Eastern African Journal of Humanities and Social Sciences
Vol. 5 No. 1 (2026): ISSN (Online): 2958-4558
DOI: https://doi.org/10.58721/eajhss.v5i1.1674
The African Politics of the Dataset: AI, Indigenous Music, and Neo-Colonial Power Dynamics

The African Politics of the Dataset: AI, Indigenous Music, and Neo-Colonial Power Dynamics

Amon Kipyegon Kirui[1] & Brian Bichanga Nyandieka[2]
[1]*University of Venda, South Africa*
[2]*Kabarak University, Kenya*

Article History
Received: 24.09.2025
Revised: 02.04.2026
Accepted: 05.04.2026
Published: 25.04.2026

Keywords
Afrobeats
Amapiano
Colonialism
Data Sovereignty
Netnography

How to cite:
Kirui, A. K., & Nyandieka, B. B. (2026). The African Politics of the Dataset: AI, Indigenous Music, and Neo-Colonial Power Dynamics. *Eastern African Journal of Humanities and Social Sciences, 5*(1), 239-251.

Abstract
The proliferation of Artificial Intelligence (AI) into the creative industries, especially in the context of Indigenous African music, is a deep-rooted political action that builds upon the historical dynamics of power of digital extraction and algorithmic colonialism. While existing scholarship predominantly critiques AI through Western intellectual property paradigms, it largely overlooks the training dataset as a primary site of epistemic violence. Based on a decolonial theoretical framework, this paper fills this gap by arguing that the AI training datasets are political artefacts that fix neo-colonial forms of oppression by freezing dynamic, living cultures to be consumed exogenously. The study utilises a triangulated research design and intentionally uses a multi-regional Pan-African scope of analysis, triangulating three analytical fields: a historical excavation of the 1939 *Mbube* (Solomon Linda) case in South Africa as an analogue blueprint of extraction; a netnographic study of five highly active industries. The findings reveal a 'dual architecture' of digital neo-colonialism, wherein external data colonialism via uncompensated scraping is facilitated by internal structural deficits and outdated post-colonial legislation. Empirically, the netnographic data reveals a phenomenon of 'silent extraction,' where grassroots resistance is actively overridden by a consumer-driven adapt or die techno-optimism. To address this perceived epistemic alienation and legal gap, the paper concludes that the way to realize ethical AI is by shifting the paradigm to Indigenous Data Sovereignty (IDSov) and the Collective Benefit, Authority to Control, Responsibility and Ethics (CARE Principles). In direct relation to the structural gaps that were revealed through the research, the paper proposes that the African Musicology Institute (AMI) should be created as a free, sovereign infrastructural centre to manage ethically sourced datasets. Finally, this study alters the academic debate of reactive copyright litigation to proactive infrastructural autonomy, which offers a critical map of computational self-determination in the Global South.

Introduction
The emergence of generative Artificial Intelligence (AI) has been a double-edged sword amid an exciting cultural milieu. Although it presents new opportunities in creating algorithms and music production unknown before (Nkasi, 2025; Olajide-Philips & Ogunbowale, 2025), it also carries grave

existential dangers related to intellectual property (IP) theft, systemic data bias, and widespread, automated culture appropriation (Dugeri, 2024; Rivera, 2025). Traditionally, music, aesthetics, and cultural innovations of black and indigenous people have been systematically stolen and deprived of the right of ownership by commercial organisations of the West; therefore, the fact that contemporary AI is largely unregulated and does not have any restrictions poses a threat of multiplying these past injustices on an industrial scale. But to place the emergence of generative AI simply as a modern threat to the individual intellectual property is to fundamentally misdiagnose the severity of the trauma; AI, in its existing structural form, is a highly sophisticated system of epistemic control, a digital machine that does not simply steal cultural material, but indeed rearranges the very terms by which African knowledge is produced, confirmed, coded in computational terms, and disseminated across the planet. While existing literature has adequately critiqued the economic and legal ramifications of AI copyright infringement, there remains a critical gap regarding the empirical analysis of the AI training dataset itself as a primary site of epistemic violence. Specifically, current literature largely fails to address how this external data extraction is actively facilitated by internal, post-colonial structural deficits and youth techno-optimism.

This paper advances a stringent decolonial perspective, arguing that absent a sovereign, robustly enforced regulatory framework, the deployment of AI will further entrench insidious modes of neo-colonial exploitation that relegate African creators to mere data-providers (Ilo, 2025; Kwet, 2019). To demonstrate this, we argue that the training dataset must not be viewed as an innocent, politically neutral container of information. Rather, it functions as a highly contested site of ontological decision-making, determining computationally what counts as legitimate music, who is legally recognised as a creator, and which specific forms of human knowing are granted computational legibility by Western developers. While AI systems continue to proliferate uncritically across the continent, they risk significantly deepening the colonial legacy of cultural extraction and psychological control (Birhane, 2020; Couldry & Mejias, 2019)—a systemic problem rooted not in the inherent nature of the technology itself, but deeply embedded within the post-colonial educational and epistemological systems that have consistently devalued Indigenous knowledge in favour of Western technological universalism (Gyamerah, 2024; Kirui et al., 2025; le Grange, 2024; Seow, 2024). Therefore, this study aims to rigorously examine digital neo-colonialism across the entire AI lifecycle, investigating training datasets as biased political artefacts through historical and contemporary case studies, evaluating the specific role of epistemic alienation in perpetuating these exploitative systems, and ultimately consolidating actionable strategies for achieving genuine technological and cultural autonomy.

