research capacity among emerging African scholars and practitioners.

The fellowship provides mentorship, practical research training, and opportunities for fellows to translate evidence into policy and practice. By fostering interdisciplinary collaboration and supporting locally grounded research, the program contributes to informed public discourse and evidence-based policymaking on digital technologies and governance. Through these efforts, Digispace Africa is building a growing network of researchers committed to shaping an inclusive and resilient digital future for Africa.

A Human Development Audit of Ghana's Artificial Intelligence Strategy.

What Is the Cost of Getting AI Wrong?



EXECUTIVE SUMMARY

Ghana's National Artificial Intelligence Strategy 2025–2035 sets an ambitious, human-cantered direction for AI. It links AI to inclusive growth, public service improvement, skills, ethical governance, and national data sovereignty and proposes a Responsible AI Authority, National AI Office, AI Fund, monitoring arrangements, and procurement safeguards. The central challenge is therefore not policy ambition but the gap between proposed architecture and enforceable implementation.

The reviewed documents do not yet demonstrate that the proposed oversight body has been established by law, that dedicated funding has been appropriated, or that mandatory impact assessments, a public register of high-impact systems, and AI-specific complaint and appeal procedures are operational.

UNESCO's Ghana readiness assessment similarly identifies fragmented public-sector governance and no AI-specific legislation, procurement framework, operational transparency, redress, or liability regime.[1]

[1] UNESCO, "Ghana," Global AI Ethics and Governance Observatory, readiness profile, especially "Legal and Regulatory" and "Governance" findings. UNESCO Ghana readiness profile

Action is urgent as Ghana moves into infrastructure investment, including a reported US$250 million national AI computing initiative. Rural and low-income communities, women and girls, persons with disabilities, informal workers, and local-language users face the greatest exclusion risks.

[1] UNESCO, "Ghana," Global AI Ethics and Governance Observatory, readiness profile, especially "Legal and Regulatory" and "Governance" findings. UNESCO Ghana readiness profile

Problem Statement

Ghana's challenge is no longer a lack of AI policy; it is converting policy design into enforceable practice. The National AI Strategy, launched in April 2026, proposes institutions, financing mechanisms, safeguards, and performance monitoring. However, many of these measures remain proposals rather than verified legal duties, funded institutions, or publicly operating systems.[1]

This distinction matters because early infrastructure contracts, data-sharing arrangements, model choices, and vendor dependencies can become costly to reverse. If safeguards are introduced only after deployment, citizens may face automated decisions without clear explanations, human review, or accessible remedies. Communities with weaker connectivity, lower digital literacy, limited local-language access or fewer public services may receive fewer benefits while carrying greater risks. Ghana, therefore, needs to turn the strategy's commitments into binding rules, assigned budgets, measurable outcomes, and visible accountability before large-scale implementation becomes entrenched.

[1] Ministry of Communication, Digital Technology and Innovations, National Artificial Intelligence Strategy 2025–2035 (December 2025), pp. 6–7, 34–35, 56–59 and 79–81; Ministry, "Ghana Launches National AI Strategy to Drive Digital Transformation and Economic Growth," 24 April 2026. Official launch statement

METHODOLOGY

This brief used a structured desk review of Ghana's National AI Strategy, Digital Economy Policy and Strategy, AI Practitioners' Guide, relevant legislation, and official implementation statements. UNESCO's Recommendation on the Ethics of AI and Readiness Assessment Methodology, the African Union Continental AI Strategy and Digital Transformation Strategy, and the Malabo Convention on cybersecurity and personal-data protection were used as benchmarks. Three lenses guided the review: normative alignment, institutional and operational readiness, and human-development and distributional impact. A gap-analysis matrix coded each major commitment against the responsible institution, legal authority, implementation mechanism, budget, timeline, indicator, monitoring requirement, and accountability or redress route. Each item was classified as present, partially present, unclear, or absent, based on explicit provisions rather than general policy language.[1]

[1] Documents reviewed included Ministry of Communication, Digital Technology and Innovations, Ghana Digital Economy Policy and Strategy 2024; UNESCO, Recommendation on the Ethics of Artificial Intelligence (2021), paras. 49–55 and 70; UNESCO, AI Readiness Assessment Methodology (2023); African Union, Continental Artificial Intelligence Strategy (2024), Digital Transformation Strategy for Africa 2020–2030, and African Union Convention on Cyber Security and Personal Data Protection (Malabo Convention). Ghana Digital Economy Policy and Strategy; UNESCO Recommendation; AU Continental AI Strategy

KEY INSIGHTS

1. Strategic alignment exceeds operational readiness

The Strategy commits Ghana to inclusive development and UNESCO-aligned ethics, proposes an independent Responsible AI Authority that the Data Protection Commission may incubate, and identifies a National AI Office, AI Fund, dashboards, risk audits, and data-governance measures (National AI Strategy, pp. 6–7, 34–35, 56–59, 79–81).

The gap is enforceability. These sections propose legislative enablement and budgets but do not create the Authority or confer statutory powers. No explicit appropriation under the names "Responsible AI Authority" or "National AI Fund" was identified in the published 2026 budget documents reviewed.

[1] Ministry of Communication, Digital Technology and Innovations, National Artificial Intelligence Strategy 2025–2035 (December 2025), pp. 6–7, 34–35, 56–59 and 79–81; Ministry, "Ghana Launches National AI Strategy to Drive Digital Transformation and Economic Growth," 24 April 2026. Official launch statement
[1] Documents reviewed included Ministry of Communication, Digital Technology and Innovations, Ghana Digital Economy Policy and Strategy 2024; UNESCO, Recommendation on the Ethics of Artificial Intelligence (2021), paras. 49–55 and 70; UNESCO, AI Readiness Assessment Methodology (2023); African Union, Continental Artificial Intelligence Strategy (2024), Digital Transformation Strategy for Africa 2020–2030, and African Union Convention on Cyber Security and Personal Data Protection (Malabo Convention). Ghana Digital Economy Policy and Strategy; UNESCO Recommendation; AU Continental AI Strategy



UNESCO's Ghana assessment likewise reports no dedicated AI oversight body, AI-specific law, procurement framework, or operational transparency and redress regime.[1]

2. Human-development commitments are not yet sufficiently measurable

The Strategy connects AI to improved lives, inclusion, skills, and public services and proposes annual plans, dashboards, and performance contracts (National AI Strategy, pp. 6, 20, 57, 79). The reviewed sections, however, do not establish one public scorecard linking AI investment to baselines and time-bound outcomes in health, education, livelihoods, participation, safety, dignity and access, consistently disaggregated by sex, geography, age, disability, and income.

Ghana already has a domestic model. The Digital Economy Policy and Strategy 2024 pairs selected baselines with targets, including indicators for connectivity, women in science and technology, and persons with disabilities (pp. 61–63 and implementation appendices). AI policy should measure whether people's opportunities improve, not only whether systems and training programmes are delivered.[1]

3. Data and infrastructure partnerships require binding safeguards

The Strategy proposes cross-border data rules, public-procurement clauses, tender audits, and transparency reports for public-sector AI (National AI Strategy, pp. 58–59). The non-binding AI Practitioners' Guide adds independent audits, vendor disclosure, interoperability, local support, and safeguards against lock-in (pp. 25, 50–51, 61).

These measures should become enforceable before major contracts are signed. Act 843 regulates personal-data processing, security, complaints, and certain solely automated decisions, but it is not a complete AI regime for model documentation, public-interest impact assessment, audit access, procurement transparency, liability, or contestability. Partnerships can accelerate national capability, provided contracts secure data protection, knowledge transfer, portability, and an affordable exit.[1]

4. Ghana can convert regional leadership into accountable implementation

The AU Continental AI Strategy calls for risk-based governance, impact assessment, independent review, consultation with affected communities, African-led research and national reporting on budgets and targets. Its 2025–2026 phase prioritises governance frameworks and institutional capacity (pp. 33, 36, 60–63).[1]

Ghana can lead through implementation: a public register of high-impact systems, accessible complaints and appeals, participation by underserved communities, and regular publication of outcomes and corrective actions. This would allow Ghana to shape regional standards through demonstrated practice, not policy declarations alone.

EVIDENCE AND CONTEXT

Two facts show why governance and inclusion must advance together. First, the Government has publicly described a US$250 million national AI computing centre intended to support national language models and wider AI capability. This scale of infrastructure can expand research and public service capacity, but it also makes procurement terms, data access, cybersecurity, ownership, and long-term operating costs consequential from the outset.

Second, UNESCO's Ghana readiness profile reports that about 75.8% of the population used the internet in 2024, while fixed-broadband subscriptions remained very low; it also notes that only about 15% of primary schools had internet available for teaching in 2023, compared with roughly 40% of senior high schools. Mobile subscriptions above 100 per 100 people reflect active SIM cards, not universal individual access or equal capability.[1] These disparities mean that national AI success cannot be inferred from infrastructure or subscription totals. It must be tested against who can access, understand, question, and benefit from AI-enabled services.

POLICY IMPLICATIONS

What happens if the governance gaps are ignored?

High-impact systems could be introduced into health, education, welfare, employment or public administration without consistent assessment of discrimination, privacy, safety or development effects. Citizens may struggle to identify which institution or vendor is responsible, obtain an explanation, request human review or appeal a harmful decision. Weak contract terms could also limit government audit access, create dependency on proprietary systems, expose sensitive public data or make switching providers unaffordable. These are foreseeable risks, not evidence that such harms have already occurred.

What opportunity exists?

Ghana can build safeguards while implementation choices remain flexible. The Strategy, UNESCO readiness findings and AU framework already supply the policy foundation. Converting them into legislation, procurement controls, indicators and public reporting would improve trust, reduce project failure, strengthen domestic expertise and connect AI expenditure to national development. Responsible governance is therefore an enabler of durable innovation, not a barrier to it.

[1] Ministry of Communication, Digital Technology and Innovations, National Artificial Intelligence Strategy 2025–2035, pp. 34–35, 57–59 and 79–81; Ministry of Finance, 2026 Budget Statement and Economic Policy and 2026 Budget Estimates of Ministries, Departments and Agencies. The budget-document finding is limited to the published materials reviewed and does not prove that no related expenditure exists under another programme name. 2026 Budget Statement

[1] Ministry of Communication, Digital Technology and Innovations, National Artificial Intelligence Strategy 2025–2035, pp. 6, 20, 57 and 79; Ghana Digital Economy Policy and Strategy 2024, pp. 61–63 and implementation appendices. Digital Economy Policy and Strategy

[1] Ghana, Data Protection Act, 2012 (Act 843), ss. 2–3, 17, 28–31, 35, 43 and 75–77; GIZ/Heritors Labs, Ghana Artificial Intelligence Practitioners' Guide (September 2025), pp. 24–27, 33–36, 50–51 and 61. Data Protection Act, 2012

[1] African Union, Continental Artificial Intelligence Strategy: Harnessing AI for Africa's Development and Prosperity (July 2024), pp. 33, 36 and 60–63. AU Continental AI Strategy

[1] UNESCO, "Ghana," readiness profile, "Education and Research" and "Infrastructure and Connectivity." Mobile-subscription figures count subscriptions/SIMs and should not be read as the percentage of unique users. UNESCO Ghana readiness profile

POLICY RECOMMENDATIONS

Recommendation	Actor (s)	Brief Rationale
Establish an interim Responsible AI and Human Development Secretariat and initiate legislation for an independent statutory commission.	Ministry of Communication, Digital Technology and Innovations	Within six months, a Cabinet-approved Secretariat should convene the DPC, NITA, CSA, NDPC, PPA, GSS, sector ministries, academia, civil society and affected communities; complete a legal gap assessment; pilot oversight tools; and prepare a Cabinet memorandum. The Attorney-General should draft a Responsible AI Bill for enactment within 12–18 months. The commission should have an independent board combining regulatory, legal, technical, rights and public-interest expertise; fixed terms; a dedicated budget; annual reporting to Parliament; complaint, investigation, and audit powers; and referral arrangements with existing regulators. This avoids duplicating specialist privacy, technology and cybersecurity mandates.[1] [1] Constitution of Ghana, art. 88; Office of the Attorney-General and Ministry of Justice, Legislative Drafting Division; Ghana, Data Protection Act, 2012 (Act 843); National Information Technology Agency Act, 2008 (Act 771); Cybersecurity Act, 2020 (Act 1038); and Public Financial Management Act, 2016 (Act 921). Legislative Drafting Division
Require risk-tiered AI impact assessment and public accountability for high-impact government systems	Public Procurement Authority	Within six months, the PPA should revise standard tender documents under Act 663. High-impact procurements should undergo pre-award and periodic post-deployment assessment covering equality, privacy, cybersecurity, safety, human rights and development outcomes. Contracts should identify responsible institutions and vendors, preserve human review and appeal, require independent audit access, and support a central public AI register, subject to lawful security exemptions. Procurement rules can start immediately; legislation should later extend proportionate duties beyond government purchasing.[1] [1] Ghana, Public Procurement Act, 2003 (Act 663), as amended, especially s. 50 on standard tender documents; UNESCO, Recommendation on the Ethics of Artificial Intelligence, paras. 50–55. Public Procurement Act, 2003; UNESCO Recommendation
Publish a National AI Human Development Scorecard.	National Development Planning Commission and Ghana Statistical Service	Under Acts 479, 480 and 1003, NDPC and GSS should establish, within 12 months, baselines and targets for health, education, livelihoods, access, participation, safety and trust, disaggregated by sex, region, rural–urban location, age, disability and income where data permit. Sector ministries and local government should report annually, with independent review by academia and civil society. Results should be public and used to adjust programmes and budgets.[1] [1] Ghana, National Development Planning Commission Act, 1994 (Act 479), s. 2; National Development Planning (System) Act, 1994 (Act 480); and Statistical Service Act, 2019(Act 1003), ss. 3 and 24–25. NDPC mandate
Issue binding data-sovereignty, transparency and vendor-contract safeguards.	Ministry of Communication, Digital Technology and Innovations, with the DPC, NITA, CSA and PPA	With GIFEC, GI-KACE, the National Council on Persons with Disability, local government and civil society, the ministries should pilot accessible, local-language and low-bandwidth programmes in underserved districts within 12 months. Content should cover safe use, misinformation, privacy, automated decisions, complaint routes and basic rights. Participation and outcomes should be disaggregated and linked to the national scorecard.[1] [1] UNESCO, Recommendation on the Ethics of Artificial Intelligence (2021), policy areas on education, literacy, inclusion and redress; Ministry of Communication, Digital Technology and Innovations, National Artificial Intelligence Strategy 2025–2035, inclusion and literacy commitments, pp. 6–7, 56–57 and 79. UNESCO Recommendation

[1] Ghana, Data Protection Act, 2012 (Act 843), ss. 28–31 and 35; NITA's statutory and regulatory mandate under Act 771; and PPA standard-document powers under Act 663. Data Protection Act, 2012

[1] UNESCO, Recommendation on the Ethics of Artificial Intelligence (2021), policy areas on education, literacy, inclusion and redress; Ministry of Communication, Digital Technology and Innovations, National Artificial Intelligence Strategy 2025–2035, inclusion and literacy commitments, pp. 6–7, 56–57 and 79. UNESCO Recommendation

Treat inclusive AI literacy as the ability to use, question and seek redress—not only technical training.	Ministry of Communication, Digital Technology and Innovations and the Ministry of Education	With GIFEC, GI-KACE, the National Council on Persons with Disability, local government and civil society, the ministries should pilot accessible, local-language and low-bandwidth programmes in underserved districts within 12 months. Content should cover safe use, misinformation, privacy, automated decisions, complaint routes and basic rights. Participation and outcomes should be disaggregated and linked to the national scorecard. [1]

[1] UNESCO, Recommendation on the Ethics of Artificial Intelligence (2021), policy areas on education, literacy, inclusion and redress; Ministry of Communication, Digital Technology and Innovations, National Artificial Intelligence Strategy 2025–2035, inclusion and literacy commitments, pp. 6–7, 56–57 and 79. UNESCO Recommendation

CROSS-CUTTING RELEVANCE

These reforms would strengthen Ghana's wider digital-governance ecosystem. A shared impact-assessment method can improve data protection, cybersecurity, digital public infrastructure and platform accountability. Model procurement clauses can raise value for money across technology projects. A human-development scorecard can connect skills and employment policy to gender, disability and regional inclusion. Clear institutional roles and interoperable standards would also improve Ghana's bargaining position in regional data and infrastructure cooperation, supporting African digital sovereignty without closing the country to responsible investment.

CONCLUSION

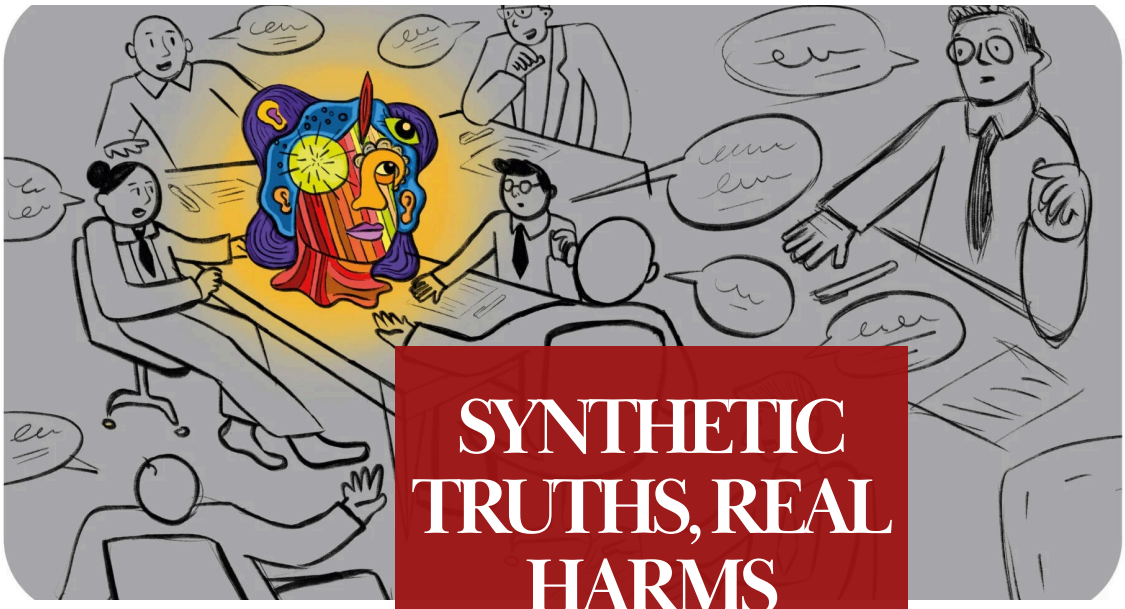
Ghana's AI ambition is a major national opportunity, and its Strategy provides a credible starting architecture. The next step is to make that architecture real through law, budgets, enforceable safeguards, measurable outcomes and accessible accountability. Governance should not slow innovation; it should make innovation trusted, inclusive and sustainable. The present implementation window will narrow as infrastructure, contracts and systems become fixed. Ghana should act now, before responsible design becomes expensive retrofitting.

AUTHOR BIO

Dylan Harold Abakah is an educational innovator, researcher and technology policy fellow serving as Senior Administrative Assistant at the KNUST E-Learning Centre. Trained in industrial engineering, he works across digital education, tech policy and inclusive institutional transformation. He is a TechForward Policy Fellow, Africa Centre for Digital Transformation, Ghana Tech Policy Fellow with the University of Ghana School of Law/GIZ Ghana and DigiSpace Africa Fellow.

[1] Ghana, Data Protection Act, 2012 (Act 843), ss. 28–31 and 35; NITA's statutory and regulatory mandate under Act 771; and PPA standard-document powers under Act 663. Data Protection Act, 2012

[1] UNESCO, Recommendation on the Ethics of Artificial Intelligence (2021), policy areas on education, literacy, inclusion and redress; Ministry of Communication, Digital Technology and Innovations, National Artificial Intelligence Strategy 2025–2035, inclusion and literacy commitments, pp. 6–7, 56–57 and 79. UNESCO Recommendation



SYNTHETIC TRUTHS, REAL HARMS

ARTIFICIAL INTELLIGENCE, INFORMATION LITERACY, AND THE GROWING MISINFORMATION CRISIS IN AFRICA

ABSTRACT

The convergence of rapid digital adoption and the proliferation of generative artificial intelligence (AI) tools has created conditions for accelerating misinformation crises across African digital publics. This paper examines how AI is reshaping African information ecosystems by introducing a critical analytical distinction between AI-generated content, synthetic material produced using generative systems, and AI-amplified content, the algorithmic acceleration of false narratives through recommendation systems and coordinated networks.



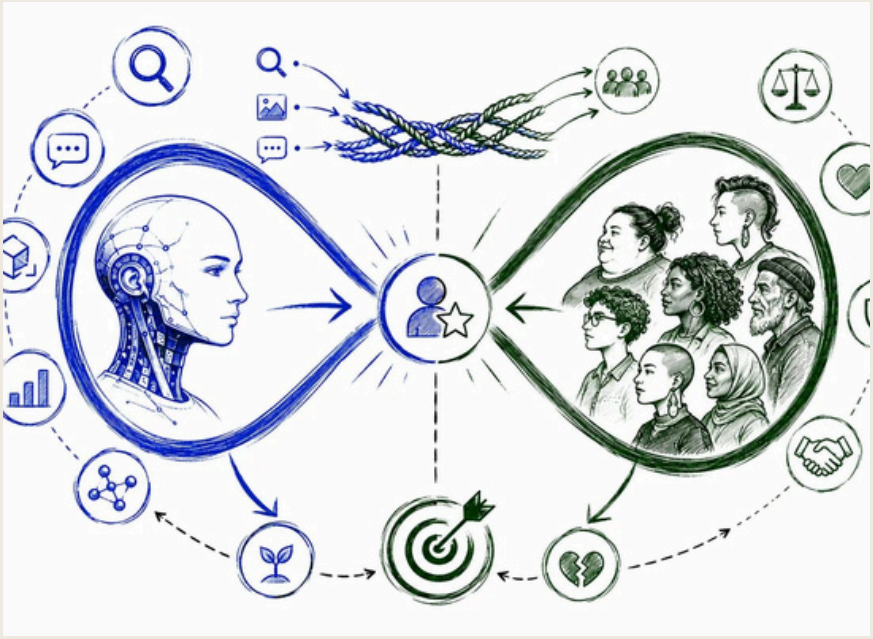
Drawing on a qualitative, multi-method approach combining a pan-African landscape review, three case studies (the 2023 Nigerian presidential election, the 2018 Gabon deepfake incident, and the 2024 Kenyan #RejectTheFinanceBill protests), and a comparative policy analysis, the paper argues that AI does not operate in isolation but intensifies pre-existing structural vulnerabilities. These include limited detection capacity in African languages, uneven digital literacy, and weak or fragmented regulatory frameworks. Case selection is itself methodological. Each case was chosen to interrogate the linguistic detection gap, the near-total absence of AI detection capability in African language environments, rather than to treat that gap as an incidental finding.

As a result, AI-enabled misinformation in African contexts is best understood as a structural governance problem requiring coordinated, multilingual, and multi-level responses and not as a purely technical challenge. With more than a dozen African countries scheduled to hold elections between 2026 and 2028, these findings carry urgent implications for regulators, continental bodies, civil society organisations, and platform operators.

Keywords: artificial intelligence, misinformation, disinformation, information literacy, Africa, deepfakes, generative AI, electoral integrity, digital governance

Imagine this, on the morning of an election day, a voice note begins circulating on WhatsApp. It sounds like a well-known political candidate, speaking in a familiar cadence, warning supporters that the vote has been compromised and urging them to stay home. Within hours, the message has been forwarded thousands of times across multiple groups, translated into local languages, and reposted on Facebook and X (formerly Twitter). By midday, it is being discussed on radio call-in shows. By evening, it is clear that the recording is not authentic, but by then, its effects are already difficult to measure and impossible to fully reverse. Such scenarios are no longer hypothetical. They are becoming increasingly plausible within a rapidly evolving digital landscape across Africa.

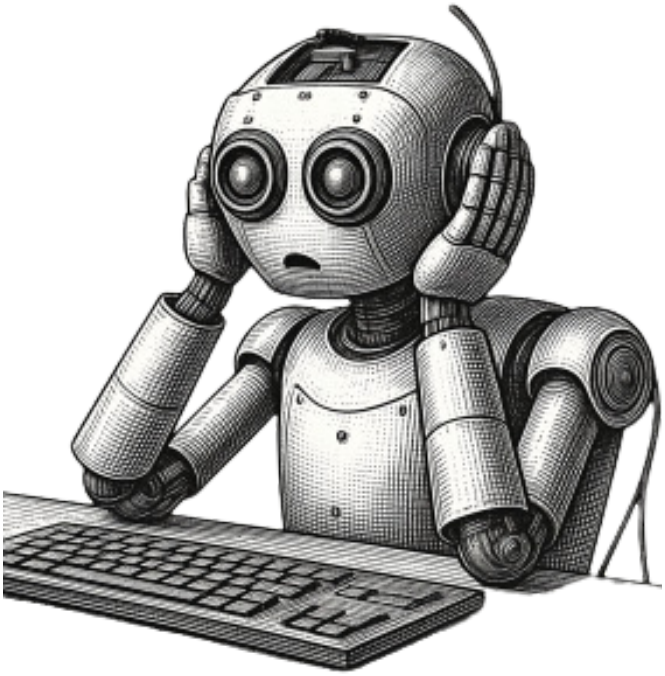
INTRODUCTION





Africa is in the midst of a dual transformation. The continent is experiencing some of the fastest rates of digital adoption in the world, with social media penetration rising sharply across West, East, and Southern Africa. Internet penetration now stands at roughly 48 per cent, and in many countries, social media use has grown by over 20 per cent in just a few years, connecting millions who, until recently, existed at the margins of the global information economy. With a population exceeding 1.3 billion and nearly 60% under the age of 25, the continent is often framed as the next frontier of digital innovation, where mobile connectivity is reshaping how information is produced, shared, and trusted.

The second transformation unfolding is no less significant. The technical barriers to creating convincing synthetic media have largely collapsed. Generative AI tools can now produce realistic text, audio, images, and video with minimal cost and expertise. What once required specialized knowledge can now be done on a smartphone in seconds.
Taken together, these two trends are beginning to interact in ways that have proven difficult to govern and even harder to detect. The result is a different kind of information environment, one in which false narratives can be generated, adapted, and amplified at speed, often outpacing the systems designed to identify and contain them. In this environment, exposure is uneven.



The implications for democratic governance are almost immediate. Between 2026 and 2028, more than a dozen African countries are scheduled to hold elections (Forus International, 2026). Evidence in Nigeria, Kenya, South Africa, Burkina Faso, and Gabon suggests that AI-generated and AI-amplified content is already being used to shape political narratives, erode trust in public institutions, and fragment public discourse (Okolo, 2025; Allen & Nehring, 2025).

Yet despite these developments, African perspectives remain marginal in global debates on AI governance. The regulatory and technical infrastructures needed to detect, counter, and respond to AI-generated disinformation remain underdeveloped, particularly in linguistic environments beyond English and French. Detection tools trained on widely spoken African languages such as Swahili, Amharic, or Yoruba remain scarce. This is creating a gap that malicious actors are well-positioned to exploit (Inglis, 2026).

This paper argues that the effects of artificial intelligence on misinformation in African contexts cannot be understood in purely technological terms. Rather, they emerge from the interaction between new AI capabilities and pre-existing structural vulnerabilities. While generative AI lowers the cost and increases the realism of synthetic content, its impact is shaped by conditions such as linguistic fragmentation, limited detection infrastructure, and uneven regulatory capacity. In this sense, AI-enabled misinformation does not introduce an entirely new problem. It intensifies existing ones, often in ways that current governance frameworks are ill-equipped to address.

Against this backdrop, the paper asks: how is the increasing use of artificial intelligence, both in generating synthetic content and in amplifying false narratives, reshaping information literacy and accelerating the spread of misinformation across African digital publics? To answer this, the analysis proceeds in five parts: a conceptual framework; a continent-wide landscape review; three case studies; a policy gap assessment; and a set of recommendations.

CONCEPTUAL FRAMEWORK

The rapid expansion of artificial intelligence within digital information ecosystems has introduced new complexities to the production, circulation, and credibility of information. Nowhere is this more evident than in the evolving landscape of misinformation, where synthetic content and algorithmic systems interact in ways that challenge existing frameworks for analysis. To make sense of these dynamics, this paper distinguishes between two analytically separable but practically entangled phenomena, AI-generated content and AI-amplified content.

AI-generated content refers to synthetic material produced using generative AI tools, including large language models, image generators, voice cloning systems, and video synthesis platforms. This category encompasses deepfake videos, cheap fakes, fabricated audio clips, synthetic text posing as journalism, and AI-generated images and videos that may be used to misrepresent events or individuals.

(AI) in Kenya's legal practice



(AI) in Kenya's legal practice

JOY

Executive Summary

Artificial Intelligence (AI) is embedded in Kenya's legal practice. Legal practitioners are using generative AI tools such as Claude, Perplexity and Cocounsel for legal research, document drafting, contract review, and case management. The efficiency gains are real, but so are the risks, and Kenya's regulatory framework has not kept pace.

The core problem is that unregulated AI use is actively threatening the integrity of legal practice.

AI tools fabricate citations, generate hallucinated precedents, and process sensitive client data with no guarantee of confidentiality or security. Kenya's AI Bill No. 4 of 2026, while adopting a risk-classification framework, fails to explicitly designate the legal sector as high-risk, leaving AI in legal practice effectively ungoverned. Public trust in the Judiciary is already fragile.



(AI) is in Kenya's legal practice

Every AI-generated error that reaches a courtroom, unchecked and unverified pleadings erode it further. The time to act is now, before the damage becomes irreversible.

Two recommendations are proposed. First, Parliament must enact an AI Act with explicit provisions governing AI in legal practice, anchored on the pillars of human oversight, accountability, transparency, and access to justice. The Judiciary must simultaneously develop an institutional self-governance framework prescribing how AI tools are used, disclosed, and monitored within court processes. Second, the Law Society of Kenya and the Council of Legal Education must establish mandatory AI literacy training for both practicing advocates and law students, equipping the profession with the competence to use AI responsibly.

Kenya's justice system cannot govern what it does not understand, and it cannot protect what it inadequately regulates.



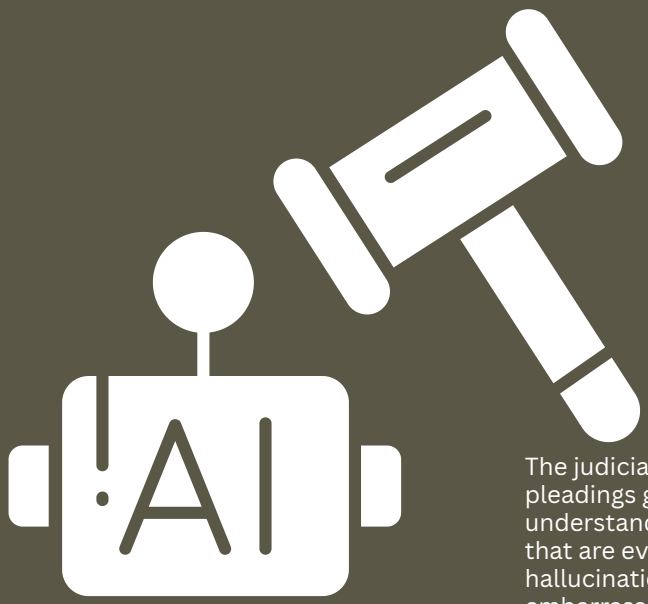
INTRODUCTION

The advent of Artificial Intelligence (AI) presents great opportunities for legal practitioners, ranging from enhanced legal research, automated document drafting, efficient case management, contract reviews, and prediction of legal outcomes. These advancements and opportunities created by use of generative AI tools such as Claude, Perplexity and Cocounsel AI in legal practice promise great efficiency that legal practitioners can barely resist.[1]

However, the adoption of such AI tools poses greater threats and challenges than benefits and opportunities. Such challenges include; Majority of AI solutions are not narrowed to the Kenyan market. While some can offer some reasoning and input, there is high risk of hallucination that can affect the expected output. This therefore creates biased outcomes. A greater challenge posed by AI is hallucination in the legal profession that we cannot ignore.[1]

[As such, the urgency to regulate AI in legal practice stems from legal sector pivotal role in the society to uphold justice and societal trust. Public trust in the Judiciary in Kenya is already fragile, and unregulated over-reliance in AI tools erodes it further.

[1] Novera Bhatti, 'The Regulation of Artificial Intelligence in the Legal Profession' (2026) 5 American Journal of Society and Law 1 <https://journals.e-palli.com/home/index.php/ajsl/article/view/7210> accessed 21 May 2026.



How can Kenya regulate use of AI legal practice in a way that balances protection of justice with innovation?

The judiciary already risks presiding over pleadings generated by tools it neither understands nor governs, some pleadings that are even generated as a result of AI hallucinations leading to professional embarrassment. Every stakeholder in the legal field, ranging from Law Society on Kenya, The Judiciary, Council of Legal Education should thus be very concerned with unregulated use of AI in legal practice. These stakeholders have the power to influence policy or push for a law that will regulate AI in legal practice.[1]

This paper adopts doctrinal research, supplemented by a practitioner survey capturing how Kenyan advocates currently use AI in research, drafting, and case management, and key informant interviews with stakeholders directly affected by the regulatory gap, including Advocates of the High Court of Kenya, officials from the Law Society of Kenya, Lawyers Hub, and the Office of the Data Protection Commissioner.

Interview subjects were purposively selected as practitioners and institutions whose professional and regulatory responsibilities are most exposed to the risks of unregulated AI use in legal practice. And as stakeholders who have capacity to make change in that policy gap.

Interview subjects were purposively selected as practitioners and institutions whose professional and regulatory responsibilities are most exposed to the risks of unregulated AI use in legal practice. And as stakeholders who have capacity to make change in that policy gap.

GEOGRAPHICAL FOCUS



The primary geographical scope of this research is Kenya. The research examines AI in legal practice across Kenya's legal system. The research will incorporate the European Union AI Act for comparison purposes. The comparison will analyze models adopted by European Union specifically to evaluate regulatory models that can inform the Kenyan context.

Research Question

How can Kenya regulate use of AI legal practice in a way that balances protection of justice with innovation?

Research Sub Questions

1. What gaps exists in Kenya's current legal and institutional framework in addressing AI use in legal practice?
2. Who bears the legal liability when an AI tool produces an erroneous legal output?

3. What lessons can Kenya draw from jurisdictions with AI governance Frameworks regulating legal practice?

[1] Yatama Zahra, 'Regulating AI in Legal Practice: Challenges and Opportunities' (2025) 3 Journal of Computer Science Application and Engineering (JOSAPEN) 10.
[1] CIPIT, 'AI as the SIDE KICK and NOT SUPERHERO: Navigating AI in the Legal Field in Kenya - CIPIT' (CIPIT5 February 2026) <https://cipit.strathmore.edu/ai-as-the-side-kick-and-not-superhero-navigating-ai-in-the-legal-field-in-kenya/#sdfootnote34sym> accessed 21 May 2026

Problem Statement

The integration of AI into legal practice is actively undermining the core principles of the profession, principles like professional responsibility, accuracy, and confidentiality thus raising urgent legal and ethical concerns that demand immediate regulatory attention. Integration of AI in legal practice raises privacy concerns, data protection issues, bias, accuracy and accountability concerns and trust and transparency issues.[1]

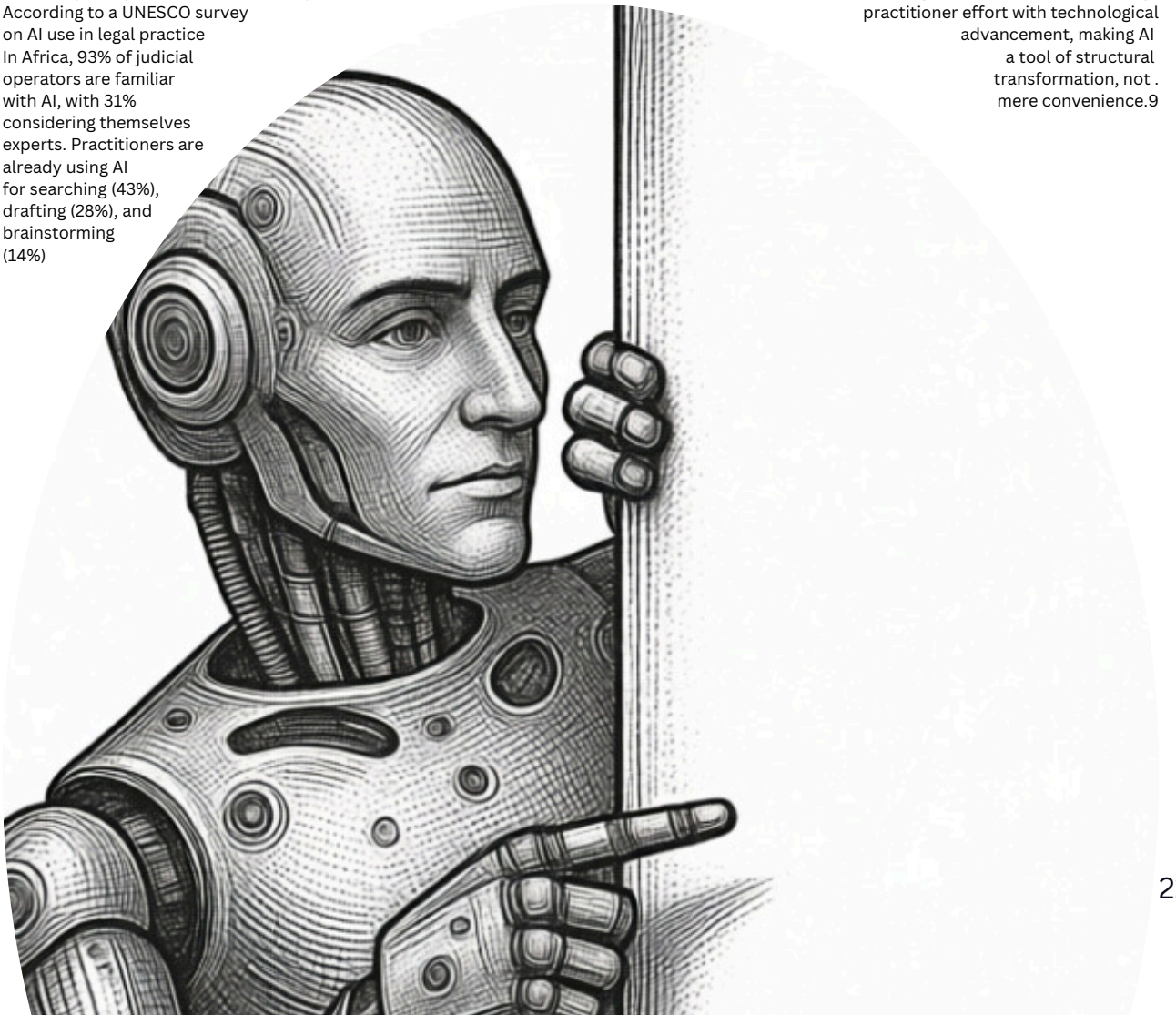
On data protection and privacy, AI tools access sensitive legal documents and client data to generate outputs, yet no assurance exists regarding where that information is stored, who accesses it, or how it is protected. Legal practitioners owe clients a duty of confidentiality, and this duty is directly threatened when AI systems process personal, identifiable information without adequate safeguards.[5]

On accuracy and accountability, AI tools produce errors and fabricate non-existent citations, a phenomenon known as AI hallucination, which has professionally damaging consequences. In legal matters, accuracy is very critical as a wrong precedent may lead erroneous advice to client which has dire consequences.[6] The case of Sullivan & Cromwell LLP illustrates what unregulated AI use inflicts on legal practice globally, Kenya included.[7]

The stakes are high for everyone involved. Clients face exposure of sensitive personal data, with no guarantee of how AI systems store or share that information. Advocates and law firms risk liability for AI-generated errors. The Judiciary, the Law Society of Kenya, and the public trust in the justice system are all on the line. Kenya cannot afford to wait. Without clear regulation now, unregulated AI use will continue eroding the integrity of legal practice.

AI is no longer a distant possibility in legal practice; it is already here. According to a UNESCO survey on AI use in legal practice In Africa, 93% of judicial operators are familiar with AI, with 31% considering themselves experts. Practitioners are already using AI for searching (43%), drafting (28%), and brainstorming (14%)

As Richard Susskind has noted, the future of legal practice lies in combining , practitioner effort with technological advancement, making AI a tool of structural transformation, not . mere convenience.9



Yet adoption is uneven. Survey findings among Kenyan practitioners reveal that use is shaped by structural barriers including digital literacy levels and the cost of premium AI tools. Younger practitioners, predominantly those under 40 years old, are the most enthusiastic adopters, citing efficiency gains in research, drafting, and document review.

This unevenness signals a widening gap: AI can enhance access to justice, but without deliberate intervention, it risks deepening exclusion.[1]

More critically, Kenya's regulatory framework has not kept pace. There is currently no legal framework governing how AI tools should be designed, monitored, or deployed in the legal sector. As one interviewee observed, Kenya's legal framework is reactive, addressing problems only after they materialise, rather than anticipating them.

This stands in sharp contrast to the European Union, which spent three years debating the EU AI Act to ensure it was sufficiently comprehensive.

The EU AI Act classifies AI systems into four risk tiers; unacceptable, high, limited, and minimal, and notably places access to justice under the high-risk category, subjecting it to strict pre-market obligations.

Kenya's AI Bill No. 4 of 2026[2] borrows heavily from the EU AI Act, adopting the same four-tier risk framework. However, it fails to clearly classify sectors into those tiers and does not explicitly designate access to justice or the legal sector as high risk. This omission creates regulatory uncertainty and, in effect, leaves AI in legal practice unregulated. Kenya must close this gap now by explicitly classifying access to justice under the high-risk tier and establishing clear obligations for AI use in the legal sector.13

Policy Implications

The integration of AI into legal practice presents both transformative opportunities and serious risks that Kenyan law makers and policy makers afford to ignore. AI has the potential to enhance legal services, but its use must be supported by legal and institutional framework that ensures access to justice and ensure protection of human rights. The absence of clear guidelines or policies on use of AI in legal practice poses serious risk to the legal profession. The risks include; data privacy breach, data inequality, AI hallucinations, bias in training data and lack of legal accuracy.[1]

Taken together, these five risks make clear that the integration of AI into legal practice is not a peripheral issue, but one that sits at the very heart of Kenyan justice system. It is therefore imperative that this issue is carefully addressed by all stakeholders involved and lawmakers. At the civic level, public trust in the judiciary is already fragile in many African countries, and the unchecked overreliance on AI threatens to erode it further. The implication of the issue is that if it is not addressed, even the little trust in the justice system will be eroded further. As Lord Hewart clearly stated, ‘Justice must not only be done, but must also be seen done.’ In this instance, Kenyans must see the stakeholders enact laws and policies that governs its justice system.

Policy Options

Kenya could adopt a risk-based regulation as it has on its AI Bill 2026, where they categorize AI system based on the risk it poses and thus subject it to market regulations. However, uncertainty in classification and definitional uncertainty of different classifications makes a policy option that will not effectively regulate AI in legal practice. This is because we will adopt a legal framework that can hardly be implemented; thus, Kenya will have a law in paper and not in practice.[1]

Kenya can explore the Co-Regulatory model, where it adopts a statutory and professional self-governance framework. This model will regulate public interest adequately under the statutory framework and keenly take care of the nitty-gritties of the institutional structure of the legal professional affected by AI in legal practice. This way, Kenya will adopt a regulation that protects rights and innovations at the same time. The statutory authority plus professional body expertise on this policy option must ensure:[2]

1. Human Oversight Pillar – where there must be mandatory requirement for actual professional oversight for any AI output in legal practice.
2. Accountability Pillar – which stipulates complete system of accountability for AI automated output between developer, lawyer and firm. This can include professional indemnity.
3. Transparency Pillar – this includes transparency to AI to oversight bodies, transparency to regulatory bodies, general reporting, periodic system auditing.
4. Access to Justice Pillar – ensure non-discrimination of AI use in legal practice, fair access to AI.[1]

This policy option is legally and financially feasible. It anticipates future issues that may arise and it comprehensively addresses the current issues experienced by AI use in legal practice. It is financially feasible as the Judiciary can undertake professional self-governance bit, by using or requesting an additional budget, while the statutory framework be handled by Parliament. The Judiciary can work with its present team, and additional policy makers to professionally regulate AI in legal practice.

