

Responsible AI for African Peace and Security: Machine Learning, NLP and Large Language Models for Regional Early Warning and the African Standby Force

Dr. Alexander Odartey Lamptey (Ph.D)

African Peace Support Trainers Association-APSTA, Greater Accra, Accra, Ghana

ABSTRACT

Artificial intelligence (AI), machine learning (ML), natural language processing (NLP), large language models (LLMs), and network analytics may strengthen African regional peace and security institutions by improving the synthesis of fragmented information, identifying patterns in conflict and political violence, supporting early warning, and informing planning for peace support operations and counter-terrorism. Yet their value depends less on technological novelty than on the quality, legitimacy, interoperability, and governance of the data and institutions through which they are used. This policy paper examines how data-driven analytical capabilities could be responsibly integrated into the African Union's Regional Mechanisms and the African Standby Force (ASF), with attention to armed conflict, political violence, violent extremism, and counter-terrorism. Using qualitative document analysis, the paper synthesises relevant scholarship, African and international policy frameworks, and public institutional material. An illustrative case study of ECOWAS's Early Warning and Response Network (ECOWARN) explores the practical relationship between field reporting, human-security indicators, regional risk analysis, and AU-regional coordination. It develops a proposed Regional Conflict Analytics and Early Warning Framework organised around six functions: shared data standards; multilingual and multimodal analysis; human-validated analytical products; regional interoperability; responsible governance; and continuous learning and evaluation. The paper argues that AI should support, not replace, political judgement, command responsibility, community knowledge, and legal review. It recommends a phased pilot focused initially on non-operational early warning and situational awareness, with clear safeguards against discriminatory profiling, unverified allegations, automated targeting, and opaque risk scoring. The paper contributes a practical governance and implementation agenda for African-led, rights-respecting use of data analytics in regional peace and security, while using ECOWARN to ground the recommendations in an existing regional early-warning architecture. (2024a, 2024b; Tabassi, 2023; UNESCO, 2021)

KEYWORDS

African Standby Force, regional mechanisms, machine learning, natural language processing, large language models, conflict early warning, counter-terrorism, political violence, responsible AI

EXECUTIVE SUMMARY

African regional institutions face a complex information environment in which conflict incidents, political violence, violent extremist activity, displacement, local grievances, financial flows, and humanitarian developments are reported through sources that vary in reliability, timeliness, language, access, and political perspective. The original research

proposal identifies the potential of machine learning and data analytics to detect patterns in terrorism-related events, analyse unstructured information, and inform policy. This paper retains that central interest but shifts the unit of analysis from a general global counter-terrorism system to the institutional needs of the African Union, RECs/RMs, and the ASF. (Lafree & Dugan, 2007; Raleigh et al., 2010; Weidmann, 2015)

The central policy argument is that the most defensible near-term role for AI is to strengthen the information-to-decision chain: collect and structure open and authorised information; detect changes and relationships; translate and summarise multilingual reporting; present uncertainty and competing interpretations; and support human-led analysis, planning, prevention, and civilian protection. Predictive outputs should be treated as hypotheses for review, not as facts or independent grounds for coercive action. (2024; Tabassi, 2023; UNESCO, 2021)

A. The paper proposes five priorities:

- Establish a federated, African-led conflict data and analytics capability connecting the AU, RECs/RMs, relevant early-warning structures, and ASF planning elements through agreed standards rather than an unrestricted centralised pool of sensitive data.
- Begin with lower-risk use cases: multilingual document triage, event coding, duplicate detection, trend summaries, source comparison, and analyst search across authorised documents. (Bommasani et al., 2021; Tabassi, 2023)
- Introduce ML and network analytics only where the use case is clear, the data are adequate, uncertainty can be communicated, and outputs can be independently validated. (Breiman, 2001; Tabassi, 2023)
- Apply safeguards grounded in international law, human rights, data protection, civilian protection, information security, and the AU's Continental AI Strategy; retain meaningful human review for consequential decisions. (2024; Tabassi, 2023; UNESCO, 2021)
- Build African institutional capacity through joint training, local-language resources, shared evaluation methods, public-interest research partnerships, and sustainable funding.

B. Drawing on the ECOWARN case, the proposed implementation pathway has three phases:

- Governance and readiness assessment;
- A bounded pilot using historical and open-source information for non-operational early warning
- Evaluated expansion to selected regional use cases. The paper recommends that the first pilot not include automated target identification, autonomous force decisions, individual-level suspicion scoring, or the ingestion of classified information into public or externally hosted LLM services. (2024; Tabassi, 2023)

An illustrative ECOWAS case demonstrates why the proposal should build on existing regional capabilities rather than assume a blank slate. ECOWAS publicly describes ECOWARN as a network combining field monitors, structured indicators, regional analysis, and preventive-action objectives; recent institutional updates also point to continued work

on data quality and AU–ECOWAS early-warning coordination (Commission, 2023, 2026a, 2026b). This case provides a concrete basis for discussing AI-assisted document triage, multilingual synthesis, source comparison, and risk-assessment support, while keeping the limits of the public evidence clear. The paper is a conceptual policy analysis and does not claim to establish that any model improves operational outcomes. Its contribution is a practical framework and research agenda that can be refined and tested through expert consultations.

