Decolonising the Algorithm: An Afrocentric Framework for AI-Powered Information Repositories in Selected African Tertiary Libraries

Munyaradzi Janyure[1], Omega Ndlovu[2], Walter Garikayi Masvikeni[3]

[1] [2]Librarians, Zimbabwe Military Academy, Gweru, Zimbabwe

[3] DARE Digital Resources, Harare, Zimbabwe

DOI: https://doi.org/10.47772/IJRISS.2026.100900719

Received: 07 October 2026; Accepted: 12 October 2026; Published: 22 October 2026

ABSTRACT

This paper explores the intersection of artificial intelligence (AI), decolonisation, and indigenous knowledge systems within selected African tertiary libraries. It argues that current AI-powered information repositories often perpetuate Eurocentric epistemologies, marginalising African perspectives and reinforcing epistemic injustice in higher education. By proposing an Afrocentric framework, the study highlights how AI can be re-engineered to respect cultural diversity, integrate indigenous knowledge, and promote equitable information retrieval. Drawing on documentary analysis of scholarly articles, institutional reports, and policy frameworks from selected African tertiary institutions, the research identifies four critical dimensions requiring transformation: metadata epistemologies, training datasets, community collaboration, and ethical safeguards. The findings reveal that existing metadata standards cannot adequately represent oral traditions and community held knowledge; AI models are trained predominantly on Western datasets, with less than 1% reflecting African linguistic contexts; communities are excluded from decisions about how their knowledge is digitized and represented and ethical frameworks remain at the level of principle without enforceable mechanisms. The paper argues that decolonising algorithms is not only a technical necessity but also an ethical imperative for African higher education. An Afrocentric framework must operate across all four dimensions simultaneously, shifting from extractive data practices toward community data sovereignty and epistemic justice. The paper concludes that epistemic justice requires moving beyond procedural inclusion toward fundamental restructuring of AI systems to centre African knowledge systems, languages, and community governance structures. The framework proposed a path toward genuinely inclusive information repositories, but its realization depends on sustained institutional commitment, adequate resourcing, and genuine collaboration among institutions, communities, and library professionals.

Keywords: Decolonisation, AI ethics, Indigenous Knowledge, Information Retrieval, Afrocentric Framework, Epistemic Justice

INTRODUCTION

Artificial intelligence (AI) has rapidly transformed information retrieval and knowledge management within tertiary institutions worldwide. However, in African contexts, AI-powered repositories often reproduce Eurocentric epistemologies, sidelining indigenous knowledge systems and perpetuating epistemic injustice (Ngubane & Omodan, 2026). This paper explores the intersection of AI, decolonisation, and indigenous knowledge within African tertiary libraries, arguing for an Afrocentric framework that re-engineer algorithms to respect cultural diversity, integrate indigenous epistemologies, and promote equitable access. By situating AI within African knowledge traditions, the study underscores the ethical imperative of decolonising algorithms to achieve epistemic justice in higher education.

This paper introduces an Afrocentric framework for AI-powered repositories in African tertiary libraries, emphasizing decolonisation, ethical AI, and indigenous knowledge integration. The framework is grounded in the recognition that epistemic justice the fair distribution of epistemic resources and recognition of diverse

knowledge systems requires not merely technical adjustments but fundamental reorientation of how AI systems are conceptualized, designed, and governed.

Background of The Study

African universities and libraries remain deeply influenced by colonial knowledge hierarchies, where Western epistemologies dominate curricula, metadata structures, and information repositories (Mkhize, 2026). Scholars argue that this epistemic dominance marginalises African perspectives and languages, reinforcing structural inequalities in knowledge production (Mabofua, 2020). The persistence of these legacies necessitates a decolonial rethinking of AI systems to ensure inclusivity and cultural relevance.

The mainstreaming of AI in higher education has introduced new challenges regarding whose knowledge is valued and whose languages are represented. Current AI models are trained predominantly on Western datasets, leading to algorithmic coloniality and data extractivism that exclude African epistemologies (Ngubane & Omodan, 2026). This exclusion perpetuates epistemic injustice, making it imperative to interrogate AI's role in reproducing colonial structures within African tertiary libraries.

Indigenous knowledge systems embody holistic, community-centred epistemologies that emphasize relationality, sustainability, and cultural continuity. Integrating these frameworks into AI design can foster abundant intelligences that respect cultural diversity (Springer Nature, 2025). An Afrocentric approach to AI in libraries would prioritise African languages, oral traditions, and community data sovereignty, thereby challenging Eurocentric dominance in information retrieval.

This study adopted a qualitative documentary analysis as its sole research method, relying on secondary sources such as scholarly articles, institutional reports, and policy frameworks. Documentary analysis is particularly suited for examining how AI intersects with decolonisation and indigenous knowledge, as it enables critical engagement with existing literature and policy frameworks (Bowen, 2009). By synthesizing insights from diverse sources, the study identified four critical dimensions requiring transformation: metadata epistemologies, training datasets, community collaboration, and ethical safeguards.

Statement of The Problem

Despite the transformative potential of artificial intelligence (AI) in enhancing information retrieval and knowledge management, African tertiary libraries continue to grapple with the problem of Eurocentric dominance embedded within AI-powered repositories. These systems, largely trained on Western datasets and guided by colonial metadata structures, marginalise indigenous knowledge systems and perpetuate epistemic injustice by excluding African languages, oral traditions, and community-centred epistemologies (Ngubane & Omodan, 2026; Hiribarren & Madore, 2025). This entrenched imbalance not only undermines cultural diversity but also reinforces structural inequalities in knowledge production and dissemination. Consequently, there is an urgent need to re-engineer AI through an Afrocentric framework that integrates indigenous epistemologies, ensures ethical safeguards, and promotes equitable access to knowledge, thereby advancing epistemic justice in African higher education.

Research Objectives

1. To examine the colonial biases embedded in AI-powered information repositories in African tertiary libraries.

2. To explore how indigenous knowledge can be integrated into AI-driven information retrieval systems.

3. To propose an Afrocentric framework for decolonising algorithms in African tertiary libraries.

Research Questions

1. What colonial legacies are reproduced in current AI-powered repositories in African tertiary libraries?

2. How can indigenous knowledge be integrated into AI-driven information retrieval systems in African tertiary libraries?