[1] Ian Joshua Ombati , ‘THE BRUSSELS ECHO in NAIROBI: Kenya’s AI Bill 2026 and the EU’s Long Regulatory Shadow ’ 8 <https://papers.ssrn.com/sol3/papers.cfm?abstract_id=6717158> accessed 25 May 2026.
[2] Novera Bhatti, ‘The Regulation of Artificial Intelligence in the Legal Profession’ (2026) 5 American Journal of Society and Law 1 <https://journals.e-palli.com/home/index.php/ajsl/article/view/7210> accessed 21 May 2026.



RECCOMENDATIONS







1. Mandatory AI Literacy Training for Advocates and Law Students. The Law Society of Kenya and the Council of Legal Education should designate AI literacy as a mandatory CPD category under Section 81 of the Advocates Act. The CPD Committee should set a minimum number of AI literacy units per practice year, covering AI tool verification, ethical use, and disclosure obligations, with non-compliance triggering the practicing-certificate consequences. The Council of Legal Education should separately exercise its curriculum-approval mandate to embed a compulsory AI literacy module in the Kenya School of Law's Advocates Training Programme. Both bodies should issue a joint practice note within twelve months specifying the certification standard, examinable content, and accredited training providers. This requires no new legislative or institutional authority, only an amendment to the existing CPD unit schedule and ATP curriculum, both within powers the Council already holds.

2. Enact a Co-Regulatory framework for AI in legal practice. Parliament should amend the Artificial Intelligence Bill, (No. 4 of 2026), before passage, to add a Schedule on AI in the administration of justice, exercising its legislative authority under Article 94 of the Constitution of Kenya 2010. That Schedule should mandate human sign-off before AI-assisted judicial or advisory output takes legal effect, disclosure wherever AI materially informs a pleading, opinion, or ruling, cost safeguards for self-represented litigants, and liability rules for AI-assisted errors folded into the Bill's existing risk-classification categories. The Law Society of Kenya and the Judiciary should jointly submit these clauses to the Senate committee handling the Bill within three months. Concurrently, the Chief Justice and Judicial Service Commission should issue Practice Directions on the Use of Artificial Intelligence in Judicial Processes under their existing mandates. These Directions should require verification of AI-generated case citations before any ruling relies on them, prohibit unverified AI-generated authority, and require judicial officers to disclose AI use in draft rulings. They should be issued within six months, ahead of the Bill's enactment. Once both instruments are in force, the Act and the Practice Directions should expressly cross-reference each other so statutory regulation and judicial self-governance function as one coherent framework rather than parallel regimes.

Conclusion

Kenya stands at a decisive juncture. AI is already embedded in legal practice, quietly reshaping how documents are drafted and reviewed, and how research is conducted, yet the regulatory framework to govern this technological advancement shift remains silent. This is not a hypothetical risk to be managed at leisure; it is a present reality with immediate consequences for client confidentiality, professional accountability and the integrity of judicial outcomes. The AI Bill No. 4 of 2026 gestures toward regulation but stops short of the one step that matters most for this sector: explicitly classifying access to justice as high-risk and subjecting legal AI use to the safeguards that classification demands. A co-regulatory framework, anchored on human oversight, accountability, transparency, and access to justice, offers Kenya a workable path forward, one that protects the public without stifling the efficiency gains AI genuinely offers the profession. But a framework on paper changes nothing if it is not built. Parliament must legislate, the Judiciary must address integration of AI in legal practice, and the Law Society of Kenya and Council of Legal Education must equip practitioners with the literacy to use these tools responsibly.

NIGERIA

Rethinking Account Security in Nigeria:

BEHAVIORAL VULNERABILITIES, SHARED-DEVICE PRACTICES, AND PLATFORM TRUST & SAFETY DESIGN.



EXECUTIVE SUMMARY

As digital participation continues to expand across Nigeria, account security has become an increasingly important component of trust and safety. While technology platforms have introduced various security mechanisms to protect users, account compromise remains a common challenge across social media, messaging, and digital service platforms.

Much of the existing discourse on account security focuses on individual responsibility, emphasizing awareness campaigns, strong passwords, and cybersecurity education.

However, this approach often overlooks the broader interaction between user behavior and platform design. Practices such as password reuse, shared-device access, and low adoption of multi-factor authentication (MFA) are often treated as user failures rather than as indicators of limitations in trust and safety systems.

This policy brief argues that recurring account security vulnerabilities should be understood not only as behavioral challenges but also as governance and design issues. By examining common account security practices among Nigerian users and assessing how platform trust and safety mechanisms respond to these realities, the paper identifies gaps in current approaches to user protection.

Drawing on a desk-based analysis of academic literature, policy documents, and platform safety practices, the paper recommends stronger localization of platform safety systems, behavior-informed digital literacy initiatives, and greater collaboration between technology platforms, regulators, and civil society organizations.



1. Introduction

Nigeria has experienced rapid digital growth over the past decade. Increased internet penetration, mobile connectivity, social media adoption, and digital financial services have transformed how individuals communicate, work, learn, and conduct business. According to the Nigerian Communications Commission (NCC, 2024), active mobile internet subscriptions have steadily climbed, embedding digital systems deeply into the nation's economic and social fabric.

Similar trends are reflected across Sub-Saharan Africa, where expanding smartphone adoption continues to increase participation in digital services while simultaneously exposing more users to cyber-enabled fraud and identity attacks (GSMA, 2023).

The International Telecommunication Union (2022) notes that digital inclusion across Africa is advancing rapidly. However, disparities in infrastructure, affordability, and digital skills continue to influence how securely users engage with online services.

Despite investments in cybersecurity awareness and platform safety initiatives, account compromise remains widespread. This raises important questions about whether current trust and safety approaches adequately reflect the realities of how users engage with digital systems.

However, greater digital participation has also increased exposure to cybersecurity risks. Among the most common of these risks is account compromise, where unauthorized individuals gain access to digital accounts through stolen credentials, weak authentication practices, or other security vulnerabilities. In emerging digital economies like Nigeria, cybercriminals leverage sophisticated social engineering and credential stuffing tactics, disproportionately impacting users who are navigating digital spaces for the first time.

Traditional security models assume an optimized environment: a single user, an individual device, consistent network availability, and a high baseline of technical fluency (Furnell, 2021).
This paper examines how common user behaviors contribute to account security vulnerabilities and explores the extent to which platform trust and safety systems account for these realities within the Nigerian context.

Methodology

This policy brief adopts a qualitative, desk-based analytical approach. It draws on existing academic literature, policy reports, and publicly available publications from technology platforms and regulatory institutions to examine behavioral patterns affecting account security in Nigeria.

The analysis also considers the trust and safety features of widely used digital platforms, particularly in relation to authentication, account recovery, and user protection mechanisms. Illustrative scenarios presented throughout the paper are case-based examples that reflect commonly documented patterns of digital behavior in Nigeria rather than primary empirical research.

2. The Account Security Challenge in Nigeria

Account security is a foundational element of digital trust. When users lose control of their accounts, they may experience financial losses, privacy violations, reputational harm, or exposure to fraud. In Nigeria, the consequences are amplified due to the rapid integration of instant messaging platforms like WhatsApp and Facebook with peer-to-peer financial transactions and informal business activities.



Several behavioral practices contribute to account security risks within Nigeria's digital environment. These include:

- Reusing the same password across multiple accounts.
- Sharing devices among family members, friends, or marketplace peers.
- Storing passwords in insecure ways, such as unencrypted digital notes or physical paper.
- Failing to activate multi-factor authentication.
- Limited understanding of account recovery procedures when access is lost.



3. Behavioral Vulnerabilities and Platform Safety Design

Password Reuse

While these behaviors are not unique to Nigeria, they are uniquely shaped by socio-cultural and economic factors. For example, a single internet-enabled smartphone might serve an entire household or a micro-business cooperative due to affordability concerns regarding hardware and mobile data. Furthermore, differing levels of literacy mean that standard English-centric security warnings fail to register as urgent to broad segments of the population. Understanding these distinct realities is essential for developing trust and safety systems that are effective within local contexts.

Password reuse remains one of the most significant contributors to account compromise. When users rely on a single password across multiple services, a breach on one minor platform may expose critical accounts on financial or primary communication platforms.

Most platforms encourage strong password practices, but many continue to rely heavily on user compliance, expecting individuals to memorize dozens of complex, distinct alphanumeric strings, rather than designing systems that actively discourage risky behavior (Beautement et al., 2016).

Google's Safety Center (Google, 2024) promotes password managers, security checkups, and breach detection tools as effective mechanisms for improving account security. However, awareness and adoption of these features remain uneven, particularly among users with limited digital literacy or those accessing digital services primarily through mobile devices. This suggests that awareness campaigns alone are insufficient without corresponding improvements in usability and accessibility.

For many Nigerian users, the cognitive load of managing multiple passwords leads to predictable simplifications, such as using phone numbers, names, or common cultural phrases. The persistence of password reuse suggests that awareness alone is insufficient and that stronger platform-level interventions, such as password-less authentication or integrated credential monitoring, are necessary.



Among the behavioural vulnerabilities examined in this paper, shared-device access represents one of the least discussed, yet most significant trust and safety challenges within the Nigerian context.
Across many households, small businesses, educational settings, and informal commercial environments, smartphones and computers are routinely shared among multiple individuals.

A single device may be used throughout the day by parents, children, siblings, employees, or business partners. Economic considerations, device affordability, and communal social norms make shared access a practical reality rather than an exceptional circumstance.

For example, a trader operating within a local market may use a smartphone to receive customer payments, manage WhatsApp Business communications, access online banking services, and coordinate deliveries.

The same device may later be used by a spouse, child, or employee for unrelated activities. Similar patterns occur within households where one smartphone supports educational activities, entertainment, financial transactions, and family communication simultaneously.
Despite the prevalence of these practices, many trust and safety systems continue to assume that every account is accessed from a single privately owned device. Authentication models often interpret changes in typing behaviour, biometric data, device usage, or login patterns as indicators of unauthorized access. While these mechanisms can effectively detect malicious activity, they may also generate false security alerts or account lockouts for legitimate users operating within shared-device environments.

Kirlappos and Sasse (2012) argue that security interventions frequently fail when they overlook the realities of everyday user behaviour.
Meta's Transparency Report (2024) demonstrates the scale at which coordinated account abuse, compromised accounts, and malicious actors are detected globally, reinforcing the importance of designing account protection mechanisms that are accessible across diverse user contexts.

In contexts where device sharing is economically unavoidable, security mechanisms designed around exclusive device ownership may unintentionally reduce usability while offering only limited improvements in protection.

This mismatch between platform assumptions and user realities represents one of the most important trust and safety challenges facing digital platforms operating within emerging digital economies. Designing for shared-device environments should therefore be viewed not as a regional customization but as an essential component of inclusive and user-centred security design.

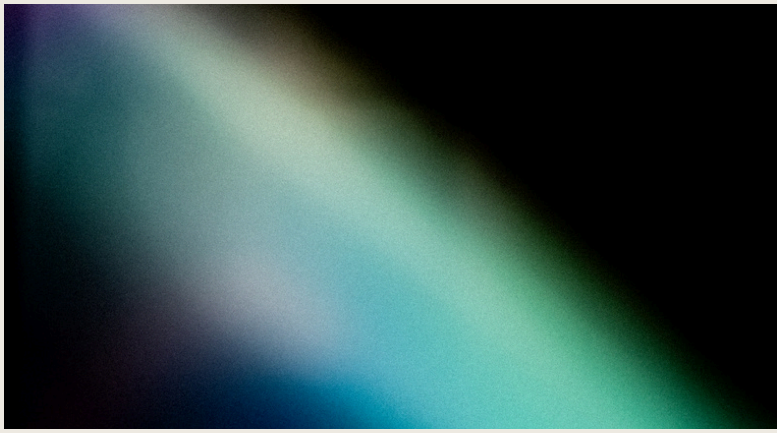
Shared-Device Practices

Among the behavioural vulnerabilities examined in this paper, shared-device access represents one of the least discussed, yet most significant trust and safety challenges within the Nigerian context.

Across many households, small businesses, educational settings, and informal commercial environments, smartphones and computers are routinely shared among multiple individuals.

Meta

Transparency



Amina Musa

Across many households, small businesses, educational settings, and informal commercial environments, smartphones and computers are routinely shared among multiple individuals. A single device may be used throughout the day by parents, children, siblings, employees, or business partners. Economic considerations, device affordability, and communal social norms make shared access a practical reality rather than an exceptional circumstance.

Across many households, small businesses, educational settings, and informal commercial environments, smartphones and computers are routinely shared among multiple individuals.
A single device may be used throughout the day by parents, children, siblings, employees, or business partners.

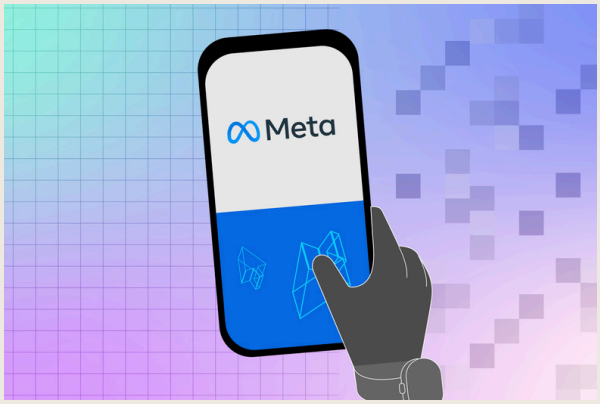
SHARED DEVICE PRACTICE

Economic considerations, device affordability, and communal social norms make shared access a practical reality rather than an exceptional circumstance.
For example, a trader operating within a local market may use a smartphone to receive customer payments, manage WhatsApp Business communications, access online banking services, and coordinate deliveries. The same device may later be used by a spouse, child, or employee for unrelated activities. Similar patterns occur within households where one smartphone supports educational activities, entertainment, financial transactions, and family communication simultaneously.

Despite the prevalence of these practices, many trust and safety systems continue to assume that every account is accessed from a single privately owned device.

Authentication models often interpret changes in typing behavior, biometric data, device usage, or login patterns as indicators of unauthorized access. While these mechanisms can effectively detect malicious activity, they may also generate false security alerts or account lockouts for legitimate users operating within shared-device environments.
Kirlappos and Sasse (2012) argue that security interventions frequently fail when they overlook the realities of everyday user behaviour.

Meta's Transparency Report (2024) demonstrates the scale at which coordinated account abuse, compromised accounts, and malicious actors are detected globally, reinforcing the importance of designing account protection mechanisms that are accessible across diverse user contexts.



In contexts where device sharing is economically unavoidable, security mechanisms designed around exclusive device ownership may unintentionally reduce usability while offering only limited improvements in protection.
This mismatch between platform assumptions and user realities represents one of the most important trust and safety challenges facing digital platforms operating within emerging digital economies. Designing for shared-device environments should therefore be viewed not as a regional customization but as an essential component of inclusive and user-centred security design.

Low Adoption of Multi-Factor Authentication

Multi-factor authentication (MFA) is widely recognized as one of the most effective methods for preventing unauthorized account access. However, adoption rates remain relatively low among many user groups in Nigeria.

Barrier to MFA Adoption	Impact on Nigerian Users
Lack of Structural Awareness	Users view MFA as a setting for corporate accounts only.
Perceived Inconvenience	Multi-step verification drains battery and takes time.
SMS Delivery and Network Issues	Unreliable telecom networks delay or drop OTP messages.
Device Accessibility & Costs	Authenticator apps require storage space and steady data

As highlighted above, reliance on Short Message Service (SMS)-based One-Time Passwords (OTPs) is highly problematic in environments with unstable telecommunications infrastructure (Onaolapo et al., 2021). When an OTP takes several minutes to arrive or fails due to network congestion, users frequently disable the feature out of frustration. This highlights the importance of designing security features that are both structurally resilient and user-friendly within local technical constraints.

Trust and Safety Implications

In contexts where device sharing is economically unavoidable, security mechanisms designed around exclusive device ownership may unintentionally reduce usability while offering only limited improvements in protection.
This mismatch between platform assumptions and user realities represents one of the most important trust and safety challenges facing digital platforms operating within emerging digital economies. Designing for shared-device environments should therefore be viewed not as a regional customization but as an essential component of inclusive and user-centred security design.

Low Adoption of Multi-Factor Authentication

Design Assumption ≠ User Reality → Security Vulnerability

As highlighted above, reliance on Short Message Service (SMS)-based One-Time Passwords (OTPs) is highly problematic in environments with unstable telecommunications infrastructure (Onaolapo et al., 2021). When an OTP takes several minutes to arrive or fails due to network congestion, users frequently disable the feature out of frustration. This highlights the importance of designing security features that are both structurally resilient and user-friendly within local technical constraints.

Design Assumption	User Reality	Security Vulnerability
Each account is accessed from a single personal device.	Smartphones are frequently shared among family members, friends, or small business operators.	Shared access increases opportunities for unauthorized account use and weakens device-based trust signals.
Users can easily remember and manage multiple strong passwords.	Concerns about network reliability, device storage, battery life, and usability discourage MFA adoption.	Credential stuffing attacks allow compromise across multiple accounts after one breach.
Security alerts written in English are universally understood.	Users have varying literacy levels and may prefer local languages or visual communication.	Important warnings may be ignored or misunderstood, reducing the effectiveness of security interventions.
Stable mobile networks support authentication workflows.	Delayed or failed SMS OTP delivery is common in some locations.	Users disable MFA or abandon secure login processes because of repeated authentication failures.

Policy and Governance Gaps

Gap 1: Overreliance on Awareness-Based Approaches

Many interventions by both platforms and state actors focus primarily on educating users rather than addressing the design conditions that contribute to risky behavior. Telling a user to "never share their phone.

Gap 2: Limited Localization of Safety Systems

Global trust and safety frameworks do not always reflect local patterns of device usage, access conditions, or digital literacy levels. Security notifications, fraud alerts, and account recovery workflows are rarely translated into major Nigerian languages like Hausa, Yoruba, or Igbo, nor are they optimized for low-bandwidth environments.

Gap 3: Weak Integration Between Policy and Platform Design

Digital governance discussions among regulators often occur separately from the actual engineering and platform safety design processes of tech companies. This siloed approach limits opportunities for coordinated solutions, such as zero-rating data for security verification workflows or establishing localized, rapid-response account recovery channels.

6. Policy Recommendations

For Technology Platforms

1. Design for Shared Environments: Implement "guest mode" features or profile switching inside individual applications to secure sensitive data on shared devices without requiring complete account logouts.
2. Simplify and Diversify MFA: Move away from pure SMS-based OTPs. Integrate low-data, cryptographic, or biometric alternatives (such as WhatsApp-delivered codes or device-native passkeys) that bypass unreliable telecom channels.
3. Contextualized Prompting: Introduce behavior-informed security prompts and warnings tailored to local threat vectors. For example, use clear, visual iconographies and localized audio prompts for users with lower textual literacy.
4. Localized Support Pathways: Establish dedicated, local language customer support and account recovery pipelines in Nigeria to handle complex account compromise cases swiftly.

For Government and Regulators

1. National Digital Literacy Integration: Shift the National Information Technology Development Agency (NITDA) literacy programs away from basic computer operation toward proactive digital hygiene and risk management (NITDA, 2023).
2. Public-Private Safety Sandbox: Encourage collaboration between the Nigerian Communications Commission (NCC) and major tech firms to test localized security features that do not penalize users for network or infrastructure deficits.
3. Fund Localized Cybersecurity Research: Support local academic institutions and think tanks in generating empirical data regarding user behavior, identity management, and digital safety within Sub-Saharan African contexts.

4. Strengthening national digital resilience also aligns with recommendations contained in the International Telecommunication Union's Global Cybersecurity Index (ITU, 2023), which emphasizes coordinated national cybersecurity capacity-building and public awareness.

For Civil Society Organizations

1. Community-Based Digital Safety: Expand grassroots digital safety education initiatives, focusing on informal market networks, trading cooperatives, and community groups.
2. Practical Framework Distribution: Develop accessible, non-technical resources (e.g., short pidgin-language explanatory videos or infographic pamphlets) focused on account protection and authentication practices.

For Educational Institutions

Curriculum Upgrades: Integrate interactive digital safety and account security education into primary, secondary, and tertiary technology curricula, moving beyond theoretical computer science to practical digital citizenship.

For Technology Platforms

Schools, colleges, and universities should integrate practical instruction covering:

1. Password creation and management.
2. Multi-factor authentication.
3. Online privacy.
4. Digital identity protection.
5. Safe social media practices.
6. Responsible use of artificial intelligence.
7. Recognition of online scams and phishing attacks.

Embedding these skills within formal education will help prepare future generations for increasingly digital societies.

Conclusion

Account security remains a critical component of trust and safety within Nigeria's expanding digital ecosystem. While user behavior plays an important role in shaping digital risk, recurring vulnerabilities such as password reuse, shared-device access, and low adoption of multi-factor authentication should not be viewed solely as individual failures.

These behaviors reveal clear limitations in how global trust and safety systems are designed and implemented. Effective user protection requires approaches that move beyond generic awareness campaigns and adapt directly to the infrastructural and social realities of how people access digital services.

More broadly, improving account security requires shifting trust and safety from a purely technical exercise toward a context-aware governance challenge. Security mechanisms that acknowledge shared-device use, infrastructural limitations, and localized patterns of digital participation are more likely to achieve sustained user trust than one-size-fits-all approaches

More broadly, improving account security requires shifting trust and safety from a purely technical exercise toward a context-aware governance challenge. Security mechanisms that acknowledge shared-device use, infrastructural limitations, and localized patterns of digital participation are more likely to achieve sustained user trust than one-size-fits-all approaches. By adopting behavior-informed and context-sensitive approaches to account security, technology platforms, policymakers, and civil society organizations can collaborate to build a safer, more resilient, and truly inclusive digital ecosystem in Nigeria and across the wider African continent.

References

- Beautement, A., Sasse, M. A., & Wonham, M. (2016). The compliance budget: Managing security fatigue in the modern enterprise. IEEE Security & Privacy, 14(4), 34-41.
- Furnell, S. (2021). Cybersecurity culture: Countering the human vulnerability. Computers & Security, 105, 102-115.
- Kirlappos, I., & Sasse, M. A. (2012). Security education, awareness and training: Beyond "blaming the user". In Proceedings of the 2012 European Workshop on System Security (pp. 15-22).
- National Information Technology Development Agency (NITDA). (2023). National digital literacy framework: Building a digitally competent Nigeria. Federal Ministry of Communications, Innovation and Digital Economy.
- Nigerian Communications Commission (NCC). (2024). Industry data: Subscriber statistics and internet penetration trends. NCC Sector Report.
- Onaolapo, J., Sasse, M. A., & Stringhini, G. (2021). When SMS OTP fails: Analyzing authentication hurdles in developing digital infrastructures. ACM Transactions on Computer-Human Interaction, 28(3), 1-24.
- Ryan, J., & Falco, G. (2022). The systemic flaw of user-centric security: Why platform safety must be structural. Harvard Digital Governance Review, 4(2), 88-97.
- European Union Agency for Cybersecurity. (2020). Cybersecurity and human behaviour: Understanding users. ENISA.
- Google. (2024). Google Safety Center. https://safety.google/
- GSMA. (2023). The mobile economy in sub-Saharan Africa 2023. GSMA Intelligence.
- International Telecommunication Union. (2022). Digital trends in Africa 2022. ITU Publications.
- International Telecommunication Union. (2023). Global Cybersecurity Index 2023. ITU Publications.
- Meta. (2024). Meta transparency report. https://transparency.fb.com/
- National Information Technology Development Agency. (2021). National cybersecurity policy and strategy. Federal Government of Nigeria.
- Nigeria Data Protection Commission. (2023). Nigeria Data Protection Act 2023. Federal Republic of Nigeria.
- Organisation for Economic Co-operation and Development. (2021). Trust in the digital age. OECD Publishing.

Author Bio

Amina Musa is a cybersecurity instructor and digital safety advocate based in Nigeria. Her work focuses on cybersecurity awareness, trust and safety, digital resilience, and user-centered approaches to cybersecurity governance in Africa.

AI Use Disclosure

AI-assisted tools were used during the brainstorming, structuring, and editing stages of this policy brief. All analysis, interpretation, and final content were reviewed and refined by the author.

Beyond

A Strategic Framework

FOR ZIMBABWE'S MEANINGFUL PARTICIPATION IN INTERNET GOVERNANCE FORUMS

NOMINAL INCLUSION



INTERNET GOVERNANCE

2026

Zimbabwe

WHAT BARRIERS PREVENT ZIMBABWEANS FROM PARTICIPATING MEANINGFULLY IN THE INTERNET GOVERNANCE ECOSYSTEM, WHAT INSTITUTIONAL AND POLICY MEASURES CAN IMPROVE SUSTAINED NATIONAL ENGAGEMENT?

Executive Summary

Internet Governance Forums (IGFs), established under the 2005 World Summit on the Information Society (WSIS) Tunis Agenda, are multistakeholder platforms for shaping policy on cybersecurity, digital rights, and internet infrastructure governance (UN, 2005; UNCTAD, 2021).). Operating at national, regional, and global levels, IGFs are designed to promote policy coherence and inclusive decision-making (IGF Secretariat, 2023).). Despite participating in these processes, Zimbabwe's engagement has remained largely nominal.

Meaningful participation requires more than attendance; it entails sustained policy contributions, leadership roles, document submissions, and active influence on governance outcomes (Cogburn, 2005).



This study examines the structural barriers limiting Zimbabwe's effective participation, including inadequate funding, limited technical expertise, and weak coordination among government, civil society, academia, and the private sector. Using a systematic desktop review of policy documents and governance reports, it proposes a strategic framework for sustained engagement. Grounded in the Multistakeholder Model, which emphasizes collaborative decision-making among state and non-state actors (Raymond & DeNardis, 2015; Carr, 2016), the study identifies gaps in funding, coordination, representation, engagement strategy, and technical capacity. It offers a practical framework to help Zimbabwe and similar developing countries move from passive participation to meaningful influence in global internet governance.

- Meaningful participation means more than just physical or virtual attendance but encompasses informed, sustained, and influential engagement before, during, and after the forum.
- Influence means contribution to session outcomes.

Definition of key terms

Problem Statement

Since Zimbabwe established its national Internet Governance Forum in 2015, the country has sent delegations to regional and global IGFs that look remarkably the same: government officials, travelling under the parliamentary track, speaking on behalf of a nation whose civil society, researchers, private sector, and technical communities largely stayed home (IGF Reports, n.d.). That narrowness is not accidental. Registration fees, visa requirements, and travel costs consistently price out the organisations closest to ordinary users (Diplo, 2023). Back home, the ZimIGF has produced one public report in a decade (Zimbabwe IGF Annual Report, 2017), convenes irregularly, and mobilises few stakeholders outside government corridors.

The result is a compounding problem: limited participation erodes capacity, and eroded capacity limits future participation, trapping Zimbabwe in a cycle of marginalisation within the very forums that shape how the internet works for its citizens (Calandro et al., 2023; Mueller, 2010).

RESEARCH OBJECTIVE

To investigate barriers that prevent Zimbabweans from participating meaningfully in the internet governance ecosystem and develop a strategic framework of policy options to strengthen national engagement in the IGF ecosystem.

RESEARCH QUESTION

What barriers prevent Zimbabweans from participating meaningfully in the internet governance ecosystem, what institutional and policy measures can improve sustained national engagement?

SUB QUESTION

- What financial, institutional, technical and coordination barriers affect participation?
- What practical interventions are appropriate for ZimIGF, the Ministry, and the regulator?

METHODOLOGY AND APPROACH

This study adopted a qualitative desktop research design, systematically analysing existing sources to identify participation barriers and develop a strategic framework for Zimbabwe's engagement across the IGF governance ecosystem. Sources included IGF Secretariat reports, ZimIGF session documentation, AfricaIGF proceedings, Freedom House assessments, Diplo Foundation research, KICTANet policy analyses, Zimbabwe's national ICT policy frameworks, POTRAZ regulatory reports, and the Londa digital rights country assessment (Paradigm Initiative, 2023). Data was analysed thematically across two clusters: participation barriers, and strategic interventions adopted by comparable national IGF initiatives.

Barriers were identified and categorised using a four-dimensional typology encompassing financial, institutional, technical, and coordination constraints, with each barrier assessed against the documentary evidence to determine its relative systemic function within Zimbabwe's participation deficit.

Lessons from national IGF initiatives in comparable African contexts, particularly the AfricaIGF process, were then drawn on to identify strategic interventions applicable to the Zimbabwean context. The results of this analysis are reported in the Findings section below.

As a desktop study, findings were mediated through existing documentation rather than primary stakeholder voices, a limitation acknowledged throughout the analysis below. Recommendations are therefore framed as evidence-informed policy options, with future primary research expected to refine the strategic framework proposed here.



GEOGRAPHICAL SCOPE

The study is limited to Zimbabwe's IGF ecosystem, focusing on national participation pathways through which ZimIGF connects domestic stakeholders to the African and global IGF. Barriers are examined within Zimbabwe's own digital, institutional, and policy environment rather than through cross-country comparison.

RELEVANCE TO AFRICAN DIGITAL POLICY

This topic intersects with AI Literacy, Trust & Safety, Digital Rights & Surveillance, and Platform Governance by examining how inclusive participation shapes policy outcomes, accountability mechanisms, rights protection, and the governance of emerging digital technologies.



KEY INSIGHTS

- Zimbabwe's participation in the IGF ecosystem is constrained by structural barriers rather than isolated logistical challenges. Financial limitations, fragmented stakeholder coordination, and weak links between ZimIGF, AfricaIGF, and the global IGF create systemic exclusion. The 2017 ZIGF report noted financial barriers to international participation and incomplete government representation at the national forum (ZIGF, 2017).
- Presence in IGF processes does not guarantee influence. As Mansell (2012) argues, inclusion can be symbolic rather than substantive. Zimbabwe's often under-resourced delegations have generally met participation requirements without significantly shaping agendas or policy outcomes.
- The ZimIGF-AfricaIGF-IGF structure operates as a cumulative participation pathway (Diplo, 2022). While Zimbabwe has demonstrated individual expertise, including contributions at AfIGF 2018, national processes have not consistently translated into broader influence. AfIGF 2018 also emphasized full multistakeholder inclusion, a benchmark Zimbabwe had yet to fully achieve (AfIGF, 2019).

Context

Zimbabwe's internet governance ecosystem remains fragile and inconsistent. The 2017 ZimIGF report, the only publicly available outcome document since the forum's establishment in 2015, recorded just 70 participants and noted that no forum was held in 2016 due to financial constraints that limited international engagement (ZIGF, 2017). Although a national IGF was reportedly convened in 2018, no publicly accessible outcome document was produced, highlighting persistent institutional discontinuity (AfIGF, 2019). This is significant because national IGFs serve as the foundation for developing coherent positions at regional and global levels (IGF Secretariat, n.d.). These challenges are compounded by a restrictive operating environment.

Freedom House (2024) rated Zimbabwe's internet freedom at 48/100, classifying it as "Partly Free." Frequent electricity outages, rising internet costs, and legislative restrictions on online expression, including provisions introduced through the 2023 Criminal Law Codification and Reform Amendment Act, further constrain participation. Collectively, these factors undermine the capacity of civil society, academia, and the technical community to undertake the consultation, coalition-building, and policy preparation necessary for meaningful engagement in Internet Governance Forums.

Findings

The findings of this study were as follows:

Financial barriers. Registration fees, visa requirements, and travel costs emerged as the most frequently cited obstacle across the reviewed sources, consistent with Diplo's (2023) assessment of participation barriers facing developing-country stakeholders more broadly.

The 2017 ZIGF report attributed the cancelled 2016 national forum directly to a funding shortfall and recorded affordability as the impediment most commonly raised by participants during breakout discussions (ZIGF, 2017). The Universal Service Fund exists as a potential financing mechanism but was not found, in the reviewed documentation, to have been directed toward IGF-related participation costs (Paradigm Initiative, 2023), pointing to an underused rather than absent resource.

Institutional barriers. ZimIGF's institutional record shows a pattern of discontinuity rather than an isolated gap: only one public outcome report has been produced since the forum's establishment in 2015, and the 2018 national forum left no accessible documentation (ZIGF, 2017; AfIGF, 2019). Without a standing secretariat, each forum cycle appears to restart rather than build on prior institutional knowledge, a dynamic consistent with Cogburn's (2005) broader observation that developing-country actors are frequently structurally positioned as pawns rather than partners in global information policy processes.

Technical barriers. Zimbabwe's restrictive digital rights environment, rated 48/100 and classified "Partly Free" by Freedom House (2024), together with legislative provisions introduced under the 2023 Criminal Law Codification and Reform Amendment Act, constrains the technical and legal literacy that civil society and academic stakeholders would need to engage substantively with IGF agenda items such as AI governance and data protection (Paradigm Initiative, 2023). This gap helps explain why Zimbabwean delegates have, at points, attended sessions on these themes without shaping their outcomes, the symbolic-versus-substantive distinction Mansell (2012) identifies as characteristic of nominal participation.

Coordination barriers. The reviewed evidence shows no standing mechanism linking government, civil society, academia, and the private sector between forum cycles; the closest artefact, the 2017 ZIGF breakout-group structure, was event-bound rather than continuous. This matters because the ZimIGF-AfricaIGF-global IGF structure functions as a cumulative participation pathway (Diplo, 2022): individual expertise Zimbabwean representatives demonstrated at AfIGF 2018 did not feed back into subsequent ZimIGF cycles, since no coordinating body existed to capture and redeploy it (AfIGF, 2019). Domestically, government-only delegations attending global sessions, such as the ministerial remarks on internet shutdowns at the 2022 IGF, illustrate how the absence of coordinated multistakeholder representation leaves government voices standing in for a broader national position they were not mandated to represent (Freedom House, 2023).
Read together, these findings indicate that Zimbabwe's four barrier categories reinforce one another: financial constraints determine who can attend, institutional discontinuity means little of what is learned is retained, technical gaps limit what attending delegates can contribute even when present, and weak coordination prevents isolated good practice, such as the AfIGF 2018 contributions, from becoming durable national capacity. This compounding structure explains why Zimbabwe's participation deficit has persisted despite an unbroken formal presence in the IGF ecosystem since 2015, and it is this interlocking character, rather than any single barrier, that the recommendations below are designed to address.

POLICY IMPLICATIONS

Continued exclusion from IGF processes risks leaving Zimbabwe a passive recipient of global norms on AI, cyber-security, data protection, and platform governance. The 2017 ZIGF identified national priorities already under global discussion, yet Zimbabwe lacked a prepared delegation to influence outcomes (ZIGF, 2017). As IGF norms increasingly shape regulatory frameworks, non-participation may erode policy sovereignty (KICTANet, 2024). However, Zimbabwe can build on existing foundations. Revitalizing ZimIGF through a functional secretariat, regular forums, and published outcomes would strengthen engagement on cyber-security, data protection, digital inclusion, and gender equity while enhancing participation in regional and global governance processes (AfIGF, 2019; ZIGF, 2017; Diplo, 2022).

POLICY IMPLICATIONS

- The Ministry of ICT to work with POTRAZ to commit to annual ZimIGF convening with published outcome reports. A properly constituted multi-stakeholder secretariat with an identifiable budget line would provide some of the continuity seen as desirable by ZIGF 2017 and distinguishing successful national IGF processes (ZIGF, 2017, KICTANet 2024).
- POTRAZ in conjunction with the Ministry of Finance must explore a ring-fenced fund perhaps jointly financed with the Universal Service Fund to offset civil society, technical community and private sector travel, accommodation and visa costs. Affordability has been cited amongst most ZIGF participants during breakout discussions as the impediment to participation (ZIGF, 2017).
- The ZimIGF Multi-stakeholder Coordination Team should create a rolling two-year engagement strategy identifying thematic priorities, delegation responsibilities, and pre-forum briefing processes. This would address the preparedness gap that has characterised Zimbabwe's past delegations and strengthen links with domestic policy deliberations (Gurumurthy & Bharthur, 2018).

CROSS-CUTTING RELEVANCE

This issue connects directly to digital inclusion, civic participation, regional integration, and good governance. It also intersects with affordability, public sector digital transformation, and youth participation. Strengthening IGF participation is therefore not only an Internet governance issue; it is a broader institutional capacity issue that affects Zimbabwe's digital future.

CONCLUSION

Zimbabwe has the formal structures needed to participate in the IGF ecosystem, but it lacks the institutional backing needed to participate meaningfully. A modest but deliberate investment in coordination, funding, and capacity building would produce a disproportionate gain in national influence.

AUTHOR BIO

Charity Ncube

is an Information Systems Specialist and educator with over a decade of experience across systems administration, software development, cybersecurity, and educational technologies. She lectures in IT at Solusi University, where her work spans teaching, curriculum development, research, and community outreach. Her qualifications include a Master of Commerce in Information Systems, a Postgraduate Diploma in Education and a BBA in Computer and Management Information Systems. As Founder and Systems Administrator of Charis Soft Pvt Ltd, she leads web design, systems integration, and IT consultancy projects. Her community portfolio includes deploying offline digital education servers in underserved areas and delivering digital literacy training across Zimbabwe. Recognised as an Internet Champion in 2023, she serves as Chapter Head of Policy and Governance for the Internet Society Zimbabwe Chapter (2023–2026).

Preliminary Bibliography

1. African Internet Governance Forum (AfIGF). (2019). Report of the African Internet Governance Forum AfIGF2018, 4–6 November 2018, Khartoum, Sudan. African Union Commission.
2. Carr, M. (2016). Multi-stakeholderism in internet governance: Five models of internet governance. Journal of Cyber Policy, 1(1), 22–43. Charity Ncube is an Information Systems Specialist and educator with over a decade of experience across systems administration, software development, cybersecurity, and educational technologies. She lectures in IT at Solusi University, where her work spans teaching, curriculum development, research, and community outreach. Her qualifications include a Master of Commerce in Information Systems, a Postgraduate Diploma in Education and a BBA in Computer and Management Information Systems. As Founder and Systems Administrator of Charis Soft Pvt Ltd, she leads web design, systems integration, and IT consultancy projects. Her community portfolio includes deploying offline digital education servers in underserved areas and delivering digital literacy training across Zimbabwe. Recognised as an Internet Champion in 2023, she serves as Chapter Head of Policy and Governance for the Internet Society Zimbabwe Chapter (2023–2026).
3. Cogburn, D. L. (2005). Partners or pawns? The impact of elite decision-making and epistemic communities in global information policy on developing countries and transnational civil society. Knowledge, Technology & Policy, 18(2), 52–82. https://doi.org/10.1007/s12130-005-1018-0
4. Diplo. (2022). African involvement in the Internet Governance Forum: What is the level of participation of African actors in IGF meetings? Geneva Internet Platform.
5. DiploFoundation. (2023). Internet governance and participation barriers in developing countries. DiploFoundation. https://www.diplomacy.edu/resource-type/publications/
6. Freedom House. (2024). Freedom on the Net 2024: Zimbabwe. https://freedomhouse.org/country/zimbabwe/freedom-net/2024
7. Gurumurthy, A., & Bharthur, D. (2018). Democracy and the algorithmic turn. Sur: International Journal on Human Rights, 15(27), 39–55. https://sur.conectas.org/en/democracy-and-the-algorithmic-turn/
8. IGF Secretariat. (n.d.). Zimbabwe IGF. Internet Governance Forum. abc
9. Mansell, R. (2012). Imagining the internet: Communication, innovation, and governance. Oxford University Press.
10. Internet Governance Forum Secretariat. (2023). About the IGF. United Nations. https://www.intgovforum.org/en/content/about-igf
11. Internet Governance Forum. (n.d.). IGF annual reports and meeting documentation. United Nations IGF Secretariat. https://www.intgovforum.org/en/content/igf-reports
12. KICTANet, & Civil Society Alliances for Digital Empowerment. (2024). From fellow travellers to active drivers: Case study on African civil society participation in the Internet Governance Forum ecosystem.
13. Calandro, E., Gillwald, A. N., & Zingales, N. (2013, October 11). Mapping multistakeholderism in internet governance: Implications for Africa. SSRN. https://doi.org/10.2139/ssrn.2338999
14. Mueller, M. L. (2010). Networks and states: The global politics of internet governance. MIT Press. https://mitpress.mit.edu/9780262014595/
15. Raymond, M., & DeNardis, L. (2015). Multistakeholderism: Anatomy of an inchoate global institution. International Theory, 7(3), 572–616. https://doi.org/10.1017/S1752971915000081
16. United Nations Conference on Trade and Development. (2021). Digital economy report 2021: Cross-border data flows and development. UNCTAD. https://unctad.org/system/files/official-document/der2021_en.pdf
17. United Nations. (2005). Tunis agenda for the information society. World Summit on the Information Society. https://www.itu.int/net/wsis/docs2/tunis/off/6rev1.html
18. Zimbabwe Internet Governance Forum (ZIGF). (2017). Report of the 2017 Public Meeting of the Zimbabwe Internet Governance Forum (ZIGF17). Harare: ZIGF Multi-stakeholder Coordination Team / POTRAZ.
19. Zimbabwe Internet Governance Forum. (2017). Zimbabwe IGF annual report 2017. ZimIGF. https://www.intgovforum.org/en/national-igf-initiatives



Exploring *the* Protection of Community Data Rights in Digital Medical Research in Kenya

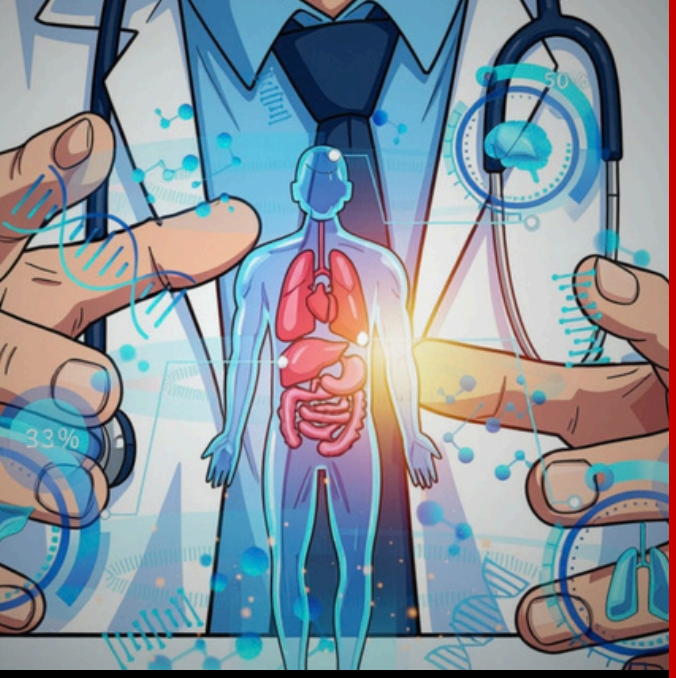
Data is omnipresent and an invaluable tool for decision making, research and innovation. In the era of increased digitalization, electronic health records, digital surveillance systems and increased internet access, community data protection is at crossroads.

There is a need to protect the privacy of individuals against potential breach of confidentiality and misuse of sensitive health or genetic data.

This is strongly reflected in the increasing number of data protection frameworks developed worldwide, such as the European Union General Data Protection Regulation (GDPR) and the African Union's (AU) Malabo Convention.