Literature Review

The academic discourse of the growth of artificial intelligence in Africa tends to increasingly construct this technological expansion as a new form of digital neo-colonialism: a highly unequal process in which technology corporations of the Global North, as well as the developing Chinese technological giants, continue to systematically extract the data, cognitive labour, and digital resources of the continent without necessarily delivering equitable economic benefits or intellectual credits (Coleman, 2019; Kwet, 2019). The key point in this critical framework is the idea of data colonialism, a process that considers the unremitting exploitation of human culture and lived experience to transform it into measurable data that can be used to generate foreign corporate wealth (Couldry & Mejias, 2019). This is a direct reflection of the old colonial scripture of taking the raw physical resources, cocoa, oil or gold, to be processed in the metropole at a very profitable price before being sold back to the African continent at a prohibitive price (Rodney, 1972). The resulting modern dynamic recreates structural power disparities strikingly reminiscent of historical colonial relations, thereby establishing Africa's dependent, subservient role in the technological world (Menon, 2023; Salami, 2024; Warganegara, 2024). Birhane (2020) calls this particular phenomenon "Algorithmic Colonisation," which is a Western

algorithmic colonisation in which the African markets fail to develop alternative software systems, which are underfunded locally, and are forced to completely depend on foreign-designed software systems that have been trained on Global North knowledge regimes almost exclusively (le Grange, 2024); furthermore, this external extraction is severely compounded by internal institutional factors, including inadequate regulatory policies, weak local technological capacity, and chronic infrastructure deficits that make it financially ruinous for African innovators to build their own sovereign AI systems, thereby cementing a vicious cycle of infrastructural dependency.

Authenticity in the African music traditions is critically deconstructed throughout the post-colonial discourse as the most controversial colonial construct that is still ingrained in modern educational and cultural management systems (Le Grange, 2024). The perpetuation of the perceived static, untouched, and completely authentic Indigenous culture is mostly an illusion codified in national curricula with the help of a lingering colonial gaze, a pattern that habitually views dynamic, living, African traditions as historical artefacts to be examined instead of as practices in the process of continuous change (Ibekwe, 2018; le Grange, 2024). This history of defining and policing Indigenous identity, which has historically been enforced in other world systems through oppressive administrative policies such as blood quantum standards (McKay, 2019), has established a profound psychological subordination, and authenticity is a key point of continued political conflict with the computerisation of reductionism. Algorithms reduce cause extreme epistemic violence by trying to reduce fluid and context-dependent cultural authenticity into a fixed, classifiable grid of mathematics (APRA AMCOS, 2024) and which is the intensification of what scholars term postdigital epistemic violence, a process that systematically disempowers the testimony and lived reality of marginalised knowers and artificial inflates the credibility of dominant, machine-generated narratives It is not just the legal identification of a musical piece that is at stake, but the ontological reality of the community; Indigenous African music is not a text to be read by a machine, but an experience to be experienced; a deep meeting of human breath, ancestral memory, temporal consciousness, and communal embodiment that is not yet fated to become the target of datafication, that is, the soul which creators fear losing to AI is an epistemological category that Western computational logic currently possesses no architecture to respectfully encode.

The training datasets of AI, in turn, should not be seen as objective reflections of the world, but rather as an inherently biased political artefact that is predetermined by the presence of long-standing colonial structures that reproduce an educational desert, which then and accordingly generates a data desert of authentic African cultural expression (Birhane, 2020; Ssekibuule et al., 2024). Since the educational systems of post-colonial nations often institutionalise Indigenous traditions as ancient myths rather than prioritising them as academic disciplines (Ibekwe, 2018; Le Grange, 2024), state-funded, sovereign records and digital preservation of these cultural resources are critically lacking. This is a particularly vulnerable situation that is exacerbated by the fact that Indigenous languages are rapidly becoming extinct and that there are completely inadequate Intellectual Property (IP) systems that are at all times unable to safeguard the commercialisation of Indigenous content by foreign companies without any prior consultation or fair sharing of benefits. What Noble (2018) refers to as "Algorithms of Oppression," in which digital discrimination is carried out by modelling data biased against members of a specific group, functions as a contemporary form of the old-time colonial practices of controlling and excluding resources. The dataset is a form of ontological reduction which reduces living, breathing cultures to the lifeless state of static numbers, devoid of the vital context, time, spirit or relational meaning of their performance, when an autonomous algorithm is fed a complex Amapiano log drum pattern or a Zulu choral harmony it captures neither the communal ceremony in which it naturally arose nor the narrow-band set of cosmic grammar by which it is strictly performed.