3. What Afrocentric framework can be proposed for decolonising algorithms in African tertiary libraries?

LITERATURE REVIEW

AI And Algorithmic Bias

The scholarly literature on algorithmic bias has established that AI systems are not neutral technical tools but rather embed and reproduce social inequalities. Noble (2018) landmark study demonstrated how search engines reinforce racism through their ranking and recommendation algorithms, while Benjamin (2019) introduced the concept of the "New Jim Code" to describe how seemingly objective technologies perpetuate racialized exclusion.

In the African context, algorithmic bias manifests in multiple domains. Buolamwini and Gebru's (2018) research on facial recognition systems showed significantly higher error rates for darker-skinned women, with implications for AI applications in security, banking, and healthcare. Birhane (2020) extends this analysis to argue that AI development in Africa constitutes a form of "algorithmic colonization," where Western epistemologies are imposed through technological infrastructure while local knowledge systems are marginalised.

Recent scholarship has begun to quantify these disparities specifically in library contexts. A 2022 audit of AI-powered discovery systems found that African-authored research articles received one-third fewer algorithmic recommendations than comparable Global North scholarship, effectively making African academic output less discoverable even when technically available. This finding underscores the urgency of addressing not merely access to AI technologies but the epistemological foundations of algorithmic relevance and ranking.

Decolonisation In Knowledge Systems

Decolonial theory provides essential analytical tools for understanding how colonial power relations persist in knowledge production and organisation. Quijano's (2007) concept of the "coloniality of knowledge" describes how Western epistemologies became universalized and positioned as superior to other knowledge systems, a process that continues to structure academic institutions globally. Mignolo (2011) advocates for "epistemic disobedience" and "delinking" from colonial knowledge structures, while Grosfoguel (2017) proposes "pluri-versalism" as an alternative to universalist claims of Western epistemology.

African scholars have made distinctive contributions to these debates. Ngugi (1986) analysis of language politics in African literature demonstrated how the choice of language for scholarly and literary expression constitutes an epistemic and political act. Mbembe's (2016) call for decolonising the African university emphasizes the need to confront the epistemic foundations of higher education institutions, not merely their demographic composition or curricular content.

In library and information science, the decolonial turn has generated critical examinations of classification systems, metadata standards, and collection development practices. The Makerere University library case study reveals how mid-twentieth-century cataloguing standards produced a "problematically narrow ontology" of African literature, an epistemic framework that continues to shape how knowledge is organised and discovered even in digital environments (Harris, 2025).

African Libraries And Indigenous Knowledge

The marginalisation of indigenous knowledge in African libraries is a well-documented phenomenon. Moahi (2012) and Ocholla's (2007) agenda-setting work on "marginalised knowledge" identified systemic barriers to integrating indigenous knowledge into library systems. Recent empirical research has confirmed and extended these findings. A 2026 study of Ghanaian public libraries found that collection development policies universally failed to adequately encompass indigenous knowledge documentation and integration, with only a

few cases offering minimal support (Dlamini, Jiyane & Mohammed, 2026). This finding underscores a critical challenge in preserving oral histories, traditional practices, and local languages within formal library systems.

The challenges of representing indigenous knowledge in digital systems are both technical and epistemological. Oral traditions resist text-based metadata schemas; community-held knowledge challenges individual authorship attribution models; and holistic knowledge forms resist disciplinary fragmentation. As one analysis notes, "indigenous knowledge systems are often relational, contextual, and embedded in practice" rather than conforming to the commodified, decontextualized forms that AI systems are designed to process (Kuranga, Khumalo & Masenya, 2026).

Promising initiatives exist, however. The Masakhane African Languages Hub operates on Ubuntu principles, positioning African communities as "co-creators rather than passive data subjects" and advocating for linguistic sovereignty the ability of African languages to thrive in digital spaces with local ownership and control. The Transformedia project in Sudan has digitized over 7,000 metadata records and 2,500 audio records of intangible cultural heritage, recruiting 18 community ambassadors to drive locally-led documentation (British Council, 2026). These examples demonstrate that community-centred, sovereignty-respecting approaches to indigenous knowledge digitization are feasible, even in challenging contexts.

AI Ethics In Africa

The emergence of AI ethics as a field has been dominated by frameworks developed in the Global North, including the NIST AI Risk Management Framework and IEEE's Ethically Aligned Design. These frameworks, while valuable, embed assumptions about infrastructure, legal systems, and individual rights that may not align with African contexts.

African scholars have proposed alternative ethical foundations rooted in indigenous philosophy. Mhlambi (2020) advances Ubuntu as a relational foundation for AI ethics, emphasizing communal welfare, human dignity, and social harmony over individual autonomy. The Ubuntu principle "I am because we are" suggests fundamentally different approaches to privacy, consent, and data governance than Western individualist frameworks.

A 2026 systematic review of AI-indigenous knowledge integration in African societies found that while Ubuntu is "widely proposed as a relational AI ethics alternative," its "operationalization is minimal," with only two prototypes documented (Goshu & Ridwan, 2026). The review identified 34 studies meeting inclusion criteria, with applications in agriculture (ITIKI drought prediction), health (AI medicinal plant mining), language, and architecture. Yet only three systems were field-validated, and governance remains fragmented: no African country has a dedicated indigenous knowledge AI policy.

This gap between conceptual richness and empirical implementation is a central concern for the present study. The Afrocentric framework proposed here responds to this gap by specifying concrete dimensions of transformation, from metadata design to community governance, that can guide practical implementation.

METHODOLOGY

Research Design And Epistemological Framing

This study adopts a qualitative research design rooted in critical documentary analysis (Bowen, 2009). Documentary analysis is uniquely suited for investigating sociotechnical phenomena where institutional policy texts, technical standards, and licensing agreements codify underlying epistemological assumptions and power relations. Documents are interrogated not as inert bureaucratic records, but as socially situated artifacts that reflect institutional commitments and systemic biases.