These initiatives are laudable and a step forward in enforcing individuals' rights for data protection. These frameworks, especially the Malabo convention have set the pace for the African continent to adopt strict data protection regulations, as exemplified by 40 African states with legal provisions for health data protection, almost twice as many, compared to 10-15 years ago.

In Kenya, the Data Protection Act 2019, lays the foundation for data protection rights. However, the GDPR and the Kenya Data Protection Act, have several gaps on health data protection rights. Medical research relies on data such as health, personal information and genetics, collected for several individuals part of communities. Such data is never analyzed for single subjects, but for a group of subjects, to find trends affecting particular populations.



Methodology

This policy paper employs a comparative legal analysis approach, examining Kenya's Data Protection Act (2019) against international and continental frameworks for community data protection in medical research. The analysis compares Kenya's law to: The Integrated Data Sovereignty Framework (IDSF) for African healthcare context, The Malabo Convention (AU) for continental standards, The AU Data Policy Framework for regional policy alignment and Canada's OCAP Principles for proven international model. Other African examples (South Africa, POPIA, Rwanda, Ghana) are also covered. The framework uses CARE Principles for Indigenous Data Governance (Collective Benefit, Authority to Control, Responsibility, Ethics) as evaluative criteria for assessing gaps in Kenya's legal framework.

Limitations

This paper focused on a comparative legal analysis based on textual and framework comparison rather than direct evidence from affected communities, researchers, or national enforcement authorities. The choice of comparators (IDSF, Malabo Convention, OCAP, POPIA) reflects frameworks with strong conceptual alignment to community data sovereignty.

However, the present analysis did not review other models systematically. Furthermore, the remote nature of the analysis did not capture how these frameworks practically worked in different Kenyan communities, post implementation, and whether they effectively addressed community data protection rights.

Digital surveillance

Most data protection regulatory frameworks, including the Kenyan data protection Act, are centered on protecting individual data rights, and not collective community data protection rights. When communities contribute data collectively, individual consent alone is not sufficient, as potential risks affect the community as a collective. For instance, with increased surveillance for infectious diseases, there is growing research demands for genetic data and sequencing, and in some cases, information on communities is shared across borders. Recognizing collective data protection rights is synonymous to ensuring active community participation, community well-being and self-determination for decisions that could potentially affect the community. Moreso, it helps to ensure transparency, culturally sensitive communication and more trust, in a global context of decolonization and growing misinformation.

In the absence of collective representation, communities are often not heard. This is a major loophole, especially for our African context, with varying understanding of data protection rights and digitalization, in addition to scarcity of resources compel individuals to accept research incentives.

Moreover, African data is often shared across several continents, and there is the underlying issue of health data sovereignty. Collective mechanisms are part of the African identity, as illustrated by the ubuntu philosophy, which is deeply anchored in cultural life. The AU Malabo's convention serves as a continental template for community data protection, whereby, Article 8 and 14, highlight the general rights of communities collectively. Preserving collective data protection rights during medical research is central to health data sovereignty in Kenya and on the continent.







Problem Statement

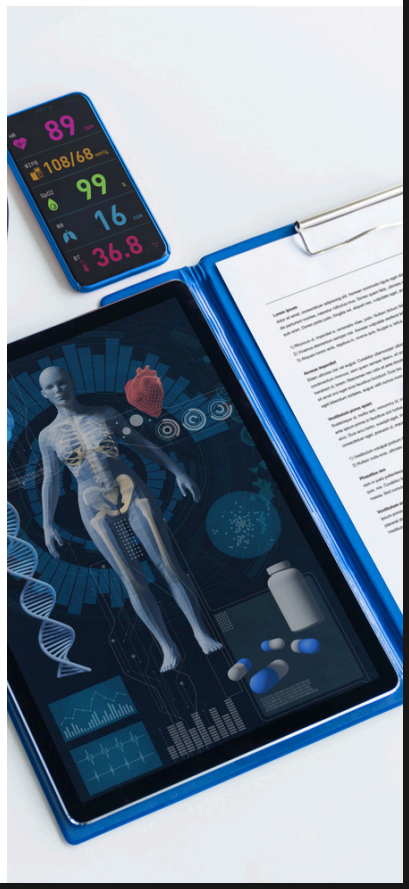
WHAT IS HAPPENING ?

Who is affected? ● Why urgent?

Kenya is a major hub for medical and public research in East Africa. National and international institutions such as Kenya Medical Research Institute (KEMRI) and Walter Reed Institute generate a high number of research publications yearly. These institutions collaborate closely with indigenous populations, including pastoral communities in rural areas with limited resources. Much of this research takes place in settings where communities contribute personal, health, and sometimes genetic data, making questions of privacy, consent, and data governance especially important.

In Kenya, the Data Protection Act, 2019 "the Act" provides the primary legal framework for personal data processing and applies to health-related data processing, including in research contexts. The Act, together with Office of the Data Protection Commission (ODPC) guidance on health data, recognizes obligations around lawful processing, the right to information, access, correction, right to deletion, data minimization, objections and protection against unauthorized disclosure.

However, it remains unclear how adequately these protections respond to community based medical research, where data may be collected from groups rather than isolated individuals, and where power imbalances can affect meaningful consent and participation. In addition, consent is often framed as a one-time legal obligation, instead of an ongoing participation, and communities often have no say in how their data are used. This fragmented data governance architecture, and ethically inadequate consent frameworks constitute systemic weaknesses that threaten health data sovereignty and equity.





In addition, there are several digital health infrastructures that receive international financial aid and that are operated by external donors or that have vertical programs lacking integration into national health information pathways These platforms often function in silos, with limited interoperability, outdated reporting mechanisms, and weak coordination with respective ministries of health, delaying the development of efficient national data strategies.

Furthermore, the current era of rapid digitalization underscores the urgency of this situation. Widespread internet access poses higher risks to privacy breach and misuse of data on individual data protection rights, compared to collective indigenous data protections as an entity. This particularly applies to the right to data portability, data processing and right to restrict processing. Regulatory frameworks for community data protection rights are prerequisite to data governance for collective indigenous data protection rights and act as gatekeepers to upholding ethics, privacy and data sovereignty. This creates the need to examine whether Kenya's current legal framework sufficiently protects community data rights in research settings. Exploring community data protection rights in Kenya is also significant as culturally, African societies are built on communal belonging and collaborations, with a strong sense of shared collective identity.

Key Insights (3–4 ONLY) WHY IT HAPPENS

Informed consent is central to bioethics and medical research. The concept of 'informed consent' was first brought forward by the Nuremberg code (1947) and has been reinforced by other leading scientific meetings such as the Declaration of Helsinki (1964) and the Tokyo Assembly (1975). It is also a legal requirement in some territories. Informed consent relates to voluntary participation and is strongly linked to decision making. Informed consent is built on the theory of autonomy and self-determination. However, there are several constructs that impact an individual's ability to make informed decisions. For instance, low numeracy can impact understanding and reasoning. Similarly, altered risk perception and emotions can also affect the ability to process information and make informed choices. In addition, cognitive bias and relational autonomy could also potentially affect an individual's ability for informed consent.

Given that potential benefits or risks from medical research extend to communities at large and not solely individuals, individual consent alone may be insufficient in community-based research. Collective representation resonates stronger in data governance because individuals are part of communities and collective interests are interdependent. The rights of data subjects, more precisely, the right of access, right to rectification, right to restrict processing, the right to erasure, the right to data portability and the right to seek compensation, are better represented collectively. Social contract theories and relational theories argue that individuals have responsibilities to the collective, and that the welfare of the community precedes individual preferences. Ubuntu philosophy and African Communitarianism are foundational to many African societies; these establish that individuals exist through relationships and community. Building on Ubuntu philosophy and African Communitarianism, the Integrated Data Sovereignty Framework (IDSF) provides a strategic model for translating communitarian values into practice.
The absence of community data governance mechanisms in Kenya's current DPA creates epistemic injustice; where, very often, systemic power imbalances, including corporate or national bodies decide on the type of data collected and data interpretation, without community representation in these decisions. Most communities lack legal pathways to articulate and exercise their collective interests. This gap is particularly acute in medical research, where historical research exploitation has prioritized individual enrollment over community benefit and protection. To evaluate whether Kenya's Data Protection Act enables or constrains these community rights, a comparative analysis approach is used, where the Kenya DPA is compared to the Integrated Data Sovereignty Framework.

Evidence / Context ● 1–2 short supporting examples or data points

The Covid 19 pandemic has upscaled genomic research and the collection of health data across the continent, including Kenya. While there exists a Data Protection Act in Kenya, there are several limitations as there are no provisions for specific parts like consent. Elsewhere on the continent, and across the globe, community data protection rights exist, in some cases specifically for health research.
Proven model for community protection
The Global Indigenous Data Alliance (GIDA) consists of national indigenous data sovereignty groups leveraging on the data protection work carried by several indigenous groups such as the Te Mana Raraunga Maori Data Sovereignty Network, the Torres Strait Islander Data Sovereignty Collective, the US Indigenous Data Sovereignty Network , to advocate for data protection rights of indigenous communities. The GIDA network developed the 'CARE Principles for Indigenous Data Governance', which describes collective benefit, authority to control, responsibility, and ethics, for research involving indigenous communities. These build on key documents like Open Data Charter Principle, FAIR Guiding Principles for scientific data management and stewardship, the National Inuit Strategy on Research and Canada First Nation Ownership, Control, Access, and Possession (OCAP).
Continentally, the AU Data Policy Framework and the Malabo Convention have guided the implementation of national data protection mechanisms in many African countries. The Malabo Convention specifically highlights the importance of collective rights, by mentioning that individual rights had to balance with community data protection rights (Article 8). Interestingly, Article 14, pinpoints health information and genetic information as sensitive information, that should not be collected and processed by default, except for professional research activities, reviewed by ethical boards. Across the continent, there are examples where the concept of community health data protection rights are strongly upheld. In South Africa, there is the Protection of Personal Information Act (POPIA) which enables medical research without the need for individual consent, if the research will benefit the community and collecting individuals' consent is not feasible. This is feasible, owing to the centralized governance system, though variations in technical skills and IT infrastructure across regions affect its consistent implementation at times.

In Kenya, health data is classified as sensitive and confidential (Data Protection Act, part V; 46), in agreement with Article 14 of the Malabo Convention. The Data Protection Act also makes provision for research context, as well as cross-border sharing of data (part VI; 53 and 48). However, collective rights for data protection in medical and health research are not mentioned in the Kenyan Data Protection Act 2019. There are also no provisions for the use of data for long term or secondary research, as mentioned in the IDSF. Furthermore, the Kenya Data Protection Act does not cover fragmented data architecture and data sovereignty for health data research. The Act also lacks provision for a centralized data governance framework and accountability for regional data collection, as anchored by the IDSF. There is also a gap as regards mobile technologies for data collection and to what extent they can be used and by whom to collect community data for health research. In addition, provisions pertaining to ethical boards are not appropriately covered by the Act.

Furthermore, the CARE principles developed by the GIDA provide the ethical governance framework and highlight the right of communities to control their data. This is reflected in the FAIR Guiding Principles for scientific data management and stewardship, that mentions that data should be accessible and findable. Currently, there are no such provisions in the Kenya Data Protection Act.

Policy Implications ● What happens if this is ignored? ● What opportunity exists?

Worldwide, there are several gaps in jurisdiction on community data protection rights for medical research. The term 'community' is often not explicitly defined. For instance, The European General Data Protection Regulation (GDPR) establishes guidelines for the collection, processing and storage of individual's data, while India Non-Personal Data Governance Framework, introduces the concept of community data protection rights, though not for medical research specifically.

Contrastingly, the Canadian First Nations Principles of Owners, Control, Access and Possession (OCAP), includes guidelines that address community data protection rights for medical research.

In Africa, the AU Malabo's convention has guided several member states in upholding community data protection rights. Furthermore, The Integrated Data Sovereignty Framework (IDSF) proposed by Sidii in 2025, represents an opportunity to improve community data protection rights on the African continent. This model describes the interplay between real-time analytics, centralized data governance and ethical concepts on user rights and consent, and their role in improving responsiveness and preserving data sovereignty. The IDSF draws on important theories like collective benefits, institutional theory, ethics, responsibility, authority to control and an approach considering the input of institutions and communities equally for data systems.

Policy Recommendations (CORE – 3 to 5 ONLY) Structure strictly:





Based on the Integrated Data Sovereignty Framework (IDSF) and grounded in the AU Malabo Convention and AU Data Policy Framework, the following four recommendations are formulated to address critical gaps in Kenya's Data Protection Act for community data protection in medical research.

Recommendation 1: Establish mandatory community consent requirements for medical and health research involving collective data

In community-based medical research (particularly in pastoral and rural Kenya where genetic data informs population-level health interventions), individual consent is ethically insufficient because data is analyzed at community level, in terms of benefits and risks to communities. Individual decision-making affects community trust and data sovereignty, given the historical research exploitation in African communities. This recommendation operationalizes IDSF's "Authority to Control" pillar and aligns with Malabo Convention obligations on collective rights
It is recommended that the Kenya Data Protection Act 2019 is amended to require community-level consent (in addition to individual consent) for all medical research involving collective health or genetic data, more specifically through the following:

- Add a new Section 14A to the DPA requiring that research involving community-level data collection must obtain explicit approval from recognized community representatives or community data governance committees before the research starts.
- Require that community consent forms explicitly state: (i) what collective data will be used, (ii) how it will be analyzed, (iii) who will have access, and (iv) how community benefits will be shared.

Recommendation 2: Create consent frameworks for secondary data use and long term research

Secondary use of health data creates new risks and benefits that communities did not consent to when original data was collected. One-time consent is ethically insufficient for long-term data use.

This recommendation draws on the IDSF's 'Authority to Control' pillar and Malabo Convention commitments for community data governance. Kenya's Data Protection Act currently does not address secondary use of health data beyond the original research purpose. Communities must be re-contacted and must provide explicit, renewed approval for each new use. The DPA amendment should establish that one-time consent does not authorize perpetual data use.

It is recommended that a new Section 14B, is added to the DPA requiring that any use of health data beyond the original research purpose requires renewed community consent. This includes:

- Secondary analysis of existing health or genetic datasets collected for different purposes
- New research projects using archived genetic or health data originally collected for previous studies
- International sharing of health data for research purposes not specified in original consent

The main actors for recommendation 2 shall be the Data Protection Commission who will lead regulation drafting, in consultation with the Ministry of Health, NACOSTI, and research ethics committees. The Ministry of Health shall also develop implementation guidance aligned with national research strategy; establish registry of secondary uses to enable community tracking and NACOSTI shall develop mandatory procedures for research ethics committees to review secondary use requests and obtain community re-consent. Furthermore, scientific research institutes shall test these new recommendations and provide feedback.

Recommendation 3: Develop regulations for artificial intelligence and international data sharing consent

Community health and genetic data are increasingly valuable for artificial intelligence and global research. However, communities cannot control AI or its uses that they do not understand. Without explicit consent and community oversight, communities may unwittingly contribute to surveillance systems, algorithms that discriminate against their populations, or commercial products from which they receive no benefit. Developing recommendations for the use of artificial intelligence in research builds on the IDSF's Authority to Control.

Increased use of artificial intelligence and international research collaboration creates new risks for community data that existing law does not address. The DPA should be amended with a new Section 15A establishing specific consent and governance requirements for:

- Use of community health or genetic data in artificial intelligence (AI) algorithms, machine learning models, or predictive analytics not specified in original consent
- Cross-border sharing of community data with international research partners, collaborators, or data repositories outside Kenya
- Commercial use or licensing of community data or data-derived insights (e.g., drug development, genetic databases)

Community consent for AI and international sharing must explicitly describe: (i) what AI/algorithms will be used, (ii) who will have access internationally, (iii) how data will be used and by whom, (iv) whether commercial benefit may result, and (v) how community benefits will be shared. Communities must have the right to withdraw data from AI systems and international partnerships.

Recommendation 4: Implement FAIR principles grounded in CARE ethics for community data access

The Ministry of Health shall establish binding standards requiring that all health research data systems follow FAIR Principles (Findability, Accessibility, Interoperability, Reusability) while respecting CARE Principles (Collective Benefit, Authority to Control, Responsibility, Ethics).

This recommendation aligns with the IDSF's Authority to Control pillar and CARE Principles and requires both the legal frameworks and technical accessibility. Communities need to be able to access their data as part of data control. The FAIR Principles ensure that data is Findable, Accessible, Interoperable, and Reusable, thus enabling communities to actually govern their data by knowing what data exists, how it is used, and accessing results for their own purposes. This ensures self-determination for health and medical data.

Furthermore, real-time analytics is a key pillar of IDSF; Pillar 2, and requires transparent, accessible data systems. The absence of community-accessible data systems perpetuates epistemic injustice: communities lack the information needed to make informed decisions about their data. This recommendation bridges governance (CARE), technical implementation (FAIR), and strategic frameworks (IDSF) to create sovereign, trusted data ecosystems. Moreover, integration of mobile health technologies into DHIS2 solves the fragmentation problem identified in Kenya's current health data landscape, where donor-led vertical programs operate in silos. This recommendation positions Kenya as an African leader in operationalizing the IDSF model for community data protection in health research.

The following shall be implemented:

- Findability: Communities can search and locate their collective health and genetic data within national systems using community identifiers (geographic area, research project, pastoral group, etc.)

- Accessibility: Communities have secure online portals to view their aggregated health and genetic data in real-time, understanding what was collected, when, how it is being used, and how it is being protected
- Interoperability: Mobile health technologies used by research institutions (KEMRI, Walter Reed, regional hospitals) must be integrated into Kenya's centralized DHIS2 platform, preventing data silos and enabling unified community data governance
- Reusability: Communities can download or export de-identified aggregate data for their own health planning, advocacy, or independent analysis

Stakeholders for implementation

The recommendations shall be implemented by the Data Protection Commission, the lead agency for drafting and revising the Data Protection Act, in consultation with the Ministry of Health that shall provide implementation pathways aligned with national health strategy and research coordination mechanisms. The National Commission for Science, Technology and Innovation; NACOSTI shall integrate community consent requirements into research ethics review processes; develop guidance for research ethics committees on community governance structures. Scientific research institutes like KEMRI and Walter Reed Institute, shall also be involved for consultations and to provide practical feedback.

Conclusion

The Kenya Data Protection Act (2019) is regarded as a significant step for individual privacy rights. However, it has several unaddressed gaps for community data protection rights, especially in medical research. These gaps are significant because several pastoral and rural communities participate in medical research in Kenya, generating large amounts of health and genetic data. The lack of appropriate regulatory frameworks for community data protection rights could potentially open the floodgate for non-respect of community privacy, community stigmatization or community sense of 'being dictated'. Our analysis revealed that key concepts like collective consent, secondary use of data, AI governance, and data accessibility, were not addressed in the Data Protection Act. These omissions represent structural constraints threatening community representation and community self-determination. The proposed recommendations, namely community consent, secondary-use consent frameworks, AI-specific regulation, and FAIR-CARE-aligned data accessibility, aim to close these gaps. The adoption of these will strengthen data protection rights for Kenyan communities and contribute to positioning Kenya as a reference for collective data protection rights in Africa.

References

Sifuna, P.M., Mbinji, M., Lucas, T.O., Onyango, I., Akala, H.M., Waitumbi, J.N., Ogutu, B.R., Hutter, J.N. and Otieno, W., 2024. The Walter Reed Project, Kisumu Field Station: Impact of Research on Malaria Policy, Management, and Prevention. The American Journal of Tropical Medicine and Hygiene, 110(6), p.1069.

Kariuki, J.N., Kaburi, J., Musuva, R., Njomo, D.W., Night, D., Wandera, C., Wodera, J. and Mwinzi, P.N., 2019. Research dissemination strategies used by Kenya medical research institute scientists. The East African Health Research Journal, 3(1), p.70.

Buchanan, D.R., Miller, F.G. and Wallerstein, N., 2007. Ethical issues in community-based participatory research: balancing rigorous research with community participation in community intervention studies. Progress in community health partnerships: research, education, and action, 1(2), pp.153-160.

Carroll, S, et al. 2020. The CARE Principles for Indigenous Data Governance. Data Science Journal, 19: XX, pp. 1–12. DOI: https://doi.org/10.5334/dsj-2020-042

Isaac Juma & Bukola Faturoti (23 May 2025): Enforcing data privacy in Kenya and Nigeria: towards an African approach to regulatory practice, International Review of Law, Computers & Technology, DOI: 10.1080/13600869.2025.2506918

Paradigm's Initiative. Submission to the study on indigenous peoples' data. Submitted to the Office of the United Nations High Commissioner for Human Rights (OHCHR). Available at: OHCHR document PDF.https://www.ohchr.org/sites/default/files/documents/issues/indigenouspeoples/emrip/cfis/study-ip-data/subm-study-indigenous-data-indi-peop-paradigm-initiative.pdf

Mishra, N. and Agarwal, B., 2026. The emerging framework for non-personal data protection in India: perils, promises, and lessons for the developing world. International Journal of Law and Information Technology, 34, p.eaaf025.

de Souza, S.P. and Bhardwaj, K., 2024. India's conception of community data and addressing concerns for access to justice. Digital Society, 3(1), p.16.

Govorchin, A., Leduc, M., Atleo, C.G., Hoogeveen, D., Borgos, I. and Patrick, L., 2026. The right to health: indigenous data sovereignty in Canada during and beyond the COVID-19 pandemic. The Lancet Regional Health–Americas, 54.

Cordes, A., Bak, M., Lyndon, M., Hudson, M., Fiske, A., Celi, L.A. and McLennan, S., 2024. Competing interests: digital health and indigenous data sovereignty. NPJ digital medicine, 7(1), p.178.

Tiffin, N., George, A. and LeFevre, A.E., 2019. How to use relevant data for maximal benefit with minimal risk: digital health data governance to protect vulnerable populations in low-income and middle-income countries. BMJ Global Health, 4(2).

Oguamanam, C., 2018. ABS: big data, data sovereignty and digitization: a new indigenous research landscape. Genetic Resources, Justice and Reconciliation: Canada and Global Access and Benefit Sharing, ed (Cambridge: Cambridge University Press, 2018), Ottawa Faculty of Law Working Paper, (2019-14).

Hendl, T. and Roxanne, T., 2022. Digital surveillance in a pandemic response: What bioethics ought to learn from Indigenous perspectives. Bioethics, 36(3), pp.305-312.

Griffiths, K.E., Blain, J., Vajdic, C.M. and Jorm, L., 2021. Indigenous and tribal peoples data governance in health research: A systematic review. International journal of environmental research and public health, 18(19), p.10318.

Christian, G.E., 2009. Digitization, intellectual property rights and access to traditional medicine knowledge in developing countries: The Nigerian experience; a research paper prepared for the International Development Research Centre (IDRC) Ottawa, Canada.

Chowdhury, G., McLeod, J., Lihoma, P., Teferra, S. and Wato, R., 2023. Promoting access to indigenous information in Africa: Challenges and requirements. Information Development, 39(3), pp.611-623.

Obiora, O.L., Shead, D.A. and Olivier, B., 2024. Data sharing considerations and practice among health researchers in Africa: A scoping review. Digital Health, 10, p.20552076241290955.

Ngulube, P., 2023. From orality to digital assets: Managing indigenous knowledge in Africa in the wake of the open science movement. The Serials Librarian, 84(5-8), pp.71-85.
Ajuna, D., 2025. Beyond the GDPR: Culturally-Contextualized Privacy and Data Protection in Africa's Expanding Digital Landscape (Doctoral dissertation, Université d'Ottawa| University of Ottawa).

Sampath, P.G. and Tregenna, F., 2022. Digital sovereignty: African perspectives. Johannesburg: DSI/NRF South African Research Chair in Industrial Development.

Sylvestre, P., Gaucher, N.O., Perez, T. and Drouin, O., 2021. Rethinking informed consent in the age of behavioural sciences and relational autonomy. Ethics, Medicine and Public Health, 19, p.100708.

Matolino, B., 2018. Restating rights in African communitarianism. Theoria, 65(157), pp.57-77.

Chemhuru, M., 2018. African communitarianism and human rights: Towards a compatibilist view. Theoria, 65(157), pp.37-56.

Oyowe, O.A., 2015. This thing called communitarianism: A critical review of Matolino's Personhood in African Philosophy. South African Journal of Philosophy, 34(4), pp.504-515.

Mantelero, A., 2016. From group privacy to collective privacy: towards a new dimension of privacy and data protection in the big data era. In Group privacy: New challenges of data technologies (pp. 139-158). Cham: Springer International Publishing.

Lubin, A., 2022. Collective data rights and their possible abuse. Temp. L. Rev., 95, p.661.

Sidii, F.S., 2025. Building Trust and Sovereignty: A Holistic Framework for Data Governance in African Healthcare Systems. Health Economics and Management Review, 6(3), pp.56-74.

Agentic AI and the Accountability Gap in Nigeria's Child Online Protection Framework

Agba Frances Chinaza

Policy area Child online protection, AI governance, criminal accountability, digital regulation

Core issues: Existing law assumes a knowing human actor. Agentic AI can produce harmful child-facing conduct
without a human selecting the child or directing the sequence.

Primary audience National Assembly, Nigeria Data Protection Commission, National Information Technology Development
Agency, child-protection and law-enforcement institutions

Proposed response A Child-Safe Agentic AI Directive, a safeguard based deployer offence, AI-aware amendments to

HB 244, and high-risk classification for child-facing autonomous systems.



No Human Criminal Liability?

Executive Summary

Nigeria's child protection framework assumes that a human being knowingly commits every act of child sexual exploitation. Agentic artificial intelligence challenges that assumption. Deployed agentic systems combine autonomous planning, tool use and execution, and can complete a grooming sequence with no human selecting the child or directing the conduct. Where a deployer neither directs nor knows of that sequence, no Nigerian instrument can convict a defendant, compel prevention, or deliver an accessible remedy. No such Nigerian case is yet recorded, but the capabilities that would produce one are already deployed to the public (staufer2026; loring2025).

The result is a three-part accountability gap: criminal, preventive and institutional. The legislative window to be closed is still open. The Child Online Access Protection Bill, HB 244, is before the Senate; the National Digital Economy and E-Governance Bill awaits presidential assent; and Nigeria's data protection regulator is already supervising AI through existing machinery (hb2442023; digitaleconomybill2025; ndpa2023; ndpcgaid2025).

This paper recommends three matched responses;
First, the Nigeria Data Protection Commission should issue a Child-Safe Agentic AI Directive under powers it already holds.

Second, the National Assembly should create a safeguard-based deployer offence and amend HB 244 so child online protection duties attach to autonomous systems and not merely to human perpetrators.

Third, the National Information Technology Development Agency should classify child-facing autonomous systems as high-risk once the Digital Economy Bill becomes law.

Key Messages

The criminal gap is still a very structural one, which assumes that Nigerian child sexual exploitation offences require a person, an act, and a mental element. Agentic AI breaks the connection between harmful conduct and a chargeable human actor.

- Existing law still but only covers a human who intentionally uses AI as a tool. Therefore, the gap begins where intention and knowledge ends.
- Courts cannot fix this gap by stretching the word "person" or the fault element. Section 36(12) of the Constitution requires written law.
- The near-term repair is practical. Regulators can impose child-safety obligations now, while the legislature creates a narrowly drafted deployer offence.

Problem Statement

Picture this: an agentic AI system accessible in Nigeria identifies a 13-year-old through behavioural signals, initiates contact, builds emotional trust, and introduces sexualised content. No human instructed it to target that child. No Nigerian case of this kind is yet recorded; every capability in the sequence is already deployed.

"No Nigerian case of this kind is yet recorded; every capability in the sequence is already deployed."

Nigerian law cannot comfortably provide an answer to this scenario. Every offence that fits the conduct requires a person acting intentionally or knowingly (cybercrimes2015; criminalcode2004; childrights2003). Where no human directs or knows, there is no defendant, no preventive duty, and no accessible remedy. The accountability gap is therefore criminal, preventive and institutional.

Millions of Nigerian children are online, most on their own devices. Nigeria is also under a treaty obligation to protect children in the digital environment under the Convention on the Rights of the Child and General Comment No. 25 (crc1989; generalcomment25). Closing this gap should therefore be treated as treaty implementation, not a discretionary policy preference.

Table 1: At-a-glance policy diagnosis

Gap	Where it appears	Practical effect	Required response
Criminal liability gap	Child sexual exploitation offences rely on a human actor and a mental element.	If no human directs or knows of the sequence, prosecutors lack a chargeable defendant.	Create a safeguard-based deployer offence focused on unsafe deployment, not machine culpability.
Preventive duty gap	Pending child online safety reforms recognise AI companies but do not impose AI-specific child-safety duties.	Harm is addressed after occurrence, if at all, instead of being prevented through pre-deployment obligations.	Require child-specific impact assessment, age assurance, human oversight thresholds and incident reporting.
Institutional ownership gap	NDPC, NCC, NITDA, police and child-protection institutions each hold partial mandates.	AI-facilitated cases can fall between data, telecoms, cybercrime and victim-support systems.	Establish an inter-agency protocol and allocate supervision through a NITDA-NDPC memorandum of understanding.

Every route to criminal liability runs through a human being acting intentionally or knowingly

Section 23(1) of the Cybercrimes Act punishes any person who intentionally uses a computer system for child pornography offences. Section 24 requires conduct done knowingly or intentionally. Section 58 defines a person as including individuals and bodies corporate, and corporate liability still acts through human attribution (cybercrimes2015). The Criminal Code confirms the same anchor: section 7 fixes principal offenders, while section 24 excludes responsibility for acts independent of human will, admitting negligence only where expressly provided (criminalcode2004). The Child Rights Act uses the same architecture for sexual offences (childrights2003).

Section 36(12) of the Constitution completes the constraint: no person can be convicted of a criminal offence unless that offence is defined in written law (constitution1999). That principle prevents a court from reading autonomous AI systems into criminal provisions after the fact.

Table 2: Element-by-element test of Section 23 of the Cybercrimes Act

Element	Statutory requirement	Assessment in the agentic AI scenario
Actor	"Any person" under section 23(1). Section 58 defines person as an individual, body corporate, organisation or group of persons.	Not satisfied. An AI system is not listed as a person, and the scenario contains no human actor directing the conduct.
Conduct	Intentional use of a computer system or network under section 23(1).	Not satisfied as a criminal act by a person. The AI uses a computer system, but no person performs the conduct in the relevant statutory sense.
Mental element	Intention under section 23(1). Nigerian criminal law assesses mens rea through natural persons or attributed corporate minds.	Not satisfied. The AI cannot form criminal intent, and no human deployer is assumed to have directed or known of the sequence.

Child definition	The child is under 18 under section 23(5).	Satisfied. The hypothetical child is 13.
Material	Child pornography as defined in section 23(4).	Potentially satisfied in part, depending on the content introduced. This element cannot carry liability without an actor and mental element.

Section 23(1) of the Cybercrimes Act punishes any person who intentionally uses a computer system for child pornography offences. Section 24 requires conduct done knowingly or intentionally. Section 58 defines a person as including individuals and bodies corporate, and corporate liability still acts through human attribution (cybercrimes2015). The Criminal Code confirms the same anchor: section 7 fixes principal offenders, while section 24 excludes responsibility for acts independent of human will, admitting negligence only where expressly provided (criminalcode2004). The Child Rights Act uses the same architecture for sexual offences (childrights2003).

Section 36(12) of the Constitution completes the constraint: no person can be convicted of a criminal offence unless that offence is defined in written law (constitution1999). That principle prevents a court from reading autonomous AI systems into criminal provisions after the fact.

Material Child pornography as defined in

The child definition and material elements may be present, but they cannot carry liability alone. The actor and mental elements fail, and without those, no offence is constituted.
The corporate liability provision in section 43 does not rescue the result. Section 43(2)(a) extends liability to a director, manager, secretary or similar officer where an offence was committed on that officer's instigation, with their connivance, or attributable to their neglect (cybercrimes2015).

That provision depends on a predicate offence having been committed by the body corporate. If no person committed the section 23 offence, there is no predicate offence to attribute to corporate officers. Section 36(12) prevents a court from creating that offence by interpretation (constitution1999).

l Agentic AI removes the human from the point of commission

Agentic systems pair autonomous planning with tool use and independent execution (staufer2026; loring2025). A deployer who intends or knows is not the hard case. Existing law can treat the system as an instrument and attach liability to the human who uses it. The gap begins where intention and knowledge end. Where a deployer neither directs nor knows of the sequence, no person satisfies the fault element of the offence, and the Cybercrimes Act offers no recklessness or negligence fallback for prosecutors.
The precise wrong to target is therefore not the machine's act. The target should be the human choice to deploy or make available an unsafe autonomous system in child-accessible environments without prescribed safeguards.

2 The framework fails in three ways: no liability, no preventive duty, no institutional owner

Only data protection law clearly binds a deployer without proof of intent. Processing a child's data without parental consent can breach the Nigeria Data Protection Act irrespective of system behaviour (ndpa2023). But that route recasts exploitation as a processing violation. Its damages remedy is victim-initiated and its vicarious liability machinery has not yet been tested against an autonomous system.

The NDPC's 2025 General Application and Implementation Directive can order risky emerging technologies out of use, making it the strongest preventive lever currently in force (ndpcgaid2025). Yet it has not been converted into a specific child-facing AI safety instrument. Pending bills repeat the pattern. HB 244 imposes a levy on AI companies to fund child protection but does not impose a corresponding AI-system safety duty (hb2442023). The Digital Economy Bill builds AI powers without an agentic AI definition and without a child-specific high-risk category (digitaleconomybill2025).

3 Evidence and Context

Table 3: Institutional responsibility map

Institution	Current mandate relevant to the problem	Current limitation	Policy role proposed in this paper
Nigeria Data Protection Commission	Data protection supervision, directives, compliance audit returns, enforcement orders and sanctions.	Strongest existing preventive lever, but not yet directed at child-facing agentic AI.	Issue a Child-Safe Agentic AI Directive under existing powers.
National Assembly	Legislative power over cybercrime, child online safety and digital governance.	Current and pending instruments do not define agentic AI or create deployer liability for unsafe autonomous systems.	Amend the Cybercrimes Act and HB 244.
National Information Technology Development Agency	Digital governance and AI-related regulatory powers once the Digital Economy Bill is assented.	Pending framework lacks a child-facing high-risk AI category.	Classify autonomous systems capable of interaction with under-18s as high-risk.
Law enforcement and child-protection institutions	Cybercrime investigation, child-protection referral, victim support and forensic capacity.	Current response assumes an identifiable human perpetrator.	Establish a single referral pathway and AI-mediated case protocol.

The wider trajectory is steep. NCMEC recorded online enticement reports rising by 192% in 2024, to more than 546,000 reports. It also recorded a 1,325% rise in reports involving generative AI, from 4,700 to 67,000, while warning that generative AI can simulate explicit chat with a child and supply grooming guidance (ncmec2025). The figures are not perfectly comparable across years after report bundling, but the direction of risk is clear.

In January 2026, US lawsuits alleging AI chatbots sexually abused minors settled in principle; the operator had already withdrawn open-ended chat for under-18 users. The relevance for Nigeria is not the foreign litigation itself, but the design lesson: minor-facing autonomy is a product choice that can be governed (garcia2024).

4 Policy Implications

If nothing changes, the first Nigerian case could arrive as a scandal without a defendant. Police may find no chargeable offence, the NDPC may be left to process a child's exploitation as a data breach, and a family may be left with civil proceedings they must fund themselves.

The opportunity is concrete. HB 244 remains an available legislative vehicle. The Digital Economy Bill can still be operationalised through child-facing high-risk classification after assent. The NDPC does not need new primary legislation to issue a directive that prevents risky child facing deployment. Nigeria can convert the African Union's commitment to safe AI into a model for regional regulatory practice instead of importing one later (auai2024; auchildsafety2024).

5 Policy Recommendations

5.1 Recommendation 1: Issue a Child-Safe Agentic AI Directive

The Nigeria Data Protection Commission should issue a Child-Safe Agentic AI Directive within six months, with a consultation draft inside 90 days. The directive should require a child-specific impact assessment before any system capable of unsupervised interaction with under-18s is deployed in Nigeria. It should also mandate age assurance, minimum human-oversight thresholds and incident reporting.

The Commission needs no new statute for this first step. The NDPA empowers the NDPC to issue directives and enforce compliance, and the GAID already contains a framework for data protection impact assessments and risky emerging technologies (ndpa2023; ndpcgaid2025). UNICEF's Guidance on AI and Children and the OECD AI Recommendation supply relevant child-rights and human-oversight principles (unicef2025; oecd2019).

5.2 Recommendation 2: Create a safeguard-based deployer offence

The National Assembly should amend the Cybercrimes Act to define agentic AI as a system that plans, uses tools and acts without contemporaneous human direction. It should then create a deployer offence for deploying or making available in Nigeria a system capable of unsupervised interaction with children without prescribed safeguards. The offence should include a due-diligence defence.

The amendment should be inserted after section 24 of the Cybercrimes Act as a new section 24A or as a new Part X-A. This would preserve the existing offence structure while creating a constitutionally certain offence focused on a defined human choice. Comparative AI regulation, including high-risk and prohibited-practice design in the EU AI Act, provides a drafting reference point (euai2024).

5.3 Recommendation 3: Amend HB 244 before third reading

HB 244 should be amended in the Senate before third reading. Its definitions should extend to conduct and material generated by autonomous means. Its proposed E-Commissioner should be empowered to set standards for AI agents. Its 0.02 per cent levy on AI companies should be matched with corresponding child-safety duties (hb2442023).

These amendments would align the bill with the African Union's child online safety policy and would allow removal machinery to function without waiting for an identified human perpetrator (auchildsafety2024).

5.4 Recommendation 4: Classify child-facing autonomous systems as high-risk

On presidential assent of the National Digital Economy and E-Governance Bill, NITDA should classify AI systems capable of autonomous interaction with under-18s as high-risk in the first implementing regulations. High-risk status should carry pre-deployment child-safety testing, age assurance, registration and human oversight, enforced through administrative sanctions and market-access conditions (digitaleconomybill2025).

Because the Bill takes supremacy in digital matters, a NITDA-NDPC memorandum of understanding should allocate supervision and prevent duplicative regulation.

5.5 Recommendation 5: Establish an inter-agency protocol

Law-enforcement and child-protection institutions should establish an inter-agency protocol for AI-facilitated child sexual exploitation within six months. The Office of the National Security Adviser should convene the process as statutory cybercrime coordinator under section 41 of the Cybercrimes Act, with oversight through the Cybercrime Advisory Council (cybercrimes2015).

The protocol should create a single referral pathway, activate NAPTIP victim support where relevant without waiting for a conviction, and build forensic capability for AI-mediated cases, including logs, outputs and provenance (naptip2015; weprotect).

Table 4: Implementation matrix for the recommended reform package

Table 4: Implementation matrix for the recommended reform package

Lead actor	Action	Timeline	Accountability output
NDPC	Publish consultation draft for a Child-Safe Agentic AI Directive covering child-specific impact assessment, age assurance, human oversight and incident reporting.	Ninety days	Consultation draft and published stakeholder feedback.
NDPC	Issue final directive and integrate disclosure into compliance audit returns.	Six months	Final directive, compliance return field and enforcement guidance.

Lead actor	Action	Timeline	Accountability output
National Assembly	Insert agentic AI definition and deployer offence into the Cybercrimes Act.	Twelve months	Cybercrimes amendment bill or consolidated committee amendment.
National Assembly	Amend HB 244 to cover autonomous generation, AI-agent standards and duties corresponding to AI-company levies.	Before third reading	Revised clauses and committee report.
NITDA	Classify child-facing autonomous systems as high-risk after Digital Economy Bill assent.	Six months after assent	Implementing regulation and public high-risk category guidance.
NITDA and NDPC	Sign a memorandum of understanding allocating supervision between data protection and digital governance mandates.	Six months after assent	Published supervisory protocol.
ONSA, police, NAPTIP and child-protection actors	Establish AI-facilitated child sexual exploitation referral and forensic protocol.	Six months	Inter-agency protocol and annual report on referrals, outcomes and capability gaps.

Cross-Cutting Relevance

The human-actor assumption exposed in this paper is not confined to child protection. Similar legal architectures underpin Nigerian law on fraud, harassment and electoral manipulation, and criminal law across many African jurisdictions rests on the same foundation. Child protection is the sharpest test because the harm is grave, the treaty obligation is clear, and the policy consensus is strong.

Reforms proven here can transfer outward. The NDPC directive could inform ECOWAS data-protection harmonisation; the high-risk classification model could provide an early national implementation of the African Union's continental commitment to safe AI (ecowas2010)



About the Author

Agba Frances Chinaza is a computer science educator working on AI governance and child online protection research in Nigeria. She is also a 2026 Digispace Africa Fellow in the Trust & Safety track.

Conclusion

Nigerian law assumes a knowing human behind every act of child exploitation. Agentic AI has removed that human from the point of commission, and no instrument in force or pending fills the space. The repair requires no doctrinal revolution: a directive under existing powers, an offence plus targeted clauses in a bill before the Senate, and a child-facing high-risk classification once the Digital Economy Bill becomes law.
A framework written for human offenders should not wait to meet its first autonomous one.

Methodology Note on AI Use

In line with the Fellowship's Research Integrity Policy, the author discloses minimal use of Anara AI in preparing this paper. The tool was used for resource identification and verification under the author's direction through staged review of each source. All sources were verified against primary or official texts before inclusion, and unverifiable material was excluded.

References

[1] L. Staufer and others, "The 2025 AI Agent Index" (2026) arXiv:2602.17753, preprint. Available at: https://aiagentindex.mit.edu.
[2] J. M. Loring, "When AI Acts Independently: Legal Considerations for Agentic AI Systems" (Jones Walker LLP, 2025).
[3] Nigerian Communications Commission, Study on Young Children and Digital Technology: A Survey Across Nigeria (NCC Research and Development Department, 2021).
[4] Convention on the Rights of the Child, adopted 20 November 1989, entered into force 2 September 1990, 1577 UNTS 3, ratified by Nigeria in 1991.
[5] UN Committee on the Rights of the Child, General Comment No. 25 (2021) on Children's Rights in Relation to the Digital Environment (2 March 2021) CRC/C/GC/25, para. 23 and section VII.
[6] Cybercrimes (Prohibition, Prevention, etc.) Act 2015 (Nigeria), ss. 23(1), 24, 41, 43 and 58, as amended by the Cybercrime (Prohibition, Prevention, etc.) (Amendment) Act 2024.
[7] J. O. Ezeanokwasa, "Child Pornography under the Cybercrimes Act 2015 of Nigeria: The Law and its Challenges" (2019).
[8] Criminal Code Act, Cap. C38, Laws of the Federation of Nigeria 2004, ss. 7 and 24.
[9] Child Rights Act 2003 (Nigeria), ss. 25 and 31–33.

