

POLICY BRIEF

Strengthening the AUDA-NEPAD Guidelines on Responsible AI in Social Protection and Welfare Systems in Africa: Recommendations from Hakakire Legal Services

Hakakire Legal Services | Abuja, Nigeria
Authors - Uche Anyamele (PhD), Christiana Joel, Ademola Idowu

Submitted in Response to the AUDA-NEPAD/APET Call for Consultations on the Guidelines on Responsible Artificial Intelligence in Social Protection and Welfare Systems in Africa

JUNE, 2026.

Key Recommendations at a Glance

- Revise the definition of Artificial Intelligence and of social assistance to reflect current OECD, EU AI Act, UNESCO, World Bank, and Asian Development Bank standards.
- Revise the language used in the description of the Guidelines — 'normative and voluntary' — to match the '*actively expected to*'/'should' posture of comparable instruments such as the ILO Social Protection Floors Recommendation, 2012 (No. 202).
- Extend the scope of the Guidelines to faith-based organisations, community savings groups, and other informal social protection actors that increasingly rely on AI-enabled beneficiary management.
- Mandate a reverse onus of proof in automated benefit suspensions, and a legally defined human intervention timeline of between 7 and 30 days for appeals against AI-mediated eligibility flags.
- Redefine 'low risk' AI applications to capture the exclusion risks inherent in administrative document processing, such as failures in parsing local names or degraded identity documents.
- Differentiate capacity-building support by institutional capacity, and add a standalone measure requiring investment in foundational data infrastructure — civil registration, vital statistics, and social registries — as a prerequisite for responsible AI deployment.
- Add a dedicated safeguard subsection addressing AI in fragile, conflict-affected, and violent (FCV) settings, including restrictions on biometric data collection and centralisation.
- Reduce the mandatory review cycle from five years to two years, introduce quarterly exclusion-rate reporting.
- Specify that meaningful beneficiary participation includes community social audits, accessible grievance mechanisms, and disaggregated satisfaction assessments.

Introduction

Artificial intelligence is already determining who receives social protection in Africa and who does not. Algorithmic eligibility screening, automated fraud detection, and AI-assisted benefit disbursement are now embedded features of social protection delivery across the continent — often reaching the most economically vulnerable populations first, and with the least margin for error.

The African Union Development Agency (AUDA-NEPAD), through the High-Level Panel on Emerging Technologies (APET), has responded to this reality by developing the Guidelines on Responsible Artificial Intelligence in Social Protection and Welfare Systems in Africa (the Guidelines) — a draft continental framework intended to guide Member States, social protection institutions, and other stakeholders in the responsible adoption of AI within welfare systems.

Hakakire Legal Services, an Abuja-based legal consultancy firm with a practice focus in artificial intelligence governance, digital policy, international trade, and technology law, welcomes the release of these Guidelines and the invitation to contribute to their refinement.

This policy brief sets out our assessment of the Guidelines, identifies gaps and areas for strengthening, and offers concrete recommendations. Our comments are offered in the spirit of constructive engagement with a document that, with targeted improvements, can serve as a meaningful continental reference for responsible AI in social protection.

What the Guidelines Get Right

Before turning to areas of concern, it is worth acknowledging what the Guidelines achieve. The document follows the same well-structured normative architecture used across the AUDA-NEPAD/APET Emerging Technology Governance Series: a foundation of cross-cutting principles addressing human rights, inclusion, data protection, transparency, accountability, safety, and sustainability, supplemented by sector-specific principles tailored to social protection. This structural continuity with companion guidelines gives the Series coherence and makes national adaptation easier.

The Guidelines also demonstrate credible engagement with the operational realities of social protection delivery. The inclusion of a dedicated fraud-detection safeguard reflects awareness that AI-enabled eligibility and benefit systems carry distinctive risks that general-purpose AI guidance does not capture. The classification of AI applications by risk level, and the attention paid to general, legal, and gender-specific measures, provide a workable skeleton for national policy development.

The document's anchoring in the African Union's wider digital governance architecture — including the Continental Artificial Intelligence Strategy (2024), the AU Data Policy Framework (2022), and the Malabo Convention on Cyber Security and Personal Data Protection (2014) — gives it normative continuity with instruments African Member States are already engaging with. This is a sound foundation. What follows is a set of targeted recommendations to strengthen its precision, enforceability, and contextual fit.