The research is theoretically anchored in Asante’s (1980) Afrocentric Theory and the Modernity/Coloniality/Decoloniality (MCD) framework. Afrocentricity positions African experiences and indigenous knowledge systems at the centre of analysis. Decolonial epistemology provides diagnostic tools to

interrogate the coloniality of knowledge (Quijano, 2007), enact epistemic disobedience (Mignolo, 2011), and achieve cognitive justice (Santos, 2014) within digital library architectures.

Document Search Strategy And Systematic Attrition Pipeline

Data collection adhered to a systematic, multi-pronged search protocol across scholarly literature, institutional records, and policy frameworks published between 2015 and 2026. This timeframe encompasses the accelerated commercial deployment of machine learning in academic libraries alongside the maturation of critical decolonial AI scholarship. Queries were executed across Web of Science, Scopus, Google Scholar, and institutional repository portals utilizing Boolean search strings combining terms: ('artificial intelligence' OR 'AI' OR 'algorithmic bias' OR 'machine learning' OR 'discovery layers') AND ('libraries' OR 'repositories' OR 'cataloguing' OR 'metadata') AND ('decolonisation' OR 'epistemic justice' OR 'cognitive justice') AND ('indigenous knowledge' OR 'African languages' OR 'Ubuntu' OR 'data sovereignty') AND ('Africa' OR 'Zimbabwe' OR 'South Africa' OR 'Uganda' OR 'Global South').

Document selection followed rigorous inclusion and exclusion boundaries. Inclusion criteria required: (a) publication between 2015 and 2026; (b) direct substantive focus on AI, metadata, or decolonisation in tertiary repositories; (c) empirical or policy grounding in African higher education or comparative Global South library contexts; (d) peer-reviewed papers, institutional policies, or official agency reports; and (e) English full-text availability. Non-English works, commercial marketing brochures, and unverified opinion essays were excluded.

Table A summarizes the systematic multi-stage document selection and attrition pipeline:

Stage	**Screening Phase & Criteria**	**Document Counts**
1. Identification	Electronic databases (n = 142); University & consortia repositories (n = 38); AI alerts & citation snowballing (n = 26).	Total: N = 206
2. Pre-Screening	Duplicate removal across indexes (n = 34); Non-English records excluded (n = 12).	Retained: n = 160
3. Title/Abstract	Assessed against eligibility criteria; excluded commercial collateral, out-of-scope non-LIS works, and ungrounded opinion pieces (n = 89).	Retained: n = 71
4. Full-Text Appraisal	In-depth qualitative review; excluded studies lacking technical repository focus, metadata depth, or empirical institutional grounding (n = 28).	Retained: N = 43
5. Final Analysed Corpus	Stratified corpus: Peer-reviewed articles/chapters (n = 19); University AI policies & roadmaps (n = 11); Consortia guidelines (n = 8); Agency reports (n = 5).	Final Corpus: N = 43

Institutional Contexts Considered

To ground documentary findings within concrete institutional realities, the documentary corpus incorporates primary institutional records and case studies from two complementary institutional tiers:

1. Primary National Tier (Zimbabwe Tertiary Institutions): Encompasses Midlands State University (MSU; repository collection policies and curriculum restructuring), University of Zimbabwe (UZ; heritage archiving and library systems), National University of Science and Technology (NUST; IT automation and staff literacy), Bindura University of Science Education (BUSE; indigenous knowledge and scientific archiving), and the Zimbabwe University Libraries Consortium (ZULC; consortium-wide licensing and vendor agreements).

2. Comparative Regional Tier (Continental Benchmarks): Encompasses the University of Pretoria (South Africa; indigenous cataloguing adaptations, color-coded tagging, and genre markers for African-language collections), Makerere University (Uganda; historical cataloguing ontology reforms), the Masakhane African Languages Hub (grassroots NLP co-creation models), and the Transformedia Project in Sudan (community-led heritage documentation under challenging environments).

Data Analysis: Deductive–Inductive Thematic Coding Scheme

Documentary data were analysed using a dual qualitative coding protocol. Deductive codes were anchored in Asante's Afrocentricity and the Modernity/Coloniality/Decoloniality (MCD) framework; inductive codes captured themes emerging from the documents. Coding units were passages, policy provisions, technical descriptions, and reported empirical results relevant to repository discovery, metadata, African-language technologies, community governance, and institutional capacity. A theme was treated as a documentary finding only when it was supported by material in the included corpus. Recommendations proposed by the authors were recorded as normative implications, not as observed outcomes.

Table B summarizes the thematic dimensions and their analytic focus:

Coding Tier	Thematic Dimension	Operational Definition & Analytic Focus
Deductive Coding (Theory-Driven: Asante & MCD Framework)	Epistemic & Algorithmic Coloniality	How discovery and ranking systems may reproduce existing visibility inequalities; claims of measured ranking disparities require a documented system audit.
	Metadata & Ontological Erasure	Whether metadata standards adequately represent oral, contextual, relational, and community-held knowledge.
	Linguistic Sovereignty	Representation and resourcing of African languages in digital and AI systems, without assuming a universal token percentage.
	Relational Ethics & Governance	Institutional application of Ubuntu ethics and the CARE Principles.
Inductive Coding (Data-Driven: Emergent Empirical Themes)	Commercial Vendor Lock-In	Institutional dependence on commercial discovery platforms and the limits of local control, where documented.
	Capacity-Policy Disconnect	Policy, skills, and training concerns reported in the included literature; no percentage is inferred unless a source reports a suitable denominator and measure.
	Ethics Washing & Tokenism	Risks of symbolic or superficial diversity interventions, where directly documented.
	Custodial vs. IP Tensions	Tensions between collective Indigenous governance and individual intellectual-property regimes.

The selection flow reports aggregate counts for identification, screening, and final inclusion. These counts provide an overview of the selection process; complete replication would additionally require search dates,

database-specific query strings, source-level retrieval counts, and a record of the reasons for each full-text exclusion.

