equalyzAI

African Language AI in 2025:

Insights, Trends and Strategic Direction

TABLE OF CONTENTS

PART III — GAPS, RISKS & CRITICAL CHALLENGES

PART IV — 2026 OUTLOOK AND PREDICTIONS

Introduction

Olubayo
Adekanmbi PhD
Co-Founder/CEO EqualyzAI

Ife
Adebara PhD
Co-Founder/CTO EqualyzAI

The year 2025 represented a defining moment for Language AI in Africa. Across the continent, advances in data creation, model development, infrastructure expansion, policy formulation, and real-world deployment converged in ways that moved African language technologies from experimentation toward operational relevance. What was once framed primarily as a challenge of low-resource languages has increasingly become a question of context-aware innovation, scale, governance, delivery, and long-term sustainability. This report documents that transition, capturing how African actors are beginning to shape Language AI systems that reflect local languages, cultural contexts, and socio-economic realities rather than relying solely on imported technologies.

From the perspective of EqualyzAI, 2025 marked a shift from isolated technical breakthroughs to the emergence of an interconnected ecosystem. Language AI is no longer only about models and datasets, but about the full stack that enables delivery. This includes compute access, edge and offline architectures, regulatory alignment, trust and safety frameworks, and sector-specific integration across healthcare, finance, education, agriculture, and public services. These layers increasingly determine whether African language technologies remain academic artifacts or mature into dependable digital infrastructure for everyday use.

This report is structured around several thematic areas that defined Language AI progress in 2025. These include the rapid expansion of speech and text datasets across African languages, the rise of African-led and African-adapted language models, improvements in compute and cloud infrastructure, advances in benchmarks and evaluation, and the growing presence of real-world deployments. Together, these themes illustrate a continental trajectory that prioritizes linguistic inclusion, practical utility, and resilience in low-connectivity and low-compute environments. They also reflect a growing recognition that Language AI in Africa must be built differently, with design choices shaped by local constraints and opportunities.

Looking ahead to 2026, several dominant issues are already emerging. Voice-first and speech-enabled systems are becoming the primary interface for digital services in many regions. Small and efficient language models are gaining prominence as a practical alternative to large, compute-intensive systems. Offline and edge

deployment strategies are becoming central to last-mile access. At the same time, governance, data ownership, evaluation standards, and cultural alignment are moving from peripheral concerns to core design requirements. These issues are not speculative; they are direct extensions of the pressures and patterns observed throughout 2025.

EqualyzAI's commitment within this evolving landscape is anchored in localisation and contextualization. We view Language AI not simply as standalone initiatives, but as interconnected digital public infrastructure. This means designing systems that operate reliably within African connectivity and building delivery rails that allow models, data, and applications to reach users at scale. Our approach emphasizes voice-first access, domain-specific intelligence, cultural groundedness and responsible deployment, with the long-term objective of enabling African institutions, businesses, and communities to own and operate their AI systems with confidence.

Finally, it is important to acknowledge the limits of any single report in a rapidly evolving ecosystem. While significant effort was made to compile developments across regions, sectors, and organizations, it is likely that some important initiatives, releases, and contributions were missed or underrepresented. This omission does not diminish their value or impact. Instead, it reflects the increasing breadth and dynamism of African Language AI. We recognize and respect the work of all contributors across the continent and view this report as a living snapshot rather than an exhaustive record. Our hope is that it serves as a foundation for continued documentation, collaboration, and shared learning as Africa shapes its own Language AI future.

Executive Summary

The year 2025 marked a watershed for African Language AI. Across the continent, advances in infrastructure, model development, dataset creation, regulation, and real-world deployments converged to create the strongest momentum Africa has seen in AI to date. This executive summary presents the main developments, investments, deployments, and emerging trends from 2025, as well as the directions that are forming the AI trajectory of Africa as it enters 2026.