In response to these systemic extractions, the global Indigenous Data Sovereignty (IDSov) movement has emerged as a powerful decolonial counter-movement rooted in the enduring aspiration of Indigenous Peoples to reclaim self-determination over their cultural outputs and combat algorithmic racism (Oguamanam, 2020; Rainie et al., 2019). IDSov explicitly asserts the inherent right of Indigenous populations to govern the collection, ownership, and application of data concerning their lived realities and knowledge systems, actively reframing them as sovereign rights-holders rather than passive subjects of technological observation. Because prevailing Western data approaches, such as the widely adopted FAIR Principles (Findable, Accessible, Interoperable, Reusable), primarily facilitate the frictionless sharing of data while deliberately ignoring historical power differentials and structural inequalities, the people- and purpose-oriented CARE Principles (Collective Benefit, Authority to Control, Responsibility, Ethics) were specifically developed to center self-determination and ethical accountability (Carroll et al., 2020; Research Data Alliance, 2019). True decolonial resistance, however, is fundamentally rooted in the urgent imperative to reform educational systems and directly address the coloniality of knowledge that makes data colonialism possible in the first place (Ndlovu-Gatsheni, 2013, 2015). A core decolonial strategy involves the radical transformation of university curricula to challenge entrenched Eurocentric epistemologies (Mapaya, 2014; Xulu & Yende, 2024; Yende & Yende, 2022), championing an activist, African-led methodological approach where future AI systems are strictly co-designed using African ethical frameworks and trained exclusively on sovereign local content to ensure epistemic justice (Kirui et al., 2025) , a process that aligns with the broader decolonial tactic of 'reverse tutelage,' where vital learning from the global peripheries actively informs and restructures central technological design (Akena, 2012; Chibike, 2024; Cruz, 2021; Mohamed et al., 2020).

Methodology

To rigorously examine the various, multi-layered dynamics of digital neo-colonialism and algorithmic extraction in the context of African indigenous music, a qualitative, triangulated research design was selected over a traditional multiple-case study, with the express purpose of enabling longitudinal and comparative analysis. This methodological approach enabled the study to carefully trace the linear, uninterrupted development of cultural extraction from the analogue colonial period to the modern digital frontier and, as such, provided a level of historical depth usually lacking in strictly technical accounts of artificial intelligence. Steering clear of conventional, researcher-dominated primary data gathering strategies, including structured human-subject interviewing, the research employed a non-obtrusive, digitally based, and structurally oriented methodology that triangulates three major analytical areas: historical document analysis, real-time Netnography of social media discourse (Kozinets, 2020), and doctrinal legal analysis of regional intellectual property laws. Such a triple, triangulated method is theoretically justified by the fact that the phenomenon under study is inherently digital; the most visceral, consequential discussions about algorithmic appropriation and the theft of culture are currently taking place on public social media platforms in real-time, whilst the structuring and legislative framework that legalizes the digital extraction offers the contextual structure and the legislative failures that legally underpin the grassroots industry reality and the macro-level policy failures that legally enable digital extraction.

Based on a purposive sampling approach, the research developed three distinct, highly representative analytical domain cases that effectively trace the historical course of neo-colonial extraction and contemporary resistance. Analytical Domain 1 is the historical map of analogue extraction, exemplified by Solomon Linda's 1939 Mbube, which serves as the baseline for the historical engineered failure of Western Intellectual Property systems to safeguard African creators against international corporations. Analytical Domain 2 looks at the current danger of algorithmic extraction by working with the current netnographic flashpoints of generative AI and voice cloning as applied

to the Afrobeats and Amapiano ecosystems in particular, because these two genres are the most influential genres globally at this moment and thus the most profitable, unregulated victims of unchecked algorithmic scraping by foreign developers. Analytical Domain 3 focuses on institutional and decolonial resistance by examining existing macro-political responses, legal frameworks, and the institutional drive towards Indigenous Data Sovereignty by regional copyright agencies. In the case of the netnographic components that involve the second and third domains, the targeted objects of study were narrowly limited to confirmed African recording artists, music producers, cultural policymakers, and recognised Collective Management Organisations (CMOs); centring these specific stakeholders represents a deliberate decolonial methodological choice designed to ensure that the ultimate epistemological authority regarding the destructive impact of artificial intelligence remains firmly with the creators and designated custodians of the culture itself.

The data harvesting process was carried out carefully across three key domains between January 2023 and March 2026 to achieve robust triangulation and to reflect the exponential growth of available generative AI tools. The X platform (previously Twitter) was chosen as the only location where netnographic observation can be conducted due to the fact that it is the current digital incarnation of the town square of the global music industry, and as the most important venue where African stakeholders can publicly discuss technological changes and complain about industry grievances in real-time. Since the first Boolean search using formal academic language (e.g., searching: AI AND Afrobeats AND appropriation) gave entirely empty results because platform algorithms obscure the formal discourse, an iterative search strategy was employed using informal industry terminology (e.g., searching: stealing beats, fake artists, or AI noise), which has proven useful in finding highly relevant and localised flashpoints in the industry. As a measure to operationalise the doctrinal legal analysis, primary statutory texts have been identified and analysed, namely the Copyright Act 98 of 1978 of South Africa and the African Union's Continental AI Strategy, which provide the stringent legal reference points required to assess structural regulatory gaps.

The analysis of the collected data was carried out according to the Reflexive Thematic Analysis (Braun & Clarke, 2021) with the purposeful end at the achievement of theoretical saturation at five of the most active flashpoints of the 'anchors' as the netnographic sampling. To ensure strict empirical rigour, this netnographic observation yielded a meticulously analysed corpus of 482 distinct user posts and complex reply threads. Theoretical saturation was officially deemed achieved at this exact corpus size because newly harvested social media replies ceased to yield any novel theoretical insights into algorithmic appropriation; instead, the discourse began to cycle repetitively through the already established themes of consumer techno-optimism and grassroots producer alienation. Stern ethical standards on digital research were followed; information found on verifiable public personalities performing in professional roles was quoted verbatim, whereas data mined on the responses of personal citizens or unverified fans was anonymised to ensure the protection of the privacy of the individuals and to reflect the general sentiment of the grassroots of the community precisely.