Analytical Consistency, Trustworthiness, And Ethical Standards

To support analytical consistency, a standardized codebook was developed and pilot-tested against five randomly selected documents across institutional categories. Three researchers independently coded the 43-document corpus using this codebook. Differences in category assignment were discussed in structured consensus meetings, with a third researcher helping to resolve ambiguous cases.

Methodological trustworthiness was ensured via data source triangulation across peer-reviewed research, university policies, commercial vendor licensing terms, and international agency reports. A comprehensive audit trail documented all keyword iterations, exclusion rationales, and analytical memos. Ethical stewardship of indigenous data was maintained in alignment with the CARE Principles for Indigenous Data Governance (Carroll et al., 2020), which build on earlier work on Indigenous data sovereignty (Carroll et al., 2019; Figueroa-Rodriguez et al., 2019), ensuring cultural knowledge was not treated as raw data for academic extraction

FINDINGS

Algorithmic Prioritisation Of Western Scholarship

Document analysis revealed systematic bias in the content that AI systems prioritise and surface. At one research-intensive Southern African university, analysis of AI-powered resource recommendation data showed that publishing in Africa is predominantly in English. The structural concentration of African research publishing in a small number of countries and the dominance of English-language journals (Ndege et al., 2025) create conditions in which AI systems trained on Global North corpora are likely to systematically under-represent African scholarship. This disparity could not be explained by collection composition alone, as the library's physical holdings included substantial African-published materials.

Research conducted by Shizha & Kariwo (2012) showed that, IT specialist at one tertiary institution in Zimbabwe described the mechanism: "The system learns from what users click on, and what users click on is what the system showed them first, which was mostly Western content because that's what the algorithm thinks is important. It's a feedback loop that excludes our own knowledge".

An article by Tajjiou (2024) noted that, when students search for African authors, they get results that frame African literature through a Western gaze; comparative studies by foreign scholars, rather than the primary texts themselves or African literary criticism. The article, it critically examined the stagnation of the comparative African literatures paradigm and advocated for an inward continental-centric framework. While Tajjiou's analysis addresses the discipline of comparative literature, the underlying epistemic mechanism of treating African knowledge forms as cultural data rather than theoretical sources applies equally to AI recommendation systems that rank African scholarship through Eurocentric relevance criteria.

Metadata That Cannot See Indigenous Knowledge

The analysis of metadata standards and practices revealed fundamental limitations in how AI systems represent indigenous knowledge. Existing classification systems lack adequate categories for oral traditions, community-held knowledge, and non-textual knowledge forms. Document analysis revealed that existing classification systems lack adequate categories for oral traditions and community-held knowledge, a limitation repeatedly noted in the literature on indigenous knowledge organisation (Nyambi & Maynard, 2012; Moahi, 2012).

The inadequacy of existing metadata is compounded by the training data used for automated classification and recommendation. AI models trained primarily on Western scholarly bodies lack the cultural knowledge to recognise and appropriately categorize African knowledge forms. Gupta, Taak & Singh (2026) found that non-

community cataloguers consistently misclassified materials, sometimes in ways that were not merely inaccurate but culturally offensive, such as categorizing ceremonial knowledge as folklore or superstition.

Critical Awareness Without Technical Capacity

Document analysis revealed a striking pattern: library professionals across institutions demonstrated sophisticated critical awareness of algorithmic bias and epistemic exclusion, yet felt powerless to address these issues. Document analysis of institutional reports and policy documents revealed a pattern in which library professionals expressed critical awareness of algorithmic bias but felt constrained by vendor relationships and lack of institutional support (see also Tsekea & Mandoga, 2025).

This capacity gap was reflected in training and policy data. The study of Zimbawean University libraries found that most libraries had not developed formal AI ethics policies, and structured training in AI literacy and ethics remained limited across participating institutions (Tsekea & Mandoga, 2025). Only a few institutions in the study had developed formal AI governance policies, and those policies focused primarily on data privacy and academic integrity rather than epistemic justice.

The vendor relationship emerged as a particular source of frustration. Most AI-powered library systems in use at participating institutions were proprietary products of North American or European companies that are not prioritising Afrocentric perspectives. Literature described limited ability to modify system behaviour, transparency into algorithmic decision-making, or influence over development priorities. The African information scientists are solely customers of digital systems not partners and they are made to adapt to systems built for Global North markets (Shizha & Kariwo, 2012).

Emergent Afrocentric Approaches To Indigenous Knowledge Integration

Despite these challenges, the research identified emergent initiatives that point toward Afrocentric alternatives and demonstrate how indigenous knowledge can be integrated into AI-driven systems. At one institution, a library had developed a locally controlled metadata schema that incorporated community-defined categories for indigenous knowledge. At another, librarians had partnered with local communities to develop a digital repository governed by community protocols, including restrictions on access to sensitive ceremonial knowledge (Kuranga, Khumalo & Masenya, 2026).

The Masakhane African Languages Hub model, documented in book: The Ubuntu Practitioner: Social Work Perspectives and referenced in several documents, exemplifies the principles of linguistic sovereignty and community co-creation. Operating on Ubuntu principles, the Hub positions. African communities as decision-makers rather than data sources, and advocates for non-extractive collaboration that ensures value generated from African language data remains accessible to African ecosystems (Balyejjusa et al., 2023).

The Transformedia project in Sudan demonstrated that community-led digitization can proceed even in contexts of active conflict, with 18 community ambassadors recruited to drive locally-led documentation (British Council, 2026). This model suggests that the capacity for Afrocentric, community-centred approaches exists, but requires institutional commitment and resource allocation to scale.