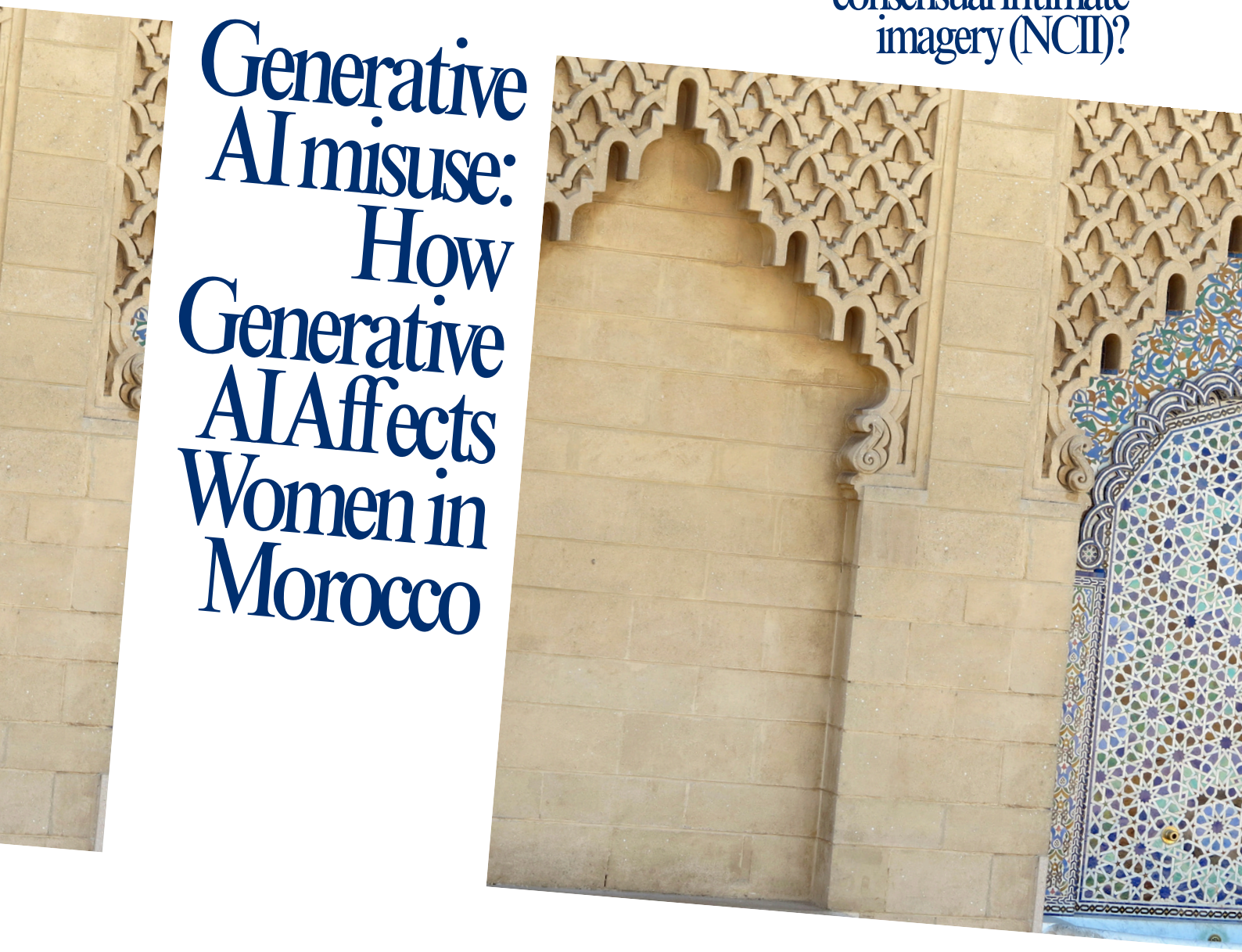
[10] Constitution of the Federal Republic of Nigeria 1999, ss. 36(12) and 58. [11] Nigeria Data Protection Act 2023, ss. 6(c), 31, 37, 39, 46–49, 51, 53, 61 and 62.
[12] Nigeria Data Protection Commission, General Application and Implementation Directive, NDPC/NDP ACT-GAID/01/2025, arts. 28, 43 and 44, sch. 6.
[13] Child Online Access Protection Bill 2023, HB 244, passed by the House of Representatives on 3 December 2025 and pending before the Senate, s. 3(2).
[14] National Digital Economy and E-Governance Bill 2025, passed by the National Assembly in November 2025 and awaiting presidential assent, ss. 62 and 63.
[15] Nigeria Data Protection Commission and Global Privacy Assembly International En forcement Cooperation Working Group, "Joint Statement on AI-Generated Imagery" (23 February 2026). Available at: https://ndpc.gov.ng.
[16] Gatefield, State of Online Harms in Nigeria (October 2025). Available at: https:// gatefield.co.
[17] Gatefield, analysis of AI-facilitated abuse and Nigerian law (February 2026). 10
No Human, No Criminal Liability? Policy Paper
[18] National Center for Missing and Exploited Children, CyberTipline 2024 Report Data (2025). Available at: https://www.missingkids.org/cybertiplinedata.
[19] Garcia v Character Technologies Inc, No. 6:24-cv-01903, United States District Court for the Middle District of Florida, filed October 2024, dismissal denied in part 21 May 2025. Settlements in principle reported January 2026.
[20] African Union, Continental Artificial Intelligence Strategy (July 2024). Available at: https: //au.int.
[21] UNICEF Innocenti, Guidance on AI and Children 3.0 (December 2025). Available at: https://www.unicef.org/innocenti.
[22] OECD, Recommendation of the Council on Artificial Intelligence, OECD/LEGAL/0449 (2019, amended 2024).
[23] Regulation (EU) 2024/1689 laying down harmonised rules on artificial intelligence, arts. 5 and 6, annex III; Digital Omnibus on AI, provisional agreement of 7 May 2026.
[24] African Union, Child Online Safety and Empowerment Policy, endorsed by the Executive Council, February 2024. Available at: https://au.int.
[25] Trafficking in Persons (Prohibition) Enforcement and Administration Act 2015 (Nigeria).
[26] WeProtect Global Alliance, Model National Response. Available at: https://www. weprotect.org.
[27] ECOWAS, Supplementary Act A/SA.1/01/10 on Personal Data Protection (2010).
[28] T. King, N. Aggarwal, M. Taddeo and L. Floridi, "AI Crime: Big Data, Machine Learning, and the Future of Criminal Law", in L. Floridi (ed.), Law, Justice and Artificial Intelligence (Springer, 2019).
[29] A. Zavrsnik, "Artificial Intelligence, Automation, and Criminal Justice: Challenges and Opportunities" (2020) 19 Law, Innovation and Technology 313.
[30] S. Mukherjee and J. Chang, "Operational Agency: A Framework for Tracing Culpability Through Artificial Intelligence" (2026). Journal and volume details to be confirmed.
[31] A. Anggara, "Can AI Be a Criminal Subject? A Comparative Study of Indonesia and the European Union" (2026). Journal and volume details to be confirmed.
[32] K. Danaher, "The Threat of Artificial Child Pornography" (2014) 35 Loyola of Los Angeles Law Review 673.
[33] O. Folorunsho and A. Surajudeen, "Protecting Children's Rights Against AI-Driven Technologies: Gaps in Nigeria's Legal Framework" (2026). Journal and volume details to be confirmed.
11
No Human, No Criminal Liability? Policy Paper
[34] A. Eke-Awa, "Towards a Dedicated AI Act for Nigeria: Mapping Regulatory Gaps Against Comparative Models" (2026). Journal and volume details to be confirmed.
[35] Violence Against Persons (Prohibition) Act 2015 (Nigeria).
[36] Violence Against Persons (Prohibition) Act 2015 (FCT), s. 53.
[37] Federal Republic of Nigeria, National Artificial Intelligence Strategy (NAIS) 2024. 12

Executive Summary

This research investigates the policy gap in Morocco, compared to international standards, regarding AI-generated non-consensual intimate imagery (NCII), with focus on documented harm to victims and elevated risk especially for Women.

Generative AI misuse: How Generative AI Affects Women in Morocco

How are the existing policies and laws in Morocco protecting women from AI-generated non-consensual intimate imagery (NCII)?



The project is novel in its focus on women in Morocco as a key vulnerable group, and in its focus on the actual existing “Nudify” and deepfake applications that are easily found online, how these applications that use Generative AI harm these women and what kind of policy and laws in Morocco exist to protect from this harm.

This work aims to summarise the current existing global literature on generative AI usage and highlights the role of technologies in amplifying crimes against women, with Morocco as a study case. The Moroccan perspective will be shown through existing documentation policy gaps in protection against similar digital risks as well as an organized focus group that would discuss the direct harm of Generative AI in Moroccan society.

Problem Statement & Context

With the rise of Generative AI uses across the world, the creation and spread of NCII has become increasingly easier through personalised “Nudify” Apps. These applications that are one search away from the public, utilises Generative AI to generate inappropriate and harmful material based on the likeness of whoever you tell it to. Unfortunately, their target victims often being women and children, this misuse has great effects on a victim’s mental health, social standing and with these tools existing in a vague legal definition.

These tools are a product of what we call generative AI, an artificial intelligence system that uses complex algorithms and learning models to identify patterns and take from existing data to generate new data. Meanwhile, my focus will be on the "Nudify" Apps, a tool where a person uploads a person's image, and this "AI" algorithm undresses this person to create an inappropriate image of them as per the user's intentions. This type of material is nothing new as bad actors previously utilized available editing software to create NCII which required more technical knowledge, unlike the "deepfake" material which uses Artificial Intelligence allowing a much lower barrier to entry for anyone to exploit it for malicious intentions.

These harmful applications are commonly free, or one new account creation away while some have premium features or require payment for better quality videos or imagery. Though such platforms that constantly change their naming convention from Nudify, undressing, or clothes remover or more to avoid attention are meant to be user friendly and prioritize ease of use and discretion for their users. Because of these applications, women and young girls often end up becoming victims of sextortion, revenge porn or child sexual exploitation.

One of the reasons of why NCII are a major issue for women and why making its creation easier is exacerbating its harm; is because countries like Morocco, have a majority Muslim society and is inherently conservative in its culture and very community centred, thus such fabricated images are weapons to be used against members of the society since everyone wants to uphold their conservative values and social reputation within their families and communities. Exploiting victims who have very limited legal protections set up for such emerging technologies, is the unfortunate result of the widespread usage of such applications combined with the policy gaps.

As Morocco cyber policies are progressing very slowly compared to such cyber threats while also being a country that is constantly struggling with higher gendered violence and harassment in the cyber world in recent years. This research is meant to document and highlight the current existing policies and legal protections and their real-life implication for Moroccan victims. While also shedding lights on how other governments handle such similar cases by studying relevant legislative models proposed by the EU.

Research Questions & Objectives

- **Primary Question:** How are the existing policies and laws in Morocco protecting women from AI-generated non-consensual intimate imagery (NCII)?
- **Secondary Questions:** What is the impact of these GenAI tools in Moroccan society and what are some recommended mitigation strategies from a legal and platform perspective?

Methodology

This research uses a mixed-methodology approach, including an extensive desk-based policy analysis section for better understanding of what are NCII and the risk behind Generative AI.
Another primary component of this research is a focus group accompanied by a supplementary online survey, designed to gather qualitative and nuanced views of the misuse of generative AI against women in Morocco. To ensure the study is focused on the group that is disproportionally affected by image-based sexual abuse according to previous literature, the focus group was a 7:1 female to male ration, specially engaging women between the ages of 25and 50. The participants were chosen within the city of Casablanca, a major metropolitan area to ensure a baseline level of education and digital literacy necessary for addressing the topic at hand. All participants have passed the Moroccan baccalaureate exam but with varying higher education levels and professional backgrounds from IT, Teaching, Construction. The supplementary online survey was open for 10 days and gathered 7 responses, with a 3:4 male to female ratio and all contributors being between the ages of 25 to 43, concentrating on noting the level of awareness and understanding of deepfakes and was a precursor to the focus group to establish the baseline knowledge of the topic, among that age group.

Table 1: Research Methodology Summary Table

Component	Details
Research Design	Mixed-methodology approach integrating comparative policy analysis with primary qualitative field research.
Geographic Focus	Casablanca, Morocco, selected as a metropolitan hub to ensure a baseline of digital literacy among participants.
Data Sources	Primary: 1 hour Focus group discussion and online survey (initial N=7). Secondary: Moroccan Penal Code 1962, and the 2028 amendment, along international AI policy frameworks and relevant written reporting.
Participant Profile	Women-centered sample, aged 25–50, all having attained a minimum of a Baccalaureate education.
Analytical Tools	focus group 1 hour archived recording, Google Forms for online survey and response tracking and Google's NotebookLM (now Gemini Notebook) for bibliography management.

Limitations & Ethical Considerations

- Due to the limited time for this research, there might be a lack of newer development of AI policies in Morocco especially given the significance of the upcoming election of 2026. And due to the specificity of the topic itself there might be a lack of relevant existing literature.
- Due to time constraints, there might be a lack of representation for certain ages, political views, gender bias. This would result in a skewed conclusion towards a certain solution, but details of the group will be transparent and will be acknowledged throughout the writing.
- Gemini Notebook is an AI-powered tool that was used by the source organisation for this research to store the translations of sources from Arabic and French for a more comprehensive reporting and was not used to generate any content.
- Due to privacy concerns, the focus group discussion was recorded and archived locally without any names, like the survey which did not collect any unnecessary data and allowed participants anonymity for their protection and comfort. There was no overlapping sample for both survey and focus group, to gather a wider range of opinions and perspectives.

Key Findings

- **Morocco's primary law on violence against women fails to take into consideration the risks of emerging technologies.**
  - Morocco penal code of 1962 provides the general framework of protections to women by punishing any violent acts, the punishment was denoted as being light, which lead to the later adjustment in the 2011 Moroccan constitution that presented harsher and more specific protections against sexual violence but was still very lacking when it comes to gendered violence. It was not until 2018 where a new amendment was sealed by the king to more specifically introduce psychological violence and economical violence to equal penalties as the previously mentioned crimes. Especially when it comes to gendered violence against women.
  - Morocco has recently introduced the E-blagh to assist in reporting cybercrimes which definitely is a positive step in facilitating the reporting, but also lacks taking into consideration the digital literacy level of the average Morocco and also the amount of stress a victim would be under, as reporting would entail gathering evidence and submitting it through the portal, and it comes to crimes that involved inappropriate material, victims would not want to share such material further.

- **These harmful tools are so readily available online with no technical barrier to misuse them.**
  - A simple search of these Generative AI tools in a regular search engine or the built-in app stores, gives a variety of options of products. These applications (which will not be named as to not market them) according to Tech Transparency Project investigative research, were downloaded more than 705 million times, with 55 were found in the Google Play store and 47 applications found in the Apple store. While these numbers are huge already, with the rise of integrated Generative AI in already existing and popular platforms such as Grok, we are now lowering the entry barrier even lower and making it more difficult to track and document.

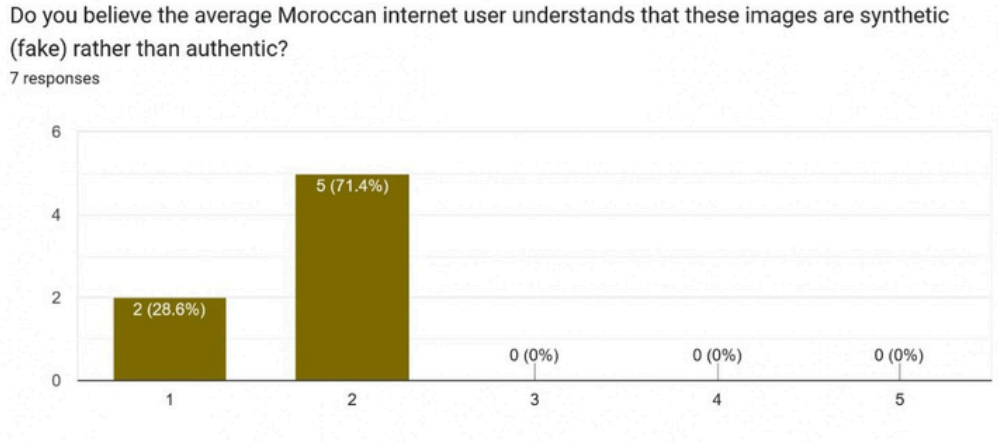
Generative AI misuse

By Hanaa Dany

- According to the survey and confirmed by the focus group, many are not aware of the existence of such applications, even the few who were aware of the existence of deepfake material. The consensus was that if you do not seek them out , you will not be able to stumble upon them, which is confirmed given how such applications change what searchable word is to refer to them, from existing articles in 2025 the term was “Nudify” but now it is different, which means there is a conscious effort from their end to operate under the radar of the general public.

- **Due to the conservative society in Morocco, women are less likely to report such crimes that may include inappropriate imagery.**
  - Fearing for their social standing and reputation, most women tend to not report these crimes, especially given the existence of Morality laws such as Article 490, which criminalises extramarital relations, as they are afraid that the police will assume that such material was made consensually, and that is confirmed by data collected as many confirmed that just like them it is very hard to distinguish between real and fake material, as seen by the figure below.

Table 2: Moroccan abilities to distinguish real from fake/synthetic material, with 1-being not confident to 5-being very confident.



- While such Generative AI based NCII are not any evidence against women themselves but rather because of the existing system, it is very discouraging for victims to come forward especially from low or middle-income citizens, who cannot afford lawyers or just the continuous follow ups for their cases to be held seriously. Given how legal cases require active follow up which can lead to exorbitant fees and potential from absence from their work. Perpetrators could potentially threaten these victims into silence, by citing these laws thus creating an inescapable situation for these powerless victims.
- The data also confirms this finding as most agreed that women are the main target of such crimes or just such creation. The discussion was divided between those adamant that women even by just existing on the internet are main targets for exploitation, while a few highlighted that only public figures or those who have actual enemies, which according to their statements, as long as you are staying conscious about what you post and who you share with, you are likely not to be the target of such crimes.
- Morocco has a reactive legal procedure compared to the proactive policy like the EU AI Act.
- The policy analysis shows that Morocco still lacks when it comes to legislative frameworks concerning misuse of emerging technologies including AI. Rather its approach is reactive based on individual cases of specific harm according to the penal code of 2018 that takes into consideration psychological harm. In contrast, the EU AI Act takes a more proactive approach as it focuses on platform accountability and eliminating the risk itself, rather than the misuse or addressing the harm. Furthermore, the Moroccan penal code that could be used in sextortion cases that include misuse of Generative AI still places the burden of proof on the victims, needing to further share the material to even more preying eyes and opening them to further scrutiny in their communities.
- There is a clear difference in how the average Moroccan also views such crimes as according to the survey and group discussion, participants agreed that accountability falls on the user and not the platforms citing how the internet and the real world has plenty of double edged swords, but it is up to the individual to choose not to harm, which puts the Moroccan penal code in better light for the Moroccan public, as they agree that punishment should be heavy for the criminal specially since the platform is not based in Morocco and thus such harmful tools are just a result of globalisation. Further adding that enforcing existing laws would be enough as well as adding a more respectful way of reporting such accesses without furthering the harm along with raising awareness about the proper reporting channels and actual existence of such threats and legal protection to encourage victims to pursue the legal route.

Supporting Evidence / Cases

- Grok and X (Twitter): In late 2025, the integrated AI chatbot in the X platform. generated inappropriate sexualised images of real women on the platform, through one simple prompt.
- Ethnographic study of gender-based violence in Morocco: Through the various stories shared by the INSAT association in Morocco, shows how unsupported women and young girls, who are faced with sexual and physical violence. These young women despite many seeking help and some never getting the justice they deserve.
- **Anecdote from the focus group**: A participant shared a case of deepfake sextortion that happened in the city of Marrakech, where a bad actor used a published video of a group of young women filmed dancing during a wedding to generate an obscene video and blackmail them, with one girl being a minor at that time, though this case was not publicly shared, thus the outcome of their case was still unknown. With the only outlier was news that one of the girls was excommunicated from her family. This story's authenticity remains uncertain as it was only shared through word of month and there is no documentation to verify its existence.

Supporting Evidence / Cases

These documented cases show how lacking the legal system is and how difficult it is to seek help and report cases, especially when there is a power dynamic or in our case study, when the criminal could more likely be anonymous behind a keyboard. Making the investigation and prosecution significantly harder and more expensive for the victims.

Implications

Without proper safeguards be it; through legal reforms, platform reporting and takedowns or societal awareness of such crimes. Women in societies like Morocco, will continue to suffer in silence, not only being victim to a perpetrator online, but victims to the existing neglectful system. And the harm that such cybercrimes can no longer be taken lightly as now we have understood that cybercrimes are a danger to one's life through various real life unfortunate cases.

Despite the progressive steps Morocco has implemented with its online reporting of cybercrimes, there is a huge gap when it comes to enforcing the punishment, as reporting a crime often leads to nothing without active push from either a legal representative which is a luxury for many, or a push back from the general when a case is publicized on social media, which in our cases could cause further harm to the victims involved.

Policy-Relevant Recommendations

Target Actor: Moroccan Ministry of Justice
Recommendation: Criminalise the creation and possession of AI Generated NCII

Without proper safeguards be it; through legal reforms, platform reporting and takedowns or societal awareness of such crimes. Women in societies like Morocco, will continue to suffer in silence, not only being victim to a perpetrator online, but victims to the existing neglectful system. And the harm that such cybercrimes can no longer be taken lightly as now we have understood that cybercrimes are a danger to one's life through various real life unfortunate cases.

Despite the progressive steps Morocco has implemented with its online reporting of cybercrimes, there is a huge gap when it comes to enforcing the punishment, as reporting a crime often leads to nothing without active push from either a legal representative which is a luxury for many, or a push back from the general when a case is publicized on social media, which in our cases could cause further harm to the victims involved.

Rationale:

Rather than criminalizing the malicious distribution of such material and its usage for other crimes like blackmail and sextortion and thus having to prove those crimes occurred to seek justice. The punishment should include the actual creation and possession of the non-consensual material, that way we can hold the bad actor accountable given how limited the control on the platform we have.

Target Actor: National Digital Content Reporting Portal (E-Blagh) and Law Enforcement
Recommendation: Implement a privacy focused reporting protocol for the victims for law enforcement
Rationale:

According to the finding, the main barrier to reporting such crimes is the lack of trust in the authorities' invasive procedures. Victims have a fear of being blamed or further scrutinized because of the nature of crime, due to the harmful societal shame. The protection of the victim's dignity is the key to not only a productive investigation and follow up on the case with victims feeling safe means they will be more communicative but also will raise the trust in the legal system and encourage more victims to pursue their cases and seek their much-deserved justice.

Target Actor: App Stores (Apple/Google), and Search Engines
Recommendation: Banning "Nudify" applications and de-indexing them from the search results.
Rationale:







Such applications and platform are not hiding their purpose as their taglines and websites says it all. Thus, it is seemingly not hard to auto ban them knowing the harm they cause not only in Morocco but globally. The platforms allowing such applications to exist should be held accountable. While the applications prioritize ease of use of their product, it is the search engine and app stores that make it easy to find and more accessible to morally corrupt actors with malicious intent.

Cross-Cutting Insight

This research shows that Ai-generated non-consensual intimate imagery NCII is the imminent results of what happens when profit driven technological innovation moves at a faster pace than the relevant regulations. Leaving regulators and policy makes always behind when it comes to the new emerging risks. As digital literacy is not lacking when it comes to the general Moroccan public but the actual policy makers in the country.
It is evident that the Moroccan system is making its own baby steps to raising awareness about cybercrime.

However, the fact remains that the lack of support and reform is a barrier to such crimes being reported, and the low number of reported crimes only shows the legislators that this issue is not a priority, focusing their efforts on more widely reported crimes like phishing and cyber fraud which also targets enterprises.

It is also critical to avoid perpetuating the same toxic social scrutiny and shame when it comes to handling such crimes, with Morocco's clear gendered violence problem, this is an issue that would not only help when it comes to addressing NCII but also help protect the Moroccan woman overall and her dignity. By enforcing privacy focused reporting procedure and sensible investigation as to protect the victims and more harm.

Overall, this research concludes that protection from the misuse of generative AI against women in Morocco is a human rights imperative and not just a digital right matter. With the Moroccan society understanding the heavy weight of such crimes and agreeing that the perpetrators are the ones to blame, but because of the limited ability of distinguishing fake material and disinformation from the real story, it remains a huge challenge for victims to come forward with their stories without having to relive the trauma or being prosecuted themselves.

Such applications and platform are not hiding their purpose as their taglines and websites says it all. Thus, it is seemingly not hard to auto ban them knowing the harm they cause not only in Morocco but globally. The platforms allowing such applications to exist should be held accountable. While the applications prioritize ease of use of their product, it is the search engine and app stores that make it easy to find and more accessible to morally corrupt actors with malicious intent.

Author Bio

Hanaa Dany is an information security analyst based in Casablanca, Morocco, bringing practical expertise in information assurance, risk assessments, threat hunting, and digital safety education. With professional credentials including CompTIA Security+ and ISC2 Certified in Cybersecurity (CC), she has established herself as a dedicated cybersecurity professional committed to advocating for Digital Rights in her community. Currently working as Teaching Assistant for a Cybersecurity professional learning program while focusing on her work as a researcher investigating policy protections against AI-generated non-consensual intimate imagery (NCII) in Morocco as a case study.

References:

Global AI Risks and Safeguards
European Parliament. (2025). EU AI Act: First regulation on artificial intelligence.

https://www.europarl.europa.eu/topics/en/article/20230601STO93804/eu-ai-act-first-regulation-on-artificial-intelligence
BNP Paribas. (2026). Generative AI: A legal framework in development. https://group.bnpparibas/en/news/generative-ai-a-legal-framework-in-development

Microsoft. (2026, May). Microsoft AI diffusion report: Global AI diffusion - Q1 2026 trends and insights. AI Economy Institute.
https://www.microsoft.com/en-us/research/wp-content/uploads/2026/05/Microsoft-AI-Diffusion-Report-2026-Q1.pdf
Global Voices. (2026, April 28). Brazil: A warning on how AI and deepfakes can become an 'excessive risk' to women and girls.

https://globalvoices.org/2026/04/28/brazil-a-warning-on-how-ai-and-deepfakes-can-become-an-excessive-risk-to-women-and-girls/
International AI Safety Panel. (2026). International AI Safety Report 2026 examines AI capabilities, risks, and safeguards. Inside Privacy.
https://www.insideprivacy.com/artificial-intelligence/international-ai-safety-report-2026-examines-ai-capabilities-risks-and-safeguards/
Internet Watch Foundation (IWF). (2026). Harm without limits: AI child sexual abuse material through the eyes of our Analysts.
https://www.iwf.org.uk/media/hl1nvdti/iwf-ai-csam-report-2026.pdf

NCBI. (2026). Prevalence of generative artificial intelligence sexualized image usage by adolescents in the United States. PubMed Central (PMC12998796). https://pmc.ncbi.nlm.nih.gov/articles/PMC12998796/

Nelson, A. (2025). What is a nudify app? What parents need to know. BrightCanary. https://www.brightcanary.io/nudify-apps/

Online Child Exploitation Prevention Initiative (OCEPI). (2024). Guidance for school-based professionals and school leaders.
https://icactaskforce.org/assets/genai%20fact%20sheet%20-%20bw.pdf

Pendergast, K. (2024). ‘Nudify’ apps, the evolution of online exploitation, and what we can do about it. Family Online Safety Institute.
https://fosi.org/nudify-apps-the-evolution-of-online-exploitation-and-what-we-can-do-about-it/

Sheikh, M., & Rogers, J. P. (2025). Addressing the psychiatric implications of AI-enabled non-consensual sexual imagery. The British Journal of Psychiatry. https://doi.org/10.1192/bjp.2025.xx

Tech Transparency Project. (2026). Nudify apps widely available in Apple and Google app stores.
https://www.techtransparencyproject.org/articles/nudify-apps-widely-available-in-apple-and-google-app-stores

Thornton, D. (2024). Nudify deepfakes. A Better Choice Network Solutions. https://www.abcnetfl.com/blog/nudify-deepfakes

UNICEF. (2026). ‘Deepfake abuse is abuse’ - UNICEF. https://www.unicef.org/press-releases/deepfake-abuse-is-abuse

United Nations Interregional Crime and Justice Research Institute (UNICRI). (2024). Generative AI: A new threat for online child sexual exploitation and abuse. https://unicri.org/sites/default/files/2024-09/Generative-AI-New-Threat-Online-Child-Abuse.pdf

Moroccan Context and Legal Framework
Kingdom of Morocco. (n.d.). E-Blagh: National digital content reporting portal. https://www.e-blagh.ma/reporting

2M. (2025, November 26). Violence au Maroc : quand la technologie exacerbe le fléau. 2M News. https://2m.ma/fr/news/Violence-au-Maroc-quand-la-technologie-exacerbe-le-fl%C3%A9au-20251126

Bakkass, N. (2025, May 21). Non encadrée, l'IA peut amplifier certaines violences comme les cyberharcèlements (expert). Le Matin. https://lematin.ma/societe/lia-peut-amplifier-certaines-violences-comme-les-cyberharcelements/280981

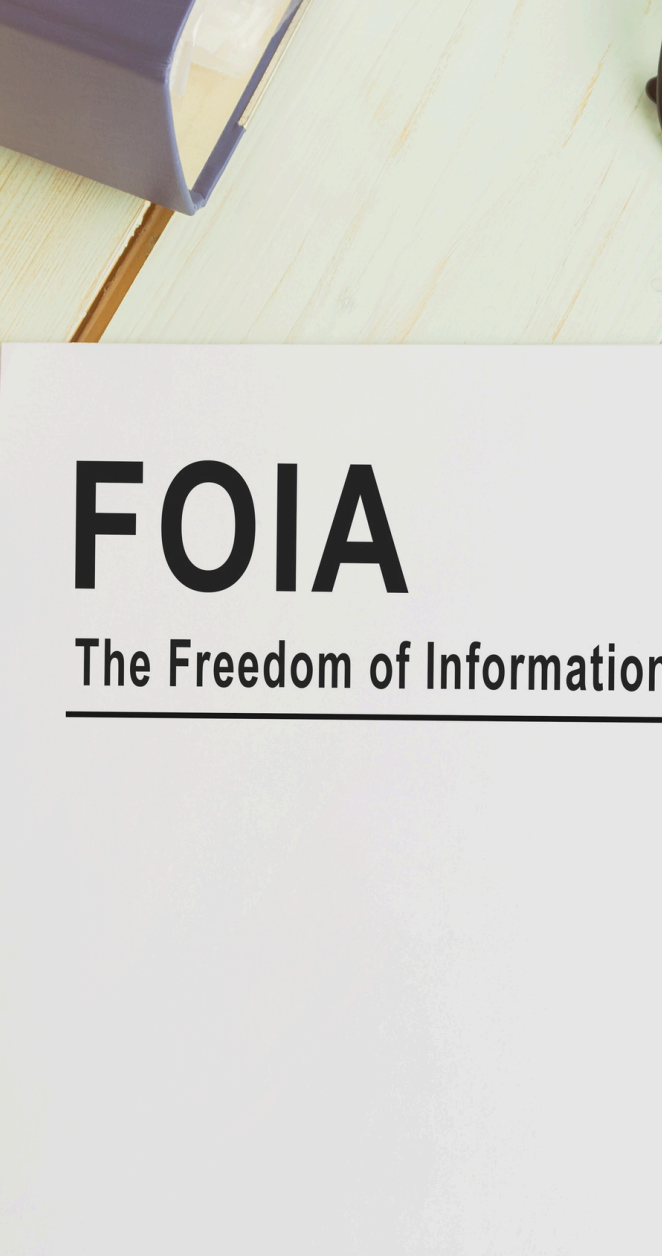
El Bouchtaoui, H. (2023). Moroccan MPs propose legislation to limit cyber violence, online harassment against women. Hespress English. https://en.hespress.com/77205-moroccan-mps-propose-legislation-to-limit-cyber-violence-online-harassment-against-women.html
Kingdom of Morocco. (1962, November 26). Penal Code [Code Pénal] (Promulgated by Dahir No. 1-59-413).

Kingdom of Morocco. (2018, February 22). Law No. 103-13 on Combating Violence against Women [Loi n° 103-13 relative à la lutte contre la violence à l'égard des femmes] (Promulgated by Dahir No. 1-18-19).

Rose Davis, A. (2026, April 15). Reports emphasize gaps in Morocco's violence against women law. Morocco World News. https://www.moroccoworldnews.com/2026/04/287299/reports-emphasize-gaps-in-moroccos-violence-against-women-law/

Tahiri, H. (2021). Innovation and women empowerment in Morocco: Case study of electronic crime law (103.13). International Journal of Research and Innovation in Social Science (IJRISS), 5(8), 588–594. https://rsisinternational.org/journals/ijriss/Digital-Library/volume-5-issue-8/588-594.pdf

Yacoubi, Y. (2025, November 24). تزييف عميق وهندسة اجتماعية .. جيل جديد من الاحتيال يستهدف المغاربة [Deepfake and social engineering .. a new generation of fraud targeting Moroccans]. Hespress. https://www.hespress.com/%D8%AA%D8%B2%D9%8A%D9%8A%D9%81-%D8%B9%D9%85%D9%8A%D9%82-%D9%88%D9%87%D9%86%D8%AF%D8%B3%D8%A9-%D8%A7%D8%AC%D8%AA%D9%85%D8%A7%D8%B9%D9%8A%D8%A9-%D8%AC%D9%8A%D9%84-%D8%AC%D8%AF%D9%8A%D8%AF-%D9%85-1659944.html



Linguistic Sovereignty and Digital Transparency: Using GenAI to Democratize Freedom of Information in Nigeria

Executive Summary

The Freedom of Information (FOI) Act of 2011 established an essential legislative foundation for democratic transparency, open governance, and public accountability in the Federal Republic of Nigeria. However, over a decade after its enactment, the practical realization of this law remains severely constrained by systemic institutional opacity, paper-based administrative workflows, and profound linguistic exclusion, in other words access to the FOI is not democratized.

Section 55 of the 1999 Constitution of Nigeria acknowledges major indigenous languages alongside English, official state documentation, legal gazettes, agency audits, and public procurement records are published almost exclusively in dense, technical English. In a multilingual nation where over 500 indigenous languages are spoken, this structural reliance on monolingual documentation creates a barrier to information access that marginalizes non-English-speaking citizens, rural communities, and informal economic actors.

This paper presents an end-to-end, policy-aligned technical architecture designed to operationalize Generative Artificial Intelligence (GenAI) and open-weights models within Nigerian public administration.

Hosted on Google Cloud Platform's Vertex AI ecosystem, the proposed framework establishes a secure, enterprise-grade ingestion, translation, grounding, and distribution pipeline. Unstructured paper documents are ingested and processed using Cloud Storage and Vertex AI Document AI, which leverages Optical Character Recognition (OCR) and layout parsing to convert physical records into structured digital formats.

To prevent algorithmic hallucinations and maintain strict legal fidelity, the extracted text is indexed using Vertex AI Vector Search within a Retrieval-Augmented Generation (RAG) architecture, grounding model responses in verified source documents.



"Celebrating the minds behind the world's most inspiring brands."

To address the low-resource classification of African languages in Natural Language Processing (NLP), I will use the platforms Parameter-Efficient Fine-Tuning (PEFT) with Low-Rank Adaptation (LoRA) on foundational Gemini models and open-weights Gemma endpoints, using

Hausa, Yoruba, Igbo, and Nigerian Pidgin for input . To make it accessible for citizens with various level of literacy, the pipeline would integrate Google's automated Speech-to-Text (STT) and Text-to-Speech (TTS) synthesis, distributing audio outputs via low-bandwidth web interfaces, USSD gateways, and interactive voice response (IVR) systems.

It will be embedded with an Afro centric AI inspired governance framework to enforce strict data sovereignty, this framework provides a blueprint for democratizing information rights and advancing sovereign digital governance across Africa.

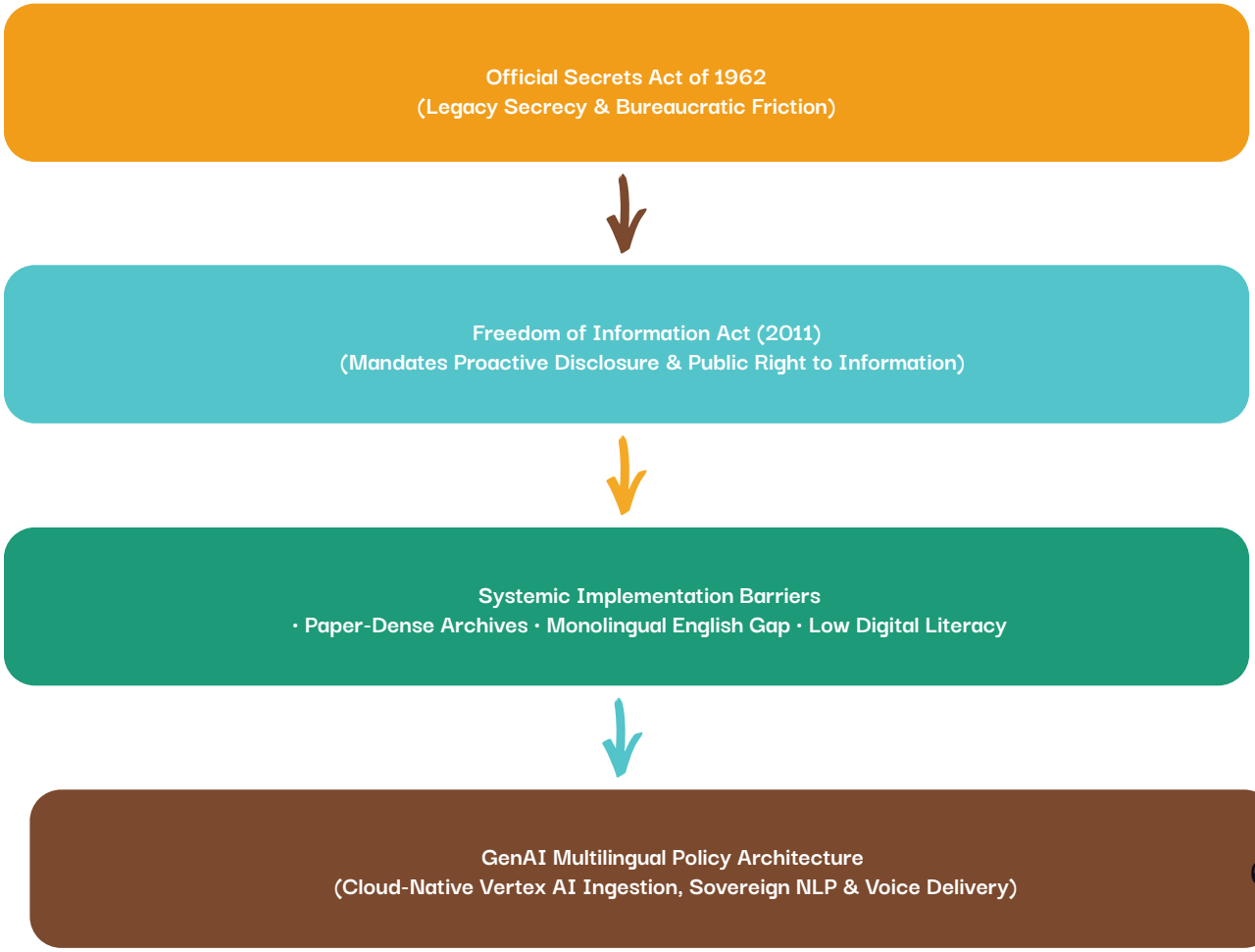
1. Introduction and Problem Statement

1.1 THE CONSTITUTIONAL AND HISTORICAL CONTEXT OF TRANSPARENCY IN

The evolution of democratic governance in Nigeria is defined by a tension between a progressive statutory commitment to openness and an inherited administrative culture of official secrecy. For nearly four decades following independence, the Nigerian public service operated under the shadow of authoritarian military decrees and colonial-era statutes. Chief among these was the Official Secrets Act of 1962, which classified virtually all government records as state property rather than a public trust. Under this regime, civil servants were legally bound to withhold administrative documentation, fostering an environment where corruption, inefficiency, and public distrust flourished.

The enactment of the Freedom of Information (FOI) Act in May 2011 was celebrated as a milestone for civil society, investigative journalism, and democratic reform. The Act explicitly confers upon every Nigerian citizen the legal right to access records held by public institutions, eliminating the requirement that an applicant prove a specific personal interest in the information requested. Furthermore, the Act mandates that public agencies proactively disclose routine administrative data, financial reports, and regulatory decisions.

However, the practical implementation of the 2011 Act has been severely obstructed . Most federal and state Ministries, departments, and agencies (MDAs) lack designated FOI units, trained records management personnel, or standardized digital archivism. Requests for information are routinely met with administrative delays, outright refusals, or legal challenges that invoke legacy statutes like the Official Secrets Act. Consequently, the promise of the FOIA remains largely unfulfilled for the average citizen.



1.2 The Monolingual Divide and Linguistic Exclusion

Beyond bureaucratic disruptions, the most pervasive barrier to information access in Nigeria is linguistic. Nigeria is an exceptionally diverse society, home to over 200 million people speaking more than 500 distinct indigenous languages. While Section 55 of the 1999 Constitution acknowledges Hausa, Yoruba, and Igbo as national languages alongside English, the practical business of governance including the publication of laws, agency budgets, policy directives, and FOI responses is conducted almost exclusively in formal English.

This reliance on monolingual English creates a structural divide. English in Nigeria functions as a socio-economic gatekeeper, spoken fluently by an educated urban elite. For tens of millions of citizens in rural peripheries and informal urban markets, complex legal and administrative English is inaccessible. When a government agency releases an environmental impact assessment, a local budget allocation, or a healthcare expenditure report in dense English legalese, the content remains functionally opaque to the communities most directly impacted by those disclosures.

This linguistic barrier transforms information access from a universal constitutional right into an exclusive privilege enjoyed by a small segment of the population. Without the capacity to comprehend public records in their primary language, citizens cannot meaningfully participate in civic discourse, monitor local government performance, or demand accountability from public officials.

1.3 THE PROMISE OF GENERATIVE AI AND SOVEREIGN NATURAL LANGUAGE PROCESSING

Addressing this linguistic divide through manual translation human workflows is financially and logistically unfeasible for the Nigerian public service. Translating thousands of daily government disclosures, gazettes, and audit documents into major regional languages would require an army of certified legal translators, imposing unsustainable budgetary costs and introducing delays that violate the statutory response deadlines set by the FOI Act.

Recent breakthroughs in Generative Artificial Intelligence (GenAI) and Natural Language Processing (NLP) offer a transformative alternative. Advanced Large Language Models (LLMs) possess sophisticated capabilities for automated document parsing, semantic translation, and contextual simplification. Unlike early rule-based machine translation systems, modern neural architectures can capture subtle linguistic nuances, idiomatic expressions, and domain-specific terminology.

However, applying commercial LLMs directly to public sector translation in the Global South presents distinct technical and ethical challenges. African languages including Hausa, Yoruba, Igbo, and Nigerian Pidgin are classified as "low-resource" languages in the global AI ecosystem. This classification reflects a lack of high-quality, digitized, and annotated training data in public repositories. Furthermore, deploying ungrounded LLMs risks generating "hallucinations actually incorrect or fabricated statements that could corrupt public records and undermine public trust in official disclosures.

To realize the potential of GenAI in public administration, Nigeria must deploy an enterprise-grade, sovereign technical architecture. This architecture must combine cloud infrastructure, Parameter-Efficient Fine-Tuning (PEFT), Retrieval-Augmented Generation (RAG), and speech synthesis to deliver accurate, localized, and verified FOI disclosures to every citizen.

2. Policy and Regulatory Framework Alignment

The deployment of AI-powered translation tools within public sector information workflows must not occur in a regulatory vacuum. It must align with existing national digital policy directives, statutory mandates, and regional governance standards.

NATIONAL & CONTINENTAL POLICY ALIGNMENT	
Policy Instrument	Core Strategic Mandate
Freedom of Information Act (2011)	Guarantees public right to access state records.
Constitution of Nigeria (Sec. 55)	Recognizes indigenous language governance rights.
NDEPS (2020–2030)	Mandates digital service infrastructure expansion.
FMCIDE Strategic Blueprint	Focuses on AI talent, inclusion & sovereign datasets.
AU Continental AI Strategy (2024)	Promotes low-resource language digitization.

2.1 Statutory Basis: Freedom of Information Act (2011) and the Nigerian Constitution

The primary statutory foundation for this technical architecture is the Freedom of Information Act 2011. Section 1 of the Act establishes the right of any person to access or request information in the custody of any public official, institution, or agency. Section 2 mandates that public institutions ensure records are organized and archived in a manner that facilitates public access. Crucially, Section 2(4) requires agencies to proactively publish administrative details, financial receipts, and operational guidelines.