Areas for Strengthening

1. Definitions Lag Behind International Standards

The Guidelines define AI as 'systems that perform tasks normally requiring human intelligence, including learning, reasoning, pattern recognition, language processing and decision-making. AI includes rule-based systems, statistical methods and machine-learning approaches.' As we have noted in our comments on sister documents in this Series, this definition omits the machine-based nature of AI systems, the varying levels of autonomy AI systems exhibit after deployment, and the input-output mechanism through which AI systems generate outputs that influence physical and virtual environments. These omissions matter directly for social protection, where the question of how much autonomy an eligibility or fraud-detection system exercises is central to risk classification and accountability.

We recommend that the definition be revised to read: *'Machine-based systems with different levels of autonomy that, for explicit and implicit objectives, perform tasks normally requiring human intelligence — including learning, reasoning, pattern recognition, language processing and decision-making — by inferring from the input received by them and generating outputs that influence both physical and virtual environments. AI includes rule-based systems, statistical methods and machine learning approaches.'* This wording draws on the updated OECD AI Principles (2023), UNESCO's definition of artificial intelligence, and Article 3 of Regulation (EU) 2024/1689 (the EU AI Act).

The Guidelines' definition of social assistance — 'non-contributory cash and in-kind transfers and social services targeted at poor and vulnerable populations' — is similarly under-specified. We recommend it be revised to read: *'Non-contributory programmes, transfers, or social services targeted at providing income and welfare support and improving the overall living standard of poor and vulnerable populations.'* This formulation aligns with the framing used by the World Bank's ASPIRE database, the Asian Development Bank, etc., and more completely captures the welfare objective that social assistance programmes serve, not merely the transfer element.

2. The 'Normative and Voluntary' Concern

As with the companion Guidelines in this Series, the description of this instrument as 'normative and voluntary' warrants reconsideration. The ILO Social Protection Floors Recommendation, 2012 (No. 202) — itself non-binding — does not describe itself as merely 'voluntary'. It consistently uses 'should' language that frames implementation as an active expectation rather than a passive option, and it backs that language with ILO monitoring mechanisms. A social protection AI governance instrument should, at minimum, match the normative tone of the ILO's foundational social protection standard, given that it governs systems determining access to the same benefits R202 is designed to guarantee.

We recommend that *'normative and voluntary'* be replaced with *'actively expected to'*. Member States and implementing actors should be called upon to adopt and implement the Guidelines and to publicly account for progress made and obstacles encountered, with a clear roadmap toward compliance where immediate implementation is not feasible.

3. Scope Gaps: Faith-Based and Informal Social Protection Actors

In many African countries, faith-based organisations, community savings/cooperative societies, and zakat committees deliver a substantial portion of social protection. As these organisations increasingly adopt digital tools for beneficiary management, they interact with AI-enabled systems in ways the Guidelines do not address.

We recommend the addition of a provision specifically addressing informal and faith-based social protection actors, with simplified guidance and support for their engagement with AI systems, and ensuring that data collected through informal channels meets the same protection standards as formal state data.

4. Due Process Gaps: Reverse Onus of Proof and Mandatory Human Intervention Timelines

Section F2 (Legal Measures) should be strengthened to provide that, where a fraud-detection AI system flags a beneficiary, benefits must not be withdrawn or suspended until the state has demonstrated that fraud actually occurred — the burden of proving that the system erred must no longer fall on the disadvantaged citizen. This reverse onus of proof is warranted by a concept known asymmetry of harm: that where the state mistakenly continues benefits to an ineligible person, the consequence is a minor and recoverable financial loss; however where the state wrongly withdraws benefits from an eligible person, the consequence can include starvation, loss of life-saving medical care, or immediate destitution.

Section F3 (Social Protection-Specific Measures) should similarly be strengthened to establish a mandatory human intervention timeline whenever an individual appeals an AI-mediated eligibility flag. A human reviewer must engage with and decide such appeals within a legally mandated window of between 7 and 30 days. This measure is necessary to protect structurally disadvantaged people in accessing benefits and to prevent starvation or serious economic shock during periods of unresolved appeal.

5. Redefining 'Low Risk' AI Applications

The current low-risk classification suggests, incorrectly, that administrative document processing and aggregate analytics are generally low risk. However, international regulatory frameworks suggest otherwise, as even administrative document processing can carry significant exclusion risk where the system fails to properly parse local names, degraded physical documents, etc.