I. INTRODUCTION: THE POLICY PROBLEM

Conflict prevention and response increasingly depend on the timely interpretation of information dispersed across official reports, local media, civil society monitoring, humanitarian assessments, open-source event datasets, and institutional reporting. In Africa, these information flows cross national boundaries and often involve multiple languages, uneven digital infrastructure, and different levels of access to reliable information. Regional mechanisms may need to understand whether a local dispute is escalating, whether political violence is spreading across a border, how violent extremist organisations are adapting, or whether civilian and humanitarian risks are changing. Such questions are political and contextual as much as technical. (Lafree & Dugan, 2007; Raleigh et al., 2010; Weidmann, 2015)

The African Peace and Security Architecture is designed to connect continental and regional capacities for conflict prevention, management, and resolution. The ASF forms part of this architecture and comprises multidisciplinary standby capabilities intended to support a range of authorised peace and security activities. The AU’s policy framework for the ASF recognises the role of the Union and RECs/RMs in operationalization and coordination (2003; Commission, 2016). This creates a policy rationale for exploring how data analytics might support shared situational awareness, early warning, mission planning, and lessons learned without undermining existing mandates or command structures.

The original research document, centred on a machine-learning approach to enhancing defence against terrorism, identifies recurring issues: fragmented event data, rare positive cases, class imbalance, changing threat behaviour, the limits of supervised models, and the potential value of anomaly detection and network analysis. It also proposes sources such as the Global Terrorism Database (GTD), the Armed Conflict Location & Event Data Project (ACLED), official reports, and news sources, alongside techniques including random forests, support vector machines, clustering, NLP, and network analysis. This policy paper preserves these themes while placing them in a regional African institutional setting and adopting a qualitative rather than model-testing design. (Breiman, 2001; Cortes & Vapnik, 1995; Lafree & Dugan, 2007; Raleigh et al., 2010). The main question is: How can ML, NLP, LLMs, and related data analytics be governed and integrated to support African regional mechanisms and the ASF in understanding armed conflict, political violence, violent extremism, and counter-terrorism while protecting human rights, civilian safety, and institutional legitimacy?

A. Three subsidiary questions guide the analysis:

- Which analytical tasks offer realistic and useful value for regional early warning and peace support planning?
- What institutional, data, technical, and legal barriers could constrain responsible use

- What governance and implementation model could support African ownership, interoperability, and accountability?

II. SCOPE, CONCEPTS, AND POLICY CONTEXT

A. Scope and terminology

The paper uses 'AI' as an umbrella term for computational systems that perform tasks associated with inference, classification, prediction, language processing, or content generation. 'Machine learning' refers to methods that identify patterns from data; 'NLP' concerns computational processing of human language; and 'large language models' are language models trained on large text corpora that can summarise, translate, extract information, and generate responses. These terms are related but not interchangeable. An LLM-generated summary is not a verified intelligence assessment, and an ML risk score is not a determination that an event or person is dangerous. (Bender et al., 2021; Bommasani et al., 2021; Vaswani et al., 2025)

'Political violence' is used broadly for organised or collective violence linked to political contestation, state repression, communal mobilisation, elections, insurgency, or other struggles over authority. 'Terrorism' is a contested legal and political category whose definition differs across legal systems and policy contexts. Data systems must preserve the source's original event classification and record the basis for any harmonised label rather than assume that all armed actors, opposition movements, insurgents, or protest groups fit a single category. This is essential to prevent analytical tools from converting disputed political labels into apparently objective facts. (Lafree & Dugan, 2007; Raleigh et al., 2010)

B. African institutional context

The AU Continental AI Strategy frames AI as a development and governance issue and calls for African-centred, ethical, responsible, and equitable approaches, with attention to infrastructure, talent, data, innovation, partnerships, and continental cooperation (2024a, 2024b). Applying this agenda to peace and security requires a sector-specific approach: operational relevance, protection of sensitive information, regional interoperability, civilian protection, and legal accountability must be built into the system from the beginning.

The ASF's regional structure makes a federated model attractive. RECs/RMs and member states may retain control over sensitive or context-specific information while sharing agreed metadata, analytical products, or authorised indicators. This would avoid assuming that all information should be centralised, while still creating opportunities for common standards, joint analysis, and coordinated learning. Any proposed system must complement, rather than bypass, established AU and regional early-warning, decision-making, and command arrangements. (2003; Commission, 2016)

C. Why the issue matters now

Data-driven tools can reduce the time required to review large volumes of reports, surface overlooked patterns, and help analysts compare developments across locations and

languages. However, the same tools can magnify reporting bias, misclassify idiomatic language, reproduce the political assumptions of their training data, or create unwarranted confidence in weak signals. The policy challenge is therefore not simply whether AI can be applied, but under what conditions it is useful, lawful, explainable, and institutionally accountable. (Bender et al., 2021; Tabassi, 2023; UNESCO, 2021)