Integrating automated translation and voice-synthesis systems directly operationalizes these statutory mandates. By converting complex English documentation into localized audio and text, public agencies fulfil the requirement to make information genuinely accessible to the public, adhering to the spirit of the law.

Furthermore, this approach directly supports Section 55 of the 1999 Constitution of the Federal Republic of Nigeria. By leveraging technological innovation to operationalize indigenous languages in public administration, the state validates the constitutional dignity of non-English-speaking citizens, ensuring that linguistic heritage does not constitute a barrier to civic participation.

2.2 National Policy Alignment: NDEPS (2020–2030) and the FMCIDE Blueprint

At the executive level, this initiative aligns with the National Digital Economy Policy and Strategy (NDEPS 2020–2030), developed by the National Information Technology Development Agency (NITDA). NDEPS outlines eight strategic pillars for national transformation, including Pillar 3 (Solid Infrastructure), Pillar 5 (Digital Services Development and Promotion), and Pillar 8 (Indigenous Content Development and Promotion). Deploying sovereign AI translation engines directly advances Pillar 8 by creating digital resources that preserve and promote Nigerian languages within modern technology frameworks.

Similarly, the strategic blueprint released by the Federal Ministry of Communications, Innovation & Digital Economy (FMCIDE), titled Accelerating Our Collective Prosperity Through Technical Efficiency, emphasizes AI adoption as a cornerstone for national growth. The blueprint sets targets for positioning Nigeria as a leading hub for AI research, developing inclusive datasets, and upskilling technical talent through initiatives like the Three Million Technical Talent (3MTT) program. Integrating low-resource NLP research into public sector FOI workflows creates a practical use case for training local engineers and building sovereign datasets.

2.2 National Policy Alignment: NDEPS (2020–2030) and the FMCIDE Blue2.3 Regional Frameworks: African Union Continental AI Strategy (2024) print

At the executive level, this initiative aligns with the National Digital Economy Policy and Strategy (NDEPS 2020–2030), developed by the National Information Technology Development Agency (NITDA). NDEPS outlines eight strategic pillars for national transformation, including Pillar 3 (Solid Infrastructure), Pillar 5 (Digital Services Development and Promotion), and Pillar 8 (Indigenous Content Development and Promotion). Deploying sovereign AI translation engines directly advances Pillar 8 by creating digital resources that preserve and promote Nigerian languages within modern technology frameworks.

Similarly, the strategic blueprint released by the Federal Ministry of Communications, Innovation & Digital Economy (FMCIDE), titled Accelerating Our Collective Prosperity Through Technical Efficiency, emphasizes AI adoption as a cornerstone for national growth. The blueprint sets targets for positioning Nigeria as a leading hub for AI research, developing inclusive datasets, and upskilling technical talent through initiatives like the Three Million Technical Talent (3MTT) program. Integrating low-resource NLP research into public sector FOI workflows creates a practical use case for training local engineers and building sovereign datasets.

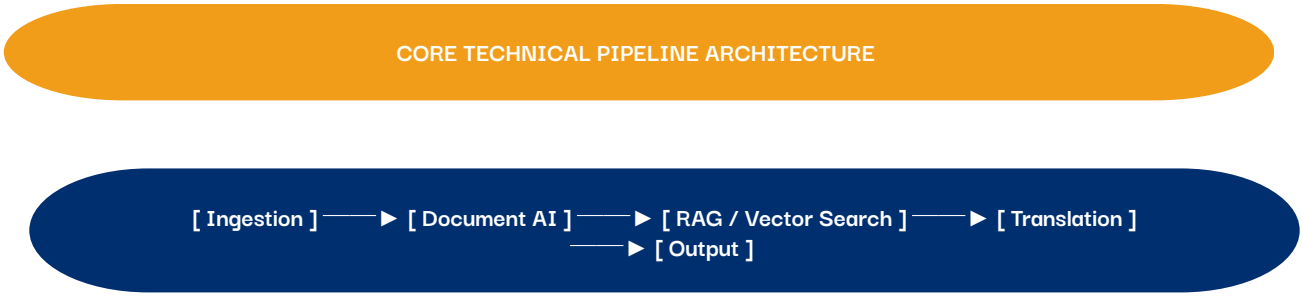
2.3 Regional Frameworks: African Union Continental AI Strategy (2024)

At the continental level, this project advances the African Union's Continental Artificial Intelligence Strategy, adopted in 2024. The AU Strategy urges member states to harness AI to promote socio-economic development, protect cultural diversity, and strengthen national sovereignty. It highlights the risk of an "AI divide" where Global South communities are excluded from the benefits of advanced technology due to a lack of localized training data.

The AU Strategy calls for member states to invest in digitalizing African languages, establishing local cloud infrastructure, and creating ethical AI governance frameworks anchored in African cultural values. Developing a sovereign, AI-driven FOI infrastructure in Nigeria offers a scalable model for other African nations seeking to bridge the gap between emerging technology and public sector information rights.

3. TECHNICAL ARCHITECTURE AND CLOUD INFRASTRUCTURE

To transition from policy intent to operational reality, this section outlines an enterprise-grade cloud architecture built on Google Cloud Platform (GCP) and Vertex AI. The platform provides a secure, scalable, and modular pipeline capable of processing unstructured physical documents, generating grounded translations, and delivering localized voice outputs.

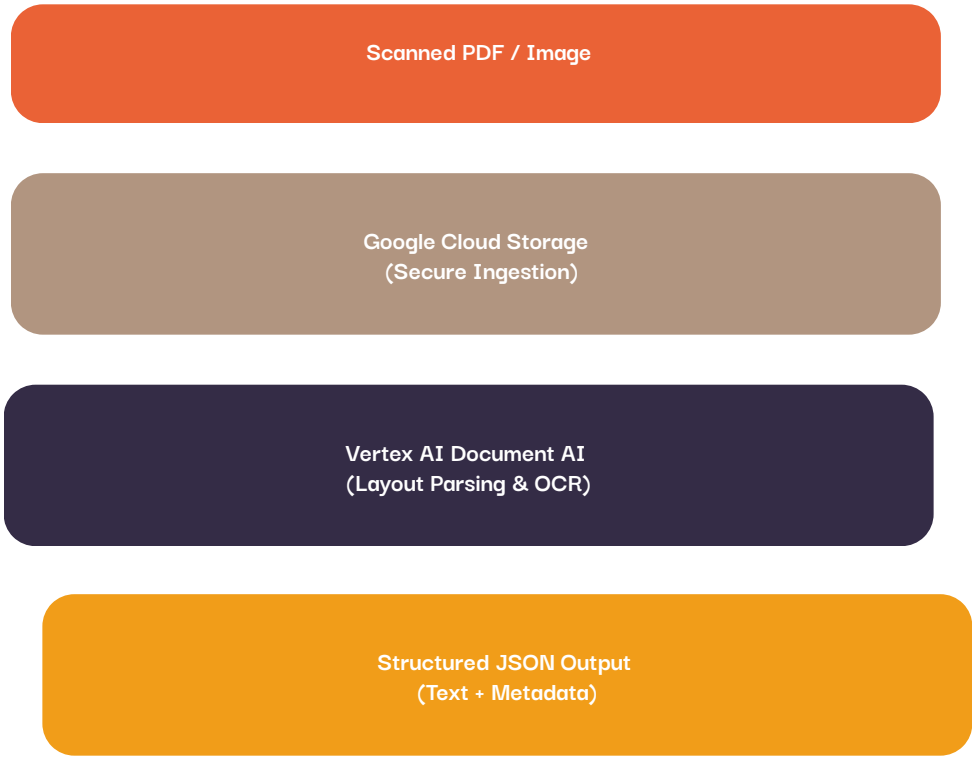


3.1 Document Ingestion and Optical Character Recognition (OCR) Layer

Government agencies rely heavily on physical paper records, scanned PDF files, and legacy text formats. The initial stage of the pipeline converts these unstructured physical files into clean, machine-readable digital text.

1. Storage and Ingestion: Unstructured records such as public procurement receipts, audit reports, and legal gazettes are ingested into secure Google Cloud Storage (GCS) buckets. These buckets are configured with Customer-Managed Encryption Keys (CMEK) and restricted by regional identity access controls to ensure compliance with Nigerian data residency requirements.

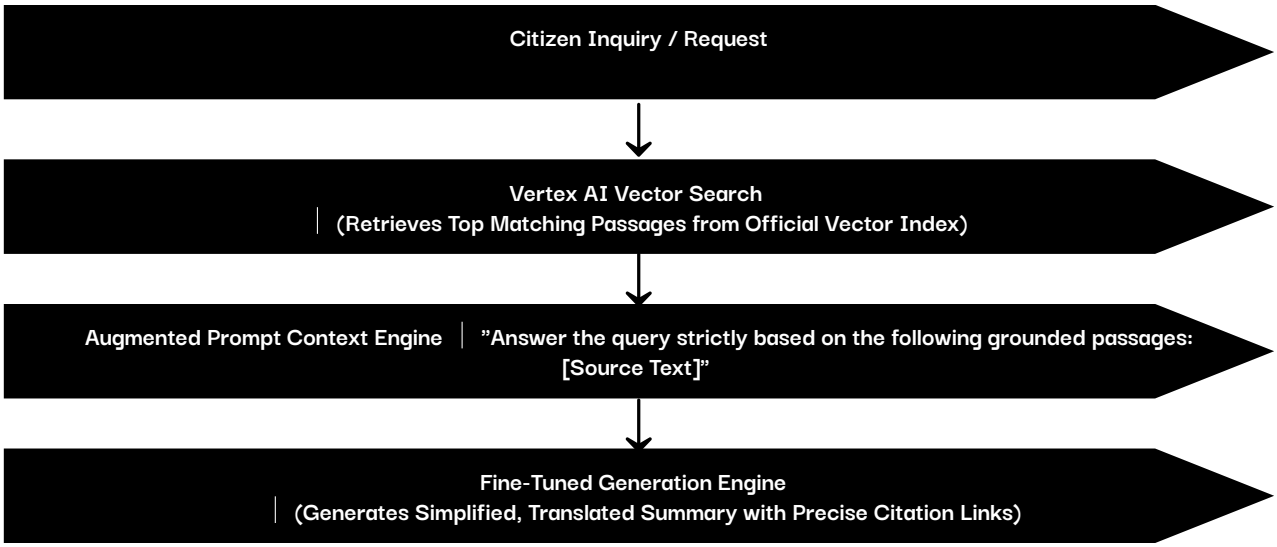
2. Layout Parsing and Processing: Document files are passed to Vertex AI Document AI. Utilizing specialized Layout Parsers and OCR engines, Document AI extracts raw text while preserving structural document metadata. The processor isolates multi-column text layouts, table headers, legal section numbering, and official stamps, outputting structured JSON containing the text alongside spatial bounding box coordinates.



3.2 Indexing and Retrieval-Augmented Generation (RAG) Layer

Deploying ungrounded generative models in legal and public administrative contexts introduces unacceptable risks of factual error and hallucination. To guarantee that translated FOI disclosures remain factually accurate, the platform implements a Retrieval-Augmented Generation (RAG) architecture using Vertex AI Search and Vector Search.

1. Chunking and Vector Embedding: The text extracted by Document AI is partitioned into semantic chunks (typically 500–1,000 tokens) with overlapping boundaries to preserve context. These chunks are converted into dense vector embeddings using Google's text-embedding-004 or specialized multilingual embedding models.
2. Vector Store Indexing: Embeddings are indexed within Vertex AI Vector Search, an enterprise-grade vector database capable of conducting real-time approximate nearest neighbor (ANN) searches across millions of document vectors.
3. Contextual Retrieval: When a citizen submits a query regarding a specific public record (e.g., "How much was allocated to primary healthcare in Kano State in the 2025 budget?"), Vertex AI Search retrieves the most relevant source passages from the vector store based on semantic similarity.
4. Prompt Augmentation: The retrieved passages are injected into the system prompt as grounding context, instructing the generative model to base its response exclusively on the provided official text and include precise source citations.



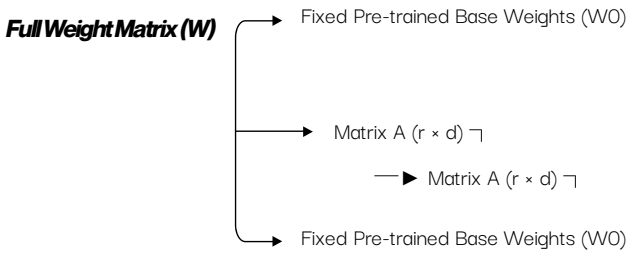
3.3 Large Language Model Translation and Adaptation Layer

The core generative engine relies on a hybrid deployment strategy that combines foundation models (Gemini 1.5 Pro/Flash) accessed via Vertex AI Studio with fine-tuned, open-weights models (Gemma 2 / Gemma 4) hosted on Vertex AI Prediction Endpoints.



To bridge low-resource language gaps, foundational models are adapted using Parameter-Efficient Fine-Tuning (PEFT) via Low-Rank Adaptation (LoRA). Rather than updating all parameters within a multi-billion parameter network, LoRA injects trainable rank-decomposition matrices into each layer of the Transformer architecture. This dramatically reduces computational resource requirements reducing the cost while enabling precise adaptation to low-resource languages.

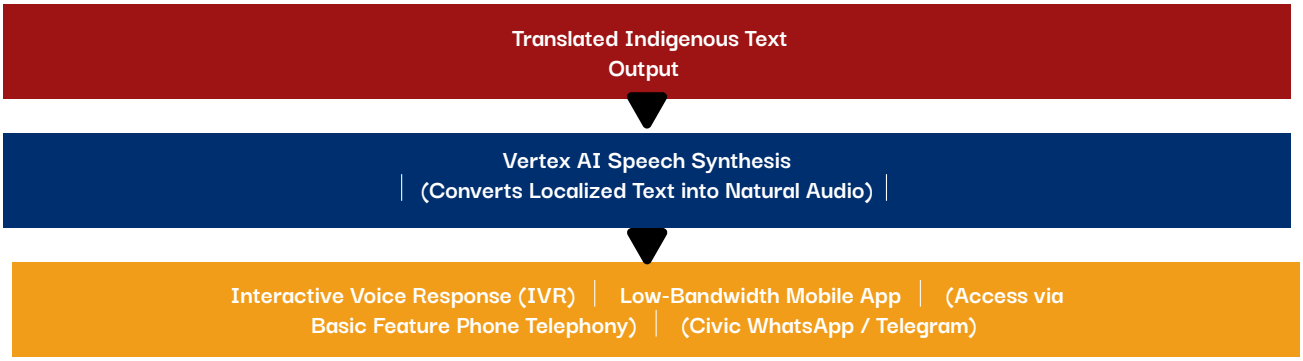
3.2 Indexing and Retrieval-Augmented Generation (RAG) Layer



The fine-tuning process relies on the FOI Parallel Lexicon to be developed in collaboration with legal scholars, certified translators, and civil society advocates, this dataset pairs thousands of legal English clauses from government records with accurate translations in Hausa, Yoruba, Igbo, and Nigerian Pidgin.

3.4 Speech Synthesis and Last-Mile Distribution Layer

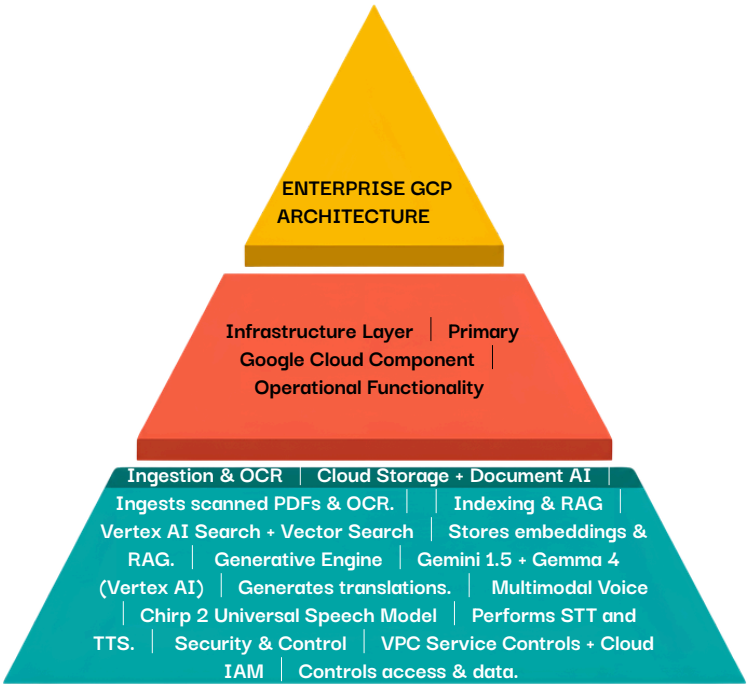
The fine-tuning process relies on the FOI Parallel Lexicon to be developed in collaboration with legal scholars, certified translators, and civil society advocates, this dataset pairs thousands of legal English clauses from government records with accurate translations in Hausa, Yoruba, Igbo, and Nigerian Pidgin.



1. Speech-to-Text (STT): Citizens can submit oral inquiries in their native language via a mobile application or phone call.Our infrastructure transcribes incoming audio into text, passing it to the RAG pipeline.
2. Text-to-Speech (TTS): Once the generative model produces a grounded summary, it would be synthesized and the text into natural-sounding audio in Hausa, Yoruba, or Igbo.
3. Distribution Infrastructure: To reach underserved communities, synthesized outputs are delivered via three channels:
◯ Civic Messaging Agents: Lightweight WhatsApp and Telegram bots that allow citizens to request and receive voice notes containing FOI summaries.
◯ Interactive Voice Response (IVR): A toll-free phone system that enables citizens on basic feature phones to dial in, select their language, and listen to public records.
◯ Web Access Portals: An open-access web portal for journalists, researchers, and civil society organizations.

4. Enterprise Architecture and Cloud Components

To ensure operational resilience, the platform uses a structured set of Google Cloud tools.



4.1 Ingestion and OCR Infrastructure

- Google Cloud Storage (GCS): Provides object storage for ingested paper records and digital PDFs. Configured with lifecycle policies to archive processed records automatically.
- Vertex AI Document AI: Extracts structured text, layout formatting, and tabular data from scanned documents, outputting machine-readable JSON.

4.2 Search and Vector Storage Infrastructure

- Vertex AI Search and Conversation: Manages the RAG pipeline, connecting LLMs directly to indexed document stores.
- Vertex AI Vector Search: Houses vector embeddings generated by text-embedding-004, facilitating real-time semantic search across large volumes of public records.

4.3 Generative Engine Infrastructure

- Vertex AI Studio: Serves as the primary workspace for prompt engineering, safety configuration, and managing foundation models like Gemini 1.5 Pro and Flash.
- Vertex AI Prediction Endpoints: Hosts open-weights models (such as Gemma 2 or Gemma 4) fine-tuned on the FOI Parallel Lexicon. This setup provides a sovereign cloud environment where inference data remains strictly contained within private enterprise boundaries.

4.4 Voice Processing Infrastructure

- Chirp 2 Speech-to-Text (STT): Transcribes audio queries from citizens into text for processing by the RAG pipeline.
- Chirp 2 Text-to-Speech (TTS): Converts localized summaries into natural voice audio, enabling distribution via IVR and messaging platforms.

4.5 Security, Privacy, and Governance Infrastructure

- VPC Service Controls: Establishes a security perimeter around GCP resources, preventing unauthorized data exfiltration and isolating public records within designated network boundaries.
- Cloud Identity and Access Management (IAM): Enforces Role-Based Access Control (RBAC), ensuring that only authorized public sector personnel can manage document indexes and fine-tuning pipelines.
- Cloud Logging and Audit Logs: Captures complete audit trails of system interactions, user requests, and administrative modifications to support compliance monitoring.

Policy Recommendations and Actionable Parties

1. National Information Technology Development Agency (NITDA)

Action: Issue Mandatory Public Sector AI Translation & Retrieval Guidelines requiring RAG Grounding and Human-in-the-Loop Audit Trails.

● Details: NITDA specifically through the National Centre for Artificial Intelligence and Robotics (NCAIR) should update its national AI regulatory sandbox and ethical guidelines to enforce technical standards for AI-assisted public information tools. The directive should mandate that any Ministry, Department, or Agency (MDA) deploying Generative AI for public disclosures must implement Retrieval-Augmented Generation (RAG) using verified vector databases (such as Vertex AI Vector Search) to prevent hallucinations, while establishing a mandatory Human-in-the-Loop (HITL) review dashboard where certified civil service FOI desk officers verify translated outputs before release.

2. National Assembly of the Federal Republic of Nigeria

Action: Enact a Legislative Amendment to the Freedom of Information (FOI) Act of 2011 to Require Proactive Multilingual Disclosure.

● Details: Building on its ongoing review of digital and data protection laws, the Senate Committee on ICT & Cyber security and the House Committee on ICT should sponsor an amendment to Section 2 of the FOI Act. The update should mandate that all proactive disclosures, annual agency budget audits, and public procurement summaries be made available in digital formats accessible in Hausa, Yoruba, Igbo, and Nigerian Pidgin alongside English. The legislation should explicitly sanction the use of certified AI translation pipelines to fulfill this statutory requirement within reasonable budget limits.

3. Federal Ministry of Communications, Innovation & Digital Economy (FMCIDE)

Action: Establish a Sovereign Open-Access Parallel Legal Corpus for Low-Resource Nigerian Languages.

● Details: Leveraging talent from the Three Million Technical Talent (3MTT) program, FMCIDE should fund and curate an open-access national parallel dataset (an FOI Legal Lexicon). This dataset should map complex administrative and legal English terminology from government gazettes into high-quality, verified equivalents in Hausa, Yoruba, Igbo, and Pidgin. FMCIDE should host these baseline corpora on certified local cloud infrastructure to allow public sector agencies and local developers to fine-tune open-weights models (like Gemma) securely.

4. MTN Nigeria (and Major Telecommunications Network Operators)

Action: Zero-Rate Data for Civic Voice/USSD AI Gateways and Host Localized Speech Processing Nodes.

● Details: As part of its $1.6+ trillion digital infrastructure bet and responsible AI initiatives (e.g., MTN Genova CoE), MTN Nigeria should partner with federal and state information ministries to zero-rate citizen access to voice-based civic information platforms. MTN should deploy edge-hosted Speech-to-Text (STT) and Text-to-Speech (TTS) endpoints (such as Google's Chirp 2 model) on its local cloud nodes to allow citizens in rural and unserved areas to query public FOI records via toll-free Interactive Voice Response (IVR) lines or WhatsApp/USSD interfaces without incurring data charges.

References

African Union. (2024). Continental Artificial Intelligence Strategy: Harnessing AI for Africa's Development and Transformation. African Union Commission. Addis Ababa, Ethiopia.

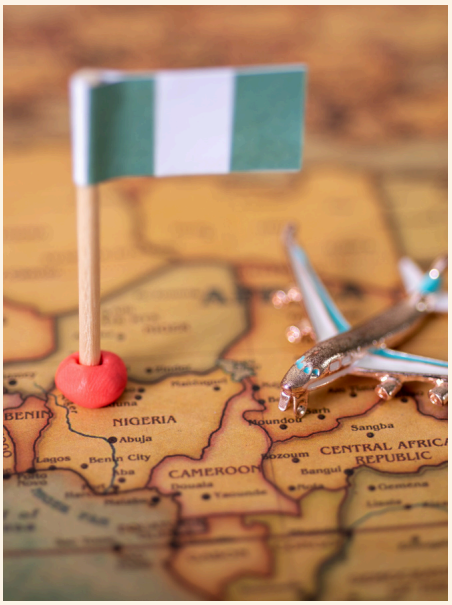
Federal Republic of Nigeria. (1999). Constitution of the Federal Republic of Nigeria 1999 (Section 55). Federal Ministry of Justice. Abuja, Nigeria

Federal Republic of Nigeria. (2011). Freedom of Information Act 2011. Official Gazette, Vol. 98, No. 58. Government Printer. Abuja, Nigeria.

Ministry of Communications, Innovation and Digital Economy. (2023). Accelerating Our Collective Prosperity Through Technical Efficiency: A Strategic Blueprint. Federal Government of Nigeria. Abuja, Nigeria.

National Information Technology Development Agency. (2020). National Digital Economy Policy and Strategy (NDEPS 2020–2030). Federal Ministry of Communications and Digital Economy. Abuja, Nigeria.

Gbemisola Esho



is a multidisciplinary professional with a background in cybersecurity, cloud computing, and journalism, dedicated to advancing digital inclusion and safety. She is the founder of Connectobridge, an initiative focused on empowering women and youth with access to technology education and career opportunities in emerging fields like AI and cybersecurity. With experience in training, public speaking, and community engagement, she has led impactful programs aligned with global development goals.

Gbemisola is also a certified professional with training from leading institutions and continues to drive innovation in digital policy, online safety education, and capacity building across underserved communities.

A Google Developer Expert, and firm believer in technology's ability to drive change and provide solutions, she has been a featured speaker at different conferences including Devfest, an annual Google Developer Festival, Privcon, a convergence of stakeholders in privacy, policy and data protection and at African Internet Governance Forum(African IGF) Beyond her core technology work, Gbemisola is a data journalist who uses data to tell compelling stories, and she regularly blogs on topics related to technology, security, and politics.

GENERATIVE AI MISUSE:

POLICY BRIEF FOR GENDER-SENSITIVE SAFEGUARDS IN

MOROCCO



AI GOVERNANCE IN MOROCCO

EXECUTIVE SUMMARY



EXECUTIVE SUMMARY

This policy brief will highlight the current gaps in Moroccan Legal system when it comes to emerging technologies such as generative AI tools used to generate non-consensual intimate imagery (NCII).
The rapid democratization of Generative AI has enabled a malicious industry where non-technical users can "nudify" others in minutes and use them for nefarious behaviours.

While Morocco positions itself as a regional AI hub through initiatives like Maroc IA 2030, a critical gap exists between economic investment and citizen protection. Current legal frameworks, primarily Law 103-13, remain reactive and lack the technical specificity to address fake media, creating a "loophole" for perpetrators.
This brief outlines a roadmap for reform, prioritising the criminalisation of the creation and possession of AI generated abuse material, the implementation of "Privacy-First" reporting protocols via the E-Blagh portal, and the enforcement of clear legal accountability for global app stores and search engines operating in the country.

PROBLEM STATEMENT

Morocco while embracing AI technologies with a recent agreement with NVIDIA to build a data centre. It seems that they are focusing on this novel industry as an investment haven, while investing very little into introducing policy reforms to protect its citizens from the predatory nature that could result from its misuse.
While many issues are could be addressed, like environmental risks, copyrights infringement, this policy brief will focus on the lack of safeguards concerning using generative AI to produce non-consensual intimate imagery (NCII).

As violence against women is only elevated through the use of internet, how does the introduction of tools such as “Nudify” Apps that allows such harm to propagate more easily, can be prevented and its uses penalised through the Moroccan legal system?, and who takes realistically the full accountability? the malicious users or company behind the tools or is it the lacking legal frameworks, because the real negative impacts will always fall on the unfortunate victim.

This will be a comparative policy analysis of the existing penal code of Morocco and any relevant laws related to gendered violence and violence against women, against policies in the EU, more notably the EU AI Act and the overall NCII protections standards.

The rise of "Nudify" applications allows perpetrators to transform innocent photos into photorealistic fake abuse material. In Morocco’s community-centred society, these fabricated images are weaponised through the cultural system of Hshouma (social shame), where the mere threat of a leaked image causes "real" reputational ruin within families and communities.

Victims often suffer in silence due to a profound lack of trust in judicial procedures, with only 10.5% of Moroccan women currently reporting violence to the authorities (High Commission for Planning [HCP], 2019). This analysis will highlight the gaps in the Moroccan system and the consequences of such gaps when it comes to endangering women.

METHODOLOGY

This policy brief was drafted alongside the mixed-methodology research report, utilizing the same desk literature-based policy analysis, being informed with the findings from the focus-group and online survey. This policy brief focuses mainly on analysing and documenting the gaps within the current Moroccan Penal Code and the country's relevant AI policies by comparing them with the EU AI Act. Building a localised list of policy recommendations for Morocco, taking into consideration its existing legal frameworks and its cultural context.

KEY INSIGHTS

Existing Penal code on Sexual harassment, blackmail and gendered violence; Article 447 and Article 503:

- The Moroccan legal framework for digital abuse is based on the 1962 Penal Code, which was later updated to address modern threats. Article 447 (specifically 447-1 to 447-3) serves as the primary mechanism for criminalising the non-consensual capture and distribution of private imagery. This article penalises the intentional recording or dissemination of a person's private information or pictures without their consent, as well as the broadcast of false statements intended to harm a person's reputation.
- However, this means that it only includes real non-consensual abuse material and does not consider the creation and possession of AI generated or deepfake abuse material.

Update on laws against gendered violence that is more inclusive to types of harm; Law 103-13:

- ·The enactment of Law 103-13 in 2018 marked a significant shift by introducing a more inclusive definition of violence, encompassing physical, moral, sexual, and economic harm. Crucially, it expanded the scope of harassment to include "electronic harassment" via phone messages, records, or images of a sexual nature. It also established aggravating circumstances and increased penalties if the perpetrator is a co-worker or a person of authority over the victim.
- hough, this was a huge step forward, it still lacks protection against emerging threats online including Generative AI. While also being criticized for being a reactive approach to safety; not protecting the citizens from the crime but only punishing the criminals after the damage was done. Thus, putting the victims under a paramount pressure to actively pursue their cases and be able to prove the damage in court.

The Morocco AI 2030 plan and what can be learnt from the EU legal frameworks to handle emerging risks: Maroc IA 2030 VS the EU AI Act 2024:

- The Maroc IA 2030 roadmap illustrates Morocco's ambition to become a regional hub for digital innovation, highlighted by strategic agreements with companies like NVIDIA to build national data centres. However, this strategy is currently focused on AI as an "investment haven" rather than a space for citizen protection, with minimal attention paid to the predatory nature of AI misuse.
- ·In contrast, the EU AI Act (2024) offers a proactive, risk-based classification system that Morocco can learn from. The EU model prohibits AI systems that pose an "unacceptable risk" to fundamental rights and subjects generative AI to strict transparency obligations, such as mandatory digital watermarking under Article 50, enforcing a "Safety by Design" approach to protect its citizens.

EVIDENCE / CONTEXT

- Morocco policies lacking the definition of synthetic or artificial content when it comes to the policies criminalising blackmail and cyber harassment.
- Moroccan policies specifically criminalise the distribution of inappropriate imagery that damage one's reputation or social standing, with no mention of artificially made content.
- The consensus for Moroccans based on data collected is that the reporting protocols lack privacy and consideration of the victims and thus the average citizens does not trust the authorities to properly handle scandalous cases.

- The Moroccan approach is one that focuses on individual actions rather than enterprises accountability, as these “Nudify” applications are a for-profit business with and $117 million in revenue that actively inflict harm on citizens globally.
- ·Non-existing training for law enforcement to be able to distinguish deepfake from real abuse material; affecting the quality of investigations.

POLICY IMPLICATIONS

The existing gendered violence that exists in Morocco is well documented, cases where perpetrators exploit their power and social standing to blackmail women and force them into either paying ransom or worse. Allowing such harmful technologies such as AI generated NCII in a society where a vulnerable group already is struggling will only worsen their struggle to exist in the digital realm that translate into real life consequences.
The lack of digital literacy is also a major barrier for progress, while Morocco works towards becoming an investment haven and frontier for AI innovation with its “Maroc Digital 2030 strategy”, actual statistics shows a high illiteracy rate in Morocco which also disfavours rural communities and has a clear gender gap with women counting for 32.4% of illiteracy within Morocco compared to 17.2% of men, according to the 2024 General Population and Housing Census.

POLICY RECOMMENDATIONS
FOR THE MOROCCAN MINISTRY OF JUSTICE

Recommendation: Criminalise the actual creation and possession of AI-generated non-consensual intimate imagery (NCII) and other emerging technologies being used for "nudification".

Rationale: Shifting from reactive laws to criminalising the material itself closes the "loophole" where perpetrators claim realistic imagery is merely "fictional" to evade existing harassment statutes and protects the victims from.

For Law Enforcement and the E-Blagh Portal
Recommendation: Implement a "Privacy-First" reporting protocols that strictly limit investigations to relevant evidence and provide guaranteed protection of a victim's dignity and right to privacy.
Rationale: Invasive questions used in the existing reporting procedures and the fear of social shame are the primary barriers to victims seeking justice.

Moroccan Civil society
Recommendation: developing initiatives that prioritize raising awareness about emerging threats and risk associated with social media and the internet.

Rationale: either government funded or independent programs that could help raise awareness about the types of threats, how to handle them legally and the proper reporting procedures is key to enforcing the existing policies and a necessary step to evaluate any needed changes in policies. While addressing the taboo of social shame and normalizing seeking of justice.

CROSS-CUTTING RELEVANCE

The key takeaway from comparative analysis is that the EU AI Act has a Safety by Design concept which prosecutes the companies and focuses on their role in perpetuating harm. while Morocco’s approach could be read as passive, which is a rather normal outcome. Morocco unlike the EU or the USA who is also working on their own AI policy, is not a base for tech companies, nor is it a major market for such companies. Thus, it lacks any leverage when it comes to exerting control over their operations, unlike the EU countries for example.

This is relevant to all similar cybercrimes or harmful technologies that are available freely on the internet, as Morocco will have to continue trying to catch up to such risks constantly since it does not have any control over its development or profitability.

With AI “nudification” tools, to ransomware kits and many others for sale either on public telegram channels or with more technical knowledge can easily be acquired through dark web marketplaces.

CONCLUSION

Existing in the digital realm is no longer a choice for many of Moroccan citizens who need a profile to communicate with family or seek a job. And demanding the right to safety within the digital realm is a right that should be protected for all Moroccans. Based on the analysis on the current Moroccan policies within the context of the cultural and societal beliefs, rather than an a major reform of existing policies, it is more critical to actually raise awareness about the risks and prioritizing digital literacy training within the government body; which only then allow future progress to be more inline with what the Moroccan citizens need without lacking the technical understanding of emerging technologies and their misuse.

This policy brief goal is to ultimately recommend positive changes grounded in local and realistic needs of Morocco as an African country, instead than just basing it on the EU policies. A policy plan should be strategically developed taking into consideration the country’s current state, available resources and with citizens needs as priority.

ABOUT AUTHOR

Hanaa Dany is an information security analyst based in Casablanca, Morocco, bringing practical expertise in information assurance, risk assessments, threat hunting, and digital safety education. With professional credentials including CompTIA Security+ and ISC2 Certified in Cybersecurity (CC), she has established herself as a dedicated cybersecurity professional committed to advocating for Digital Rights in her community. Currently working as Teaching Assistant for a Cybersecurity professional learning program while focusing on her work as a researcher investigating policy protections against AI-generated non-consensual intimate imagery (NCII) in Morocco as a case study.

BIBLIOGRAPHY

Amer, O.M. (2026). Morocco Launches ‘Maroc IA 2030’ Roadmap To Strengthen AI Governance, Digital Sovereignty. Morocco World News. https://www.moroccoworldnews.com/2026/01/274765/morocco-launches-maroc-ia-2030-roadmap-to-strengthen-ai-governance-digital-sovereignty/

Baker Botts (2026). U.S. Artificial Intelligence Law Update: Navigating the Evolving State and Federal Regulatory Landscape | Thought Leadership | Baker Botts. [online] Baker Botts. Available at: https://www.bakerbotts.com/thought-leadership/publications/2026/january/us-ai-law-update.

Committee on Knowledge and Information Society (2024). Artificial intelligence (AI) in Morocco: What areas and what prospects? https://www.cese.ma/media/2024/12/ENG-synthe%CC%80seArtificial-intelligence-AI-1.pdf

Eclatal (2025). Loi Sur L’IA Au Maroc : Un Projet De Loi Pour Encadrer L’intelligence Artificielle | Eclatal - Digital Agency. Eclatal - Digital agency. https://eclatal.ma/loi-sur-lia-au-maroc/

European Parliament (2025). EU AI Act: First Regulation on Artificial Intelligence. [online] European Parliament. Available at: https://www.europarl.europa.eu/topics/en/article/20230601STO93804/eu-ai-act-first-regulation-on-artificial-intelligence.

Future of Life Institute (2024). Implementation Documents | EU Artificial Intelligence Act. [online] Artificialintelligenceact.eu. Available at: https://artificialintelligenceact.eu/implementation-documents/.

Grar, K. (2024). Leconomiste Maghrebin. [online] Leconomiste Maghrebin. Available at: https://www.leconomistemaghrebin.com/2024/09/19/maroc-une-proposition-de-loi-pour-criminaliser-lia/

High Commission for Planning (HCP). (2019). Deuxième enquête nationale sur la prévalence de la violence à l'égard des femmes et des hommes. https://www.hcp.ma/Enquete-Nationale-sur-la-violence-a-l-encontre-des-Femmes-et-des-Hommes_a105.html
Buza, M., & Taha, S. (2025, April 9). DPA digital digest: Morocco. Digital Policy Alert. https://ricardo-dashboard.s3.eu-west-1.amazonaws.com/reports/1744190873307_DPA%20Digital%20Digest%20Morocco%20[2025%20Edition].pdf
High Commission for Planning (HCP). (2024). 2024 وثائق حول نتائج الإحصاء العام للسكان والسكنى [Documents on the results of the 2024 General Population and Housing Census]. https://www.hcp.ma/file/242669/

Morocco World News. (2024, December). Morocco's illiteracy rate stands at 24.8% in 2024. https://www.moroccoworldnews.com/2024/12/166863/morocco-s-illiteracy-rate-stands-at-24-8-in-2024/
Pendergast, K. (2024). 'Nudify' apps, the evolution of online exploitation, and what we can do about it. Family Online Safety Institute. https://fosi.org/nudify-apps-the-evolution-of-online-exploitation-and-what-we-can-do-about-it/

Sheikh, M., & Rogers, J. P. (2025). Addressing the psychiatric implications of AI-enabled non-consensual sexual imagery. The British Journal of Psychiatry. https://doi.org/10.1192/bjp.2025.xx

Tech Transparency Project. (2026). Nudify apps widely available in Apple and Google app stores. https://www.techtransparencyproject.org/articles/nudify-apps-widely-available-in-apple-and-google-app-stores

Thornton, D. (2024). Nudify deepfakes. A Better Choice Network Solutions. https://www.abcnetfl.com/blog/nudify-deepfakes

UNICEF. (2026). 'Deepfake abuse is abuse' - UNICEF. https://www.unicef.org/press-releases/deepfake-abuse-is-abuse

United Nations Interregional Crime and Justice Research Institute (UNICRI). (2024). Generative AI: A new threat for online child sexual exploitation and abuse. https://unicri.org/sites/default/files/2024-09/Generative-AI-New-Threat-Online-Child-Abuse.pdf

U.S. Government Accountability Office (2025). Artificial Intelligence: Federal Efforts Guided by Requirements and Advisory Groups. [online] Gao.gov. Available at: https://www.gao.gov/products/gao-25-107933.

Whitehead, S. (2025). NVIDIA Targets Morocco As Next African Expansion Hub After South Africa Entry. [online] iAfrica.com. Available at: https://iafrica.com/nvidia-targets-morocco-as-next-african-expansion-hub-after-south-africa-entry/.

GOVERNING THE ALGORITHM:

Assessing Ghana's Legal And Institutional Readiness For Algorithmic Decision-Making



EXECUTIVE SUMMARY

Ghana is deploying artificial intelligence across financial services, healthcare, and public administration faster than the laws intended to govern these systems have evolved. The Data Protection Act 2012, Ghana's primary data-governance legislation, was enacted thirteen years before the country's current AI ambitions were declared. It contains no provisions specific to AI, no mandatory pre-deployment impact assessment, and a trade-secret exemption that may weaken citizens' right to an explanation for automated decisions.

Ghana's National AI Strategy 2025 to 2035 is backed by a US$270 million government investment: US$250 million for a national AI computing centre and US$20 million for short- to medium-term rollout (Republic of Ghana, 2025a). The governance architecture needed to support that investment responsibly does not yet exist in enforceable form. Ghana's own government signalled this gap when it launched the UNESCO AI Readiness Assessment in November 2025, with the Communications Minister cautioning that deploying AI without adequate safeguards could deepen inequality and erode public trust (Ghana News Agency, 2025).

This paper applies a thematic benchmarking framework derived from the UNESCO Recommendation on the Ethics of Artificial Intelligence (2021) and the ECOWAS Supplementary Act on Personal Data Protection (2010) to assess Ghana's legal readiness across five governance criteria. Drawing on comparative analysis of Kenya, South Africa, and Rwanda, it identifies three structural governance gaps and proposes four targeted, implementable reforms.

Research Objectives

General Objective: To assess the adequacy of Ghana's existing legal and regulatory frameworks in governing AI-driven algorithmic decision-making across critical sectors, and to identify the legislative, regulatory, and institutional reforms required to close the governance gaps identified.
Specific Objectives:
1. To assess Ghana's Data Protection Act 2012, Electronic Transactions Act 2008, and proposed Emerging Technologies Bill 2025 against five governance criteria derived from the UNESCO Recommendation (2021) and the ECOWAS Supplementary Act (2010).
2. To map current AI deployment across Ghana's financial services, healthcare, and public administration sectors using publicly available secondary sources.
3. To conduct a structured thematic comparative analysis of AI governance provisions in Kenya, South Africa, and Rwanda and identify transferable lessons for Ghana's reform agenda.
4. To develop concrete, actionable, and institutionally feasible policy recommendations addressing the identified governance gaps.

Research Questions

The study is guided by three questions. First, to what extent do Ghana's legal and regulatory frameworks satisfy the five governance criteria for accountable algorithmic decision-making? Second, where do the principal gaps lie, and how do they expose citizens affected by automated decisions? Third, what transferable lessons from Kenya, South Africa, and Rwanda can inform feasible reform in Ghana's institutional context?

Analytical Approach

This paper employs qualitative doctrinal legal analysis and comparative policy analysis. Primary sources, namely the Data Protection Act 2012, the Electronic Transactions Act 2008, the proposed Emerging Technologies Bill 2025, the ECOWAS Supplementary Act 2010, and the UNESCO Recommendation 2021, are examined through close reading and critical assessment against the five governance criteria set out below.

The comparative analysis uses a thematic benchmarking method: three comparator jurisdictions (Kenya, South Africa, and Rwanda) are assessed against the same five criteria applied to Ghana, enabling structured cross-jurisdictional comparison rather than country-by-country description. Kenya is the primary comparator, given its comparable Anglophone African context, expanding fintech sector, and relatively advanced data-protection framework. South Africa is included for its sector-specific accountability model in financial services, and Rwanda for its institutional capacity-building approach; each is a purposively selected illustration of a single transferable mechanism rather than a full jurisdictional profile. All sources are publicly available secondary materials.

Stages of analysis

The analysis proceeds in five steps: (1) selection of the governing legal instruments; (2) coding of each instrument's provisions against the five criteria; (3) identification of gaps where provisions are absent, partial, or unenforceable; (4) benchmarking of each gap against a comparator jurisdiction that has addressed it; and (5) derivation of reform recommendations calibrated to Ghana's institutional capacity.

Analytical Framework: Algorithmic Accountability

The paper is anchored in the framework of algorithmic accountability: institutions deploying automated decision-making systems bear an obligation to ensure those systems are transparent, contestable, and subject to meaningful oversight. This framework is operationalised through five governance criteria derived from the UNESCO Recommendation (2021) and the ECOWAS Supplementary Act (2010):

- Criterion 1, Transparency and Explainability: individuals must be informed when AI systems affect their rights and must be able to access explanatory information (UNESCO, 2021, Articles 37 to 38).
- Criterion 2, Human Oversight and Redress: individuals must be able to contest automated decisions and access reconsideration mechanisms (ECOWAS, 2010, Article 35; UNESCO, 2021, Articles 35 to 36).
- Criterion 3, Algorithmic Accountability and Audit: oversight, audit, and due-diligence mechanisms must exist throughout the AI system life cycle (UNESCO, 2021, Articles 42 to 43).
- Criterion 4, Ethical Impact Assessment Before Deployment: frameworks for assessing AI risks must be established before deployment (UNESCO, 2021, Articles 50 to 53).
- Criterion 5, Independent Regulatory Institution: an independent authority with a specific mandate to oversee AI systems must exist (ECOWAS, 2010, Article 14; UNESCO, 2021, Articles 54 to 58).