Proposed Afrocentric Framework For Decolonising Algorithms

The analysis identified four interconnected dimensions that any Afrocentric framework for decolonising algorithms must address:

Dimension One: Metadata Epistemologies

Existing classification systems embody Western worldviews and cannot adequately represent indigenous knowledge. An Afrocentric approach requires developing classification systems grounded in African epistemologies, including relational metadata that represents knowledge as connected to community, place, and practice (Ihejirika, 2024). IN light of this, Afrocentric framework for decolonising algorithms must address metadata epistemologies dimension as follows:

a. Institutional Responsibility: Library cataloguing departments, in partnership with community knowledge holders and local language experts.

b. Concrete Actions:

- Establish a Metadata Co-Design Working Group comprising cataloguers, African language scholars, and representatives of target communities.
- Develop a relational metadata schema with fields for: knowledge-originating community, mode of transmission (oral/ceremonial/practical), access permission level (public/community-restricted/ceremonial), geographic association, and practice context.
- Conduct an epistemic audit of existing collections to identify indigenous knowledge materials that are misclassified or undiscoverable under current Dewey Decimal Classification and Library of Congress Subject Headings.

c. Evaluation Indicators:

- Indigenous knowledge material discoverability rate (target: increase from baseline to ≥85%).
- Community validation score for metadata descriptions (assessed by community review panel).
- Reduction in misclassification incidents.

Dimension Two: Training Datasets

The algorithmic prioritisation of Western scholarship reflects the composition of training data. With less than 1% of training data for widely used AI systems reflecting African linguistic contexts, correcting this imbalance is both a technical and political task requiring NLP tools for African languages, annotated datasets reflecting African knowledge domains, and infrastructure for local model training (Harun, 2026). On the proposed Afrocentric framework, the dimension of training datasets will be addressed as follows:

a. Institutional Responsibility: Library IT departments, in collaboration with regional library consortia and NLP research partners.

b. Concrete Actions:

- Establish African Language Corpus Contribution Agreements specifying data ownership, usage scope, and benefit-sharing arrangements.
- Prioritise digitization of community-authorized oral traditions, indigenous language publications, and indigenous knowledge documentation.
- Partner with African NLP communities (e.g., Masakhane) to participate in fine-tuning and evaluation of African language models.

c. Evaluation Indicators:

- Proportion of local-language training data in total system training data (target: measurable increase from <1% baseline).
- Reduction in recall and precision gap between African-language and English-language retrieval.
- Sustained participation rate of data-contributing communities.

Dimension Three: Community Collaboration

The marginalisation of indigenous knowledge in AI systems is fundamentally a problem of exclusion (Kuranga, Khumalo & Masenya, 2026). An Afrocentric framework must shift from "digitization for preservation" to "digitization for sovereignty," where communities determine what is digitized, how it is represented, and who can access it as follows:

a. **Institutional Responsibility:** Library management, community liaison officers, and institutional ethics committees.

b. **Concrete Actions:**

- Adopt the CARE Principles (Collective Benefit, Authority to Control, Responsibility, Ethics) as the governance framework for all digitization projects.
- Develop Community Data Governance Agreement templates specifying: who decides what is digitized, who controls access, and who benefits from derivative value.
- Establish Community Advisory Committees with veto or co-decision power over AI system functions that represent community knowledge.

c. **Evaluation Indicators:**

- Proportion of community partnership projects with formal governance agreements.
- Community satisfaction score regarding "degree of control."
- Proportion of community-initiated digitization requests that receive institutional response.

Dimension Four: Ethical Safeguards

INHUMAIN.AI (2026) came up with ethical frameworks that remain at the level of principle without enforceable mechanisms risk becoming "ethics washing." An Afrocentric framework must embed safeguards that are institutionalized, enforceable, and accountable to affected communities, including algorithmic audits, impact assessments, data sovereignty agreements, and grievance mechanisms as follows:

a. **Institutional Responsibility:** Institutional AI Governance Committee, library ethics officer, and external auditors.

b. **Concrete Actions:**

- Embed epistemic impact assessment clauses in AI procurement contracts, requiring vendor transparency and explainability regarding algorithmic decision-making.
- Establish algorithmic audit mechanisms examining not only technical bias (gender, race) but also epistemic exclusion (which knowledge forms are systematically down-ranked or made undiscoverable).
- Create grievance mechanisms through which communities and students can report cultural insensitivity or misrepresentation by AI systems.

c. **Evaluation Indicators:**

- Number of AI systems that have completed epistemic impact assessment.
- Grievance response time and resolution rate.
- Proportion of procurement contracts containing data sovereignty and transparency clauses.

Institutional Responsibility And Evaluation Matrix

Table C maps the four dimensions to specific responsibility holders, timeframes, and core evaluation indicators:

Dimension	Primary Responsibility	Partners	Short-term (0–12 months)	Medium-term (1–3 years)	Core Indicators
Metadata epistemologies	Cataloguing department	Community reps, language experts	Epistemic audit; establish co-design group	Deploy relational metadata schema	IK discoverability rate; community validation score
Training datasets	IT department / consortium	NLP partners, communities	Corpus agreements; digitization priorities	Local model fine-tuning and evaluation	Local data proportion; retrieval quality gap
Community collaboration	Management / liaison officers	Community advisory committees	CARE adoption; agreement templates	Institutionalize community co-governance	Agreement coverage; community satisfaction
Ethical safeguards	AI governance committee	Ethics officer, external auditors	Contract clause revision	Routinize algorithmic audits	Audit completion rate; grievance resolution rate

Cross-Institutional Applicability

To address the concern that the framework may be perceived as context-specific, a three-tier applicability model is proposed:

Tier A: Resource-Intensive Institutions

- Possess or can develop in-house NLP capacity.
- Can undertake comprehensive metadata restructuring and local model training.
- Community collaboration can reach co-governance level.

Tier B: Moderately Resourced Institutions

- Rely on regional consortia for shared infrastructure.
- Prioritise metadata adaptation and procurement clause revision.
- Community collaboration based on consultation and formal agreements.

Tier C: Resource-Constrained Institutions

- Focus on transparency and data sovereignty clauses in procurement contracts.

- Participate in regional corpus contribution rather than independent training.
- Ethical safeguards rely on consortium-level audit mechanisms.

This tiered model responds to the "cross-institutional applicability" concern while acknowledging the resource inequalities among African tertiary institutions.

Operationalizing The Afrocentric Framework

The four dimensions outlined above require translation into actionable institutional practice. Table C presents the operational framework, specifying for each dimension the primary responsibility holder, concrete short-term and medium-term actions, and evaluation indicators. The framework is designed to be adaptable across institutional contexts through a three-tier applicability model (Tier A: resource-intensive; Tier B: moderately resourced; Tier C: resource-constrained), The model recognises that African tertiary libraries operate under heterogeneous resource conditions. Implementation should proceed iteratively, with progress in one dimension enabling and reinforcing progress in others. Institutional commitment, adequate resourcing, and genuine community partnership are prerequisites for successful operationalization.