Findings and Discussion

The findings derived from this qualitative multiple-case study provide essential empirical grounding for the theoretical themes identified in the literature, meticulously mapping the trajectory of cultural extraction from its analogue colonial roots through to contemporary algorithmic threats and concluding with emerging institutional pushback. Rather than artificially separating the raw data from its broader theoretical implications, this section presents the findings chronologically across the three designated cases, immediately and critically synthesising the empirical evidence with the decolonial framework to explicitly illustrate the highly effective dual architecture of digital neo-colonialism.

Case 1: The Analogue Blueprint of Extraction (The 'Mbube' Precedent)

To properly understand the deeply political nature and extractive logic of contemporary AI training datasets, it is absolutely imperative to establish the historical baseline of cultural extraction that set the precedent for modern intellectual property frameworks. The scholarly discourse increasingly frames the rapid expansion of AI in Africa as an aggressive form of "data colonialism," a dynamic that directly and intentionally mirrors the historical colonial script of extracting raw materials from the global periphery for highly lucrative processing and refinement in the metropole (Couldry & Mejias, 2019; Kwet, 2019). The findings of this historical analysis demonstrate unequivocally that this 'extractive supply chain' was perfected conceptually and legally in the analogue era, most notably through the tragic 1939 case of the Zulu composition *Mbube* by Solomon Linda, which serves as the quintessential, enduring baseline for cultural dispossession on the continent. The documentary analysis of the *Mbube* case reveals a systemic, legally engineered failure of Western Intellectual Property frameworks to protect Indigenous creators from well-resourced external entities; Linda, a brilliant Zulu musician, recorded the vocal composition and subsequently sold his worldwide copyright to a local recording studio for a trivial, exploitative amount of approximately ten Kenyan Shillings (Dugeri, 2024; Schwenzer & Leisinger, 2007). Stripped entirely of its proprietary protections and divorced from its original creator, the song's core musical data was exported to the Global North, where it was extensively reworked into *Wimoweh* and eventually the globally ubiquitous pop hit, *The Lion Sleeps Tonight*, generating an absolute avalanche of royalties for American publishers—particularly through Disney's immensely profitable *The Lion King* franchise—while Linda's cultural origins were systematically obscured and his immediate heirs lived in extreme poverty until a posthumous 2006 legal settlement finally awarded them partial compensation (Schwenzer & Leisinger, 2007).

The theoretical implication of this historical finding is profound, as it clearly demonstrates that the ongoing exploitation of African cultural expression is not a novel, accidental consequence of modern computational technology, but rather a structural, historical feature embedded in the global intellectual property regime itself. Conventional copyright laws have historically privileged the Western 'explorer' or commercial processor who formalised, transcribed, or commodified the culture over the original Indigenous creator who organically birthed it through lived experience and communal heritage. In the *Mbube* case, Linda's profound lived experience, communal ancestral heritage, and individual musical genius were legally reduced to the status of a cheap, raw natural resource—treated conceptually much like extracted cocoa, oil, or gold—which was then repackaged, refined, and sold back to the global consumer market as an expensive, finished commercial commodity (Rodney, 1972). This precise analogue blueprint maps perfectly onto the immediate existential risks posed by contemporary generative Artificial Intelligence models that dominate the digital landscape. The extraction of Linda's music for a mere ten shillings is the exact historical equivalent of the modern, automated "data scraping" process, where foreign technology conglomerates unilaterally harvest the continent's rich, diverse cultural data to endlessly train their proprietary AI models without offering equitable compensation, algorithmic transparency, or securing prior informed consent from the originating communities. Therefore, the current 'data desert' and the extreme vulnerability of African artists to algorithmic appropriation are not accidental technological glitches; they are the highly efficient modern digital manifestations of a long-standing neo-colonial architecture designed specifically to extract cultural value from the periphery in order to concentrate technological wealth and epistemological power in the Global North (APRA AMCOS, 2024).

Case 2: The Modern Threat and 'Silent Extraction'

While the historical *Mbube* precedent perfectly illustrates the analogue roots of cultural dispossession, the rapid proliferation of generative Artificial Intelligence represents a terrifying quantum leap in the speed, massive scale, and deeply covert nature of this extraction. To accurately capture the real-time, visceral collision between global AI models and localised African genres—specifically focusing on the highly lucrative Afrobeats and Amapiano ecosystems—this study conducted a rigorous netnography of the X (formerly Twitter) platform, achieving theoretical saturation across five distinct "anchor" flashpoints that comprehensively summarise the spectrum of contemporary industry resistance and complicity. The ethnographic data forcefully reveal that algorithmic extraction is not solely a top-down corporate action carried out by faceless developers; rather, it is actively infiltrating and poisoning peer-to-peer African collaboration at the ground level. For example, South African music producer MMINO (@BrazoWaAfrika) highlighted a critical flashpoint in which another creator used AI deceptively in a proposed musical partnership, publicly stating, "*Some Of Us Still Respect The Art Of Actually Making Music & Creativity... Keep Your AI*". This deep frustration illustrates the modern, highly contested collision over cultural authenticity, proving that when generative AI is passed off as genuine human collaboration, it severely exacerbates 'postdigital epistemic violence,' deflating the vital testimony and labour of marginalised human knowers while artificially inflating the credibility of machine-generated outputs (MacKenzie, 2023). This specific threat reaches an absolute existential peak when AI successfully mimics the emotive, spiritual qualities of African genres, as seen when cultural commentator Múyìwá Mátùlúkò (@MuyoSan) documented a viral AI-generated Afro Soul remix, expressing deep psychological conflict: "*I am conflicted because this rendition actually has soul... This is why we should be very afraid of what's to come,*" proving that the colonial notion of "staged authenticity" has now been entirely automated, allowing machines to stage the 'soul,' of African music without requiring the lived experience of the African creator (MacCannell, 1973/2023).