We recommend that an AI application in social protection should only be considered low risk if it has no direct impact on the individual and does not automate the primary critical channels through which citizens interact with the state, so as to ensure it does not deny access to important activities such as redress and registration. This recommended definition aligns with international regulatory frameworks and standard national approaches to AI risk classification, such as the EU AI Act which recognises that consequential automated processes carry elevated risk regardless of their administrative framing.

6. Capacity and Data Infrastructure Gaps

The Guidelines' capacity-building measures do not address the profound variation in institutional capacity across AU Member States. Rwanda's IremboGov platform, for example, now hosts more than 200 digital government services and has saved citizens an estimated hundreds of millions of working hours that would otherwise have been spent travelling to government offices, whereas the Democratic Republic of Congo still lacks a functioning digital ID system despite repeated legislative and procurement efforts spanning over a decade. Capacity-building designed for moderately functional states will neither reach nor address those that need it most.

We recommend the inclusion of a capacity-building section that explicitly addresses differentiated support, with dedicated technical assistance for Member States with low institutional capacity, providing simplified guidance that does not presuppose existing and functional national AI regulation or data protection authorities.

A related and equally significant gap is the absence of a data infrastructure investment imperative. Nearly 500 million people in sub-Saharan Africa still lack any form of legal identification, and African governments have in several cases deployed digital ID systems ahead of robust data governance frameworks, putting vulnerable groups at heightened risk of exclusion. We recommend a standalone general measure requiring that Member States invest in foundational data infrastructure — civil registration, vital statistics, and social registries — as a prerequisite for responsible AI deployment, and that AI pilots not be launched where underlying data quality cannot support the accuracy requirements of the intended application.

7. Safeguards for AI in Conflict, Insurgency, and Fragile Settings

As of 2026, armed conflict is simultaneously active across the Great Lakes, the Horn of Africa, and the Sahel, with the Lake Chad Basin functioning as a zone of near-permanent insurgency spanning four AU Member States. The specific risks of deploying AI in social protection in these settings go beyond what the Guidelines' current safeguards contemplate.

Biometric data collected from conflict-affected populations can be seized by armed actors to identify, target, or extort individuals, particularly where the data reveals political affiliation, religion, or ethnicity. Cash transfer records can similarly expose beneficiaries to kidnapping and extortion, a risk underscored by recurrent cases of mass kidnapping in parts of Nigeria.

We recommend the addition of a dedicated subsection within Section F6 (Safeguards and Exemptions) specifically addressing AI in fragile, conflict-affected, and violent (FCV) settings, covering: a presumption against collecting or centralising biometric data in FCV-affected areas until adequate security can be guaranteed; data self-destruction or emergency anonymisation protocols for beneficiary data in areas at risk of armed group control; and a requirement that international humanitarian actors adhere to ICRC/UNHCR data protection standards as a minimum baseline.

8. Strengthening Enforcement: Review Cycles and Fraud Detection Safeguards

The Review Period Is Too Long

A five-year review cycle is inadequate in a sector where AI systems can cause mass exclusion events within months of deployment. We recommend reducing the mandatory review cycle to two years for all AI systems in social protection, with continuous monitoring and mandatory quarterly exclusion-rate reporting for systems used in eligibility determination, benefit disbursement, and fraud detection. Annual monitoring reports by AUDA-NEPAD/APET on AI in social protection should be a baseline expectation.

The Fraud Detection Safeguard Is Incomplete

The existing fraud-detection safeguard does not acknowledge the systemic tendency of such systems toward over-detection of marginalised populations. We recommend strengthening the safeguard to require mandatory demographic disaggregation of fraud-flag rates, published by gender, age, geography, and disability status, alongside an accessible, time-sensitive appeal process with independent review.

9. Beneficiary Participation Mechanisms Are Underspecified

In social protection, meaningful participation often has specific operational meaning such as community score cards, social audits, beneficiary satisfaction surveys, community monitoring committees, and grievance desk systems

We recommend that the implementation section specify that meaningful beneficiary participation includes, at minimum: community-level social audits of AI-driven decisions; accessible grievance mechanisms in local languages with defined response timelines; community monitoring committees with escalation authority; and periodic beneficiary satisfaction assessments disaggregated by gender, age, disability, and geographic location.