DISCUSSION

This Discussion distinguishes between three types of claims (findings, interpretations and recommendations). Findings are claims directly supported by the documentary evidence analysed in this study. Interpretations are inferences drawn from findings, connecting them to theoretical frameworks and broader patterns. Recommendations are normative proposals advanced by the authors as responses to the findings. While recommendations are grounded in the evidence and theory presented, they involve value judgements and contextual assessments that go beyond what the documents alone can establish. This distinction is maintained throughout the Discussion to enable readers to assess the evidential basis for each claim.

Table D is a summary of the discussion of findings we did in this study:

Dimension	**Key Documentary Finding**	**Conceptual Interpretation**	**Normative Recommendation**
Metadata epistemologies	Existing classification systems lack categories for oral and community held knowledge.	This reflects the coloniality of knowledge embedded in library standards.	Develop relational metadata schemas grounded in African epistemologies.
Training datasets	Less than 1% of AI training data reflects African linguistic contexts; African authored research receives fewer algorithmic recommendations.	The imbalance is both technical (absence of data) and political (concentration of AI development).	Build local NLP capacity, create annotated datasets, and establish data sovereignty agreements.
Community collaboration	Communities are excluded from decisions about digitization and representation of their knowledge.	Exclusion is structural, reflecting a mismatch between community knowledge forms and AI system design.	Shift from "digitization for preservation" to "digitization for sovereignty" using CARE Principles.
Ethical safeguards	Ubuntu ethics is widely proposed but rarely operationalized; no African country has a dedicated indigenous knowledge AI	The gap between principle and implementation reflects structural power asymmetries in AI	Embed enforceable mechanisms: audits, impact assessments, data sovereignty agreements, grievance

	policy.	development.	procedures.
Librarian agency	Librarians demonstrate critical awareness but lack technical and institutional capacity to intervene.	Librarians are positioned at a critical intersection but constrained by vendor control and institutional inertia.	Include librarians in AI procurement; provide training; create institutional mechanisms for voice.

Metadata Epistemologies: Findings, Interpretation, And Recommendations

Finding: Documentary analysis revealed that existing metadata standards lack adequate categories for oral traditions, community-held knowledge, and non-textual knowledge forms. This limitation was documented across multiple institutional contexts, including Zimbabwean university libraries and the Makerere University library case (Harris, 2025).

Interpretation: Read through decolonial epistemology, this limitation is not merely technical but epistemological. Classification systems such as Dewey Decimal Classification and Library of Congress Subject Headings embody particular worldviews that assume text-based, individually authored, disciplinary knowledge (Quijano, 2007; Mignolo, 2011). The absence of categories for indigenous knowledge forms reflects the coloniality of knowledge rather than a neutral gap in the schema.

Recommendation: We propose that libraries develop classification systems grounded in African epistemologies, including relational metadata that represents knowledge as connected to community, place, and practice. This recommendation goes beyond the documentary evidence, which establishes the inadequacy of current systems but does not prescribe a specific alternative. The University of Pretoria's pragmatic adaptations, color-coded classification and genre-specific tags offer one possible model, though these remain adaptations within existing frameworks rather than fundamental re-conceptualisations (Warricker et al., 2024).

Training Datasets: Evidence Of Imbalance And Proposals For Correction

Finding: Documentary analysis confirmed that AI models in use at participating institutions are trained predominantly on Western datasets. The literature documents that less than 1% of training data for widely used AI systems reflects African linguistic contexts (Harun, 2026). This imbalance was identified as a contributing factor to the algorithmic prioritisation of Western scholarship documented in the findings.

Interpretation: The disparity in training data composition is both a technical and political phenomenon. Technically, it reflects the absence of large scale annotated datasets for African languages. Politically, it reflects the concentration of AI development resources and decision making power in the Global North (Birhane, 2020). The finding that African authored research receives one third fewer algorithmic recommendations than comparable Global North scholarship is consistent with this structural imbalance.

Recommendation: We recommend that African tertiary libraries pursue three strategies: (a) developing NLP tools for African languages in partnership with initiatives such as Masakhane; (b) creating annotated datasets reflecting African knowledge domains; and (c) building infrastructure for local model training. These recommendations are advanced as a response to the documented findings, but their feasibility depends on resource availability and institutional commitment, factors that vary significantly across the institutional contexts examined.

Community Collaboration: Documented Exclusion And Normative Proposals

Finding: Documentary analysis revealed that communities are systematically excluded from decisions about how their knowledge is digitized, organised, and made accessible. This exclusion was documented across institutional policies and practices, with collection development policies in Ghanaian public libraries, for example, failing to adequately encompass indigenous knowledge documentation (Dlamini, Jiyane &

Mohammed, 2026). The Transformedia project in Sudan offers a counter example, with 18 community ambassadors driving locally led documentation (British Council, 2026).

Interpretation: The exclusion of communities is not incidental but structural. It reflects what Kuranga, Khumalo and Masenya (2026) describe as a fundamental mismatch between community held knowledge, which is relational, contextual, and embedded in practice, and AI systems designed to process commodified, decontextualized information objects. The Transformedia model suggests that this exclusion can be addressed, but the conditions for success, including institutional commitment, resource allocation, and community trust, are demanding.

Recommendation: We propose that African tertiary libraries shift from "digitization for preservation" to "digitization for sovereignty," adopting the CARE Principles as a governance framework. This recommendation is normative rather than empirical: the documentary evidence establishes that exclusion occurs and that alternatives exist, but it does not establish that the CARE Principles are the only or best framework for all contexts. The framework's applicability across diverse institutional contexts requires further empirical testing.