III. ANALYTICAL APPROACH AND METHODOLOGY

This paper adopts a qualitative, policy-oriented design based on structured document analysis and conceptual synthesis. It draws on selected academic literature on conflict-event data and machine learning, and public policy documents relevant to the AU, the ASF, and responsible AI. The analysis organises the material around institutional use cases, enabling conditions, risks, and policy options. (2024; Lafree & Dugan, 2007; Raleigh et al., 2010). The method has four steps. First, it identifies the operational and analytical challenges: incomplete and inconsistent data, rare events, changing patterns, multilingual information, and the need to combine structured and unstructured sources. Second, it maps those challenges to potential capabilities of ML, NLP, LLMs, and network analysis. Third, it assesses the governance requirements that follow from the risks of bias, error, privacy intrusion, opacity, and misuse. Fourth, it develops a proposed framework and phased recommendations for the AU, RECs/RMs, ASF-related structures, and research partners. (Breiman, 2001; Cortes & Vapnik, 1995; Tabassi, 2023)

The paper does not report original interviews, focus groups, field observation, or a new computational experiment. A key limitation is that public documents cannot fully establish current internal data flows, technical maturity, procurement constraints, or classified operational requirements across all RECs/RMs. The proposed framework should therefore be treated as a decision-support model for consultation and piloting, not as a description of an existing AU capability.

IV. WHAT AI AND DATA ANALYTICS CAN—AND CANNOT—CONTRIBUTE

A. Machine learning for event patterns and anomaly detection

ML can assist analysts in identifying recurring event patterns, shifts in frequency, changes in location or reported tactics, and relationships between conflict indicators. Supervised methods such as random forests or gradient boosting learn from labelled examples; unsupervised methods such as clustering and anomaly detection can identify observations that differ from an established baseline. These techniques may help prioritise analyst attention, but they do not explain causality by themselves. A statistical association between an event and later violence is not proof that one caused the other. (Breiman, 2001; Cortes & Vapnik, 1995)

Conflict and terrorism datasets present particular methodological challenges. Violent events are unevenly reported; definitions change; data may be revised retrospectively; and the most severe incidents are not always the most completely documented. Rare-event imbalance can make a model appear accurate while missing the events of greatest policy concern. Evaluation should therefore report precision, recall, false-positive and false-negative rates, calibration, geographic and temporal performance, and the consequences of

errors—not accuracy alone. Event-data research highlights both the value of systematic coding and the importance of understanding how reporting and event definitions shape the resulting dataset (Lafree & Dugan, 2007; Raleigh et al., 2010). (Lafree & Dugan, 2007; Raleigh et al., 2010; Weidmann, 2015)

B. NLP for multilingual open-source analysis

NLP can support document classification, named-entity extraction, topic discovery, event coding, duplicate detection, translation assistance, and the comparison of narratives across reports. These functions may be particularly useful when information appears in English, French, Arabic, Portuguese, and African languages or dialects. Yet performance varies substantially across languages, spelling conventions, transliteration systems, dialects, and code-switching. A system trained mainly on high-resource languages can overlook local reporting or misread culturally specific expressions. (Bender et al., 2021; Bommasani et al., 2021; Vaswani et al., 2025)

A responsible workflow should retain the source text, date, language, provenance, translation status, and confidence level. Extracted entities and event labels should be reviewable by a human analyst. NLP should not infer that a person supports violence simply because the person is quoted in a report, belongs to a community mentioned in an incident, or uses language that a model associates with extremism. Research on online extremist content detection demonstrates that automated classification is possible, but it also raises questions about context, platform dependence, and the limits of treating language signals as reliable indicators of intent (Conway et al., 2017; Scrivens et al., 2018). (Bender et al., 2021; UNESCO, 2021)

C. Large language models as analyst assistants

LLMs may be useful for summarising long documents, generating first-pass timelines, comparing official statements, extracting claims for verification, drafting structured situation reports, and allowing analysts to search authorised document collections in natural language. They may lower the administrative burden of reviewing large volumes of material, but they can also generate plausible yet unsupported statements, omit contradictory evidence, misattribute claims, or conceal uncertainty behind fluent prose. (Bender et al., 2021; Bommasani et al., 2021)

LLM outputs should therefore be treated as drafts that require verification against source documents. Each factual statement in an analytical product should be traceable to a cited source, and the interface should distinguish direct evidence, model-generated synthesis, analyst interpretation, and unresolved uncertainty. Sensitive or classified information must not be entered into public or unapproved third-party systems. Retrieval-augmented generation over an approved document repository may improve traceability, but it does not eliminate errors or guarantee that sources are accurate. (Bender et al., 2021; Tabassi, 2023; UNESCO, 2021)

D. Network and geospatial analytics

Network analysis can represent relationships among reported events, organisations, locations, resources, and public claims. Geospatial and temporal analysis can help analysts compare incident sequences and identify changes in patterns across borders. These methods are most defensible when used to understand event dynamics, information flows, humanitarian exposure, or institutional coordination. They are much more sensitive when applied to infer individual culpability or to identify people for coercive action. A network connection is not proof of operational membership, and proximity is not proof of participation. (Raleigh et al., 2010; Weidmann, 2015)

E. Decision support, not automated command

AI may support planning by helping staff assemble a common operating picture, compare scenarios, identify information gaps, and track the assumptions underlying a warning. It should not independently determine whether force is deployed, who is designated a terrorist, who is detained, or who may be targeted. Such decisions require lawful authority, contextual judgement, accountability, and human responsibility. The central policy principle is augmentation rather than substitution. (2024; Tabassi, 2023; UNESCO, 2021)