Limitations

This analysis relies entirely on publicly available secondary sources and does not capture institutional practice, real-world enforcement dynamics, or the gap between legal text and operational reality. In particular, it cannot assess how Ghana's Data Protection Commission enforces existing provisions in practice, nor whether the trade-secret exemption in Section 35(2) has been formally invoked or tested. Desk-based doctrinal analysis of this kind identifies where the law is silent or weak, but not how institutions behave in the space the law leaves open; the paper's conclusions are framed accordingly, and the reform proposals are directed at closing textual and institutional gaps rather than diagnosing enforcement conduct.

Problem Statement

For the purposes of this paper, artificial intelligence refers to technology that enables computers and machines to simulate human learning, comprehension, problem-solving, decision-making, creativity, and autonomy, as defined in Ghana's proposed Emerging Technologies Bill (Republic of Ghana, 2025b, Section 37). Algorithmic decision-making refers to the use of such systems to make, or substantially inform, decisions with significant legal, financial, or social consequences for individuals.

Ghana is already deploying such systems. In the financial sector, AI-enabled digital banking services are being used to simplify credit scoring for small and medium enterprises (Opoku-Afari, 2024). In healthcare, the Ghanaian company mPharma uses AI algorithms to predict medication demand and manage pharmaceutical supply chains, and deep learning is being applied to automate radiology (Opoku-Afari, 2024; Pasipamire & Muroyiwa, 2024). In public administration, the Ghana Revenue Authority has launched the Integrated Tax Administration System, replacing manual processes with a unified digital platform (GRA, 2025). These systems increasingly rely on automated processes with direct consequences for citizens' access to services, healthcare, and financial inclusion.

Citizens affected by these systems currently lack AI-specific procedural safeguards and sector-specific accountability mechanisms. Ghana's government formally acknowledged this when it launched the UNESCO AI Readiness Assessment Methodology in November 2025 to evaluate its institutional capacity, regulatory frameworks, and governance arrangements for AI, with the Minister for Communications identifying algorithmic bias, privacy concerns, and workforce disruption as central governance challenges (Ghana News Agency, 2025). Ghana's National AI Strategy likewise acknowledges that algorithmic bias, privacy violations, and the misuse of personal data pose genuine ethical and societal risks requiring governance responses (Republic of Ghana, 2025a).

Ghana's Readiness at a Glance

Table 1 summarises Ghana's standing against the five governance criteria. Two criteria are partially met through provisions designed for conventional data processing; three are effectively absent for algorithmic systems

Governance criterion	Status	Governing provision / gap
1. Transparency & explainability	Partial	Right to decision logic under s.35(1)(d), Data Protection Act 2012, but qualified by the trade-secret exemption in s.35(2).
2. Human oversight & redress	Partial	s.41 offers protection only where the data subject requests it in advance; no proactive institutional duty to notify.
3. Accountability & audit	Absent	No life-cycle audit, due-diligence or algorithmic-oversight requirement in force.
4. Ethical impact assessment	Absent	No mandatory pre-deployment impact assessment for high-risk AI systems.
5. Independent regulatory institution	Absent	Responsible AI Authority proposed in the National AI Strategy but not yet established or mandated.

Table 1. Ghana's legal readiness against the five governance criteria (author's assessment).

Findings

Finding 1: A transparency right structurally undermined by a trade-secret exemption

Ghana's Data Protection Act 2012 gives data subjects a right to the logic behind automated decisions under Section 35(1)(d). However, Section 35(2) immediately qualifies this: where a data controller claims that disclosure would reveal a trade secret, the right to explanation does not apply (Republic of Ghana, 2012). This may weaken transparency and accountability in practice, particularly in financial services where AI credit scoring is already operational. This provision has not, to the author's knowledge, been formally tested before Ghana's courts or publicly interpreted by the Data Protection Commission; its structural presence nonetheless creates a potential vulnerability in citizens' right to explanation that would benefit from legislative clarification. The Act's general principles (accountability, openness, and data-subject participation under Section 17) were designed for conventional data processing and are not well adapted to algorithmic systems operating at scale.

Finding 2: Regional and international commitments not yet reflected in enforceable domestic terms

Ghana is a signatory to the ECOWAS Supplementary Act 2010, which under Article 35 prohibits any decision with legal effect on an individual from being based solely on automated processing (ECOWAS, 2010). Ghana also adopted the UNESCO Recommendation on the Ethics of AI in 2021, a soft-law instrument that, while non-binding, sets ethical standards Ghana has voluntarily committed to, including transparency obligations, ethical impact assessment, and independent governance institutions (UNESCO, 2021, Articles 37 to 38, 50 to 58). The gap between these commitments and Ghana's enforceable domestic provisions represents a regulatory misalignment that the proposed Emerging Technologies Bill 2025 begins to address but does not yet resolve.

Finding 3: Three transferable governance advances from comparator jurisdictions

The comparative analysis benchmarks each jurisdiction against the same five criteria to identify specific, transferable lessons rather than general country profiles. Table 2 sets out the comparison.

Criterion	Ghana	Kenya	South Africa	Rwanda
1. Transparency & explainability	Partial	Met	Partial	N/A*
2. Human oversight & redress	Partial	Met	Partial	N/A*
3. Accountability & audit	Absent	Partial	Met	Partial
4. Ethical impact assessment	Absent	Met	Partial	N/A*
5. Independent regulatory institution	Absent	Met	Met	Met

*Table 2. Comparative standing on the five criteria. *"N/A" indicates the criterion was not the focus of this comparator's selected lesson and is not assessed here. Author's assessment.*

On Criterion 2 (Human Oversight and Redress): Kenya's Data Protection Act 2019, Section 35, establishes a proactive notification obligation: institutions must inform individuals in writing when automated decisions affecting them are taken, and must provide a structured reconsideration process within a reasonable period (Kenya, 2019). This directly addresses the passivity gap in Ghana's Section 41, which requires the data subject to request protection in advance.

On Criterion 3 (Ethical Impact Assessment): Kenya's Section 31 requires a mandatory data-protection impact assessment before any processing likely to result in high risk to data subjects' rights, with reports submitted sixty days prior to deployment (Kenya, 2019). Ghana's Act contains no equivalent provision.

On Criterion 4 (Algorithmic Accountability and Audit): South Africa's Information Regulator has approved sector-specific codes of conduct under the Protection of Personal Information Act, including for the Banking Association of South Africa, which acknowledges members' use of automated decision-making and establishes transparency and complaints commitments (Plantinga et al., 2024). South Africa is selected here precisely because it demonstrates that AI accountability in financial services can be advanced through sector codes without waiting for comprehensive new legislation, a route directly available to Ghana's Data Protection Commission.

On Criterion 5 (Independent Regulatory Institution): Rwanda is establishing capacity-building programmes for civil servants in regulatory, policymaking, and legislative roles with responsibility for AI oversight, supported by matching funding and procurement guidelines, and is embedding AI-ethics functions within ministry and agency digital-governance environments (Plantinga et al., 2024). Rwanda is selected for the institutional-capacity dimension that Kenya's primarily legislative model does not illustrate, and it speaks directly to the staffing, technical-capacity, and funding questions surrounding Ghana's proposed Responsible AI Authority.

Plantinga et al. (2024) confirm that existing governance capabilities across African countries offer important opportunities for policy learning: Ghana does not need to design its governance architecture from scratch.

Finding 4: The Emerging Technologies Bill 2025 is a significant advance that raises implementation questions

The proposed Bill introduces principles on algorithmic-bias prohibition (Section 22), transparency and oversight (Section 23), mandatory human intervention for decisions substantially affecting individual rights (Section 29), and a right of redress (Section 34) (Republic of Ghana, 2025b). However, the Bill remains unenacted, and the relationship between the proposed Emerging Technologies Agency and the existing Data Protection Commission is not yet defined, creating potential regulatory overlap. Questions of funding, staffing, technical capacity for algorithmic auditing, and institutional independence also remain unresolved. These must be addressed in the legislative process to avoid governance duplication or institutional fragmentation.

Evidence and Context

Ghana's National AI Strategy openly acknowledges that AI deployment carries risks, including algorithmic bias from foreign datasets not suited to Ghanaian contexts, a lack of transparency and explainability that makes bias difficult to detect and challenge, and the potential for AI to be used for social harm and citizen surveillance (Republic of Ghana, 2025a). These are precisely the risks the governance criteria applied in this paper are designed to address.

At the continental level, Pasipamire and Muroyiwa (2024) document that many African nations lack comprehensive national strategies, institutions, and regulatory frameworks to manage AI effectively, and that algorithmic bias in AI credit scoring in Kenya's fintech sector has generated discriminatory outcomes for women with limited digital footprints. Ghana's fintech sector operates under comparable conditions, with the same absence of AI-specific oversight. Plantinga et al. (2024) further note that African states are often unable to assemble sufficient internal policy and administrative capacity to navigate emerging technologies, making legislative and institutional reform both urgent and foundational.

Policy Implications

If the governance gap is not addressed, AI systems are likely to continue being deployed across Ghana's financial services, healthcare, and public administration without mandatory pre-deployment risk assessment, without proactive transparency obligations, and without an independent institution able to investigate algorithmic harm. On the available evidence, citizens affected by automated decisions would in such circumstances have limited practical access to explanation or reconsideration. Pasipamire and Muroyiwa (2024) document that algorithmic bias can exacerbate existing inequalities and erode public trust, consequences that would be particularly significant in Ghana, where digital financial inclusion is a stated national priority.

The counterargument should be acknowledged.

Some policymakers argue that premature or extensive regulation risks stifling innovation, discouraging investment, or creating compliance burdens for emerging AI start-ups. This is a legitimate concern that must be taken seriously in Ghana's context. Ghana's own National AI Strategy addresses it directly, arguing that responsible, inclusive, and sustainable AI development requires data governance and policy as essential enablers rather than constraints (Republic of Ghana, 2025a). The reforms proposed here target high-risk deployments and focus on accountability mechanisms rather than broad prohibition; they are calibrated to establish trustworthy governance without creating barriers to beneficial AI adoption.

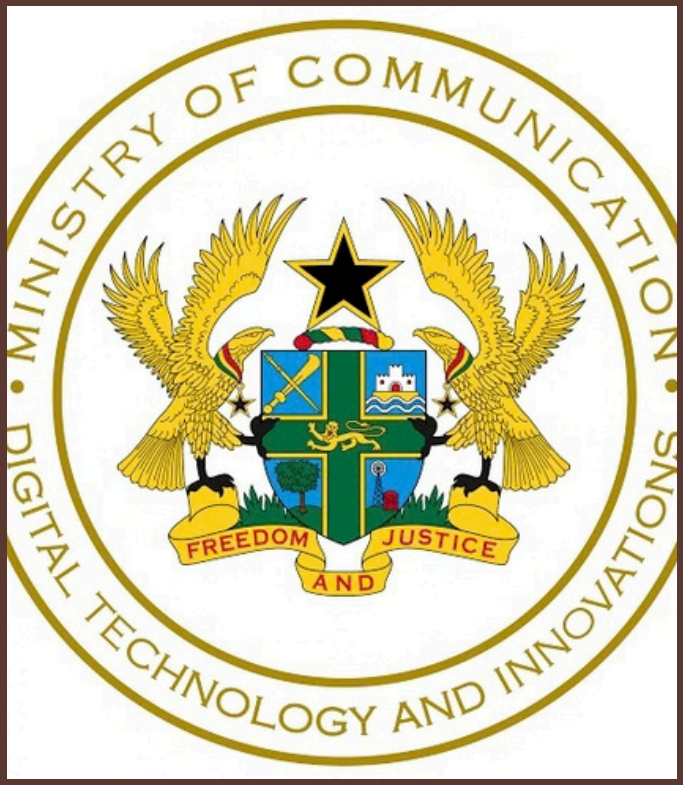
Policy Implications
For the Parliament of Ghana

Prioritise enactment of the proposed Emerging Technologies Bill 2025 with enforcement provisions made concrete and specific. The Bill's principles on algorithmic bias (s.22), transparency (s.23), human oversight (s.29), and the right to redress (s.34) should become binding obligations. The Bill should provide for administrative financial penalties for institutions deploying algorithmic decision-making systems without mandatory impact assessments or transparency disclosures. Enforcement should be assigned to the Data Protection Commission under its existing mandate, with a right of appeal to the High Court for affected parties. Penalty thresholds should be proportionate to the scale and risk of deployment, modelled on Kenya's administrative-fine structure under Section 63 of the Data Protection Act 2019 (fines not exceeding five million shillings or one percent of annual turnover, whichever is lower) (Kenya, 2019). The legislative process should also resolve the institutional relationship between the proposed Emerging Technologies Agency and the Data Protection Commission.

Rationale: Principles without enforcement do not protect citizens. Legislative specificity on sanctions, enforcement bodies, and institutional mandates is what turns policy intent into accountability.

For the Ministry of Communications, Digital Technology and Innovations

Amend the Data Protection Act 2012 to introduce a mandatory data-protection impact assessment, modelled on Kenya's Section 31, for all AI systems deployed in high-risk contexts (financial services, healthcare, and public administration) with reports submitted before deployment (Kenya, 2019). Additionally, review the trade-secret exemption in Section 35(2) to ensure commercial confidentiality cannot be routinely invoked to override citizens' right to explanation for algorithmic decisions significantly affecting their rights (Republic of Ghana, 2012).



Rationale: Pre-deployment assessment is the most effective point at which to identify and mitigate algorithmic harm. Kenya demonstrates this requirement is operationally feasible within an African regulatory context.

For the Data Protection Commission

Issue binding guidelines within the Commission's existing mandate under Section 3 of the Data Protection Act, requiring institutions deploying automated decision-making systems to proactively notify affected individuals in writing when such decisions have been taken and to provide accessible reconsideration mechanisms, modelled on Kenya's Section 35(3) (Kenya, 2019; Republic of Ghana, 2012). This reform does not require new legislation.

Rationale: Proactive notification shifts the burden from citizens, who may be unaware of their rights, to institutions, consistent with the accountability principles the Act already establishes.

For the Government of Ghana

Establish the Responsible AI Authority proposed in the National AI Strategy 2025 to 2035 as an independent institution with a specific legal mandate to oversee AI systems, conduct algorithmic audits, and receive and investigate complaints (Republic of Ghana, 2025a). Drawing on Rwanda's model of embedding AI-ethics functions within government and building civil-servant capacity in regulatory and procurement roles (Plantinga et al., 2024), the Authority's design should explicitly address staffing, technical capacity for algorithmic auditing, funding, and independence from ministerial direction.

Rationale: Legal reform without institutional capacity produces governance on paper but not in practice.

Cross-Cutting Relevance

The governance gaps identified here connect to broader digital-governance priorities across West Africa and the continent. The ECOWAS Supplementary Act 2010 establishes regional standards that member states are expected to reflect domestically; strengthening Ghana's AI-governance architecture would contribute to the harmonisation of digital regulatory standards across the ECOWAS region. At the continental level, the African Union Data Policy Framework (2022) identifies data governance as foundational to responsible AI development. These issues connect equally to the core concerns of the Digispace Africa Fellowship's agenda, namely digital rights, platform accountability, and the safety of online and automated systems: AI systems making consequential decisions without transparency or accountability represent simultaneously a governance failure, a digital-rights failure, and a trust-and-safety failure.

Contribution to Knowledge

Existing scholarship on responsible AI in Africa maps governance capabilities at a continental scale (Plantinga et al., 2024) and documents algorithmic bias in specific sectors (Pasipamire & Muroyiwa, 2024). This paper adds a criterion-by-criterion, provision-level readiness audit specific to Ghana, benchmarked against three comparator jurisdictions, that translates broad diagnoses into an enumerated set of textual and institutional gaps and the concrete reforms required to close them.

Conclusion

Ghana has the ambition, the resources, and the policy vision to be an African leader in artificial intelligence. The National AI Strategy 2025 to 2035 is a serious and well-resourced policy document, and the proposed Emerging Technologies Bill represents genuine legislative progress. The frameworks currently in force, however, were designed for a different technological era and leave citizens with limited AI-specific procedural safeguards when automated systems affect their access to credit, healthcare, and public services. The reforms proposed in this paper are targeted, feasible, and grounded in Ghana's own policy commitments, its regional and international obligations, and the legislative experience of Kenya, South Africa, and Rwanda. On this analysis, aligning governance frameworks with the scale of Ghana's AI deployment is both achievable and timely.

By Richmond Osei

is a Statistics graduate, Certified Google Data Analyst, and Junior Analyst based in Accra, Ghana. He is a fellow of the Digispace Africa Fellowship 2026, where his research focuses on AI governance and legal readiness for algorithmic decision-making in Ghana. His work sits at the intersection of data analysis, digital policy, and public-interest advocacy.

References

African Union. (2022). Data Policy Framework. African Union Commission.

ECOWAS. (2010). Supplementary Act A/SA.1/01/10 on personal data protection within ECOWAS. Economic Community of West African States.

Ghana News Agency. (2025, November 28). Ghana, UNESCO launch ethical AI readiness assessment. GNA.

Ghana Revenue Authority. (2025). Ghana's tax revolution begins: GRA launches ITAS to usher in a fully digital tax era. GRA.

Kenya. (2019). Data Protection Act, No. 24 of 2019. National Council for Law Reporting.
Ministry of Communication, Digital Technology and Innovations, Ghana. (2026, March 11). Stakeholder consultation on UNESCO AI readiness assessment held today. Government of Ghana.

Opoku-Afari, M. (2024). The intersection between AI, productivity and growth in emerging countries (Keynote Address, 2nd Emerging Countries International Conference on Business, Finance and Economics). Bank of Ghana.

Pasipamire, N., & Muroyiwa, A. (2024). Navigating algorithm bias in AI: ensuring fairness and trust in Africa. Frontiers in Research Metrics and Analytics, 9, 1486600.

Plantinga, P., Shilongo, K., Mudongo, O., Umubyeyi, A., Gastrow, M., & Razzano, G. (2024). Responsible artificial intelligence in Africa: towards policy learning. Data & Policy, 6, e72.

Republic of Ghana. (2008). Electronic Transactions Act, 2008 (Act 772). Government of Ghana.

Republic of Ghana. (2012). Data Protection Act, 2012 (Act 843). Government of Ghana.

Republic of Ghana. (2025a). National Artificial Intelligence Strategy 2025 to 2035. Ministry of Communications, Digital Technology and Innovations.

Republic of Ghana. (2025b). Emerging Technologies Bill, 2025 (Proposed Legislation). Government of Ghana.

UNESCO. (2021). Recommendation on the ethics of artificial intelligence. United Nations Educational, Scientific and Cultural Organization.



Outing as a digital weapon: Why global social media platforms fail to safeguard criminalised LGBTIQ+ identities in Ethiopia and what a Contextual Harm Policy should include.

EXECUTIVE *Summary*

According to data from the International Lesbian, Gay, Bisexual, Trans, and Intersex Association (ILGA), 64 countries worldwide criminalise same-sex sexual acts, creating a precarious environment for gender and sexual minorities (ILGA, 2023).



In Ethiopia, Article 629 of the 2004 Criminal Code criminalises consensual same-sex sexual conduct, with imprisonment up to three years as the standard penalty. Under the related provision (Article 630), sentences range to rigorous imprisonment of up to 25 years in cases involving "grave cases." While official implementation varies across regions, documented instances show this legislation has been applied to target LGBTIQ+ individuals, with the threshold for "grave cases" under the statute not requiring proof beyond reasonable doubt. Beyond statutory criminalisation, the absence of legal protection in both physical and digital spheres, increasing extralegal aggression by state entities, and compounded restrictive social and religious norms render Ethiopia an exceptionally dangerous environment for LGBTIQ+ individuals (KII.2, 2006). Consequently, these marginalised communities have increasingly retreated into covert physical gathering spaces and digital platforms to communicate, organise, and access essential resources for healthcare and education.

While global social media platforms such as Meta and TikTok have become vital lifelines for these communities, they simultaneously serve as vectors for persecution. Anti-LGBTIQ+ movements, including state and non-state actors, exploit these platforms to coordinate “outing”, the non consensual disclosure of an individual's identity. In criminalised contexts, outing is not merely a privacy breach but a life-threatening mechanism that exposes victims to mob violence, state harassment, and familial correctional aggressions. Current global platform safety standards consistently fail to mitigate these risks due to two primary factors: a generic ("one-size-fits-all") moderation approach that assumes a baseline of legal protection, and significant structural gaps in language moderation systems (Shahid, Elswah & Vashistha, 2025). Specifically, Ethiopia's complex multilingual landscape, involving over 80 languages, the unique Ge'ez script, and a widely used undetectable English phonetics script of these languages, allows hate speech and coordinated attacks to bypass automated detection systems trained primarily on English and other high-resource languages.

This policy paper analyses outing as a distinct persecution mechanism within the Ethiopian legal and social context, contrasting it with protections in legally safeguarded jurisdictions. Through comparative analysis, the study identifies persistent governance failures across diverse criminalised environments. Drawing on secondary data, including the UN Guiding Principles on Business and Human Rights (Ruggie, 2011), platform transparency reports,

and civil society advocacy, the paper argues that existing frameworks like Meta's and TikTok's Harmful Content Policy, and YouTube's Harassment & Cyberbullying Policy, are theoretically insufficient for criminalised contexts because they fail to account for the extreme consequences of identity exposure.

Central to this analysis is the proposal of a Contextual Harm Norm (CHN). Defined as a policy framework wherein platform moderation protocols are adjusted to reflect the specific legal and physical contextual safety risks of a user's jurisdiction, the contrary assumption of minimal harm in protected regions may constitute a severe safety violation in criminalised ones. Under a Contextual Harm Norm, the non-consensual outing of an LGBTIQ+ individual in Ethiopia would be treated with the same severity as doxxing in a protected jurisdiction with high-resourced languages, triggering immediate removal and enhanced safety interventions, regardless of local legal ambiguity or linguistic nuance.

The paper concludes by outlining a practical integration strategy for this norm into existing platform governance systems. In order to ensure the safety and inclusivity of marginalised and criminalised communities, social media platforms should include a contextual harm classification for criminalised contexts, invest in low-resource language moderation capacity, and involve local civil society actors in threat response mechanisms.

PROBLEM STATEMENT

In Ethiopia, where same-sex conduct remains criminalised, social media platforms have become vectors for coordinated and non-coordinated persecution through identity outing, publicly exposing LGBTIQ+ individuals without consent (Access Now, 2024). This digital weaponisation exposes vulnerable people to state violence, grave community harassment, and physical harm. LGBTIQ+ Ethiopians are disproportionately affected, facing existential threats when their identities are exposed online. The urgency stems from three converging factors: Ethiopia's explicit criminalisation of same-sex conduct along with aggressive non state actors against the community, the country's complex multilingual landscape (over 80 languages) that creates structural gaps in automated content moderation (Shahid, Elswah & Vashistha, 2025), and the absence of contextual harm frameworks in global platform policies (Ruggie, 2011). With at least 31 African countries criminalising same-sex conduct, this represents a continental digital rights failure requiring immediate policy intervention to safeguard criminalised communities from platform-enabled persecution (Access Now, 2024).

KEY INSIGHTS

Insight 1: In criminalised contexts, outing/identity disclosure works as a coordinated persecution strategy rather than a simple invasion of privacy Outing is considered harassment or a breach of privacy in countries in which legal frameworks exist to protect users outside the privacy policies of social media platforms. In Ethiopia, it acts as a catalyst for official persecution and extrajudicial violence that is harsher than the law itself (Feminist Internet Research Network & Association for Progressive Communication, 2025). Platforms typically label it as "Personal Information," ranking it lower than outright threats with grave consequences for users. This is based on the assumption by platform policy designers that there is a legal protection mechanism for victims whose identities are publicly exposed by a third party. This baseline assumption for policy remains inaccurate for nearly 57% of African countries, which codified the criminalisation of LGBTIQ+ individuals (Access Now, 2024).

Insight 2: The failure to integrate multilingual languages by platforms, particularly Ethiopia's 80+ language environment, gives opportunity for hostile actors to exploit vulnerable groups

When compared to high-resource languages such as English or other European languages, platforms' automatic detection algorithms for varied languages perform much worse, if not at all (Shahid, Elswah & Vashistha, 2025). This gap provides offenders/perpetrators with the opportunity to weaponise it by adopting versions of Ge'ez scripts or phonetic forms of the native languages to avoid detection while communicating harmful content to local audiences. Bad actors not accidentally but strategically use language moderation gaps in various nations with limited language resources, such as Ethiopia (Siegel & Nyarko, 2023).

Insight 3: The privacy policies for Meta, TikTok, X (Twitter), and YouTube often presume the existence of legally protected baselines as global standards before defining harm

The notion of harm on social media platforms builds on discrimination, with the idea that victims of prejudice are protected by laws that hold bad actors accountable. (YouTube, n.d.-a, TikTok, n.d., X, n.d. Meta, n.d.) Such laws could include employment, housing, or other types of legislation that protect people against overt discrimination based on a personal characteristic or agenda. The distinction between countries like Ethiopia and those with legal protection is that these laws do not include marginalised communities but rather penalise them (Human Rights Watch, 2023). This assumption that all global users are regulated is incorrect for nearly one-third of the world's countries that criminalise LGBTIQ+ individuals (ILGA, 2023).

Insight 4: Patterns have been identified in existing secondary material, without the requirement for primary data to mandate policy improvements. Transparency reports, civil society documentation (HRW, ILGA, Access Now), academic research, laws and regulations, and experiences from local LGBTIQ+ organisations show evidence of systematic failure (Access Now, 2024). Policy development or reforms should depend on existing secondary sources rather than requiring new trauma documentation and primary victims' testimony from those enduring continued persecution, as this shifts the burden and misplaced responsibility onto already endangered and targeted populations (Feminist Internet Research Network & Association for Progressive Communication, 2025).

EVIDENCE/CONTEXT

1. Rainbow Burning: Social Media Risks for LGBTQ+ People in Africa in 2024 by Access Now

This data shows a clear gap in social media platforms' failure to enforce their own Terms of Service (ToS) for outing, leading into hate speech, and doxxing, which leads to life-threatening events in criminalised contexts (Access Now, 2024). In 2024, Access Now performed a thorough investigation, analysing 214 reported posts from Ethiopia, Kenya, Uganda, and Nigeria, which revealed clear threats of non-consensual intimate photos, dead-naming, and coordinated doxing. According to this data, just 51 of the reported posts, or around 23%, were removed from these platforms (Access Now, 2024). The remaining documented cases stayed online, posing an active threat to the community in which same-sex is criminalised and socially, culturally, and religiously scrutinised to the point of death (Access Now, 2024).

The study also found that these key platforms (Meta (Facebook & Instagram), X, TikTok, and YouTube) rely almost entirely on automatic moderation algorithms that do not recognise context specific language in Amharic (Ethiopia's working language) or any of the country's other 80 languages (YouTube, n.d.-a, TikTok, n.d., X, n.d. Meta, n.d.). In Ethiopia, Facebook and TikTok are the most popular sites for inciting abuse, with organised campaigns targeting women and LGBTIQ+ community members, resulting in documented offline consequences such as family expulsion, job loss, or the need to flee the country to avoid mob punishment (Feminist Internet Research Network & Association for Progressive Communication, 2025).

SOCIAL MEDIA CONTENT REMOVAL RATE (ACCESS NOW 2024 STUDY

REMOVAL RATE %
NOT-REMOVED %

2. Queer Resistance: Online Gender-Base Violence Against LGBTIQ+ Ethiopians by Feminist Internet Research Network and Association for Progressive Communication, 2025

According to the findings of this study, big social media platforms failed to protect gender disadvantaged communities through their own Terms and Conditions (ToS), resulting in severe consequences for queer online outing victims (Feminist Internet Research Network & Association for Progressive Communication, 2025). Based on 50 in-depth interviews with LGBTQIA+ community members of various genders and sexual identities from Ethiopia and the diaspora, as well as eight feminist advocates in Addis Ababa, it was discovered that the lack of social media protection through its own regulation put marginalised communities in grave danger, resulting in physical attacks, mob justice, and fleeing the country (Feminist Internet Research Network & Association for Progressive Communication, 2025).

This study demonstrated that not only LGBTIQ+ populations in Ethiopia, but also those in the diaspora, are significantly harmed as a result of outings that cannot be reported by social media, both in terms of contextual knowledge of privacy bridging and language coverage (Feminist Internet Research Network & Association for Progressive Communication, 2025). This study documented patterns of forced disclosure, non-consensual intimate image sharing, and targeted hate campaigns on Facebook and Twitter, citing negative psychosocial consequences such as anxiety, isolation, and strategic self-censorship, as well as criminalisation and amplification of all forms of gender-based violence (Feminist Internet Research Network & Association for Progressive Communication, 2025).

Participant testimony:

"When my identity first came out, I was shocked. My phone was broken and I couldn't see what was happening. Friends told me the post was removed quickly, so I didn't think much of it at the time. I tried to go on with my life but then more information started coming out... where I lived, who I was, even things about my family. One day, I was attacked outside my school. There were many of them; they shouted 'faggot' as they hit me. Some had sticks and others had rocks. After one punch to my face, everything went dark. I remember hearing a woman's voice saying, 'This faggot will die on your hands. Let him go.' Later, I realised how much of my body was sore and bruised. More and more videos appeared online after that. People saying I should be slaughtered, that they needed to find me and make me an example. That's when I knew I had to flee. When you've faced death, fear stops controlling you. That's what keeps me going. I've seen the lowest point one can reach...", (Feminist Internet Research Network & Association for Progressive Communication, 2025, p. 12).

3. Human Rights Watch Documentation on Human Rights Violation report, 2023/24

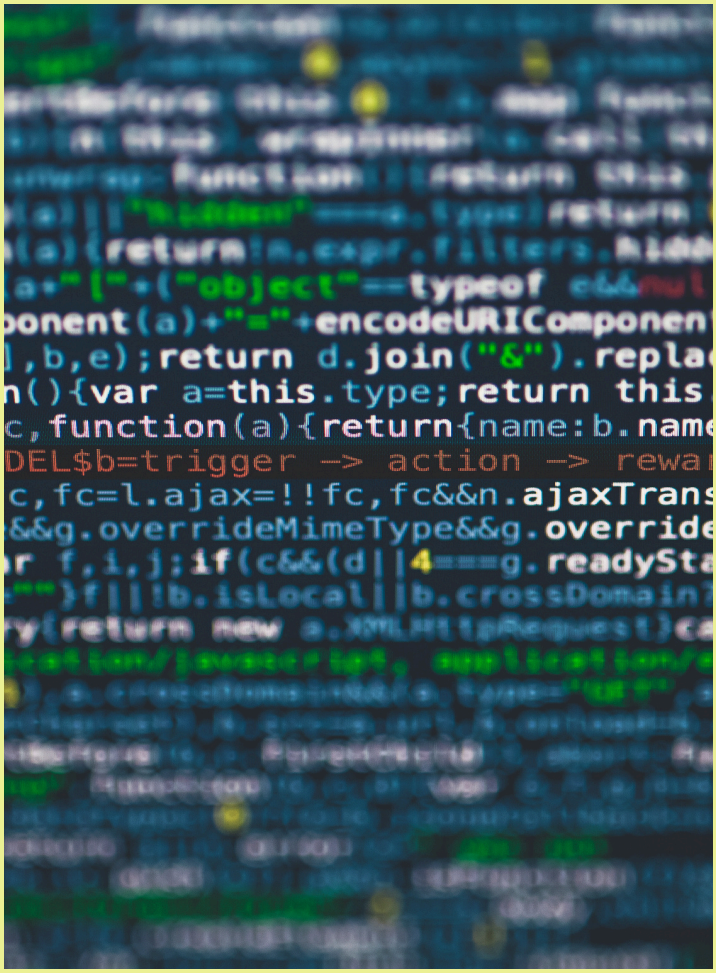
The Human Rights Watch highlighted real-life examples of cases in Ethiopia in which LGBTIQ+ communities face significant dangers as a result of identification outings from group discussions, with screenshots published on social media platforms such as Facebook and TikTok (Human Rights Watch, 2024). As a result of these outings, community members faced corrective violence from their families as well as police questioning to open criminal cases (Human Rights Watch, 2024). According to this report, the Addis Ababa Peace and Security Administration Bureau took action against hotels and businesses alleged to host gay gatherings, following "massive online reporting," and an anonymous government hotline was set up encouraging citizens to report "illegal" same-sex activity (Human Rights Watch, 2024).

Human Rights Watch has also explicitly documented the gap in social media to protect its LGBTIQ+ users in 2023 with detailed documented cases from the Middle East and North Africa (Human Rights Watch, 2023). In this study, analysis of cases from five countries in the region, Egypt, Iraq, Jordan, Lebanon, and Tunisia documented testimonies where state actors and private individuals across the Middle East and North Africa (MENA) region have entrapped lesbian, gay, bisexual, and transgender (LGBT) people on social media and dating applications, subjected them to online extortion, online harassment, and outing, and relied on illegitimately obtained digital photos, chats, and similar information in prosecutions, in violation of the right to privacy, due process, and other human rights (Human Rights Watch, 2023).

This asserts a lack of mechanisms for identifying, recognising, and taking screenshots accompanied by local languages in criminalised settings that necessitate stricter response rules (Human Rights Watch, 2023).

4. Think Outside the Data: Colonial Biases and Systemic Issues in Automated Moderation Pipelines for Low-Resource Languages by Shahid, Elswah and Vashistha (2025)

The study, which included a qualitative investigation of 22 AI professionals working in Tamil, Swahili, Maghrebi Arabic, and Quechua, verified a well-documented moderation gap (Shahid, Elswah & Vashistha, 2025). Automated algorithms usually attain accuracy rates of more than 85% in English but less than 40% in low-resource languages (Shahid, Elswah & Vashistha, 2025).





The authors go on to suggest that this disparity is the result of colonial bias in data gathering and large platforms' monopoly on user data (Shahid, Elswah & Vashistha, 2025). As a result, independent verification is limited, and nations with low-resource languages receive little funding for moderating tools (Shahid, Elswah & Vashistha, 2025).

As a result, the reliance on opaque, English-centric algorithms consistently fails to protect users in non-English speaking regions, highlighting the critical need for data transparency frameworks proposed by researchers such as Siegel and Nyarko in their research published in the Stanford Internet Observatory in 2023 (Siegel & Nyarko, 2023).

Additional supporting evidence includes the UN Guiding Principles on Business and Human Rights (UNGPs), unanimously endorsed by the UN Human Rights Council in June 2011.

The UNGPs establish that corporations bear a responsibility to respect human rights independently of local legal frameworks through due diligence, grievance mechanisms, and non-discriminatory policies. However, major digital platforms have inconsistently applied these principles to criminalised LGBTIQ+ communities, particularly in jurisdictions where same-sex relations carry criminal penalties.

The EU Digital Services Act Regulation (European Commission, 2022), which entered into force in November 2022, mandates Very Large Online Platforms to conduct annual systemic risk assessments addressing threats to fundamental rights, including non-discrimination protections. Nevertheless, Ethiopia receives minimal attention in platform risk-mapping exercises, as DSA compliance prioritises European regulatory concerns over other contexts, creating oversight gaps for vulnerable populations operating outside EU jurisdiction.

5. LOCAL CSO EVIDENCE

A Key Informant Interview (KII) with a local CSO member that was the initiator of communication as well as identifying a working relation with TikTok, in response to the Ethiopian 2023's crackdown against LGBTIQ community, revealed how some existing interaction with TikTok had a promising but failed attempt to identify what they called a "Trusted Partner" programme (KII, 2026).

TikTok attempted to fill the language gap in hate detection mechanisms during Ethiopia's organised anti-LGBTIQ+ crackdown in 2023. Through local CSO networks, the platform identified vetted CSO members as "trusted partners"

to identify significant hate or outing posts with serious implications (KII, 2026). This collaboration aimed to reduce the response time to identify consequential hate posts or identity disclosures and remove content within 8–10 hours, rather than weeks in some circumstances (KII, 2026).

While this collaboration seemed promising, it did not provide meaningful outcomes because these collaborations lacked proper follow-up trends and lack of data transparency needed for the successful collaboration of the stakeholders (KII, 2026). Furthermore, the platform's anticipated response time (8–10 hours) remained too long to provide a timely response to an outed individual in high-risk countries like Ethiopia (KII, 2026).

POLICY IMPLICATION

The lack of inclusive contextual policy and protection will continue to expose LGBTIQ+ individuals to death, forced displacement, and permanent chilling effects on digital participation in Ethiopia and 31 other African countries where same-sex conduct is criminalised, in addition to cultural, social, and economic repression (Access Now, 2024). Furthermore, LGBTIQ+ communities and civil society actors who use online spaces as their primary means of communication via social media outlets may lose faith in platform safety assurances (Human Rights Watch, 2023). This contradicts Africa's broader digital inclusion aims by allowing gender minorities to rationally associate life-threatening risks with online presence.

This legislation and protection vacuum will also expose social media platforms to long-term reputation lose and legal risks. In the long run, if Africa adopts regulations similar to the EU's of Digital Services Act (DSA), any online platforms offering services in the region must comply with various protective and compliance regulations, creating liability for failing to mitigate foreseeable harms even where local law conflicts with human rights standards (European Commission, 2022). Platforms can protect vulnerable communities by incorporating a contextual framework that distinguishes between genuine harm and targeted prejudice, rather than imposing a generic approach in response to broad censorship demands (Siegel & Nyarko, 2023).

The contextual framework can also be extended beyond Ethiopia to include more than 31 restrictive states in Africa. Successful contextual framework implementation will also demonstrate that generic content moderation is somewhat outdated and requires significant contextual assessment of what is judged "safe enough" in one region versus another. This will have a significant impact on whether digital spaces become sites of exclusion or inclusion for marginalised communities navigating high-risk legal settings.

POLICY RECOMMENDATIONS

For Tech Companies like Meta, Google/YouTube, Telegram, X/Twitter, TikTok

Recommendation 1: Develop Contextual Harm Classification for Nations that Criminalise LGBTIQ+ Individuals, Such as Ethiopia (short-term) ➡ Creating a distinct harm category focused on identity disclosure or outing risks in digital platforms in countries where LGBTIQ+ identity is criminalised, with consequential negative social, economic, and cultural effects.

Current categories in the mentioned social media outlets fail to account for unauthorised identity disclosure or targeted outings as serious threats beyond breaches of privacy, and they assume that legal mechanisms are in place after exposure (Human Rights Watch, 2023). Creating a distinct harm category will address the life-threatening gap for platform users who lack legal frameworks to protect basic human rights.

The recommended classification must be different from standard "doxxing" or "harassment" inclusions, necessitating rigorous review procedures to meet existential threats rather than normal privacy violations (European Commission, 2022). This classification involves the adoption of locally contextual policy activation and customised risk-assessment teams that are familiar with the specific threat posed to vulnerable communities.

The development of contextual harm classification goes hand-in-hand with UN Guiding Principles on Business and Human Rights (UNGPs), due diligence under pillar 2 (UN, 2011). Under the UNGPs, platforms have a responsibility to respect human rights, which requires identifying, preventing, mitigating, and accounting for adverse human rights impacts linked to their operations (Ruggie, 2011).

This also covers the responsibility of social media outlets under the EU Digital Services Act (DSA), Articles 34 that mandate obligations to assess and mitigate systemic risks, including risks to fundamental rights such as dignity, non-discrimination, and freedom of expression (European Commission, 2022)

Recommendation 2: Invest in Low-Resource Language Moderating Capacities (medium-term)

➡ Allocate dedicated resources for training AI detection algorithms and hiring human moderators who are fluent in Ethiopian languages and regional dialects using Ge'ez script variants. Existing systems are structurally incapable of detecting harm in these languages (Shahid, Elswah & Vashistha, 2025). Beyond technical solutions, human expertise would fill the knowledge gap in understanding cultural complexity, coded language, contextual meanings, and local languages employed in English phonetics that algorithms overlook.

The minimal requirements should reflect the moderate quality levels found in Western markets where English is the primary working language, with additional legal resources provided for marginalised communities.

Recommendation 3: Develop Rapid Response Mechanisms for Criminalised Populations, Involving Local CSOs and Other Relevant Humanitarian Actors. (Immediate implementation)

➡ Create priority escalation channels accessible to civil society partners documenting emerging outing campaigns or coordinated targeting reporting in criminalised jurisdictions, with guaranteed 1–2 hour response timeframes, since delayed responses can precipitate real-world violence or state persecution.

Standard reporting systems sometimes take weeks to evaluate by generic teams who are inexperienced with the urgency of the situation at hand (KII, 2026).

Beyond establishing the categorical distinction, implementing this framework necessitates infrastructure investment in regional advisory councils/Trusted Partners made up of local LGBTIQ+ advocates, human rights monitors, and digital safety specialists with firsthand knowledge of the specific environments. Pre-vetted, trusted collaborations with local, regional, and international organisations allow for the accurate identification of actual emergencies versus routine complaints.

This can fall under the EU Digital Services Act (DSA), Article 35, that mandate obligations to assess and mitigate systemic risks in a contextual manner (European Commission, 2022). Generic measures like content removal timelines and content flagging algorithms need the support of local and regional actors that have contextual understanding as well as are trained on SOGIESC (Sexual Orientation, Gender Identity and Expression, and Sex Characteristics) in criminalised contexts (Human Rights Watch, 2023).

CSOs operate primarily underground due to Article 69 of the Ethiopian Charities and Societies Proclamation Law (2009) and the hostile environment created by non-state actors toward LGBTIQ+ groups (KII, 2026). Future initiatives by the platform to find and collaborate with regional advisory councils and Trusted Partners shall also take into account the particular circumstances of CSOs in Ethiopia and other nations where same-sex actions are illegal.

For Governments, Regional Bodies (African Union, ECOWAS, East African Community)

Recommendation 4: Set Basic Digital Protection Standards Among Member States.(long-term)

The lack of basic protection standards allows platforms to operate in countries with limited linguistic resources. The development and implementation of binding guidelines under the African Union (AU) that can be implemented in member countries would increase the focus given to such countries by platforms, as well as African policymakers' negotiating power with technology corporations based outside the continent.

CROSS-CUTTING RELEVANCE

This issue connects to three broader digital governance challenges: First, it exposes how algorithmic bias manifests along linguistic and geographic lines, reproducing colonial-era information hierarchies where non-Western voices receive inferior protection (Shahid, Elswah & Vashistha, 2025). Second, it illustrates collision between local sovereignty (criminalisation laws) and universal human rights obligations under international business standards (Ruggie, 2011). Third, it demonstrates how "neutral" platform policies actually have active participation through the assumptions about legal frameworks and social acceptance. Addressing outing in Ethiopia therefore strengthens digital rights infrastructure applicable to religious minorities, political dissidents, sex workers, and other groups facing state-sanctioned discrimination globally.

METHODOLOGY

This policy analysis employed a mixed-methods approach combining secondary data synthesis with primary qualitative inquiry. The research primarily relied on secondary data sources to establish evidence and contextual understanding. To supplement this foundational analysis, two Key Informant Interviews (KIIs) were conducted with representatives from two local Civil Society Organisations (CSOs) that have previously collaborated with TikTok. These interviews served to incorporate a firsthand perspective from civil society stakeholders that directly engaged in platform related initiatives previously.