The immediate consequence of this unchecked data colonialism and algorithmic mimicry is severe, structural economic displacement that threatens the livelihoods of thousands of creative professionals across the continent. Moving from the perspective of the recording studio to the executive boardroom, A&R strategist Freyy (@Freyy_is) explicitly outlined the terminal, downward trajectory of algorithmic adoption within the corporate music sector: "*The threat to artists is not that AI music is too good. It's that the industry will use it as an excuse to stop paying for human creativity... they will take the cheaper option if the quality is close enough*". This vital netnographic evidence reveals that the primary danger of AI is not merely aesthetic or philosophical, but deeply structural; Global North distribution networks, major record labels, and advertising agencies are actively weaponising generative AI to systematically cut African human labour out of the supply chain entirely, cementing Africa's highly dependent, subservient position within the global technological landscape (Salami, 2024). In response to this multi-tiered, existential threat, localised decolonial resistance is beginning to emerge through strict cultural gatekeeping mechanisms, notably when THE LEGISLATURE (@pianoparliament), a highly prominent Amapiano media entity, issued a definitive, sovereign boundary on the platform: "*AI music does not have a space in Amapiano,*" representing a desperate attempt by cultural custodians to draw a hard, uncompromising line against further algorithmic appropriation.

Crucially, however, the netnographic analysis revealed a highly disturbing absence of mass, viral mobilisation against AI among the broader African public, suggesting a deep psychological disconnect between the culture's creators and its consumers. In fact, the community's replies to @BrazoWaAfrika's initial complaint about deception were overwhelmingly characterised by harsh consumer pragmatism and a deep, systemic techno-optimism that actively defended algorithmic extraction. Anonymised users repeatedly attacked the producer's stance; User E argued aggressively, "*if they don't get on the bandwagon they might lose like BlackBerry and Nokia,*" while User A completely

flattened the philosophical distinction between a basic musical instrument and an autonomous generative model, stating, "*AI is technology, FL is technology, CDJ is technology. Same same.*" This aesthetic devaluation and extreme techno-optimism are profound, highly visible internal consequences of a post-colonial educational system that has historically and systematically alienated African youth from their own heritage, teaching them to view traditional practices as hopelessly 'archaic' while simultaneously embracing Western technological paradigms as the ultimate zenith of human progress (Ibekwe, 2018; Kanu, 2007). Therefore, digital neo-colonialism in the African music sector currently operates as a highly insidious 'Silent Extraction': the genuine economic and cultural threats articulated by producers and executives are actively masked, diluted, and shouted down by a techno-optimistic public narrative that wrongly equates algorithmic extraction with democratised innovation. This is arguably the deepest, most devastating wound of colonial logic; it is not merely that the system extracts valuable resources, but that it successfully teaches the colonised population to narrate their own dispossession and cultural erasure as inevitable, desirable progress, revealing a psychic architecture wherein the subject actively consents to the destruction of their own cultural grammar in the blind pursuit of modernity.

Case 3: Institutional Pushback and the Legal Vacuum

While grassroots digital resistance remains highly fragmented and frequently stifled by the internal techno-optimism observed in Case 2, formal institutional pushback against algorithmic extraction is rapidly crystallising among regional regulatory bodies. However, the findings from this study's rigorous doctrinal legal analysis reveal that African institutions are currently forced to fight this critical decolonial battle on severely compromised legal ground, hamstrung by massive, localised regulatory voids dating back to the immediate post-colonial era. As digital neo-colonialism accelerates at an exponential pace, national legal frameworks across the continent remain perilously anchored in the analogue era, wholly unequipped to manage the complexities of machine learning and large language models. South Africa's Copyright Act 98 of 1978 perfectly exemplifies this internal infrastructural deficit; the Act fundamentally operates on the historical assumption of human authorship, requiring a strict standard of "originality" derived explicitly from human skill, effort, and judgment. While Section 1(1) of the Act does technically contain a provision for 'computer-generated works', vesting legal authorship in "*the person by whom the arrangements necessary for the creation of the work were undertaken*"—contemporary legal scholars note that this archaic framework has rapidly devolved into a massive legal lacuna in the era of highly autonomous generative AI (Fisher, 2024). As Fisher (2024) convincingly argues, it is practically impossible for a court to logically assign the necessary arrangements to a specific human individual when foreign, completely black-box AI systems autonomously generate complex musical output from massively scraped, uncredited datasets; a modern prompt-engineer simply typing *make an Amapiano beat* cannot legally or ethically fulfill the rigorous threshold of authorship envisioned by legislators in 1978, leaving African creators in a continuous state of unprotected ambiguity.