V. ILLUSTRATIVE REGIONAL CASE STUDY: ECOWAS AND ECOWARN

The Economic Community of West African States (ECOWAS) provides a useful institutional case for considering how AI-enabled analytics might complement an established regional early-warning system. ECOWAS describes the ECOWAS Warning and Response Network (ECOWARN) as a mechanism established under the 1999 regional conflict-prevention protocol and operated by the Commission's Early Warning Directorate. Its public descriptions emphasise field reporting, regional monitoring, and the use of information to support prevention and response (ECOWAS Commission, n.d.; ECOWAS Commission, 2023) (Commission, 2023). This makes ECOWARN a relevant case for a policy paper concerned with strengthening—not replacing—human-led early warning.

A. Institutional features and policy relevance

ECOWAS reporting identifies a network of 77 field monitors and a framework using 55 predefined human-security indicators to support regional risk assessment. Public material also describes the system as combining qualitative and quantitative approaches, including the collection and verification of information and analysis by field monitors and headquarters-based analysts (Commission, 2023, 2024). These features are important because they show that regional early warning already depends on structured data, contextual interpretation, and human judgement. The policy question is therefore how AI could improve the speed, consistency, traceability, and accessibility of information moving through this workflow.

In June 2026, ECOWAS and the African Union undertook a technical engagement intended to strengthen cooperation between ECOWARN and the AU Continental Early Warning System (CEWS), including information sharing, joint conflict analysis, and governance monitoring. ECOWAS has also reported a programme to strengthen early-warning capacity through

modernised data collection, analytical capability, interoperability with AI tools, and analyst capacity-building (Commission, 2025, 2026). These publicly documented developments make interoperability and institutional ownership practical policy issues, rather than abstract design principles.

B. Responsible AI opportunities for ECOWARN

A bounded AI pilot could support ECOWARN analysts in at least four ways. First, NLP could help classify and summarise incoming reports while preserving links to the original source, language, date, location, and confidence assessment. Second, duplicate detection and entity resolution could flag reports that may describe the same event, reducing avoidable double counting while leaving final decisions to analysts. Third, multilingual search and translation assistance could help analysts compare relevant reporting across French, English, and other locally used languages; all translations and interpretations would require verification by competent speakers. Fourth, statistical or machine-learning tools could help identify changes in aggregate indicator patterns or unusual reporting shifts, provided the system distinguishes changes in violence from changes in media coverage or monitor activity.

These are proposed use cases, not claims that ECOWAS currently deploys all these functions. The available public material supports the existence of ECOWARN's early-warning architecture and ongoing capability-strengthening efforts, but it does not establish the accuracy or operational impact of any particular AI model. Any pilot should therefore be designed and evaluated jointly with authorised ECOWAS stakeholders, using approved data and clear safeguards.

C. Lessons, risks, and transferability to the AU and ASF

The first lesson is to augment the field-monitor and analyst network rather than automate away its contextual role. The quality of local reporting, source verification, and interpretation is an asset that AI should support. The second is that interoperability requires agreed definitions, metadata, access controls, and procedures for resolving conflicting assessments between national, regional, and continental systems. The ECOWAS–AU engagement indicates that coordination is already an institutional priority; any AI layer should reinforce those arrangements and the principles of subsidiarity and complementarity, not create a parallel reporting chain (Commission, 2026).

The third lesson is that early warning must be linked to early response. Better summaries or risk scores will not by themselves resolve political reluctance, mandate limitations, resource constraints, or coordination failures. AI-supported alerts should therefore include a traceable evidence trail, uncertainty labels, a named analyst responsible for validation, and a documented route to the appropriate decision-making body. Evaluation should measure not only technical accuracy and processing time, but also usefulness to analysts, false-alert burden, and inclusion of community and civil-society perspectives, protection of sensitive data, and whether findings inform lawful and timely preventive action.

The ECOWARN case is analytically illustrative rather than a full institutional audit. It relies on public documentation and cannot determine internal system performance, current technical architecture, or the causes of any warning-response gaps. Nevertheless, it grounds

the paper's central recommendation: African regional AI initiatives should be built around existing institutions, human expertise, and shared governance, beginning with low-risk analytical assistance and scaling only after independent evaluation.

VI. PRIORITY USE CASES FOR RECS/RMS AND THE ASF

Table 1: Priority Use Cases For Recs/Rms and the Asf

Use case	Potential contribution	Conditions and safeguards
Regional early-warning synthesis	Combine authorised event reports, official statements, humanitarian updates, and public reporting into a structured regional picture.	Preserve source provenance; flag contradictory reporting; show uncertainty and missing data.
Multilingual document triage	Translate, classify, summarise, and retrieve relevant documents across languages.	Human review by language-competent analysts; record original and translated text; test local-language performance.
Conflict-event data harmonisation	Support common event schemas, duplicate detection, and consistent metadata across regional datasets.	Do not erase source-specific definitions; document coding decisions and version history.
Trend and anomaly monitoring	Highlight unusual changes in event frequency, geography, reported tactics, or cross-border spillover indicators.	Treat anomalies as prompts for review, not proof of a threat; evaluate false alarms and missed events.
Peace support planning support	Summarise context, lessons learned, logistics constraints, civilian-protection reports, and scenario assumptions.	Use only authorised information; keep command decisions human-led; protect operational security.
Counter-terrorism policy analysis	Compare public reporting on recruitment narratives, financing risks, and prevention interventions at an aggregate level.	Avoid individual suspicion scoring from weak proxies; comply with the law and financial-data access controls.
After-action learning	Structure lessons from public or authorised mission reports and track recurring institutional recommendations.	Protect sensitive records; validate summaries; avoid treating generated text as the official record.