1.1 Key Developments in African Language AI

African language AI saw unprecedented progress in 2025 in three major domains:

- Data ecosystems, with new speech, text, translation, and benchmark datasets released across North, West, East, Central, and Southern Africa.
- Model innovation, including the launch of multilingual LLMs in Nigeria and Uganda, edge-optimized models in South Africa, and domain-specific healthcare models in East Africa.
- Infrastructure expansion, with new GPU access programs, compute accelerators, local cloud regions, and regional distribution platforms moving training and deployment closer to African users.

These developments reduce long-standing barriers from data scarcity to compute access and indicate that African NLP and speech technology are transitioning from basic building to scalable deployment.

1.2 Major Releases, Investments and Deployments

The year also provided many high-impact releases and investments that moved the ecosystem from experimentation toward operationalization:

- Language Models: N-ATLAS (Nigeria), Sunflower (Uganda), InkubaLM (South Africa), and UlizaLlama (East Africa) increased the availability of African-language AI systems, each optimized for different use cases.
- Industry Deployments: Healthcare (e.g., PROMPTS, clinical ASR tools), financial services (AI fraud detection, AI credit scoring), and telecom operations (AI-driven customer care) leveraged AI directly into service delivery.
- Infrastructure Investments: Corporate and philanthropic commitments, including Google.org funding, Masakhane research grants, and accelerator-backed compute access improved the capacity for local innovation.
- Government & Policy Action: New national AI strategies, privacy-law updates, and regional declarations reinforced governance frameworks and Africa's position in global AI discourse.

1.3 Top 5 Regional Trends

Across the continent, several cross-cutting trends emerged from the year's advances:

1. Acceleration of African-language LLMs: Countries invested in models designed for local languages, which shows that linguistic inclusion is essential for digital participation.
2. Expansion of community-led data creation: Hackathons, RFPs, and dataset initiatives encouraged local researchers and citizens to develop linguistic resources for underrepresented languages.
3. Growth of compute accessibility: Regional GPU hubs, cloud-region expansion, and subsidized access programs reduced barriers to model training and fine-tuning.
4. Sectoral integration of AI: Healthcare, financial services, education, and telecoms adopted AI tools to address many challenges like clinical shortages, fraud risk, and literacy constraints.
5. Strengthened governance and safety frameworks: Regional declarations, national strategies, and new privacy laws created a more coherent regulatory environment for responsible AI use.

These trends highlight a continental ecosystem that is neither peripheral nor reactive; Africa is actively shaping how AI technologies evolve and are governed.

1.4 High-level Predictions for 2026

Building on 2025's momentum, several strategic shifts are likely to define 2026:

- Voice AI becomes mainstream: Increased speech datasets and ASR/TTS deployments make voice the default interface for many services.
- Offline and low-compute models proliferate: Edge-optimized LLMs and small-model architecture expand reach in low-bandwidth regions.
- Local-language AI assistants scale across sectors: Education, agriculture, and public health benefit from culturally rooted conversational agents.
- Benchmark standardization takes hold: African-led evaluation frameworks become widely adopted across industry and research.
- Regional fine-tuning hubs emerge: Countries with growing compute and model expertise become centers for continental model adaptation.

These predictions reflect the continent's shift toward African-owned and African-optimized AI systems.

1.5 Visual Infographics of Key Trends

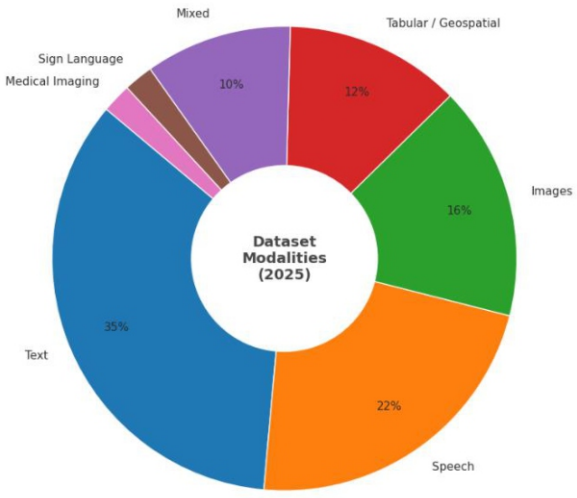


Figure 1. Share of Dataset Modalities Across Africa (2025)