This legislative paralysis is most damaging and extractive during the crucial "input phase" of AI development, the exact moment when the actual extraction of the African 'data desert' occurs. Mikioni (2024) asserts definitively that the mass, automated scraping of African text and audio catalogues to indiscriminately train foreign AI models constitutes a direct, egregious violation of the exclusive right to reproduction explicitly outlined under Section 6(a) of the Copyright Act. Furthermore, foreign technology conglomerates cannot legitimately seek legal refuge under South Africa's highly specific fair dealing doctrine (Section 12), as the extraction of cultural data to build commercial, highly lucrative, subscription-based AI models fundamentally violates the strict statutory requirements of private study, research, or personal use (Mikioni, 2024). Despite these glaring, easily identifiable statutory violations, the abject failure of national governments to proactively update intellectual

property legislation means that foreign AI developers can seamlessly exploit these post-colonial legal grey areas, extracting vast African catalogues with near-total impunity while local creators have no viable legal recourse. Because national governments have thus far failed to close these disastrous regulatory voids, African Collective Management Organisations (CMOs) have been forcefully transitioned from their traditional role as administrative royalty collectors into acting as the frontline, militant defenders of (IDSov. In mid-2025, the Southern African Music Rights Organisation (SAMRO) issued stark, highly publicised warnings regarding the proliferation of generative AI, specifically highlighting the extractive threat to localised, highly lucrative genres like Amapiano, and aggressively advocating for mandatory algorithmic transparency—demanding that AI developers publicly disclose their training datasets and ensure equitable compensation for artists whose catalogues were ingested without consent, which represents a direct, structural institutional challenge to the ongoing mechanics of data colonialism.

To support these vital but highly localised institutional defences, ideological resistance is finally manifesting at the macro-policy level across the continent. The official endorsement of the African Union's *Continental Artificial Intelligence Strategy* (2024) signifies a critical, desperately needed pivot toward genuine technological self-determination and sovereign data governance (African Union, 2024). The strategy's explicit, pan-African mandate to cultivate Africa-centric datasets and fiercely protect cultural heritage serves as a direct legislative counter-measure to the encroachment of algorithmic colonialism. By explicitly prioritising data governance models that serve African socio-economic needs rather than Western corporate interests, the African Union is officially and structurally rejecting the highly extractive open data (FAIR) paradigms championed by the Global North, firmly asserting that data derived from African cultural expressions is not a free global resource, but rather sovereign property that must grant 'Authority to Control' back to the originating communities in strict alignment with the CARE Principles (Carroll et al., 2020). However, as this specific legal case powerfully demonstrates, relying solely on reactive copyright reform and broad policy declarations remains an inherently reactionary defence; it only penalises the act of extraction long after the cultural data has already been permanently compromised and ingested into the machine learning model. True decolonial resistance requires highly proactive epistemological autonomy; to effectively combat the neo-colonial dominance deeply embedded within foreign algorithms, African institutions must completely reclaim the foundational means of technological production, which necessitates a fundamental, root-and-branch restructuring of post-colonial educational systems to prioritise African musicologies and cultivate sovereign, ethically sourced digital archives that cannot be accessed by foreign scraping bots (Mapaya, 2014).

Conclusion

The aggressive expansion of generative Artificial Intelligence within the African music sector is unequivocally not a politically neutral technological advancement; rather, it is fundamentally and intentionally structured as a highly efficient system of digital neo-colonialism. Through a rigorous qualitative multiple-case study encompassing historical precedents, contemporary netnographic discourse, and doctrinal legal analysis, this study demonstrates conclusively that algorithmic extraction relies on a deeply entrenched Dual Architecture that marries external corporate greed with internal legislative and epistemic failures. By articulating this Dual Architecture, this research moves significantly beyond existing concepts of 'data colonialism,' which have traditionally focused almost exclusively on the external actions of Western corporations. Instead, this framework introduces a critical new institutional dimension, proving empirically that algorithmic exploitation is a symbiotic process that relies equally on internal, post-colonial apathy and domestic regulatory voids. Furthermore, it establishes a unique, unbroken theoretical link between historical analogue dispossession and modern digital extraction, proving that AI is not a novel disruption, but rather the

ultimate scaling of an old colonial script. Externally, Global North technology corporations are actively automating the historical, analogue blueprint of cultural dispossession—first perfected in the tragic 1939 *Mbube* case—by unilaterally harvesting vast amounts of African musical data to endlessly train their highly lucrative, proprietary algorithms.

This mass, uncredited scraping directly induces a severe algorithmic flattening, permanently stripping complex genres like Afrobeats and Amapiano of their localised soul and lived experience, producing sterile, uncredited, and commodified outputs that displace African human labour. Internally, this massive external extraction is critically facilitated by profound structural and epistemological deficits across the continent; the expansion of AI operates largely as a 'Silent Extraction,' heavily masked by a deeply ingrained domestic techno-optimism that incorrectly equates algorithmic appropriation with inevitable, desirable modernisation. This epistemic apathy is further weaponised by severe, ongoing legislative paralysis, as outdated, post-colonial statutory frameworks, such as South Africa's Copyright Act of 1978, create vast regulatory voids that allow foreign technology firms to exploit the African "data desert" with near-total legal impunity. Ultimately, the question before African scholars, policymakers, and creators is not whether AI will reshape the continent's musical heritage—it already has permanently altered the landscape; the true question is whether Africa will willingly permit its sonic civilisation to be compressed into free training data for foreign corporate profit, or whether it will finally reclaim cultural authorship at the foundational level of infrastructure, epistemology, and code. In this regard, the proposed African Musicology Institute (AMI) serves not merely as an educational recommendation, but as a vital structural counter-measure—an institutional embodiment of Indigenous Data Sovereignty designed to permanently dismantle the Dual Architecture from within.