The recommended starting point is to prioritise administrative and analytical tasks with clear benefits and comparatively manageable risks. More sensitive use cases should require a separate legal basis, documented impact assessment, higher-level authorisation, independent review, and explicit criteria for stopping the system if harms exceed benefits. (Tabassi, 2023; UNESCO, 2021)

A. *A Proposed Regional Conflict Analytics and Early Warning Framework*

The proposed Regional Conflict Analytics and Early Warning Framework (R-CAEW) is a governance-and-workflow model, not a claim that a new platform already exists. It is designed to connect regional information needs with data quality, analytical methods, human validation, and accountable decisions. (2024; Tabassi, 2023)

Table 2: A Proposed Regional Conflict Analytics and Early Warning Framework

Layer	Core function	Policy design requirement
1. Mandate and task definition	Specify the decision or analytical question the system is intended to support.	A named institutional owner; clear lawful purpose; prohibited uses; review date.
2. Data stewardship and access	Catalogue authorised sources and define who can access, share, correct, and retain data.	Data-sharing agreements; provenance; classification; retention limits; audit logs.
3. Interoperability and data quality	Use common metadata, event definitions, timestamps, geospatial references, and version control.	Federated architecture; preserve source labels; record missingness and confidence.
4. Analytical services	Apply NLP, ML, network analysis, and LLM-assisted search to suitable tasks.	Model cards or equivalent documentation; language testing; reproducibility; version tracking.
5. Human validation and interpretation	Analysts check outputs against source evidence, local context, and alternative explanations.	No automated high-impact decisions; second review for consequential assessments.
6. Decision support and feedback	Deliver concise alerts, dashboards, and assessments to authorised decision-makers.	Explain uncertainty; record whether and how advice was used; allow corrections and challenge.
7. Independent governance and evaluation	Review legality, fairness, security, effectiveness, and unintended consequences.	Periodic audits; incident reporting; external scrutiny where feasible; stop/rollback authority.

The architecture should be federated by default. Regional bodies may retain source data under their own mandates while sharing agreed metadata, aggregate indicators, or approved analytical products. Where a shared repository is necessary, access should be tiered by role and purpose. Interoperability should mean that institutions can exchange agreed information safely—not that all data become universally accessible. (2024; Tabassi, 2023)

A minimum viable pilot should focus on a small number of public or explicitly authorised sources and a defined regional question, such as improving the speed and traceability of monthly conflict trend synthesis. It should not begin with a continent-wide predictive system. A narrow pilot allows the institutions to test data quality, user needs, language performance, governance procedures, and the real-world usefulness of outputs before expanding. (Tabassi, 2023)

VII. GOVERNANCE, LAW, ETHICS AND OPERATIONAL RISK

A. *Human rights, civilian protection and due process*

AI-supported security analysis can affect the rights and safety of individuals and communities even when a system is described as decision support. A false association, mistranslated statement, or unverified label may influence how a community is perceived or how resources are deployed. Governance should reflect applicable international human rights law, international humanitarian law where relevant, refugee law, national law, and

the mandates governing each institution. UNESCO's Recommendation on the Ethics of Artificial Intelligence emphasises human dignity, rights, transparency, proportionality, privacy, accountability, and human oversight (UNESCO, 2021). (2024; UNESCO, 2021)

The system should prohibit decisions about detention, designation, targeting, or the use of force from being made solely on the basis of AI output. Individuals should not be assigned risk labels based only on ethnicity, religion, nationality, language, location, political affiliation, family connections, or other protected or weakly correlated characteristics. Aggregate trend analysis may still create indirect harms; teams should therefore assess how indicators could stigmatise communities or divert resources unfairly. (Tabassi, 2023; UNESCO, 2021)

B. Bias, missing data and representativeness

Open-source conflict data are not a neutral mirror of violence. Events in areas with better connectivity, more media attention, or easier access for monitors may be more likely to appear in a dataset. State and non-state actors may also strategically distort reporting. A model can learn the geography of reporting rather than the geography of violence. Analysts should record source coverage, distinguish 'no reported event' from 'no event', and test performance across countries, languages, rural and urban settings, and different types of violence. (Raleigh et al., 2010; Weidmann, 2015)

C. Explain ability and uncertainty

Decision-makers should be able to understand why an alert was generated, what evidence supports it, what evidence contradicts it, and how confident the system is. A dashboard should not display a single risk score without explaining its meaning, limitations, and threshold. Confidence estimates should not be confused with the probability that a political event will occur unless the model has been appropriately calibrated and evaluated for that purpose. (Tabassi, 2023)