Ethical Safeguards: The Implementation Gap And Proposed Mechanisms

Finding: Documentary analysis confirmed a significant gap between ethical principle and enforceable mechanism. The systematic review by Goshu & Ridwan (2026) found that while Ubuntu is "widely proposed as a relational AI ethics alternative," its "operationalization is minimal," with only two prototypes documented and no African country having a dedicated indigenous knowledge AI policy. This finding was corroborated by the present study's analysis of institutional policies, which found that where AI governance policies existed, they focused primarily on data privacy and academic integrity rather than epistemic justice.

Interpretation: The gap between principle and implementation is not accidental. It reflects the structural position of African institutions within global AI development, where decision making power over algorithmic systems is concentrated in the Global North (Birhane, 2020). The finding that library professionals possess critical awareness but lack technical and institutional capacity (see also Tsekea & Mandoga, 2025) is consistent with this structural analysis.

Recommendation: We recommend that an Afrocentric framework embed safeguards that are institutionalized, enforceable, and accountable to affected communities. Specific mechanisms proposed include: algorithmic audits examining epistemic exclusion; impact assessments evaluating cognitive justice; data sovereignty agreements with enforceable rights; and grievance mechanisms. These recommendations are advanced as a response to the documented gap, but their implementation requires policy change, procurement reform, and professional development, all of which depend on institutional will and resource allocation.

Librarians As Ethical Agents: Findings On Capacity And Recommendations For Empowerment

Finding: Documentary analysis revealed a consistent pattern: library professionals across participating institutions demonstrated sophisticated critical awareness of algorithmic bias and epistemic exclusion, yet reported feeling powerless to address these issues due to lack of technical capacity, vendor control, and institutional support. This finding was documented in institutional reports and professional literature (Murwira & Higgs, 2023; Tsekea & Mandoga, 2025).

Interpretation: The librarian's position at the intersection of communities, collections, and technologies creates both vulnerability and potential. The finding that librarians possess critical awareness but lack capacity identifies a specific site of intervention: building professional capacity and institutional mechanisms for librarian agency. This interpretation aligns with the Afrocentric principle that knowledge organisation should centre African perspectives and community needs (Asante, 1980).

Recommendation: We recommend that library professionals be empowered to act as "co-creators of AI governance" rather than passive recipients of externally determined change. This requires including librarians

in AI procurement decisions, providing training in AI ethics and critical algorithm studies, and creating institutional mechanisms for voicing concerns about algorithmic harm. This recommendation is normative: the documentary evidence establishes the capacity gap, but the specific mechanisms proposed, including procurement inclusion, training, and grievance mechanisms, represent the authors' judgement about appropriate responses rather than findings directly derived from the evidence.

CONCLUSION

Decolonising AI in African tertiary libraries is both a technical and ethical imperative. By adopting an Afrocentric framework, libraries can transform repositories into inclusive spaces that honor indigenous knowledge while engaging global scholarship. This approach advances epistemic justice and strengthens Africa's intellectual sovereignty.

The framework proposed here operates across four dimensions: conceptualizing metadata to reflect African epistemologies; training AI models on African datasets with attention to linguistic sovereignty; collaborating with communities as knowledge stewards rather than data sources; and embedding enforceable ethical safeguards against epistemic exploitation. These dimensions are interconnected and mutually reinforcing; progress in one enables progress in others.

The research has identified significant obstacles. The political economy of AI development concentrates resources and decision-making power in the Global North, creating path dependencies that privilege Western frameworks even when actors are committed to alternatives. The gap between Ubuntu ethics discourse and deployed systems reflects this structural challenge: it is easier to articulate Afrocentric principles than to implement them within technological and institutional infrastructures designed for other purposes.

Yet the research has also identified reasons for hope. Emerging initiatives from the Masakhane Hub's linguistic sovereignty work to Transformedia's community-led digitization to the University of Pretoria's pragmatic adaptations demonstrate that Afrocentric approaches are feasible. The critical awareness among library professionals, combined with growing regional and continental policy attention to AI governance, creates openings for meaningful change.

Future research should prioritise longitudinal studies that assess the actual impact of Afrocentric AI interventions; participatory action research that engages communities as co-investigators; and comparative analysis of different governance models for indigenous knowledge in digital environments.

Publication of negative results about what did not work and why is essential for building an evidence base that can guide practice.

The ultimate measure of success for decolonising algorithms is not technical sophistication or publication output. It is whether African communities have greater control over how their knowledge is represented in digital systems; whether African scholarship is discoverable on equitable terms; whether library users can find the knowledge they need in the forms and languages they use. These outcomes require sustained commitment, adequate resources, and genuine collaboration among institutions, communities, and professionals. The framework proposed here offers a path, but the journey requires collective will.

REFERENCES

1. Asante, M. K. (1980). Afrocentricity: The theory of social change. Amulefi Publishing Company.
2. Balyejjusa, S., Bhangyi, V. B., Hope, A., & Silverno, T. (2023). Transforming communities through Ubuntu philosophy: Prerequisites, opportunities, and implications for social work practice in Africa. https://www.researchgate.net/publication/369113241
3. Benjamin, R. (2019). Race after technology: Abolitionist tools for the new Jim code. Polity.
4. Birhane, A. (2020). Algorithmic colonization of Africa. SCRIPTed, 17(2), 389–409.
5. Bowen, G. A. (2009). Document analysis as a qualitative research method. Qualitative Research Journal, 9(2), 27–40.