To effectively dismantle the dual architecture of algorithmic coloniality, this study proposes the following targeted, structural interventions across the educational, governmental, and corporate sectors. The primary recommendation is the immediate establishment and robust funding of the African Musicology Institute (AMI) as a sovereign, decolonial, pan-African academic hub, partnered with institutions such as the University of Venda in South Africa and Kabarak University in Kenya. AMI must forcefully lead the continent's charge for epistemological autonomy; its core mandate should be to permanently reverse the systemic neglect that created the data desert by actively overseeing the development of vast, ethically sourced, and entirely sovereign datasets of African music. By rigorously implementing IDSov frameworks, AMI will provide the crucial epistemological guidance necessary to ensure that any AI models trained on African soil are infused with deep cultural context, actively preventing the algorithmic flattening of local genres while training a new generation of African technologists to co-design AI tools firmly grounded in local ethical frameworks, effectively repurposing the technology for cultural resurgence rather than extraction.

References

African Union. (2024). *Continental artificial intelligence strategy*. African Union Commission. https://au.int/en/documents/20240809/continental-artificial-intelligence-strategy

Akena, F. A. (2012). Critical analysis of the production of Western knowledge and its implications for indigenous knowledge and education in Africa. In J. P. L. & J. K. (Eds.), *The politics of education in Africa: Challenges and encounters* (pp. 53–71). Palgrave Macmillan.

APRA AMCOS. (2024). *AI and music report*. https://www.apraamcos.com.au/about/supporting-the-industry/research-papers/aiandmusic

Birhane, A. (2020). Algorithmic colonisation of Africa. *SCRIPTed: A Journal of Law, Technology & Society*, *17*(2), 389–409. https://doi.org/10.2966/scrip.170220.389

Braun, V., & Clarke, V. (2021). *Thematic analysis: A practical guide*. Sage Publications.

Carroll, S. R., Garba, I., Figueroa-Rodríguez, O. L., Holbrook, J., Lovett, R., Materechera, S., Parsons, M., Raseroka, K., Rodriguez-Lonebear, D., Rowe, R., Sara, R., Walker, J. D., Anderson, J., & Hudson, M. (2020). The CARE principles for Indigenous data governance. *Data Science Journal*, *19*(1), 43. https://doi.org/10.5334/dsj-2020-043

Chibike, J. J. (2024). Beyond the neocolonizer's appreciation of climate change: Consumption of Zimbabwean tabloid Umthunywa. In U. S. Akpan (Ed.), *De-neocolonizing Africa: Harnessing the digital frontier* (pp. 273–286). Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-40439-9_16

Coleman, D. (2019). Digital colonialism: The 21st-century scramble for Africa through the extraction and control of user data and the limitations of data protection laws. *Michigan Journal of Race & Law*, *24*(2), 417–439. https://doi.org/10.36643/mjrl.24.2.digital

Couldry, N., & Mejias, U. A. (2019). *The costs of connection: How data is colonising human life and appropriating it for capitalism*. Stanford University Press.

Cruz, C. C. (2021). Decolonising philosophy of technology: Learning from bottom-up and top-down approaches to decolonial technical design. *Philosophy & Technology*, *34*, 1847–1881. https://doi.org/10.1007/s13347-021-00486-z

Dugeri, M. (2024). Generative AI and the risk of cultural appropriation of Indigenous African musical works. *Journal of Intellectual Property and Information Technology*, *4*(1), 17–34.

Fisher, A. (2024). *Copyright law as it pertains to machine learning as an artificial intelligence pillar in South Africa* [Master's mini-thesis, University of the Western Cape]. UWC Scholar Repository.

Gyamerah, K. (2024). *The role of African Indigenous knowledge systems and pedagogies in decolonising and transforming mathematics and science learning in Ghana* [Doctoral dissertation, Queen's University]. QSpace.

Ibekwe, E. U. (2018). Africa and cultural authenticity. *Online Academic Press*, *1*(2), 50–59.

Ilo, E. C. (2025). *Africa's struggle for sovereignty: Neo-colonialism and external control in the 21st century*. Pan-African University Press.

Kanu, Y. (2007). Tradition and educational reconstruction in Africa in postcolonial and global times: The case for Sierra Leone. *African Studies Quarterly*, *9*(3), 65–84.

Kirui, A. K., Wanyama, M. N., & Mapaya, M. G. (2025). Epistemics of African music performance: Navigating knowledge, tradition, and the digital frontier. *African Musicology Online*, *14*(2), 14–24. https://doi.org/10.58721/amo.v14i2.1107

Kozinets, R. V. (2020). *Netnography: The essential guide to qualitative social media research* (3rd ed.). Sage Publications. https://doi.org/10.4337/9781800377486.netnography

Kwet, M. (2019). Digital colonialism: US empire and the new imperialism in the Global South. *Race & Class*, *60*(4), 3–26. https://doi.org/10.1177/0306396818823172

le Grange, L. (2024). Authenticity as colonial and curricular construct. *Journal of Decolonial Education*, *3*(1), 12-28.