D. Cyber security, data sovereignty and procurement

Conflict-related data can expose sources, communities, operational plans, or institutional vulnerabilities. Data minimisation, encryption, access controls, logging, secure backups, retention rules, and incident response are therefore core requirements. Procurement contracts should clarify ownership of data and models, permitted reuse, hosting location, audit access, model updates, incident notification, and exit arrangements. African institutions should avoid creating critical dependencies on external vendors without transfer of skills, continuity plans, and meaningful control over sensitive data. (2024; Tabassi, 2023)

E. Generative AI-specific risks

LLMs can hallucinate, reproduce biased content, leak information through prompts or logs, and generate summaries that remove caveats. Controls should include approved models and environments, restricted data categories, source-grounded outputs, mandatory verification, clear labelling of generated content, and staff training. The NIST AI Risk Management Framework offers a useful lifecycle structure—govern, map, measure, and

manage risks—although it should be adapted to African mandates and legal contexts rather than adopted mechanically (Bender et al., 2021; Tabassi, 2023).

F. Governance principles for practical adoption

- Purpose limitation: each system must have a documented use case, lawful basis, owner, and expiry or review date.
- Human accountability: a named official remains responsible for decisions; AI output cannot transfer or dilute responsibility.
- Traceability: preserve source citations, data lineage, model versions, prompts or workflow settings where appropriate, and analyst changes.
- Proportionality: use the least intrusive method that can reasonably meet the legitimate objective.
- Contestability and correction: create a route for analysts and affected institutions to challenge errors and correct records.
- Security by design: protect sensitive data and ensure that external services are approved before use.
- Independent evaluation: test for false positives, false negatives, disparate performance, and unintended consequences before deployment and periodically thereafter.

VIII. IMPLEMENTATION ROADMAP

Table 3: Implementation Roadmap

Phase	Indicative period	Main actions	Gate before proceeding
Phase 1: Readiness and co-design	0–6 months	Map mandates, current workflows, data sources, languages, legal constraints, infrastructure, and user needs. Establish a multidisciplinary steering group.	Approved use case, risk assessment, data inventory, governance charter, and named accountable owner.
Phase 2: Bounded pilot	6–12 months	Prototype a document-triage and event-synthesis workflow using historical and open/authorised data. Compare AI-assisted work with existing analyst practice.	Independent evaluation of quality, time saved, false alerts, traceability, user trust, and security.
Phase 3: Regional interoperability	12–24 months	If pilot evidence is positive, agree common metadata, secure exchange protocols, training, and cross-regional analytical exercises.	Signed data-sharing arrangements, budget, incident response, audit plan, and clear exit criteria.
Phase 4: Sustained capability	24 months onward	Institutionalise training, model monitoring, lessons learned, research partnerships, and periodic policy review.	Regular reporting to authorised governance bodies; continued evidence of net benefit and compliance.

These periods are indicative planning windows, not a claim that implementation can be completed within them regardless of resources. Progress should be determined by readiness and evaluation gates rather than a fixed technology procurement calendar.

A. *Institutional roles*

The AU Commission and relevant peace and security structures could convene policy alignment and set continental interoperability principles, subject to existing mandates. RECs/RMs would identify regional priorities, steward locally relevant data, and lead contextual validation. ASF-related planning and training structures could specify operational information needs and incorporate approved analytical products into exercises and lessons-learned processes. Member states would determine lawful access to national data and ensure domestic legal compliance. Universities and research institutes could support independent evaluation, language-resource development, and capacity-building. Civil society and affected communities should be consulted on risks, terminology, and safeguards where appropriate. (2003; Commission, 2016)

B. *Skills and institutional capacity*

Training should not be limited to technical staff. Analysts need skills in source evaluation, uncertainty, data bias, and verification of generated summaries; commanders and policymakers need to understand what models can and cannot infer; legal advisers and data-protection officers need to review proposed uses; and technical teams need secure engineering and audit capabilities. A regional community of practice could share evaluation protocols, anonymised lessons, and language resources while respecting institutional restrictions. (Tabassi, 2023; UNESCO, 2021)

IX. MONITORING, EVALUATION AND LEARNING

A pilot should be assessed against the quality of the decision-support process, not only the technical performance of a model. Baseline measures should be collected before deployment and compared with the pilot workflow. The evaluation should include both quantitative indicators and qualitative feedback from intended users. (Tabassi, 2023)

Table 4: Evaluation Dimension

Evaluation dimension	Possible indicators	Interpretation caution
Timeliness	Time from source publication to reviewed analytical product; time spent on document triage.	Speed is not a benefit if verification quality falls.
Analytical quality	Source traceability; rate of corrected factual errors; agreement between reviewers; usefulness ratings.	Reviewer agreement is not proof of truth; preserve dissenting views.
Detection performance	Precision, recall, false-positive rate, false-negative rate, calibration where applicable.	Use appropriate base rates and report uncertainty; accuracy alone is insufficient.
Coverage and inclusion	Language coverage; geographic/source coverage;	More data do not automatically mean more representative data.

	missing-data patterns; performance gaps.	
Governance and security	Access violations; audit completeness; incidents; retention compliance; completed impact reviews.	Zero recorded incidents may indicate weak reporting rather than zero risk.
Institutional value	Analyst workload, user confidence, decisions informed, lessons incorporated.	Do not attribute changes in conflict outcomes to the tool without a credible evaluation design.