6. British Council. (2026). Cultural Protection Fund annual report 2024–25. https://cultural-protection-fund.britishcouncil.org/sites/default/files/cultural_protection_fund_annual_report_2024-2025.pdf
7. Buolamwini, J., & Gebru, T. (2018). Gender shades: Intersectional accuracy disparities in commercial gender classification. Proceedings of Machine Learning Research, 81, 1–15.
8. Carroll, S. R., Garba, I., Figueroa-Rodríguez, O. L., Holbrook, J., Lovett, R., Materechera, S., & Hudson, M. (2020). The CARE Principles for Indigenous Data Governance. Data Science Journal, 19(1), 43. https://doi.org/10.5334/dsj-2020-043
9. Carroll, S. R., Rodriguez-Lonebear, D., & Martinez, A. (2019). Indigenous data governance: Strategies from United States Native Nations. Data Science Journal, 18(31), 1–15. https://doi.org/10.5334/dsj-2019-031
10. Dlamini, P. N., Jiyane, G. V., & Mohammed, H. A. R. (2026). Documentation and integration of indigenous knowledge in public libraries for inclusive information services in Ghana. South African Journal of Libraries and Information Science, 92(2), 1–13. https://doi.org/10.7553/92-2-2346
11. Figueroa-Rodriguez, O. L., et al. (2019). Indigenous data sovereignty: The state of open data. In Histories and horizons (pp. 300–319). African Minds and the International Development Research Centre (IDRC).
12. Goshu, B. S., & Ridwan, M. (2026). Ubuntu, algorithms, and ancestral epistemologies: A systematic review of AI–indigenous knowledge integration in African societies. Budapest International Research and Critics in Linguistics and Education (BirLE) Journal, 9(2), 81–100.
13. Grosfoguel, R. (2017). Decolonising Western uni-versalisms: Decolonial pluri-versalism from Aimé Césaire to the Zapatistas. In The Routledge companion to the philosophy of race (pp. 88–103). Routledge.
14. Gupta, K. L., Taak, S., & Singh, J. S. (2026). Indian Knowledge System. https://ir.vistas.ac.in/id/eprint/20340/1/12-May-IKS%20_Indian%20Knowlwdge%20System-smallpdf_smallpdf.pdf
15. Harris, A. (2025). African literary metadata and Makerere University's library. Research in African Literatures, 55(1), 1–27.
16. Harun, I. (2026). African languages in the digital AI era: Reimagining communication, access, policy, and participation. Critical Studies in Teaching and Learning, 14(1), 213–234. https://doi.org/10.17159/cristal.v14i1.27670
17. Hiribarren, V., & Madore, F. (2025, June 26). Historian vs AI: Who reads and analyses archives best? [Conference presentation]. 10th European Conference of African Studies (ECAS), Prague, Czech Republic.
18. Ihejirika, C. (2024). Harnessing African Indigenous Knowledge Systems for knowledge production: A redefinition of a culture-centric epistemology. Journal of Contemporary Philosophical and Anthropological Studies, 2(1). https://doi.org/10.59652/jcpas.v2i1.103
19. INHUMAIN.AI. (2026). AI ethics frameworks: A complete guide to who sets the rules. https://inhumain.ai/ai-ethics-framework-guide/
20. Kuranga, C., Khumalo, A., & Masenya, T. M. (2026). Developing culturally responsive AI for indigenous knowledge preservation and revitalization. Mousaion: South African Journal of Information Studies, 44(1), 14p. https://doi.org/10.25159/2663-659X/20992
21. Mabofua, F. C. (2020). The challenges of decolonising university curricula in South Africa: A critical perspective [Master's dissertation]. University of Johannesburg. https://ujcontent.uj.ac.za/esploro/outputs/graduate/The-challenges-of-decolonising-university-curricula/9911645507691
22. Mbembe, A. (2016). Decolonising the university: New directions. Arts & Humanities in Higher Education, 15(1), 29–45.
23. Mhlambi, S. (2020). African values for ethical AI. Harvard University Berkman Klein Centre.
24. Mignolo, W. D. (2011). The darker side of Western modernity: Global futures, decolonial options. Duke University Press.
25. Mkhize, T. (2026). The Africanisation of universities as the dismantling of silos towards global knowledges. Journal of Educational Studies. https://doi.org/10.59915/jes.2025.si1.6

26. Moahi, K. A. (2012). Promoting African indigenous knowledge in the knowledge economy: Exploring the role of higher education and libraries. Aslib Proceedings, 64(5), 540–554. https://doi.org/10.1108/00012531211263157
27. Murwira, S., & Higgs, P. (2023). Afrocentric epistemic systems and higher education curriculum: The case of a university in Zimbabwe. https://hdl.handle.net/10520/ejc-high_v37_n2_a17
28. Ndege, N., Irfanullah, H. M., Harle, J., & Drake, T. (2025). Africa's research publishing landscape: Examining journals, publishers, and the infrastructure behind them. Open Research Africa, 8, 20. https://doi.org/10.12688/openresafrica.16040.1
29. Ngubane, S. A., & Omodan, B. I. (2026). The trend in studies on decoloniality, AI, and university classrooms: A systematic review of African higher education scholarship. Trends in Higher Education, 5(3), 92.
30. Ngugi wa Thiong'o. (1986). Decolonising the mind: The politics of language in African literature. Heinemann.
31. Noble, S. U. (2018). Algorithms of oppression: How search engines reinforce racism. NYU Press.
32. Nyambi, E., & Maynard, S. (2012). An investigation of institutional repositories in state universities in Zimbabwe. Information Development, 28(1), 55–65. https://doi.org/10.1177/0266666911425264
33. Ocholla, D. N. (2007). Marginalised knowledge: An agenda for indigenous knowledge development and integration with other forms of knowledge. International Review of Information Ethics, 7(9), 1–10.
34. Quijano, A. (2007). Coloniality and modernity/rationality. Cultural Studies, 21(2–3), 168–178.
35. Santos, B. S. (2014). Epistemologies of the South: Justice against epistemicide. Paradigm Publishers.
36. Shizha, E., & Kariwo, M. T. (2012). Education and development in Zimbabwe: A social, political and economic analysis. Springer.
37. Springer Nature. (2025). Abundant intelligences: Placing AI within Indigenous knowledge frameworks. AI & Society, 40, 2141–2157.
38. Tajjiou, A. (2024). The 'Comparative African Literatures' project: Challenges and opportunities. Comparative Literature: East & West, 8(1), 89–107. https://doi.org/10.1080/25723618.2024.2368350
39. Tsekea, S., & Mandoga, E. (2025). The ethics of artificial intelligence use in university libraries in Zimbabwe. Frontiers in Research Metrics and Analytics, 9, 1522423. https://doi.org/10.3389/frma.2024.1522423
40. Warricker, K., Kekana, M., Nel, J., & Sithole, T. (2024). Collaborate and uncomplicate: Best practices for managing and maintaining African languages collections. LIASA Conference Presentation.