MacCannell, D. (2023). Staged authenticity: Arrangements of social space in tourist settings. *American Journal of Sociology*, *79*(3), 589–603. (Original work published 1973) https://doi.org/10.1086/225585

MacKenzie, A. (2023). Postdigital epistemic violence. In P. Jandrić (Ed.), *The encyclopedia of postdigital science and education* (pp. 1–5). Springer Cham. https://doi.org/10.1007/978-3-031-35469-4_7-1

Mapaya, M. G. (2014). The study of indigenous African music and lessons from ordinary language philosophy. *Mediterranean Journal of Social Sciences*, *5*(20), 2007–2014. http://dx.doi.org/10.5901/mjss.2014.v5n20p2007

McKay, D. L. (2019). Real Indians: Policing or protecting authentic Indigenous identity? *Sociology of Race and Ethnicity*, *7*(1), 12–25. https://doi.org/10.1177/2332649218821450

Menon, S. (2023). Postcolonial differentials in algorithmic bias: Challenging digital neo-colonialism in Africa. *SCRIPTed: A Journal of Law, Technology & Society*, *20*(2), 383–405. https://doi.org/10.2218/scrip.20.2.2023.8980

Mikioni, T. (2024). The clock is ticking for the Sustainable Development Goals: A geographical indications-based intervention in the quest to end poverty. *South African Intellectual Property Law Journal*, *13*(2), 108–124. https://doi.org/10.47348/SAIPL/v13/i2a5

Mohamed, S., Png, M. T., & Isaac, W. (2020). Decolonial AI: Decolonial theory as sociotechnical foresight in artificial intelligence. *Philosophy & Technology*, *33*, 659–684. https://doi.org/10.1007/s13347-020-00405-8

Ndlovu-Gatsheni, S. J. (2013). *Coloniality of power in postcolonial Africa: Myths of decolonisation*. CODESRIA. https://muse.jhu.edu/book/25053

Ndlovu-Gatsheni, S. J. (2015). Decoloniality as the future of Africa. *History Compass*, *13*(10), 485–496. https://doi.org/10.1111/hic3.12264

Nkasi, E. (2025). *AI and Afrobeats: Can machines make our music?* The BFA Agency. https://www.thebfaagency.com/post/ai-and-afrobeats-can-machines-make-our-music

Noble, S. U. (2018). *Algorithms of oppression: How search engines reinforce racism*. New York University Press. https://doi.org/10.2307/j.ctt1pwt9w5

Oguamanam, C. (2020). Indigenous Peoples, data sovereignty, and self-determination: Current realities and imperatives. *African Journal of Information and Communication*, *26*, 124–146. https://doi.org/10.23962/10539/30360

Olajide-Philips, P., & Ogunbowale, S. (2025). *KorinAI launches Africa's first AI music lab*. Tech in Africa. https://www.techinafrica.com/korinai-launches-africas-first-ai-music-lab-redefining-music-creation-and-distribution/

Quijano, A. (2007). Coloniality and modernity/rationality. *Cultural Studies*, *21*(2–3), 168–178. https://doi.org/10.1080/09502380601164353

Rainie, S. C., Kukutai, T., Walter, M., Figueroa-Rodriguez, O. L., Walker, J., & Axelsson, P. (2019). Issues in open data: Indigenous data sovereignty. In T. Davies, S. Walker, M. Rubinstein, & F. Perini (Eds.), *The state of open data: Histories and horizons* (pp. 300–319). African Minds and International Development Research Centre. https://doi.org/10.5281/zenodo.2677801

Research Data Alliance. (2019). *CARE principles for Indigenous data governance*. The Global Indigenous Data Alliance. https://doi.org/10.5334/dsj-2020-043

Rivera, C. (2025). *AI, cultural appropriation, and the orange economy*. Cultural Policy Institute.

Rodney, W. (1972). *How Europe underdeveloped Africa*. Bogle-L'Ouverture Publications.

Salami, A. O. (2024). Artificial intelligence, digital colonialism, and the implications for Africa's future development. *Data & Policy*, *6*, e15. https://doi.org/10.1017/dap.2024.12

Schwenzer, I., & Leisinger, B. (2007). The Lion Sleeps Tonight – Award of royalties to the heirs of Solomon Linda. *Loyola of Los Angeles Entertainment Law Review*, *28*(2), 87–103.

Seow, E. (2024). *Decolonising music education across diverse institutions*. Music In Africa.

Ssekibuule, R., Quinn, J. A., & Leyla, A. (2024). Mapping the African data landscape: Challenges and opportunities for sovereign AI development. *Journal of Information Policy*, *14*(1), 45–68. https://doi.org/10.5325/jinfopoli.14.1.0045

Warganegara, M. R. R. (2024). Shifting from 'AI solutions' to 'AI coloniality': Resignification of artificial intelligence and digital apartheid. *Global South Review*, *4*(2). https://doi.org/10.22146/globalsouth.93722

Xulu, Z. C. S., & Yende, S. J. (2024). Decolonising sound: The scholarship of teaching and learning for 21st-century music students in South Africa. *International Journal of Learning, Teaching and Educational Research*, *23*(12), 62–77. https://doi.org/10.26803/ijlter.23.12.4

Yende, S. J., & Yende, N. E. (2022). The quest for curricularisation of uMaskandi Zulu traditional music in higher education in the context of Africanisation. *Journal of African Education*, *3*(1), 103–125. https://doi.org/10.31920/2633-2930/2022/v3n1a5

Yin, R. K. (2018). *Case study research and applications: Design and methods* (6th ed.). Sage Publications.