The evaluation should include a pre-agreed stop or rollback process. A pilot should be paused if it repeatedly generates serious uncorrected errors, exposes sensitive information, creates discriminatory outcomes, cannot provide traceable evidence, or is repurposed beyond its approved mandate. Findings should be documented in a learning report that distinguishes demonstrated benefits from hypotheses and unresolved questions. (Tabassi, 2023; UNESCO, 2021)

X. DISCUSSION: TOWARD AN AFRICAN REGIONAL MODEL

The proposed framework rests on three arguments. First, the most immediate value of AI for regional security institutions may lie in better information management rather than automated prediction. A well-governed workflow that makes evidence easier to find, compare, translate, and verify may be more useful than a complex model built on inconsistent data. Second, regional interoperability is a governance challenge as much as a technical one. Shared standards require agreement on definitions, access, responsibility, and how uncertainty is communicated. Third, legitimacy is operationally important: systems that are perceived as opaque, discriminatory, externally controlled, or politically instrumentalised may weaken trust and reduce the quality of information available to institutions. (Bender et al., 2021; Bommasani et al., 2021; Tabassi, 2023)

These arguments caution against treating AI as a shortcut around political analysis or institutional coordination. Armed conflict and political violence arise from interacting historical, economic, political, social, and security dynamics. Data analytics can help identify signals and structure evidence, but they cannot independently explain grievances, determine political intent, or establish the legitimacy of a response. Local knowledge and expert judgement remain indispensable. (2024; UNESCO, 2021)

The African Union's Continental AI Strategy provides a relevant policy foundation for an Africa-centred approach, while the ASF and regional mechanisms provide an institutional setting in which interoperability and shared learning can be explored. A successful model should build local capability and not merely import tools. This includes African control over priorities, locally meaningful evaluation, language inclusion, transparent procurement, and sustainable financing. (2024; Commission, 2016)

The policy framework is deliberately cautious about person-level profiling and operational targeting. Such applications carry especially high risks of error, rights violations, escalation, and loss of accountability. The proposed initial focus is aggregate and institutional: understanding patterns in reported events, synthesising public or authorised information, supporting planning, and improving learning. More sensitive uses, if ever considered, would

require a separate legal and ethical assessment and cannot be justified merely by technical feasibility. (Tabassi, 2023; UNESCO, 2021)

XI. CONCLUSION

Machine learning, NLP, LLMs, and related analytics could help African regional mechanisms and the ASF strengthen early warning, situational awareness, cross-border analysis, and institutional learning. Their contribution will depend on the quality of source data, the fit between a model and a real policy question, regional interoperability, local-language capability, and robust safeguards. A technically sophisticated system may still fail if it confuses reported absence with absence of violence, treats disputed labels as facts, or hides weak evidence behind a confident score. (2024a, 2024b; Tabassi, 2023)

This paper proposes a Regional Conflict Analytics and Early Warning Framework that links mandate definition, data stewardship, interoperability, analytical services, human validation, decision support, and independent evaluation. It recommends beginning with bounded, non-operational use cases and expanding only when evaluation demonstrates clear benefit and governance controls work in practice. The core principle is that, AI should strengthen African-led judgement and civilian protection, not replace lawful authority, political analysis, or human responsibility. (2024; Tabassi, 2023; UNESCO, 2021)

The next research stage should test and refine the framework through expert consultation and additional regional case studies, subject to access and ethics requirements. The present paper offers a focused conceptual and policy contribution to debates on responsible AI, conflict analysis, and African regional security.

POLICY RECOMMENDATIONS

- Adopt an African-led policy position on responsible AI for regional peace and security. The AU and RECs/RMs should translate the principles of the Continental AI Strategy into a sector-specific governance note for conflict early warning, peace support planning, and counter-terrorism policy analysis. (2024a, 2024b)
- Create a federated regional conflict-data interoperability arrangement. Agree minimum metadata, event-coding guidance, provenance requirements, confidence labels, access tiers, and secure exchange procedures while preserving institutional authority over sensitive information. (2024; Tabassi, 2023)
- Start with low-risk, high-utility analytical support. Prioritise document triage, multilingual retrieval, source comparison, duplicate detection, event coding assistance, and analyst-facing summaries before considering higher-risk predictive applications.
- Require human verification and evidence traceability. AI-generated claims must be traceable to source material; uncertainty and contradictions must remain visible; no coercive or use-of-force decision should be based solely on automated output. (2024; UNESCO, 2021)
- Build regional capacity in local languages and contextual validation. Fund language resources, analyst training, multidisciplinary teams, and partnerships with African universities and research centres.