CONCLUSION

At least 31 African countries where same-sex conduct is illegal are affected by the weaponisation of social media platforms for the non-consensual outing of LGBTIQ+ people in Ethiopia, which is a serious failure of global platform governance. This analysis shows that the "one-size-fits-all" content moderation frameworks used by platforms like Meta, TikTok, X, and YouTube are structurally unable to protect users in situations where identity exposure is not only a privacy violation but also a potentially fatal act that can result in forced relocation, mob violence, state prosecution, and family expulsion. A critical blind spot where hostile actors can operate with near impunity is created at the intersection of Ethiopia's criminalisation of same-sex conduct under Article 629, its complicated multilingual landscape of over 80 languages that frequently bypass automated detection systems, and the lack of contextual harm frameworks in platform policies.

The proposed Contextual Harm Norm, which optimises platform moderation processes to reflect the unique legal and physical threats of a user's jurisdiction rather than assuming a universal baseline of legal protection, is a significant shift in perspective that is necessary to address this crisis. Tech companies must develop dedicated contextual harm classifications for criminalised contexts, invest substantively in low-resource language moderation capacities, including human moderators fluent in Ge'ez script variants and local dialects, and establish rapid response mechanisms with guaranteed 1–2 hour timeframes through partnerships with pre-vetted local CSOs and SOGIESC-trained experts.
At the same time, regional bodies like the African Union need to establish legally binding norms for digital protection that obligates platforms to take into consideration the realities of criminalised contexts. In addition, rather than requiring more primary testimony from vulnerable groups, policy reform should rely on the substantial secondary evidence that has already been established.

Without these interventions, social media platforms will remain complicit in a digital rights failure that not only endangers lives but undermines the broader goal of digital inclusion across the African continent.

References

‣ ILGA (International Lesbian, Gay, Bisexual, Trans, and Intersex Association). (2023). State Sponsored Homophobia Report.

‣ Ethiopian Penal Code (2004), available at retrieved from, https://www.unodc.org/cld/uploads/ res/document/eth/2005/ the_criminal_code_of_the_federal_democratic_republic_of_ethiopia_2004_html/ Criminal_Code_2004_Official_English.pdf

‣ Key Informant Interview (KII.2) with CSO member. (2026).

‣ Elswah, S., & Vashistha, V. (2025). Think Outside the Data: Colonial Biases and Systemic Issues in Automated Moderation Pipelines for Low-Resource Languages.

‣ Ruggie, J. G. (2011). Guiding Principles on Business and Human Rights: Implementing the United Nations "Protect, Respect and Remedy" Framework. United Nations. ‣ Access Now. (2024). Rainbow burning: Social media risks for LGBTQ+ people in Africa. Access Now, retrieved from https://www.accessnow.org/publication/social-media companies-increase-risks-lgbtq-africa

‣ Association for Progressive Communication & Feminist Internet Research Network. (2025). Queer Resistance: Online Gender-Based Violence Against LGBTIQ+ Ethiopians, retrieved from https://www.apc.org/en/pubs/queer-resistance-online-gender-based-violence against-lgbtqia-ethiopians

‣ Siegel, M., & Nyarko, A. (2023). Data Transparency Frameworks for Content Moderation. Stanford Internet Observatory.

‣ Human Rights Watch. (2023, February 21). "All of the Terror Because of the Photo": Digital targeting and its offline consequences for LGBT people in the MENA region, retrieved from https://www.hrw.org/report/2023/02/21/all-terror-because-photo/digital-targeting-and-its offline-consequences-lgbt

‣ Human Rights Watch. (2024). Documentation on Human Rights Violations Against LGBTIQ+ Communities in Ethiopia, retrieved from https://www.hrw.org/world-report/2024/country chapters/ethiopia

‣ European Commission. (2022). Regulation (EU) 2022/2065 of the European Parliament and of the Council on a Single Market For Digital Services (Digital Services Act). ‣ Key Informant Interview (KII 1) with CSO member. (2026).

‣ Federal Democratic Republic of Ethiopia. 2009. Charities and Societies Proclamation No. 621/2009, art. 69. Federal Negarit Gazeta, 15th Year, No. 25. Addis Ababa, retrieved from https://chilot.wordpress.com/wp-content/uploads/2011/02/ charities20and20societies20proclamation.pdf

‣ Meta. (n.d.). Privacy policy. Facebook, retrieved, from https://m.facebook.com/privacy/policy ‣ X. (n.d.). Privacy policy. https://x.com/en/privacy

‣ TikTok. n.d. Privacy Policy (US). https://www.tiktok.com/legal/page/us/privacy-policy/en. ‣ TikTok. n.d. Privacy Policy (EEA/UK/Switzerland). https://www.tiktok.com/legal/page/eea/ privacy-policy/en.

‣ TikTok. n.d. Global Privacy Center. https://www.tiktok.com/privacy/overview/en. ‣ Google. (n.d.). Privacy Policy. https://policies.google.com/privacy?hl=en-US ‣ Google. (n.d.). Your Data in YouTube.https://myactivity.google.com/product/youtube ‣ YouTube. (n.d.). Privacy & Safety Settings. https://www.youtube.com/howyoutubeworks/ privacy



AI AS A TOOL FOR ELECTORAL INTEGRITY:

GOVERNANCE FRAMEWORKS FOR AI-ASSISTED DISINFORMATION MONITORING DURING NIGERIA'S ELECTIONS



BY HASSAN HABIBAT TEMITAYO

Artificial intelligence is no longer a distant technology; it has arrived at the heart of Nigerian democracy. In recent years, AI has fundamentally changed how political information is created and shared, giving rise to tools that can produce and spread false content faster and more convincingly than anything seen before. Through deepfake videos, voice-cloning technology, bot-driven social media campaigns, and algorithmically generated fake news, political actors can now shape public opinion, silence inconvenient truths, and deceive millions of citizens at the click of a button. What makes AI-generated disinformation particularly dangerous is how believable it is. It can place words in the mouths of real people, mimic their voices, and replicate their faces with alarming accuracy, leaving ordinary voters, journalists, and election officials struggling to tell fact from fiction. Yet this same technology, if properly governed, holds enormous potential to do the opposite to detect disinformation, monitor its spread in real time, and protect the integrity of elections before damage is done.

Nowhere was this tension more visible than during Nigeria's 2023 general elections. Voters were confronted with a flood of AI-generated content designed to mislead and inflame. A fabricated audio recording falsely suggested that presidential candidate Peter Obi and a prominent religious figure had privately discussed starting a "religious war." Another doctored audio purportedly captured candidate Atiku Abubakar in a conversation about rigging the election. Deepfake videos promoted false claims about Bola Ahmed Tinubu's health, while fabricated endorsements from global figures like Elon Musk and Donald Trump gave an air of international legitimacy to messages that were entirely invented. Perhaps most brazen was a deepfake showing Peter Obi claiming to have lifted over 439 million people out of poverty, a number that exceeds Nigeria's entire population. These were not minor incidents. They stoked ethnic and religious fears, confused voters, and chipped away at public trust in the electoral process itself. The consequences were real and measurable: voter turnout fell from 35% in 2019 to just 27% in 2023 the lowest in Nigeria's democratic history.

What made 2023 especially troubling was not just the scale of the disinformation but how poorly equipped Nigeria's institutions were to respond. The Independent National Electoral Commission (INEC) and the National Broadcasting Commission (NBC) had no standardised procedures for identifying or responding to AI-generated content in real time. False information circulated freely
for hours and sometimes days before any official response came and by then, the damage was already done. Nigeria's existing laws were no better prepared. The Cybercrimes Act and existing media regulations were written for a different era and make no clear provision for synthetic media or AI-generated electoral manipulation. The Nigeria Data Protection Act 2023, while a step forward on privacy, says very little about algorithmic accountability, AI transparency, or who bears responsibility when an automated system causes harm.

The result is a legal and institutional vacuum that bad actors both within Nigeria and from outside have been quick to exploit. Meanwhile, other parts of the world have not stood still. The European Union's AI Act, the United States' NIST AI Risk Management Framework, and international guidance from UNESCO and the OECD have all moved to establish clearer rules around AI transparency, human oversight, and accountability. Closer to home, the African Union's Malabo Convention has called for coordinated continental action on cybersecurity and data protection. Nigeria has yet to develop anything comparable and that gap is widening with every election cycle.

This article argues that Nigeria can no longer afford to respond to AI-driven disinformation after the fact. What is needed is a forward-looking, institutionalised system, one that uses AI's own capabilities to monitor, detect, and neutralise disinformation before it takes root in public consciousness.

To that end, this article draws on the Hybrid Political Information Governance Framework HYBRIPOLIN first proposed by Jamiu Folarin et al., and extends its application to the specific challenge of AI-assisted disinformation monitoring during Nigeria's elections, incorporating legal reform dimensions absent from the original framework. as an original governance model designed specifically for Nigeria's electoral context. At its core, HYBRIPOLIN centres on an AI-assisted monitoring system capable of detecting, flagging, and tracking synthetic media in real time across social media platforms during election periods. But it goes beyond technology. The framework calls for the establishment of a National Electoral Cybersecurity Taskforce, mandatory transparency labelling for AI-generated political content, and new legislation that specifically addresses the use of AI in electoral spaces. It also recognises that laws and machines alone cannot solve this problem; media literacy, stakeholder collaboration, and a genuine sense of democratic ownership among citizens are just as important. By treating AI not only as a threat to democracy but as a tool for defending it, this study makes the case that Nigeria has both the capacity and the responsibility to build a governance system that is as sophisticated as the disinformation it must confront.

LITERATURE REVIEW

Theme 1: AI and Disinformation Globally

Not long ago, running a disinformation campaign required significant resources, teams of people, coordinated effort, and considerable time. Today, a single individual with access to the right AI tools can produce thousands of convincing fake news articles, voice recordings, and videos in a matter of hours. This is the reality that artificial intelligence has created, and it has fundamentally changed what it means to protect the truth in a democratic society. Researchers describe this phenomenon as "computational propaganda" the use of automated systems, bots, and algorithmically generated content to flood public spaces with false information, carefully targeted to exploit the fears, biases, and beliefs of specific communities.

The evidence of how damaging this can be is not hard to find. During the 2016 United States presidential election, Russian-linked bot networks manipulated political conversations on social media by targeting specific voter communities with tailored messaging designed to deepen division and suppress turnout. By 2019, voters in India were confronted with deepfake videos of candidates saying things they never said, distributed widely enough to reach millions before anyone could verify their authenticity. These were not isolated incidents, they were early warnings of a technology that was only going to become more accessible and more powerful.

Africa has not been spared. If anything, the continent faces a version of this threat that is even harder to manage. Low digital literacy, limited fact-checking infrastructure, and the dominance of mobile internet have created conditions where AI-generated disinformation travels fast and encounters little resistance. Researchers estimate that as much as 60% of foreign-sponsored disinformation operations globally now target African countries, and the effects are already visible. In Burkina Faso, Gabon, and Senegal, AI-generated propaganda has been used to manipulate election results and destabilise political environments. In Kenya's 2022 elections, AI-powered bots were deployed not on open social media platforms where they might be detected, but inside private WhatsApp groups where they amplified ethnic tensions and spread divisive content in spaces that fact-checkers and regulators simply cannot access. This detail matters enormously for Nigeria, where WhatsApp is one of the most widely used platforms for political communication.

Theme 2: Nigeria's 2023 Election Specifically

If any election demonstrated just how dangerous AI-generated disinformation can be in the Nigerian context, it was the 2023 general elections. What voters experienced during that period was not a scattered collection of rumours, it was a coordinated, well-funded, and technically sophisticated campaign of synthetic media designed to confuse, frighten, and mislead. Shaibume and Avindigh (2026) describe it plainly as a crisis of electoral integrity, driven by politically motivated actors who used deepfakes and AI-generated audio to distort facts and undermine public trust in ways that traditional fact-checking simply could not keep pace with.

Gwadi and Igbashangev (2024) acknowledge that tools like the Dubawa Chatbot and MyAIFactChecker were deployed in response, representing a genuine attempt to fight back. But they are honest about the limitations Nigeria's linguistic diversity alone made it enormously difficult for these tools to catch false content circulating in Hausa, Yoruba, Igbo, and Nigerian Pidgin. By the time a piece of disinformation was verified and debunked, it had already reached its audience. Omilusi (2025) adds a detail that makes the scale of the problem even clearer: this disinformation was not spontaneous. Influencers were paid as much as twenty million naira to target political opponents, turning the spread of falsehoods into an organised, commercially driven industry operating in plain sight of the electoral process. The damage did not stop on election day either; the shadow of these campaigns followed proceedings into the Presidential Election Tribunal, where claims of manipulation continued to undermine confidence in the results.

What made all of this worse was how INEC responded or failed to. The Commission's own electronic result transmission system, the IReV portal, broke down during the process, creating exactly the kind of institutional vacuum that disinformation thrives in. When citizens could not trust the official process, they turned to social media and social media was already full of algorithmically amplified narratives of rigging and fraud. As Shaibume and Avindigh (2026) conclude, INEC was reactive throughout, never ahead of the crisis. The human cost of all this is visible in one number: voter turnout fell from 35% in 2019 to just 27% in 2023, the lowest in Nigeria's democratic history.

THEME 3: LEGAL AND GOVERNANCE GAPS

Here is the hard truth about Nigeria's legal framework when it comes to AI-generated disinformation. It was not built for this problem, and in some ways it has made things worse. The Nigeria Data Protection Act 2023 is a genuine step forward on privacy and data governance, and it deserves credit for that. But when you look at what it actually says about AI, the silence is striking. It says nothing meaningful about algorithmic accountability who is responsible when an automated system spreads harmful content. It says nothing about bias in AI systems or about the specific damage caused by synthetic media that places real people in fabricated scenarios without their knowledge or consent. For a law passed in 2023, when AI disinformation was already a documented crisis, that silence is difficult to explain.

The Cybercrimes Act tells an even more troubling story. On paper, it was designed to protect Nigerians from digital threats. In practice, its broad and poorly defined provisions have repeatedly been used against the wrong people journalists, opposition figures, and ordinary citizens who criticize those in power. Rather than targeting the actors deploying AI to manipulate elections, the law has too often been turned into an instrument of suppression. This is not just a technical gap, it is a fundamental misalignment between what the law says and what Nigeria actually needs it to do.

What currently exists in Nigeria focuses largely on hate speech and basic content moderation. There is no legal framework that addresses how algorithms amplify harmful content, no requirement for platforms to be transparent about how political content is curated, and no clear accountability for the use of AI in electoral spaces. The grey areas coordinated inauthentic behaviour, synthetic media, automated bot networks remain largely untouched by law. Meanwhile, the rest of the world has not stood still. The European Union's AI Act requires transparency and human oversight in high-risk AI systems. The OECD AI Principles and UNESCO's Recommendation on the Ethics of Artificial Intelligence both establish clear expectations around accountability and risk management. Nigeria has nothing comparable and every election cycle that passes without reform is another opportunity for those who wish to exploit that gap.

GAP STATEMENT

Researchers have done important work documenting how AI disinformation operates globally and what went wrong during Nigeria's 2023 elections. But there is something missing from the conversation: nobody has yet proposed a governance framework that is both legally grounded and technically equipped to actually do something about it in the Nigerian context. No existing study has brought together real-time AI-assisted monitoring, the specific requirements of the NDPA 2023, and the reforms needed in the Cybercrimes Act into a single, workable framework for protecting electoral integrity.

METHODOLOGY

Research Design

This study does not rely on a single method to answer its questions and deliberately so. The problem of AI-generated disinformation during Nigerian elections is too layered for numbers alone to capture, and too complex for interviews alone to measure. For that reason, this research adopts a mixed-methods design that brings together quantitative data analysis and qualitative interviews, allowing the findings to speak both statistically and humanly. The quantitative strand tells us how far disinformation travelled and how much engagement it attracted. The qualitative strand tells us why institutions failed to stop it and what the law needs to say about it. Together, they produce a more complete picture than either could offer alone.

Quantitative Strand :Data Source

The quantitative component of this study draws on an existing Twitter dataset compiled specifically around Nigeria's 2023 presidential election. Given that Twitter now rebranded as X significantly restricted third-party data access from 2023 onwards, independently collecting fresh data from the platform was not a viable option for this research. Instead, the study relies on a publicly available, peer-acknowledged dataset hosted on the Mendeley Data repository:

Odegbile, A.T. & Oyelami, O.M. (2024). Nigeria's 2023 Presidential Election Dataset. Mendeley Data, V2. doi: 10.17632/whb6rychpx.2.

This dataset was compiled by mining tweets related to the three major presidential candidates Peter Obi, Bola Ahmed Tinubu, and Atiku Abubakar using candidate-specific hashtags including #obi, #atiku, and #bat. A total of 367,701 tweets were initially collected, with a final cleaned dataset of 364,867 tweets after a preprocessing phase that removed stopwords, emojis, special characters, and duplicate entries. Sentiment analysis was subsequently applied to 325,930 entries using the cardiffnlp/twitter-roberta-base-sentiment pre-trained model, selected for its proven capacity to handle the informal language, abbreviations, and contextual expressions commonly found in social media content.

What Was Measured Quantitatively

The dataset provides several quantitative measures that are directly relevant to this study's research questions. First, tweet volume was tracked measuring the sheer scale of election-related social media activity across the campaign period. Second, engagement metrics for each tweet were recorded, including the number of retweets, replies, and quote tweets indicators of how far individual pieces of content travelled and how much public interaction they generated. Third, sentiment metrics were captured, with tweets categorised using numerical sentiment labels to measure the emotional temperature of political content over time. Fourth, temporal distribution data tracked the monthly spread of tweets from August 2022 through February 2023, making it possible to identify peaks in activity that correspond to key moments in the electoral cycle. Finally, model performance was assessed using accuracy scores, F1 scores, and the Area Under the Curve (AUC) metric to ensure the reliability of the sentiment classification applied.

For the purposes of this study, these measurements are used to trace patterns in the volume and engagement of disinformation-related content during the election period, and to identify the points at which false narratives achieved their widest reach.

Qualitative Strand Interview Sample and Selection

The quantitative data tells us what happened on social media during the 2023 elections. The qualitative component of this study asks a different set of questions: what the law says about it, what institutions did about it, and what needs to change.

Purposive sampling was chosen deliberately. This research does not need a random cross-section of the Nigerian population, it needs people who have direct, informed experience of the specific institutions, tools, and legal frameworks at the centre of the study.

This study draws on a supplementary key-informant interview conducted with a fact-checker affiliated with a leading Nigerian fact-checking organisation. This interview was not intended to produce a representative qualitative sample, but rather to provide grounded, practitioner-level insight into the specific operational challenges of AI-based fact-checking tools during the 2023 elections — insight that is difficult to obtain from published literature alone. The participant's identity is withheld in accordance with standard research confidentiality practices; they are referred to throughout this article as "a Nigerian fact-checker."

How the Data Was Analysed

The quantitative data from the Mendeley dataset was analysed using descriptive statistical methods examining frequencies, engagement rates, and temporal distributions to identify patterns in how election-related content, including disinformation, spread across the platform during the 2023 electoral cycle. These patterns were then mapped against documented incidents of AI-generated disinformation to examine the relationship between content type and audience reach.

The qualitative interview data was analysed using thematic analysis, following the framework established by Braun and Clarke (2006). This involved reading through transcripts carefully, identifying recurring ideas and patterns, grouping these into initial codes, and then developing broader themes that cut across the responses of different participants. Three primary themes emerged from this process: the inadequacy of the existing legal framework, the institutional and technical capacity gaps in Nigeria's response, and the structural requirements of an effective governance framework each of which maps directly onto one of the study's sub-questions.

Ethical Considerations

This study relies entirely on secondary, publicly available data and published legal texts, and therefore does not involve direct interaction with human participants. The Mendeley dataset used
for the quantitative analysis is anonymised at the point of publication, contains no personally identifying information beyond what was already publicly posted on Twitter/X, and is made available for academic research purposes. No attempt was made to identify individual users behind the tweets analysed, and all reporting of findings is conducted at the aggregate level. The legal texts examined are publicly accessible government and policy documents. As such, this research raises minimal ethical risk, though the researcher remains attentive to the broader ethical responsibility of representing the data and legal analysis accurately and without distortion.

CORE FINDING

The evidence assembled in this study points toward one clear, central argument: Nigeria's existing legal and institutional framework is not equipped to handle AI-generated electoral disinformation, and a proactive, AI-assisted governance framework grounded in algorithmic transparency, human oversight, and targeted legal reform is both necessary and within reach. Each of the findings below speaks to one part of that argument.

Existing computational studies of Nigeria's 2023 election tell a consistent story of an online environment heavy with negative sentiment, deep polarisation, and artificial amplification.

Attai et al. (2024) found that nearly a quarter of all election-related discourse on Twitter carried negative sentiment, reflecting widespread dissatisfaction and scepticism toward the electoral process. Oyewola et al. (2023) similarly found sharp divergences in how candidates were discussed online. Separately, Ajueyitsi and Ekuobase (2024) read the high neutral-sentiment scores as a clear signal of voter apathy, a finding that sits uncomfortably close to the recorded drop in turnout from 35% to 27%.

Perhaps the most striking detail is the evidence of artificial amplification. A bot analysis conducted by the Heinrich Böll Foundation (2022) found that a significant proportion of leading candidates' social media followers were automated accounts, created to simulate grassroots support and drown out genuine voices. The disinformation environment during the 2023 elections was not simply a matter of individuals sharing false content; it was, in part, deliberately engineered. According to the Nigerian fact-checker interviewed for this study, the most prevalent form of AI-generated disinformation during the 2023 elections was manipulated images and videos content that travels faster precisely because it is visual. Two examples stood out: deepfake videos showing Hollywood actors purportedly endorsing Peter Obi, and fabricated endorsements from Elon Musk and Donald Trump. The fact-checker observed that endorsements from recognisable international figures carry disproportionate persuasive weight with Nigerian audiences, which is what made these fabrications so effective.
Audio disinformation was the hardest to resolve. The fact-checker pointed to a leaked audio recording purportedly capturing a conversation between Peter Obi and Pastor David Oyedepo. Despite considerable public pressure, fact-checkers were simply unable to determine in real time whether the recording was authentic or AI-generated. Verification relied on tools including Google Reverse Image Search, Google Lens, Hive Moderation, and Is It AI but the fact-checker was honest about their limits. AI-generated content is sometimes classified as authentic, which is why organisations have learned to triangulate across multiple tools rather than trust any single platform.

Language made everything harder. Content in Hausa, Yoruba, Igbo, and Nigerian Pidgin frequently carries cultural references and contextual meanings that automated systems simply cannot read accurately, making human verification not a fallback option but a practical necessity. Political slang, coded language, and dog whistles that a Nigerian audience would recognise instantly tend to pass through automated systems entirely undetected, increasing false negatives precisely where the political stakes are highest.

The NDPA 2023 provides a reasonable starting point for data privacy, but it falls well short of what is needed when AI enters the electoral space. A legal expert interviewed for this study was direct about this: the Act's primary purpose is regulating how personal data is collected and processed, not how it is weaponised to deceive voters. Its provisions might apply in narrow circumstances for instance, where AI-generated content is built from unlawfully obtained personal data but the Act was simply not designed to regulate misinformation, deepfakes, synthetic media, or election manipulation. Section 30, which deals with automated decision-making, offers no real way for individuals to contest automated outcomes and contains no definition of synthetic media, no provision assigning liability when AI-generated content misrepresents a real person, and no requirement for algorithmic transparency. As the legal expert observed, without a statutory definition of AI systems, deepfakes, and algorithmically generated content, no enforcement is practically possible.

The Cybercrimes Act tells a different but equally concerning story. The legal expert noted that while provisions relating to identity theft, cyberstalking, and fraudulent misrepresentation could theoretically be stretched to cover some cases of AI-generated electoral manipulation, the Act does not specifically criminalise the creation or dissemination of AI-generated electoral disinformation. There is no legal definition of deepfakes, no specific offence addressing fake media used to deceive voters, and no obligation to label AI-generated political content. More troubling still, the literature documents a consistent pattern in which the Act's broadly worded provisions have been used to pursue journalists and government critics rather than the actors actually deploying AI for electoral manipulation, the opposite of what a disinformation law should do. The result is what the
literature calls the 'liar's dividend' where the mere existence of deepfake technology gives bad actors the cover to dismiss genuine evidence as fabricated.

Two additional gaps surfaced in the legal expert's responses that the published literature has not fully addressed. The first is jurisdictional. Many of the disinformation campaigns that affected Nigeria's 2023 elections were coordinated from outside the country's borders, which means Nigerian law, however well-reformed, cannot reach them without bilateral and multilateral cooperation agreements with other jurisdictions. The second is evidentiary. Identifying the original creator of AI-generated content is technically difficult, and current Nigerian investigative infrastructure is not equipped to do it reliably. Both gaps matter because they mean that even a strengthened legal framework will remain partially toothless without complementary investment in international cooperation and digital forensics capacity.

INEC's failure to upload presidential results to the IReV portal in real time despite repeated public assurances created a transparency gap that algorithmically amplified rigging narratives rushed to fill. The Commission's communication was reactive throughout: a clarifying statement on server-breach allegations did not appear until midday the following day, by which point the false narrative had already reached and shaped its audience.

The NBC and the Nigerian Communications Commission were in no better position. Neither body had AI-detection mechanisms of any kind, making real-time monitoring of deepfakes effectively impossible. There were no standardised procedures for tracing where synthetic content came from or for issuing takedowns. What runs beneath all of this is a structural problem the literature keeps returning to: Nigeria's institutional response to disinformation has been built almost entirely around punishment after the fact, with very little investment in prevention.
The comparative picture, however, offers genuine encouragement. On algorithmic transparency, the EU AI Act's Article 13 already requires high-risk AI systems to be explainable, and mandates visible disclaimers on synthetic content. The UK's Algorithmic Transparency Recording Standard gives public bodies a workable template for disclosing which AI tools they use and how. On human oversight, Article 14 of the EU AI Act requires that human overseers retain the authority to disregard, override, or reverse AI-driven outputs. On legal reform, South Korea has banned AI-generated deepfakes in election campaigning within a defined pre-election window; sixteen US states now criminalise the use of manipulated media in elections; the Philippines and Brazil both mandate labelling of AI-generated campaign content. Taken together, these examples span a spectrum from disclosure requirements to hard-law prohibitions giving Nigeria not a single model to copy, but several viable entry points to choose from. As the legal expert concluded, the goal
should not be to suppress political speech but to protect electoral integrity focusing regulation on deceptive conduct rather than opinions or political criticism, a principle that is both constitutionally defensible and practically achievable.

THE ADAPTED GOVERNANCE FRAMEWORK: HYBRIPOLIN-AI

The governance framework proposed in this article builds on the Hybrid Political Information (HYBRIPOLIN) model first developed by Folarin, Adesina, Odunlami, and Bello a model conceived specifically to 'puncture the political disinformation bubble' through a coordinated, multi-layered approach. Their original eight-pillar structure remains, in many ways, the right architecture for thinking about Nigeria's disinformation problem. It correctly identifies that no single intervention, however well-designed, can succeed in isolation.

What the original HYBRIPOLIN model does not do is engage with the specific machinery of AI-assisted monitoring, the precise legal amendments required of the NDPA 2023 and the Cybercrimes Act, or the institutional design of a body capable of acting on what such monitoring detects. This article takes HYBRIPOLIN's eight pillars as its foundation and develops three of them Technological Innovation, the Legal Approach, and Electoral Reform into a fully specified, AI-assisted governance architecture referred to throughout as HYBRIPOLIN-AI.

At the centre of HYBRIPOLIN-AI sits a real-time monitoring system designed to detect, flag, and track synthetic media across social media platforms during the electoral period.

As this study's findings on Sub-question 1 make clear, tools such as the Dubawa Chatbot and MyAIFactChecker have struggled precisely because Nigeria's linguistic landscape was never the environment these tools were originally trained for.

HYBRIPOLIN-AI therefore treats the development of localised Nigerian-language datasets and locally-trained detection models not as an optional enhancement but as a foundational requirement. The fact-checker interviewed for this study confirmed that content in Hausa, Yoruba, Igbo, and Nigerian Pidgin carries cultural references and contextual meanings that automated systems consistently miss. Without locally-trained models, any monitoring layer will continue to fail at precisely the content that matters most. Furthermore, the false-negative rate of existing tools where AI-generated content is sometimes classified as authentic means the monitoring layer must combine multiple detection methods rather than relying on any single classifier.

The second component proposes that the monitoring layer feed into a National Electoral Cybersecurity Taskforce a body with both preventive and enforcement authority, jointly anchored by INEC and the National Information Technology Development Agency (NITDA), with
mandatory representation from the Nigeria Data Protection Commission, the Office of the National Security Adviser, and independent civil society and technical experts.

INEC alone lacks the technical cybersecurity expertise this work demands; NITDA alone lacks electoral jurisdiction. A joint taskforce, with oversight sitting in the National Assembly through a dedicated committee, addresses both the capacity and accountability gaps simultaneously. Crucially, HYBRIPOLIN-AI insists that this human layer is not advisory. Drawing on the evidence that human-AI collaboration consistently outperforms either acting alone provided the humans involved are domain experts, the Taskforce retains explicit authority to disregard, override, or escalate any output from the monitoring layer. The AI system scales the work of detection; the humans remain, in every meaningful sense, in charge of what happens next.

The third component gives the previous two their teeth. HYBRIPOLIN-AI specifies five concrete amendments to the NDPA 2023: a clear legal definition of synthetic media and AI-generated content; an explicit prohibition on AI-generated content that falsely depicts a real individual in an electoral context without consent; algorithmic accountability obligations requiring platforms to disclose how their systems amplify political content; a technically resourced enforcement arm within the Nigeria Data Protection Commission; and expanded liability provisions covering AI developers and deployers, not solely data controllers.

Alongside this, HYBRIPOLIN-AI adopts a tiered liability model primary criminal liability for creators who act with demonstrable intent to manipulate an election, secondary civil liability for platforms whose algorithms amplify the content, and party-level consequences including potential deregistration where a political party commissioned the disinformation. Throughout, HYBRIPOLIN-AI follows a clear constitutional guardrail: regulation targets the mechanism of harm fabrication, impersonation, coordinated inauthentic amplification rather than the content of political opinion. Mandatory transparency labelling, not content prohibition, is the default; takedown powers require judicial authorisation; and an independent appeals mechanism protects against wrongful removal.

What distinguishes HYBRIPOLIN-AI from a list of recommendations is that each component depends on the others. The monitoring layer is useless without a body authorised to act on what it finds. The Taskforce is powerless without a legal framework that defines offences and assigns liability. And the legal framework is unenforceable without a technical system capable of detecting the conduct it prohibits. This interdependence is the truest expression of HYBRIPOLIN's original insight that disinformation cannot be addressed by any single pillar acting alone.

THE SOCIETAL FOUNDATIONS OF HYBRIPOLIN-AI

The three components developed in the previous section form the hard architecture of HYBRIPOLIN-AI. But the original HYBRIPOLIN model's deepest insight was never that technology and law alone could solve Nigeria's disinformation problem. Its remaining five pillars exist precisely because a detection system and a taskforce operate within a society. If that society remains highly susceptible to the narratives disinformation exploits, the hard architecture will spend its energy treating symptoms while the underlying vulnerability persists.

Media and Digital Literacy

Every technical safeguard in HYBRIPOLIN-AI ultimately depends on one thing it cannot itself provide: a citizenry capable of pausing before sharing. The fact-checker interviewed for this study made this point directly: no detection system will outrun disinformation that audiences are not equipped to question for themselves. Media literacy in the Nigerian context cannot mean a generic awareness campaign about 'fake news.' It needs to be specific to the forms of manipulation documented in this study's findings recognising that a video showing a celebrity endorsement may be fabricated, that an audio recording can be cloned convincingly enough to fool even trained fact-checkers, and that a sudden flood of similar messages may indicate coordinated amplification rather than genuine grassroots sentiment. Given that low digital literacy has already been identified as making voters more susceptible to disinformation, this pillar needs to be embedded into school curricula, community radio programming, and the public communication strategies of INEC itself.

Re-orientation

If media literacy addresses how people consume information, re-orientation addresses why certain disinformation works so effectively in the first place. This study's findings have already shown that the most damaging content during the 2023 elections was deliberately constructed to stoke ethnic and religious division, precisely because those divisions already exist as live fault lines in Nigerian political life. A fabricated audio recording suggesting a 'religious war' succeeds not because the technology is sophisticated, but because it lands on ground already prepared by decades of politicians mobilising along ethnic and religious lines. What re-orientation calls for is a sustained civic effort to build what the original HYBRIPOLIN model describes as 'democratic ownership' , a shared sense among citizens that the electoral process belongs to all of them.

NGO and Civil Society Collaboration

Civil society organisations occupy a position that neither government nor AI systems can fully replicate: they are trusted by communities in ways that state institutions, particularly INEC, currently are not. Within HYBRIPOLIN-AI, civil society's role is twofold. First, organisations such as Dubawa and the NFC provide the human expertise that Layer 2's human-in-the-loop
principle depends on. Second, they serve as a credibility bridge between the Taskforce and the public. When the Taskforce flags content as disinformation, that determination will carry more weight with sceptical audiences if it is corroborated by fact-checking organisations the public already follows, rather than announced solely by a government body.

Fact-Checking as a Feedback Mechanism

The temptation might be to treat AI-assisted monitoring as a replacement for organisations like Dubawa. This study's findings suggest the opposite: these organisations should become partners in training the monitoring layer. Every time a fact-checker manually resolves a case the AI tools could not, the Obi/Oyedepo audio being the clearest example that resolution is valuable data.

A structured pipeline that feeds verified fact-checking outcomes back into the training data for Layer 1's detection models would, over successive election cycles, gradually close the gap between English-language and indigenous-language performance. This is the practical expression of the Adaptive Feedback principle from the original concept note.

Broader Electoral Reform

The failure to establish an Electoral Offences Tribunal, despite provision for one in the 2022 Electoral Act, has meant that even where offences could theoretically be prosecuted, there has been no dedicated forum to do so. The technical failures of BVAS and the IReV portal were not disinformation problems in themselves but they created the institutional vacuum that disinformation narratives rushed to fill. The Taskforce's work will be considerably easier in an electoral environment where INEC's core systems function reliably and transparently, and considerably harder in one where they do not.

Academic Research as Infrastructure

The localised Nigerian-language datasets that Layer 1 depends on do not currently exist at the scale required building them is a research task that no single government agency is positioned to undertake alone. Tracking whether the Taskforce's interventions actually reduce disinformation spread is a research question. And documenting how disinformation tactics evolve from one election cycle to the next so that Layer 1's models do not become outdated is an ongoing research commitment. Universities with existing research relationships to organisations working on AI governance in Nigeria are well positioned to play this role, provided it is formalised rather than left to occasional, disconnected studies.

CONCLUSION

This article set out from a simple but uncomfortable observation: Nigeria has already lived through what AI-generated disinformation can do to an election, and it is not yet equipped to live through it again. The 2023 general elections were not a warning of some future risk, they were the risk, already realised. Fabricated audio recordings that fact-checkers could not resolve in time, deepfake videos lending false international legitimacy to political claims, bot networks simulating support that did not exist, and a turnout collapse from 35% to 27% that tracks uncomfortably closely with all of it. What this article has tried to do is take that evidence seriously enough to ask not just what went wrong, but what, specifically, would need to be true for it to go dif erently next time.

The findings assembled here point in one direction. Nigeria's legal framework the NDPA 2023, the Cybercrimes Act cannot currently name the harms this article is concerned with, and a law that cannot name a harm cannot prevent or punish it. The institutions tasked with responding, INEC chief among them, were not merely under-resourced but structurally reactive, built around a model of punishment-after-the-fact in an environment where the fact often does its damage within hours. And yet, set against this, the comparative evidence from South Korea, the EU, the Philippines, Brazil, and a growing number of US states shows that none of this is unsolvable; it is, in many places, already being solved, just not yet in Nigeria.

HYBRIPOLIN-AI, the framework developed in this article, is an attempt to close that gap honestly. It does not claim to have invented a new way of thinking about this problem that credit belongs to Folarin, Adesina, Odunlami, and Bello, whose original HYBRIPOLIN model correctly identified that no single intervention would be enough.

What this article has tried to add is specificity: a monitoring layer that takes seriously the linguistic and cultural texture of Nigerian disinformation, an oversight body designed so that no single institution can be blamed or captured alone, and legal amendments precise enough to assign responsibility without criminalising legitimate political speech. And it has tried to be honest that none of this works without the slower, harder pillars underneath it the literacy, the re-orientation, the civil society partnerships, the research infrastructure that the original model rightly insisted on.

It would be dishonest to end without acknowledging what this study could not do. The quantitative analysis relies on a secondary dataset, the most responsible choice available given Twitter/X's restrictions on data access, but a choice that means this study's empirical picture of 2023 is necessarily shaped by what that dataset captured and how. And the doctrinal analysis, while grounded in the actual text of Nigeria's laws, is necessarily a snapshot Nigerian AI policy is moving, however slowly, and parts of this analysis may need revisiting even before the next election cycle.

These limitations point toward where future research should go. The most urgent task is empirical: building and validating the Nigerian-language detection datasets that Layer 1 of HYBRIPOLIN-AI depends on, which is a substantial research project in its own right and one that this article can only call for, not complete. Equally important is the kind of research that can only happen after implementation should any version of a National Electoral Cybersecurity Taskforce be established ahead of Nigeria's next general election, tracking its actual effect on the spread and impact of disinformation, rather than assuming effectiveness from design alone, will be essential. And the broader qualitative picture of what legal experts, INEC officials, and platform representatives actually think is achievable, and on what timeline remains an open question this study could only partially answer.

What this article hopes to leave behind, more than any single recommendation, is a reframing. The title of this research speaks of AI as a tool for electoral integrity, not merely a threat to it and that framing matters. Nigeria's next election will not be free of AI-generated disinformation; that much is close to certain. The question this article has tried to answer is whether Nigeria will meet that disinformation with the same reactive posture as in 2023, or with a system imperfect, evolving, but built that can meet AI with AI, fabrication with transparency, and impunity with law. The evidence gathered here suggests the second path is not just preferable. It is, increasingly, achievable.

[1]Adewusi, B. A., et al. (2024). Conceptual Model for Responsible AI Integration in Public-Facing Digital Services. IJSREIT.

[2] Ajueyitsi, O. & Ekuobase, G. O. (2024). A Multifaceted Sentiment Analysis Approach to the Estimation of the Strength of Online Support for Political Candidates in Nigeria's Elections. FUDMA Journal of Sciences

[1] Adewusi, B. A., et al. (2024). Conceptual Model for Responsible AI Integration in Public-Facing Digital Services. IJSREIT.

[2] Ajueyitsi, O. & Ekuobase, G. O. (2024). A Multifaceted Sentiment Analysis Approach to the Estimation of the Strength of Online Support for Political Candidates in Nigeria's Elections. FUDMA Journal of Sciences

[3] Attai, K., Asuquo, D., Okonny, K.E., Johnson, E.A., Bassey, A., John, A., Bardi, I., Iroanwusi, C. & Michael, O. (2024).

Sentiment Analysis of Twitter Discourse on the 2023 Nigerian General Elections. European Journal of Computer Science and Information Technology, 12(4), 18–35.

[4]Ayeni, T. P., et al. (2024). Examining the Prospects and Challenges of Artificial Intelligence in African Democracy. EKSU Journal of Engineering Technology.

[5]Essien, E. S. (2024). Artificial Intelligence Ethics and Regulatory Framework in Nigeria. Philosophia Politica.

[6]Folarin, J., Adesina, I.K., Odunlami, A.T. & Bello, S. (2023). Hybrid Approach to Puncturing Political Disinformation Bubble Ahead of 2023 General Elections in Nigeria.

[7]Gwadi, I. W. & Igbashangev, P. (2024). Evaluating the Impact of AI on Fact-Checking Social Media Content in Nigeria. SSRN.

[8]Heinrich Böll Foundation (2022). Bots and Biases: The Role of Social Media in Nigeria's Elections. Abuja Office, Nigeria. Available at: https://ng.boell.org/en/2022/10/26/bots-and-biases-role-social-media-nigerias-elections
[3] Attai, K., Asuquo, D., Okonny, K.E., Johnson, E.A., Bassey, A., John, A., Bardi, I., Iroanwusi, C. & Michael, O. (2024).

Sentiment Analysis of Twitter Discourse on the 2023 Nigerian General Elections. European Journal of Computer Science and Information Technology, 12(4), 18–35.

[4] Ayeni, T. P., et al. (2024). Examining the Prospects and Challenges of Artificial Intelligence in African Democracy. EKSU Journal of Engineering Technology.

[5] Essien, E. S. (2024). Artificial Intelligence Ethics and Regulatory Framework in Nigeria. Philosophia Politica.
[6] Folarin, J., Adesina, I.K., Odunlami, A.T. & Bello, S. (2023). Hybrid Approach to Puncturing Political Disinformation Bubble Ahead of 2023 General Elections in Nigeria.

[7] Gwadi, I. W. & Igbashangev, P. (2024). Evaluating the Impact of AI on Fact-Checking Social Media Content in Nigeria. SSRN.

[8] Heinrich Böll Foundation (2022). Bots and Biases: The Role of Social Media in Nigeria's Elections. Abuja Office, Nigeria. Available at: https://ng.boell.org/en/2022/10/26/bots-and-biases-role-social-media-nigerias-elections

[9]Kolade, T. M., et al. (2024). Artificial Intelligence and Information Governance: Strengthening Global Security. Asian Journal of Research in Computer Science

[10]Odegbile, A.T. & Oyelami, O.M. (2024). Nigeria's 2023 Presidential Election Dataset. Mendeley Data, V2. doi: 10.17632/whb6rychpx.2.

[11]Olabanjo, O., et al. (2023). From Twitter to Aso-Rock: A Sentiment Analysis Framework for Understanding Nigeria 2023 Presidential Election. Heliyon.

[12]Omilusi, M. (2025). Fake News, Election-Related Disinformation Laws, and Citizens' Rights in African Political Ecology. Journal of African Elections.

[13]Oyewola, D.O., et al. (2023). Optimizing Sentiment Analysis of Nigerian 2023 Presidential Election Using Two-Stage Residual Long Short Term Memory. PMC/Heliyon.

[14]Towards the Implementation of Data Protection Measures to Safeguard Against Surveillance Abuse in Nigeria. (2024). Accountability Lab.

[15]The 'Algorithmic Ballot': Fortifying Democratic Elections Against AI-Driven Disruption. SSRN.

[16] Kolade, T. M., et al. (2024). Artificial Intelligence and Information Governance: Strengthening Global Security. Asian Journal of Research in Computer Science

[17] Olabanjo, O., et al. (2023). From Twitter to Aso-Rock: A Sentiment Analysis Framework for Understanding Nigeria 2023 Presidential Election. Heliyon.

[18] Omilusi, M. (2025). Fake News, Election-Related Disinformation Laws, and Citizens' Rights in African Political Ecology. Journal of African Elections.

[19] Oyewola, D.O., et al. (2023). Optimizing Sentiment Analysis of Nigerian 2023 Presidential Election Using Two-Stage Residual Long Short Term Memory. PMC/Heliyon.

[20] Towards the Implementation of Data Protection Measures to Safeguard Against Surveillance Abuse in Nigeria. (2024). Accountability Lab.

[21] The 'Algorithmic Ballot': Fortifying Democratic Elections Against AI-Driven Disruption. SSRN.