Priority Recommendations

Drawing on the analysis above, Hakakire Legal Services recommends the following priority actions:

Short-term (within 12 months):

- Revise the definitions of AI and social assistance, redefine 'low risk' AI applications to reflect current international standards, and mandate the reverse onus of proof in automated benefit suspensions.
- Develop differentiated capacity-building programmes and a foundational data infrastructure investment requirement, recognising the wide variation in institutional capacity across Member States.

Medium-term (1–3 years):

- Establish a defined implementation and transition period of between one and five years, depending on the nature of the measure, for Member States to achieve compliance with the Guidelines, with priority given to due-process and FCV safeguards.

- Pursue treaty-level or binding continental instruments on AI in social protection, building on the foundation that these Guidelines establish.

Conclusion

Social protection is, by definition, the system of last resort for Africa's most vulnerable populations. When artificial intelligence determines who receives that protection, the margin for error narrows considerably, because a wrongly denied benefit is not an inconvenience but, in the most severe cases, a matter of survival. We believe that with the implementation of these recommendations, the Guidelines can become a framework that genuinely protects the people social protection systems exist to serve.

References

- AUDA-NEPAD / APET, Guidelines on Responsible Artificial Intelligence in Social Protection and Welfare Systems in Africa (Draft, 2025), Emerging Technology Governance Series.
- African Union, Continental Artificial Intelligence Strategy (2024).
- African Union, AU Data Policy Framework (2022).
- African Union Convention on Cyber Security and Personal Data Protection (Malabo Convention, 2014).
- Abdul Mohammed and others, 'The Gathering Storm Facing Africa in 2026: Entrenching Conflicts, Fractured Order, and Eroding Agency' (Amani Africa, 14 January 2026) <<https://amani-africa-et.org/the-gathering-storm-facing-africa-in-2026-entrenching-conflicts-fractured-order-and-eroding-agency/>> accessed 2 June 2026.
- Asian Development Bank, 'Social Protection and Labor' <<https://www.adb.org/what-we-do/themes/social-development/overview/social-protection-labor>> accessed 2 June 2026.
- European Parliament and Council, Regulation (EU) 2024/1689 of 13 June 2024 laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) [2024] OJ L 1689/1 <<https://artificialintelligenceact.eu/article/3/>> accessed 1 June 2026.
- GSDRC, 'Types of Social Protection' <<https://gsdrc.org/topic-guides/social-protection/types-of-social-protection/>> accessed 30 May 2026.
- International Committee of the Red Cross, 'Data Protection in Humanitarian Action' (ICRC, 2025) <<https://www.icrc.org/en/article/data-protection-humanitarian-action-publication>> accessed 2 June 2026.
- International Labour Organization, Social Protection Floors Recommendation, 2012 (No. 202), adopted by the International Labour Conference, June 2012. https://normlex.ilo.org/dyn/nrmlx_en/f?p=NORMLEXPUB:12100:0::NO::P12100_INSTRUMENT_ID%2CP12100_LANG_CODE:3065524%2Cen
- iubenda, 'OECD Updates AI Definition: A Step Forward in Shaping EU's AI Law' (iubenda Blog) <<https://www.iubenda.com/en/blog/oecd-updates-ai-definition-a-step-forward-in-shaping-eus-ai-law/>> accessed 1 June 2026.
- OECD, 'AI Principles Overview — Updated Definition of an AI System' (OECD.AI, November 2023) <<https://oecd.ai/en/ai-principles>> accessed 1 June 2026.
- Rwanda Dispatch News Agency, 'Irembo Platform Transforms Access to Public Services' (5 November 2025) <<https://rwandadispatch.com/irembo-platform-transforms-access-to-public-services/>> accessed 2 June 2026.
- UNESCO, 'Artificial Intelligence (AI)' <<https://www.unesco.org/en/query-list/a/ai>> accessed 1 June 2026.
- UNESCO, Recommendation on the Ethics of Artificial Intelligence (2021).
- European Centre for Development Policy Management, 'Digital ID Systems in Africa: Challenges, Risks and Opportunities', ECDPM Discussion Paper No. 360 (2023) <<https://ecdpm.org/application/files/5517/0254/4789/Digital-ID-systems-in-Africa-ECDPM-Discussion-Paper-360-2023.pdf>> accessed 2 June 2026.
- World Bank, 'ASPIRE: The Atlas of Social Protection — Indicators of Resilience and Equity' <<https://www.worldbank.org/en/data/datatopics/aspire/indicator/social-assistance>> accessed 2 June 2026.