- Establish independent risk and performance reviews. Require pre-deployment impact assessments, ongoing bias and security testing, audit logs, model documentation, and public or institutional reporting appropriate to the sensitivity of the work. (Tabassi, 2023; UNESCO, 2021)
- Use staged pilots and evidence-based scale-up. Each pilot should have an accountable owner, defined baseline, success criteria, stop conditions, realistic budget, and independent evaluation.
- Develop responsible procurement and data-sovereignty clauses. Contracts should protect data ownership, restrict vendor reuse, specify hosting and access controls, provide audit rights, and enable institutions to exit without losing essential records or capabilities.
- Keep prevention and civilian protection central. Evaluate whether analytics improve understanding of community concerns, humanitarian risks, and conflict drivers—not only whether they increase security monitoring.
- Commission a follow-on empirical study. Test the proposed framework through structured expert interviews and comparative analysis of ECOWARN and at least one other regional mechanism, subject to ethics approval and access permissions.

REFERENCES

- [1] African Union Commission. (2016). the Peace Support Operations Division (PSOD). Peace and Security Department. <https://archives.au.int/handle/123456789/5016>
- [2] African Union Peace and Security Council. (2024, June 13). Communiqué of the 1214th meeting of the Peace and Security Council held on 13 June 2024, on artificial intelligence and its impact on peace and security in Africa. <https://www.peaceau.org/en/article/communique-of-the-1214th-meeting-of-the-peace-and-security-council-held-on-13-june-2024-on-artificial-intelligence-and-its-impact-on-peace-and-security-in-africa>
- [3] African Union. (2003). Policy framework for the establishment of the African Standby Force and the Military Staff Committee. Document adopted by the Third Meeting of African Chiefs of Defence Staff, 15–16 May 2003. African Union. <https://library.au.int/policy-framework-establishment-african-standby-force-and-military-staf-committee>
- [4] African Union. (2024). Continental artificial intelligence strategy. African Union Commission. <https://au.int/en/documents/20240809/continental-artificial-intelligence-strategy>
- [5] Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the dangers of stochastic parrots: Can language models be too big? Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency, 610–623. <https://doi.org/10.1145/3442188.3445922>
- [6] Bommasani, R., Hudson, D. A., Adeli, E., Altman, R., Arora, S., von Arx, S., Bernstein, M. S., Bohg, J., Bosselut, A., Brunskill, E., Brynjolfsson, E., Buch, S., Card, D., Castellon, R., Chatterji, N., Chen, A., Creel, K., Davis, J. Q., Demszky, D., . . . Liang, P. (2021). On the opportunities and risks of foundation models. arXiv. <https://arxiv.org/abs/2108.07258>
- [7] Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5–32. <https://doi.org/10.1023/A:1010933404324>

- [8] Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning*, 20, 273–297. <https://doi.org/10.1007/BF00994018>
- [9] ECOWAS Commission. (2023, June 13). ECOWAS Early Warning Directorate holds retreat to strengthen competences of field monitors. <https://www.ecowas.int/ecowas-early-warning-directorate-holds-retreat-to-strengthen-competences-of-field-monitors/>
- [10] ECOWAS Commission. (2024, March 15). ECOWAS strengthens analytical capabilities of ECOWARN field monitors. <https://www.ecowas.int/ecowas-strengthen-analytical-capabilities-of-ecowarn-field-monitors/>
- [11] ECOWAS Commission. (2025, November 13). ECOWAS and AfDB join forces to further strengthen the regional organisation's early warning capacity. <https://www.ecowas.int/ecowas-and-afdb-join-forces-to-further-strengthen-the-regional-organisations-early-warning-capacity/>
- [12] ECOWAS Commission. (2026b, June 9). ECOWAS and AU deepen cooperation on early warning and conflict prevention mechanisms. <https://www.ecowas.int/ecowas-and-au-deepen-cooperation-on-early-warning-and-conflict-prevention-mechanisms/>
- [13] ECOWAS Commission. (2026a, February 27). ECOWARN annual retreat to strengthen data collection capacities and support early warning systems and preventive measures in West Africa. <https://www.ecowas.int/ecowarn-annual-retreat-to-strengthen-data-collection-capacities-and-support-early-warning-systems-and-preventive-measures-in-west-africa/>
- [14] ECOWAS Commission. (n.d.). ECOWAS Warning and Response Network (ECOWARN). <https://ecowarn.ecowas.int/>
- [15] LaFree, G., & Dugan, L. (2007). Introducing the Global Terrorism Database. *Terrorism and Political Violence*, 19(2), 181–204. <https://doi.org/10.1080/09546550701246817>
- [16] Raleigh, C., Linke, A., Hegre, H., & Karlsen, J. (2010). Introducing ACLED: An armed conflict location and event dataset. *Journal of Peace Research*, 47(5), 651–660. <https://doi.org/10.1177/0022343309355235>
- [17] Tabassi, E. (2023). Artificial intelligence risk management framework (AI RMF 1.0) (NIST AI 100-1). National Institute of Standards and Technology. <https://doi.org/10.6028/NIST.AI.100-1>
- [18] UNESCO. (2021). Recommendation on the ethics of artificial intelligence. <https://www.unesco.org/en/legal-affairs/recommendation-ethics-artificial-intelligence>
- [19] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is all you need. *Advances in Neural Information Processing Systems*, 30, 5998–6008. <https://papers.nips.cc/paper/7181-attention-is-all-you-need>
- [20] Weidmann, N. B. (2016). A closer look at reporting bias in conflict event data. *American Journal of Political Science*, 60(1), 206–218. <https://doi.org/10.1111/ajps.12196>