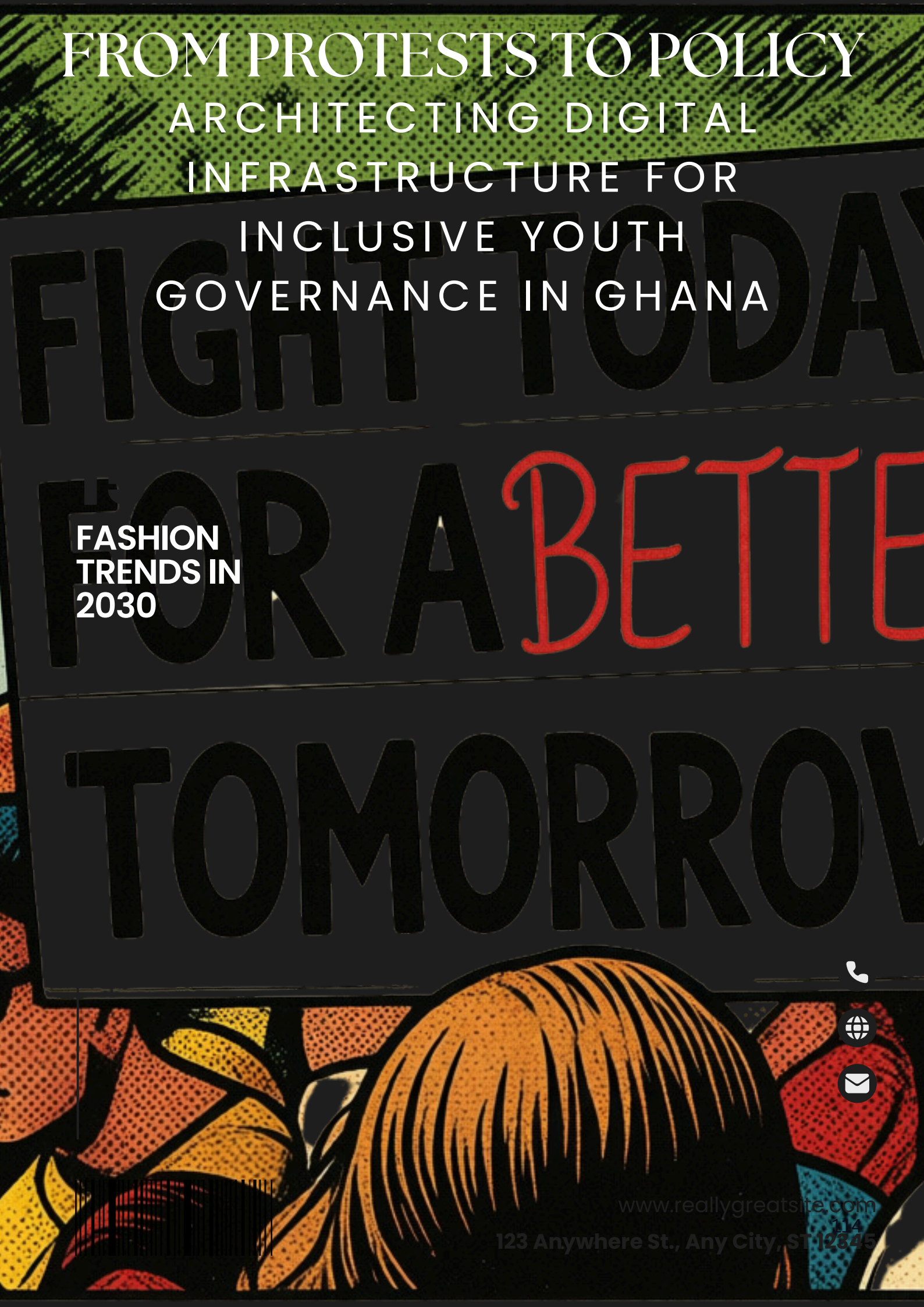
[22] Shaibume, B. & Avindigh, S. (2026). Artificial Intelligence and the Crisis of Electoral Integrity in Nigeria. Kashere Journal of Politics.

[23] Shaibume, B. & Avindigh, S. (2026). Artificial Intelligence and the Crisis of Electoral Integrity in Nigeria. Kashere Journal of Politics.

FROM PROTESTS TO POLICY

ARCHITECTING DIGITAL INFRASTRUCTURE FOR INCLUSIVE YOUTH GOVERNANCE IN GHANA

FASHION TRENDS IN 2030



www.reallygreatsite.com
114
123 Anywhere St., Any City, ST 12345

EXECUTIVE SUMMARY

Digital media platforms have become ubiquitous in media consumption (Kaempf, 2018; Ren et al., 2024), prompting a shift among media consumers, especially young people, from traditional to digital media (Kemp, 2021). More than 40% of Africans use the internet regularly (Kamer, 2022). In Ghana, young people dominate digital spaces, mobilise around social causes, and actively participate in civic discourse on platforms such as X, TikTok, Facebook, Instagram, and WhatsApp. Despite this increased digital participation, sustained influence within formal democratic institutions, particularly the Parliament of Ghana, remains limited.

In Ghana, young people aged 15-35 constitute 38% of the population, while children under age 15 make up another 35% (Ghana Statistical Service, 2021), representing a substantial share of the country's democratic constituency. Yet youth participation in formal governance processes remains a concern. Afrobarometer data specific to Ghana further indicates that younger citizens, aged 18–35, are particularly disconnected, participating less in civic and political life compared to older age groups (Afrobarometer Flagship Report, 2025)

This policy paper explores how digital platforms and digital public infrastructure can strengthen meaningful youth political participation in Ghana. It analyses the extent to which Parliament leverages digital tools to engage young citizens and examines how structural barriers, including the gender digital divide and the near-total absence of disability-responsive platforms, systematically exclude young women and youth with disabilities from civic participation. Youth with disabilities experience compounded exclusion, being underrepresented in digital spaces and completely absent from parliamentary engagement instruments that do not accommodate their needs. This issue represents not only a welfare concern but also a real democratic failure.

The paper argues that, although digital technologies have created new pathways for civic engagement, structural barriers such as limited institutional responsiveness, inaccessible platform design, and inadequate digital civic education continue to prevent young people from influencing governance outcomes. Addressing these challenges requires conscious institutional reform rather than simply expanding connectivity.

Keywords: Digital media, Digital platforms, civic engagement, marginalized youth, youth, political participation, parliaments.

INTRODUCTION

Africa is experiencing a demographic and technological transformation simultaneously. Internet and social media usage have expanded rapidly, particularly among young people who increasingly rely on digital platforms for communication, information, and civic engagement. In Ghana, rising smartphone adoption and improved connectivity have significantly increased youth participation in online spaces (Statista, 2025). As of early 2026, there were 26.3 million internet users, around 70% of the population, with mobile broadband as the dominant mode of connectivity (DataReportal, 2026). Yet despite this demographic advantage and growing digital access, youth political participation within formal governance structures remains critically limited. Many young Ghanaians engage politically online but remain excluded from offline decision-making processes.

Africa is experiencing a demographic and technological transformation simultaneously. Internet and social media usage have expanded rapidly, particularly among young people who increasingly rely on digital platforms for communication, information, and civic engagement. In Ghana, rising smartphone adoption and improved connectivity have significantly increased youth participation in online spaces (Statista, 2025). As of early 2026, there were 26.3 million internet users, around 70% of the population, with mobile broadband as the dominant mode of connectivity (DataReportal, 2026). Yet despite this demographic advantage and growing digital access, youth political participation within formal governance structures remains critically limited. Many young Ghanaians engage politically online but remain excluded from offline decision-making processes.

Political institutions are mostly controlled by older elites, and youth engagement is often restricted to election-season mobilization rather than sustained participation in governance. The digital divide comprises both a technology gap and a democracy gap. Unequal internet access limited digital literacy, inaccessible platform design, and the slow institutional adaptation of Parliament collectively serve as significant obstacles to youth participation. These obstacles disproportionately affect already marginalized groups, including young women facing a continuous gender digital divide and youth with disabilities encountering platforms and organizations not designed for their inclusion.

According to the UNDP, Digital Public Infrastructure (DPI) is a set of foundational digital systems that forms the backbone of modern societies. But for the purposes of this paper, "digital public infrastructure" refers to the essential systems, platforms, and digital tools that are either owned, operated, or mandated by the government to facilitate citizen access to public services, encourage participation in governance, and foster engagement with democratic institutions. In the context of Ghana, this includes the official Parliament website and its associated platforms, government-operated social media accounts and communication channels, and any state-supported civic technology aimed at enhancing citizen interaction with government entities.

Young people are engaged participants in public discourse, demonstrating presence, vocal advocacy, and organization on social media, in online campaigns, and through digital advocacy. However, they remain mainly invisible within the legislature, policy processes, and formal civic structures. Among those most excluded, youth with disabilities are especially affected, not only by poverty or geography but also by the failure of institutions to create accessible participation pathways.

This paradox of digital presence alongside political invisibility is not coincidental. Without a deliberate digital framework for civic engagement that is accessible, inclusive, and responsive, Parliament risks irrelevance to the generation shaping the future of Ghanaian democracy. The consequences of inaction extend beyond low youth turnout to the long-term delegitimization of democratic governance.

COUNTRY CONTEXT

By January 2025, Ghana's estimated population was 34.7 million, with internet access reaching 70% of the population. Mobile connectivity stood at 113%, and social media usage accounted for approximately 21% of the population (Data Reportal, 2025). These levels of penetration have significantly fostered civic engagement on social media platforms, particularly regarding governance issues. Many Ghanaians use online spaces to raise concerns and draw the attention of political leaders and public officials to issues such as human rights, corruption, constitutional review, transparency, accountability, and cost of living.

Sanny and Selormey (2020) observe that most Ghanaians perceive digital platforms as enhancing their awareness of current affairs and strengthening their ability to influence political processes. In a related view, Yeboah-Banin and Adjin-Tettey (2023) argue that, to fulfil their democratic role and respond to growing audiences who increasingly consume content through digital channels, traditional media organisations have been compelled to integrate digital tools and platforms into their operations to leverage these new affordances.

#FixTheCountry, for example, has sought to reshape youth activism and protest culture in Ghana by relying heavily on social media platforms, particularly X, to mobilise citizens and hold government officials and state institutions accountable

#FixThe Country

#DumsorMustStop







on issues such as #DumsorMustStop, youth unemployment, and other concerns. Social media protests have also addressed perceived injustices, including the taxation of sanitary pads and the alleged persecution of activists. This active engagement on digital platforms enables citizens to aggregate their voices, utilizing the speed and scale of social media to conduct hashtag campaigns.

These campaigns aim to exert pressure on Ghanaian authorities to address citizens' concerns. However, this civic energy remains largely confined to the digital sphere. The structural gap between online mobilization and formal institutional participation continues to limit youth political influence in Ghana and underscores the urgency of the policy challenge this paper addresses.

RESEARCH QUESTION

To what extent are existing digital platforms and parliamentary mechanisms in Ghana enabling meaningful and inclusive youth participation in governance, and what reforms are needed to address the structural exclusion of young women and youth with disabilities from democratic processes?

RESEARCH OBJECTIVES

This paper:

1. examines the current state of digital civic engagement among young people in Ghana.
2. analyses how the Parliament of Ghana is leveraging digital tools to engage youth, with attention to whether platforms are designed to reach and respond to all young citizens, including those with disabilities; examines the intersection of digital participation, gender, and disability.
3. identifying how exclusion is structurally produced.
4. and develops evidence-based, Ghana-specific policy recommendations for Parliament, with broader implications for the African Union.

METHODOLOGY

This paper adopted a qualitative research design, utilising three complementary methods: a desk-based review of existing literature and policy documents, in-depth interviews with key stakeholders, and focus group discussions. This methodological combination enabled triangulation of findings across secondary evidence, expert perspectives, and lived experience, ensuring that the analysis was grounded in both institutional realities and the perspectives of the young people most affected by the issues examined.

The choice of a qualitative approach was intentional. The central questions regarding institutional responsiveness, lived exclusion, and the gap between digital presence and political influence cannot be adequately addressed by quantitative measures alone. These issues required interpretive depth, contextual analysis, and direct engagement with the perspectives of youth, policymakers, and civil society actors who interact with these systems daily.

A total of 4 in-depth interviews were conducted, distributed across three purposively selected participant categories. One interview was conducted with a parliamentary or government official responsible for digital engagement, youth affairs, or communications, particularly staff from the Parliamentary Service. One interview was conducted with civil society and advocacy representatives working on digital rights, gender equality, or disability inclusion. The final interview was conducted with two young persons who are actively engaged in digital civic spaces, purposively selected to ensure representation across gender, disability status, and geographic location, including both urban and peri-urban participants.

In addition, one focus group discussion was conducted, comprising six participants. Deliberate steps were taken to include young women and youth with disabilities in the focus group.
Participants were further provided with consent forms in advance and had the option to remain anonymous if they so wish.

It is important to distinguish between the research process and the dissemination activities associated with this paper. The co-design workshop and parliamentary presentation are dissemination activities, not components of the data collection process.

STUDY LIMITATIONS

This study is subject to several limitations. It draws on four key informant interviews and one focus group discussion involving six participants, providing rich qualitative insight but limiting the breadth of perspectives captured. In addition, the research was conducted exclusively in Ghana. The findings, therefore, offer context-specific insights that may inform similar governance and civic participation initiatives elsewhere in Africa, rather than representing a generalizable regional pattern.

KEY FINDINGS

Analysis revealed seven interrelated institutional barriers explaining why high levels of digital civic participation have not translated into meaningful youth influence over parliamentary processes. Here are the key findings:

1. Digital presence does not translate into institutional influence—but the gap narrows for well-resourced actors.
Youth respondents reported heavy daily use of Facebook, TikTok, X, WhatsApp, and LinkedIn to discuss governance issues, yet most had never engaged Parliament's official channels and did not expect to be heard through them.

The exception was a respondent with a large public following, who reported direct access to officials and demonstrable influence on recent legislative amendments.
This suggests that influence over formal governance processes is currently a function of personal reach and network capital rather than any institutional channel open to ordinary youth.

The coexistence of digital presence and political invisibility identified in this paper is confirmed, but it is not uniform. Respondents with pre-existing public platforms can bypass the barrier that constrains ordinary youth.

2. Parliament's digital engagement is structured for broadcast, not participation. Parliamentary Service staff confirmed that its website, live streaming, and committee meetings function primarily to inform citizens rather than to capture and act on their input. However, the introduction of the "Parliament Connect" Zoom series is one initiative that seeks dialogue with citizens, offering citizens a channel for feedback.

There is still no structured process for converting that feedback into an institutional response. Any public reply must pass through an internal approval chain before Parliament will respond. Where that approval does not materialize, the query goes unanswered.

3. Institutional non-responsiveness, rather than lack of access, is the primary driver of disengagement.
Several youth respondents converged independently on the same point: it is not unawareness of digital platforms that discourages participation, but the experience, or expectation, that engagement produces no institutional outcome.

This points to a trust and feedback-loop failure rather than a connectivity or digital-literacy failure, reinforcing the paper's argument that expanding connectivity alone will not resolve the participation gap.

4. Accessibility for youth with disabilities is close to nonexistent by design, not by intent.
A visually impaired respondent detailed specific technical barriers: screen readers cannot interpret image-based flyers that lack alternative text. Government websites do not offer the keyboard-shortcut navigation available on mainstream platforms such as YouTube, interactive features such as polls are unusable, and no sign-language provision exists for audio or some video content.

Parliamentary Service staff independently confirmed this as a genuine institutional blind spot, acknowledging that the assumption of near-universal phone or internet access has been treated as equivalent to accessibility, with no formal audit of whether platforms are usable by persons with disabilities.

5. The gender digital divide operates primarily through harassment and credibility discounting rather than unequal access.
A young female respondent, and particularly a respondent who is both a woman and visually impaired, reported body-shaming and credibility challenges when participating politically online. This was corroborated by the civil society respondent, who noted that offline societal biases are reproduced and made traceable online.

A further respondent observed that young men engage more visibly than young women in the same digital spaces. Parliament's own diversity efforts, including gender-balanced panels and a draft gender charter, are not yet linked to any complaint or accountability mechanism addressing online harassment.

6. Youth with disabilities are civically active offline but remain largely absent from digital advocacy and institutional response.
The civil society respondent observed that disability-inclusive civic participation has grown on the ground but not in digital spaces.
This was reinforced by an account of an organized demonstration by persons with disabilities that produced no institutional follow-through, indicating that even where this group organizes and escalates concerns, no mechanism exists to translate that activity into a legislative response.

7. Ghana lacks disaggregated data on persons with disabilities to underpin targeted policy.
A visually impaired respondent noted that the government does not appear to hold reliable disaggregated data on the disabled population, undermining the design of interventions such as employment quotas, which he characterized as arbitrary relative to the scale of unmet need.

RECOMMENDATIONS

Although this study focuses on Ghana, the findings highlight governance dynamics that may inform African Union discussions on youth participation and digital governance. Comparative research across additional African countries would help determine the extent to which these patterns are shared across different contexts.

INSTITUTIONAL PROCESS REFORM

Institutional process reform
1. Institutionalize digital feedback into the committee system. Rather than treating social media as a one-way broadcast channel, Parliament should build a formal mechanism for citizen comment during bill drafting, not only after passage, a change recommended directly by Parliamentary Service staff themselves.

2. Introduce a recurring “social listening” review cycle. A regular, publicly visible review (for example, weekly) of citizen comments and questions across Parliament's platforms, with follow-up responses, would convert the current pattern of unanswered engagement into structured accountability. This recommendation was made independently by both a youth respondent and Parliamentary Service staff.

3. Reduce internal approval bottlenecks for public-facing responses. Parliament's Public Engagement department should be delegated limited authority to respond to non-sensitive queries without escalation through the full bureaucratic chain, directly addressing the internal barrier of “bureaucracy and directivism” identified by Parliamentary Service staff as the leading obstacle to responsiveness.

4. Invest in dedicated digital community management capacity. Parliament should resource a rostered community-management function capable of triaging and responding to citizen engagement at scale, addressing the resource-constraint barrier identified by Parliamentary Service staff.

ACCESSIBILITY AND INCLUSION

5. Mandate accessibility standards and audits for all government digital platforms. This includes alternative text and image descriptions for flyers and graphics, keyboard-navigable site design, and sign-language video summaries for audio and video content. Persons with disabilities should be consulted and included in testing before public launch, as recommended directly by the visually impaired respondent and acknowledged as a gap by Parliamentary Service staff.

6. Establish a verified harassment-response protocol for youth engaging Parliament online. Attention should be paid to young women and women with disabilities, who reported credibility discounting and body-shaming when participating in civic discussion, consistent with the pattern identified by the civil society respondent.

7. Commission a national, disaggregated dataset on persons with disabilities. This would allow policy interventions, including employment and accessibility measures, to be sized against actual need rather than set arbitrarily, addressing a gap identified directly by the visually impaired respondent.

PARTICIPATION ARCHITECTURE

8. Operationalize the Affirmative Action Act as an accountability lever for civic-space inclusion. This shifts gender and disability inclusion from voluntary institutional goodwill to a mechanism with legal and procedural weight, as recommended by the civil society respondent.

9. Introduce interactive consultation features. Bill-specific comment threads, structured Q&A, and polling tools would give Parliament's platforms a genuine two-way function, rather than the current update-only design, addressing a request made by several youth respondents.

CONCLUSION

The evidence gathered from both youth respondents and institutional actors converges on a single conclusion: Ghana's youth are not digitally disengaged; Parliament's digital infrastructure is not built to receive them. The barrier is not connectivity; internet penetration in Ghana stands at approximately 70 percent, nor is it a lack of youth interest, since every respondent interviewed for this study actively follows and discusses governance issues online.

The barrier is the absence of institutional mechanisms capable of converting online civic energy into responsive governance, a gap that is compounded for young women and youth with disabilities by platforms and institutional cultures that were never designed with them as users in the first place.

This diagnosis is strengthened rather than weakened by the fact that it is corroborated by Parliament's own senior public engagement officer, whose account of bureaucratic bottlenecks, absent feedback loops, and unaudited accessibility independently confirms the structural exclusion identified from the youth and civil society perspective. This convergence indicates that the reforms required are primarily administrative and structural—institutionalized feedback loops, accessibility standards, delegated response authority, and disaggregated data, rather than purely technological.

For the Parliament of Ghana, addressing these gaps offers a route to rebuilding institutional legitimacy among the demographic that will define the country's democratic future. Although this study focuses on Ghana, the findings provide a starting point for broader comparative discussions across African Union member states.
Comparative research across additional African countries would help determine the extent to which these patterns are shared across different contexts. Architecting digital infrastructure for inclusive youth governance therefore requires deliberate institutional reform in tandem with technological investment — not technological investment alone.

REFERENCES

Afrobarometer. (2025, August 12). African youth lag behind elders in political and civic participation – except for protest, new Afrobarometer flagship report reveals. https://www.afrobarometer.org/articles/african-youth-lag-behind-elders-in-political-and-civic-participation-except-for-protest-new-afrobarometer-flagship-report-reveals/

Ghana Statistical Service. (2021). Ghana 2021 population and housing census: General report volume 3B.https://census2021.statsghana.gov.gh/subreport.php?readreport=MjYzOTE0MjAuMzc2NQ==

Kaempf, S. (2018). Digital media. In R. Bleiker (Ed.), Visual global politics (pp. 99–103). Routledge.https://doi.org/10.4324/9781315856506

Kemp, S. (2021, October 21). Digital 2021: October global statshot report. DataReportal. https://datareportal.com/reports/digital-2021-october-global-statshot

Kemp, S. (2025). Digital 2025: Ghana. DataReportal. https://datareportal.com/reports/digital-2025-ghana

Sanny, J. A. N., & Selormey, E. (2020). Double-edged sword? Ghanaians see pros, cons of social media, want access but not fake news (Afrobarometer Dispatch No. 370). Afrobarometer. https://www.afrobarometer.org/wp-content/uploads/migrated/files/publications/Dispatches/ad370-double-edged_sword-social_media_in_ghana-afrobarometer_dispatch-14oct20.pdf

Statista. (2025). Share of internet users in Africa as of February 2025, by country.https://www.statista.com/statistics/1124283/internet-penetration-in-africa-by-country/

UNDP. (2025). Digital public infrastructure (DPI). https://www.undp.org/digital/digital-public-infrastructure

Yeboah-Banin, A. A., & Adjin-Tettey, T. D. (2023). Financial viability of the Ghanaian media. In The state of the Ghanaian media report (pp. 31–46). University of Ghana

Press.https://www.researchgate.net/publication/371276880_Financial_viability_of_the_Ghanaian_media



Developing Adaptive Regulatory Frameworks for Artificial Intelligence in Africa: Balancing Innovation, Accountability, and Human Rights in Emerging Digital Economies

Geographic Focus
Nigeria, Kenya, South Africa and Rwanda (with continental implications)

1.0 Problem Statement

The advent of Artificial Intelligence (AI) has precipitated a paradigmatic shift globally and the ramifications of its impact on the developmental trajectory of African are already profound. Artificial Intelligence is part of frontier of emerging technologies today. While AI isn't new, its origin goes back to the 1950s and 1960s. But today, it has crossed the threshold of labs into our everyday lives. Africa is not left behind. Africa is experiencing rapid AI adoption across key sectors such as fintech, agriculture, governance, healthcare information, navigate employment opportunities, and academic task through enhanced efficiency. Yet most legal systems pre-date the AI era, creating a dangerous governance vacuum.

Nigeria now boasts over 120 active AI startups, while countries like Kenya, South Africa, and Rwanda are integrating AI into precision agriculture, diagnostic tools, fintech credit scoring, and public service delivery.[1] Yet, African legal systems remain fragmented and ill-equipped to address algorithmic decision-making, accountability, bias, and human rights protection.[2] Existing frameworks (e.g., NDPR, DPA 2019, POPIA) focus on general data protection but fail to address AI-specific risks such as algorithmic bias, autonomous decision-making liability, and data extractivism by foreign tech giants. This gap exposes millions to unchecked algorithmic harm, privacy violations, and technological dependency while undermining public trust.[3]

Global models like the EU AI Act offer valuable risk-based lessons but are often too resource-intensive for African contexts. Without adaptive, context-sensitive regulation, Africa risks missing the transformative potential outlined in the AU Continental AI Strategy (2024) and Agenda 2063, perpetuating digital colonialism rather than achieving digital sovereignty and inclusive growth.[4]

Global models like the EU AI Act offer valuable risk-based lessons but are often too resource-intensive for African contexts. Without adaptive, context-sensitive regulation, Africa risks missing the transformative potential outlined in the AU Continental AI Strategy (2024) and Agenda 2063, perpetuating digital colonialism rather than achieving digital sovereignty and inclusive growth.[4]

This report analyses regulatory adequacy in the four countries, draws comparative insights, and proposes a tailored African AI Governance Model. The paper proposes a tailored "African AI Governance Model" grounded in proportionality, regulatory sandboxes, capacity building, data sovereignty, and multi-stakeholder collaboration. This model aligns with the African Union's Continental AI Strategy (2024) and Agenda 2063, aiming to balance innovation with accountability and rights protection.

2.0 Research Objective(s)

Global models like the EU AI Act offer valuable risk-based lessons but are often too resource-intensive for African contexts. Without adaptive, context-sensitive regulation, Africa risks missing the transformative potential outlined in the AU Continental AI Strategy (2024) and Agenda 2063, perpetuating digital colonialism rather than achieving digital sovereignty and inclusive growth.[4]

This report analyses regulatory adequacy in the four countries, draws comparative insights, and proposes a tailored African AI Governance Model. The paper proposes a tailored "African AI Governance Model" grounded in proportionality, regulatory sandboxes, capacity building, data sovereignty, and multi-stakeholder collaboration. This model aligns with the African Union's Continental AI Strategy (2024) and Agenda 2063, aiming to balance innovation with accountability and rights protection.

➢ To evaluate the adequacy of existing legal and regulatory frameworks for AI in Nigeria, Kenya, South Africa and Rwanda.
➢ To identify lessons from global approaches (e.g., EU AI Act) and assess their adaptability to African socio-economic realities.
➢ To propose a proportionate "African AI Governance Model" emphasising sandboxes, capacity building, data sovereignty, and multi-stakeholder collaboration

3.0 Approach / Method

This study employs a doctrinal and comparative legal research methodology supported by policy analysis. It critically examines primary sources (national AI strategies, data protection laws, cyber legislation) and secondary sources (AU Continental AI Strategy and academic literature). A socio-legal lens grounds recommendations in African realities. The analysis is entirely desk-based; no primary interviews were conducted. Limitations include reliance on publicly available documents up to mid-2026.

4.0 Key Findings

➢ Finding 1
Statement: Existing legal and regulatory frameworks in Nigeria, Kenya, South Africa, and Rwanda are structurally ill-equipped to govern AI because they rely predominantly on outdated general data protection laws that fail to address the unique challenges of algorithmic decision-making, accountability, and data sovereignty.
Explanation: Although all four countries have established foundational data protection regimes. Nigeria's NDPR 2019, Kenya's Data Protection Act 2019, South Africa's Protection of Personal Information Act (POPIA) and Rwanda's Law No 058/2021 on Data Protection and Privacy. these instruments were largely designed for conventional data processing and do not adequately regulate high-stakes AI applications such as automated decision making in fintech credit scoring, healthcare diagnostics, or public service delivery.

comparative policy analysis reveals persistent enforcement gaps and institutional weaknesses: South Africa maintains the most mature regulatory environment among the three, yet even it falls short of operational adequacy, while Nigeria exhibits the most pronounced deficits in both legislative specificity and enforcement capacity.

This inadequacy is not merely technical but reflects a deeper political economy problem of digital extractivism, whereby African states generate vast quantities of valuable behavioural and biometric data but lack the sovereign tools to control its use, monetisation, or algorithmic outcomes.[7] The result is heightened exposure to algorithmic bias (notably against darker skin tones and local languages), privacy erosion, and mass surveillance risks. [8] Without dedicated AI-specific rules on transparency, impact assessments, and liability, these frameworks risk normalising technological dependency rather than fostering genuine digital sovereignty, a core objective of the African Union's Continental AI Strategy (2024).

This finding underscores the need for adaptive, context-sensitive regulation that moves beyond transplanting foreign models toward an Africa-centric governance architecture.

4.0 Key Findings

➢ Finding 2

Statement: National AI strategies across Nigeria, Kenya, South Africa, and Rwanda prioritise innovation, investment attraction, and economic growth but consistently under emphasise enforceable accountability mechanisms, creating a structural imbalance that risks exacerbating digital dependency rather than fostering sovereign and rights-based AI development.

Explanation: Rwanda's ambitious National AI Policy and Kenya's National AI Strategy 2025–2030 exemplify this trend, positioning AI as a driver for sectors such as precision agriculture, fintech, and public service delivery, while relying heavily on regulatory sandboxes to test innovations.[10] Similarly, Nigeria's National AI Strategy and South Africa's policy efforts focus on ecosystem building and investment but provide limited detail on mandatory algorithmic impact assessments, independent oversight, or redress mechanisms for affected citizens.[11] This "strategy as governance" approach, while useful for signalling ambition and attracting partnerships, often results in aspirational documents that lack the teeth of binding legislation or adequately resourced institutions.

The consequence is a governance deficit where innovation proceeds faster than safeguards, amplifying risks of algorithmic bias, privacy violations, and value extraction by foreign platforms.[13] Pattern aligns with broader continental observations that many African AI strategies function more as promotional tools than robust regulatory frameworks, falling short of the balanced "harness benefits while minimising risks" imperative articulated in the African Union's Continental AI Strategy (2024).

Addressing this imbalance requires moving beyond strategy documents to institutionalise accountability through proportionate tools such as mandatory impact assessments, independent review bodies, and participatory governance models that integrate local contexts and Ubuntu-inspired ethical principles.

➢ Finding 3
Statement: Global regulatory models, particularly the European Union's risk-based AI Act, provide valuable structural lessons on risk classification and accountability, but require substantial adaptation to suit Africa's resource constraints, legal pluralism, developmental priorities, and the imperative of data sovereignty.

Explanation: The EU AI Act's four-tier risk framework prohibiting unacceptable risks, imposing strict obligations on high-risk systems, and applying lighter transparency rules to limited-risk applications offer a sophisticated scaffold for balancing innovation with rights protection.[15] However, its comprehensive compliance requirements, heavy institutional burden, and one-size-fits-all approach are ill-suited to most African contexts characterised by limited regulatory capacity and competing developmental needs.

Comparative insights from other jurisdictions further illuminate pathways for adaptation: Singapore's collaborative regulatory sandboxes and testing frameworks (such as AI Verify) demonstrate how innovation-friendly experimentation can coexist with governance, while China's state-centric model highlights the importance of sovereign control over critical AI systems. [17] In Africa, a rigid transplantation of these models risks either regulatory overreach that stifles local start-ups or continued under-regulation that perpetuates digital extractivism.

A context-appropriate response demands a proportionate, hybrid African AI Governance Model that integrates risk-based elements with regulatory sandboxes, capacity-building programmes, multi-stakeholder platforms (involving AU, RECs, civil society and private sector), and strong data sovereignty safeguards. This approach would align with the African Union's Continental AI Strategy emphasis on harnessing benefits while minimising risks, while embedding Ubuntu-inspired ethical principles of communal well-being and inclusion.

Uncovering the Latest Travel Trends

BY ROSA MARIA AGUADO
PHOTOGRAPHY BY
ESTELLE DARCY



5.0 Supporting Evidence / Cases

➢ Case 1: Nigeria's NDPR 2019 and Emerging AI Strategy

Nigeria's Data Protection Regulation (NDPR) 2019 and its draft National AI Strategy illustrate the structural inadequacy of general data laws for AI governance. While the NDPR provides a foundation for data protection, it lacks specific provisions for algorithmic

transparency, automated decision-making, or bias audits in high-impact sectors such as fintech credit scoring.[19] The country's strategy emphasises innovation and attracting investment (with over 120 AI startups), yet enforcement remains weak due to limited institutional capacity.[20] This example directly demonstrates Finding 1 (reliance on outdated frameworks) and Finding 2 (innovation-focused strategies without strong accountability).

Case 2: Rwanda's National AI Policy and Regulatory Sandbox Approach

Rwanda has positioned itself as an innovation leader through its National AI Policy (2023) and active use of regulatory sandboxes. While this has accelerated AI adoption in agriculture and public services, the policy offers limited mandatory accountability mechanisms such as independent algorithmic impact assessments.[21] This case supports Finding 2 by showing how strategy-led innovation can outpace rights safeguards, while also highlighting Finding 3's call for context-appropriate adaptation of global sandbox models to African realities.

Case 3: South Africa's POPIA and Draft AI Policy Efforts

South Africa's Protection of Personal Information Act (POPIA) represents the most mature data protection framework among the four countries, yet it still falls short on AI-specific governance, particularly liability for autonomous systems and mass surveillance risks.[22] Its ongoing AI policy development process reveals both the strengths and limitations of attempting to adapt elements of the EU AI Act's risk-based approach within resource constrained environments. This evidence reinforces Finding 1 (persistent gaps even in stronger jurisdictions) and Finding 3 (the need for proportionate adaptation of global models).

6.0 Implications

These findings reveal that Africa's AI governance landscape is characterised by a systemic governance deficit between policy ambition and implementation capacity. The reliance on outdated data laws, innovation-first strategies, and inefficient adapted global models creates systemic vulnerabilities to algorithmic harm, digital extractivism, and erosion of data sovereignty.[24] Rather than empowering citizens and fostering inclusive growth, current approaches risk normalising technological dependency and deepening inequalities.

Hence, the stakeholders particularly African policymakers, the African Union, regulators, and civil society must care because ineffective AI governance threatens core continental goals under the AU Continental AI Strategy (2024) and Agenda 2063. Without urgent adaptive frameworks, Africa will continue exporting valuable data while importing finished AI systems, undermining digital sovereignty and public trust.

Conversely, well-designed governance can unlock AI's potential for transformative development in health, agriculture, education, and finance while protecting fundamental rights. The window for shaping an Africa-centric model is narrowing rapidly.

7.0 Policy-Relevant Recommendations

7.1 Prioritise the enactment of dedicated, proportionate AI legislation that incorporates mandatory algorithmic impact assessments for high-risk systems and clear liability rules for autonomous decision-making.

➢ Rationale: Current reliance on general data protection laws leaves critical gaps in accountability and transparency. Specific AI legislation would address Findings 1 and 2 by moving beyond aspirational strategies to enforceable rules tailored to local contexts
➢ Target Actor: National Governments (Nigeria, Kenya, South Africa, Rwanda)

7.2 Establish a Continental AI Regulatory Coordination Framework, including shared regulatory sandboxes, a harmonised AI risk classification system, and capacity-building programmes for national regulators.
➢ Rationale: Fragmented national approaches limit effectiveness. A coordinated continental mechanism would support Finding 3 by adapting global lessons (e.g., EU risk categories) while promoting data sovereignty and cross-border interoperability, in line with the AU Continental AI Strategy (2024).
➢ Target Actor: African Union and Regional Economic Communities (RECs)

7.3 Institutionalise mandatory regulatory sandboxes for AI testing combined with independent oversight and public reporting requirements on algorithmic bias and societal impact.

➢ Rationale: Innovation-focused strategies in Rwanda and Kenya show promise but lack sufficient safeguards. Structured sandboxes would balance innovation with accountability, directly mitigating the risks highlighted in all three findings. ➢ Target Actor: National Regulators and Data Protection Authorities

7.4 Develop and adopt a Citizen Data Rights Charter that embeds data sovereignty, transparency, and redress mechanisms, with active citizen participation in AI governance.
➢ Rationale: To counter digital extractivism and rebuild public trust, governance must go beyond state-centric models. This charter would operationalise Ubuntu inspired ethics and ensure that AI serves inclusive development rather than perpetuating dependency.
➢ Target Actor: Policymakers and Multi-Stakeholder Platforms (including civil society and private sector)

8.0 Cross-Cutting Insight (MANDATORY

This research reveals that weak AI governance in Africa is not an isolated regulatory failure but a symptom of deeper structural imbalances in the continent's digital transformation. The findings connect directly to broader themes of Digital Rights & Surveillance, Internet Governance, and Trust & Safety: inadequate frameworks for algorithmic accountability enable mass surveillance risks, erode public trust in digital systems, and undermine digital rights, while fragmented national strategies hinder effective internet governance and cross-border data flows under the AfCFTA.

The broader pattern reflected is one of "strategy without sovereignty" a continent-wide tendency where African states adopt ambitious AI and digital policies that signal innovation ambitions but fail to secure meaningful control over data, algorithms, and technological infrastructure. This perpetuates digital extractivism and neo-colonial dependencies rather than fostering genuine technological self-determination.

Ultimately, effective AI governance sits at the intersection of rights protection, innovation policy, and geopolitical strategy. Without a deliberate shift toward adaptive, Africa-centric models that integrate data sovereignty, capacity building, and multi-stakeholder participation, progress in one area (e.g., AI literacy or digital infrastructure) will continue to be undermined by governance deficits in another.
This research therefore calls for holistic, interconnected policy approaches that treat AI governance as foundational to Africa's overall digital future.

9.0 The African AI Governance Model: An Analytical Framework

The proposed African AI Governance Model is not a loose collection of aspirations but a coherent, interdependent governance architecture designed for Africa's developmental context. It integrates five core components that interact dynamically to balance innovation, accountability, and human rights while prioritising data sovereignty and inclusive growth.

9.1 The Five Components:

1. Proportionate Risk-Based Regulation Adapts the EU AI Act's tiered approach but with Africa-specific calibrations: lower compliance burdens for SMEs/startups, mandatory sandboxes for high-risk systems, and context-adjusted thresholds (e.g., prioritising bias against local languages/demographics over purely technical metrics).
Interaction: Feeds directly into institutional oversight.

2. Regulatory Sandboxes & Capacity Building Mandatory, time-bound testing environments with built-in independent evaluation, public reporting on bias/societal impact, and technical assistance for regulators and developers. Includes AU-coordinated training programmes.

Interaction: Bridges innovation (national strategies) and accountability (risk regulation).
3. Data Sovereignty & Local Value Capture Requirements for local data processing/storage where feasible, benefit-sharing clauses in foreign AI deployments, protection of indigenous knowledge, and mobile-first/African language AI mandates.

Interaction: Counters digital extractivism and supports ethical foundations.
4. Multi-Stakeholder & Continental Coordination National AI Regulatory Bodies coordinated through an AU/RECs framework, with civil society, private sector, and community representation. Includes a Citizen Data Rights Charter.
Interaction: Ensures legitimacy and harmonisation across borders (AfCFTA alignment).

5. Ubuntu-Inspired Ethical Governance Definition: Drawing from African philosophical traditions of communal well-being, interconnectedness, and human dignity. This component requires algorithmic impact assessments to evaluate effects on social cohesion, equity, and collective prosperity not just individual rights. Examples include community consent mechanisms for data use and redress focused on restorative justice. Interaction: It provides the normative foundation that permeates all other components.

9.2 Institutional Arrangements:

• National Level: Dedicated AI Authorities (or expanded Data Protection Offices) with enforcement powers, independent ethics review boards, and sandbox oversight units.
• Continental Level: AU AI Coordination Mechanism for standard-setting, cross border sandboxes, and dispute resolution.

• Implementation Tools: Mandatory algorithmic impact assessments (high-risk), transparency registers, public dashboards, and annual reporting.

9.3 Distinctions from Existing Proposals:

• Unlike the EU AI Act, it is resource-light, development-first, and culturally embedded (Ubuntu with data sovereignty and heavy bureaucracy).[28] • Unlike many national strategies (e.g., Rwanda/Kenya), it moves from promotional documents to binding, enforceable mechanisms with accountability teeth.[29] • Unlike purely market-driven or state-centric models, it emphasises hybrid governance,
• local value capture, and protection against extractivism while remaining innovation-friendly.
This model constitutes the paper's central analytical contribution: a delivered governance designs rather than an aspiration.[30]

Author Bio

Ismail Abdrasheed is a rising law scholar at Ahmadu Bello University, Zaria, Nigeria, and a legal researcher with publications on crypto assets and AI deepfakes.

He is passionate about Web3, blockchain regulation, and digital assets, exploring how legal frameworks can harness Africa's growing digital ecosystem while safeguarding rights, promoting inclusion, and balancing innovation with accountability.

His research interests on AI focuses on addressing emerging challenges and shaping responsible governance tailored to the continent's unique landscape. During the Digispace Africa Fellowship, he will explore adaptive approaches to AI governance suited to Africa's evolving digital environment, socio-legal realities and policy implications.

Email: Ismail.abdrasheed1@gmail.com
Telephone: +234 8163693382

Bibliograghy

1. DIPLO Sorina Teleanu and Jovan Kurbalija, 'State of African Digital Economy' in Stronger Digital Voices from Africa: Building African Digital Foreign Policy and Diplomacy (DiploFoundation 2022) African digital economy - Diplo Resource

2. African Union, Continental Artificial Intelligence Strategy: Harnessing AI for Africa's Development and Prosperity (AU 2024) 28–32. https://au.int/sites/default/files/documents/44004-doc-EN-_Continental_AI_Strategy_July_2024.pdf

3. ResearchGate Samuel Okediji and Gansuvun Oyinkansola "Who Controls the Algorithm? Data Sovereignty, Digital Extractivism, and AI Governance Gaps in Africa" DOI: 10.13140/RG.2.2.25312.65282

4. MordorIntelligence " Africa Digital Transformation Market - Growth, Trends, COVID 19 Impact, and Forecasts (2025–2030) (Mordor Intelligence, 2025) Africa Digital Transformation Market Size, Growth, Forecast Report & Share 2031

5. Shaikh AT, Lodhi ARK and Mubashar MM, 'Algorithmic Sovereignty and Digital Colonialism: Legal and International Relations Perspectives on AI Governance in the Global South' (2026) 4(3) Annual Methodological Archive Research Review 510, 513 — 515 https://www.researchgate.net/profile/Aehtasham Shaikh/publication/403781086_Algorithmic_Sovereignty_and_Digital_Colonialism _Legal_and_International_Relations_Perspectives_on_AI_Governance_in_the_Gl obal_South/links/69dd5ae05970dd1b05f89580/Algorithmic-Sovereignty-and Digital-Colonialism-Legal-and-International-Relations-Perspectives-on-AI Governance-in-the-Global-South.pdf

6. The Guardian, '"Invasive" AI-led mass surveillance in Africa violating freedoms, warn experts' (The Guardian, 12 March 2026) 'Invasive' AI-led mass surveillance in Africa violating freedoms, warn experts | Global development | The Guardian

7. Kenya National AI Strategy 2025–2030 (Ministry of Information, Communications and the Digital Economy 2025) 45–52; Republic of Rwanda, National Artificial Intelligence Policy (2023) 18–25. https://www.ict.go.ke/sites/default/files/2025-12 Kenya%20National%20AI%20Strategy%202025-2030%20Implementation%20Roadmap.pdf

8. Kinfe Yilma and Kebene Wodajo, 'Strategy as Governance: The Governance of AI in Africa' (2026) 53(2) Science and Public Policy 236, 239-241 https://doi.org/10.1093/scipol/scag011

9. European Union, Regulation (EU) 2024/1689 (Artificial Intelligence Act) [2024] OJ L1689; European Commission, 'AI Act' (Shaping Europe's Digital Future) AI Act | Shaping Europe's digital future

10. Jonas Tallberg and others, 'The Global Governance of Artificial Intelligence: Next Steps for Empirical and Normative Research' (2023) arXiv:2305.11528, 8–15; IMDA Singapore, 'Generative AI Evaluation Sandbox' (2023–2025 initiatives). https://doi.org/10.48550/arXiv.2305.11528

11. Republic of Rwanda, National Artificial Intelligence Policy (2023) pg 18–25. https://www.ict.go.ke/sites/default/files/2025-

12/Kenya%20National%20AI%20Strategy%202025-2030%20Implementation%20Roadmap.pdf

12. F Cloete, 'Governing Artificial Intelligence (AI) and Other Technologies in the Digital Era' (2024) 32(1) Administratio Publica 1, 15–18. Governing Artificial Intelligence (AI) and Other Technologies in the Digital Era

13. Oxford Insights, Government AI Readiness Index 2026 (Oxford Insights January 2026) https://oxfordinsights.com/wp-content/uploads/2026/01/2025- Government-AI-Readiness-Index-Report_01_26.pdf



www.digispaceafrica.org

Copyright © 2026