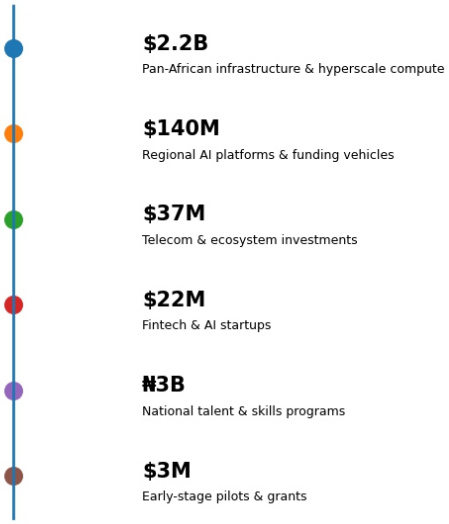


Figure 2. AI Investment in Africa (2025)

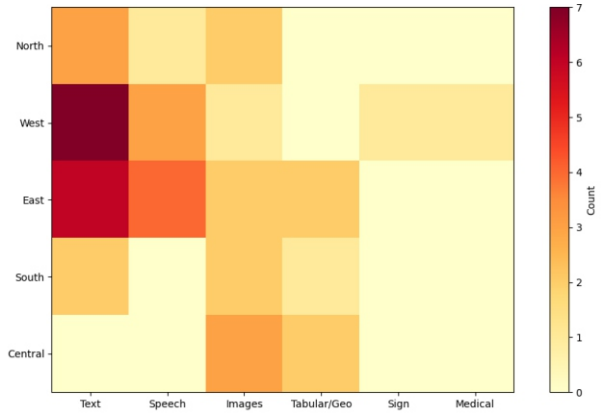


Figure 3. Dataset Modality Heatmap by Region (2025)

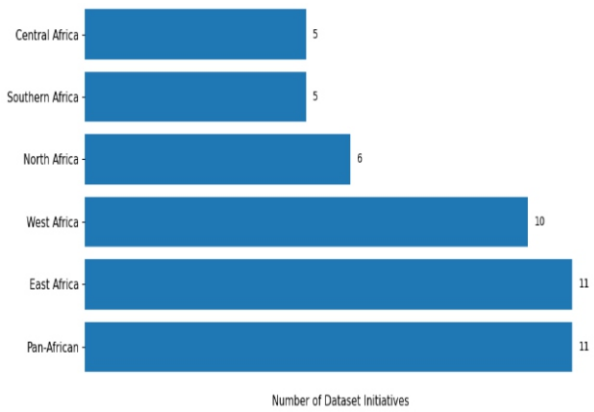


Figure 4. Dataset Initiatives by African Region (2025)

2.0

Introduction & Methodology

This report provides a review of African Language AI progress in 2025, including developments across model releases, dataset creation, national strategies, benchmark innovations, sector deployments, and ecosystem investment. As the use of AI is increasing across the continent, understanding how language technologies evolve has become essential to form the digital future of Africa.

The following subsections outline the purpose, methodological foundations, and boundaries of this study.

2.1 Purpose and Significance

The main objective of this report is to document and analyze advances in African Language AI during 2025. Specifically, the significance of this report lies in three main contributions:

1. Providing a continent-wide view of the progress in language AI by highlighting regional variation and shared trends.
2. Situating technical achievements within broader socio-economic contexts, including public trust, data governance, and digital inclusion.
3. Offering researchers and practitioners a consolidated evidence base for policy formulation, investment planning, and model development.

By framing African AI developments on their own terms, the report emphasizes that Africa is not simply adapting global AI tools; it is shaping AI trajectories rooted in African languages, realities, and priorities.

2.2 Key Definitions

For clarity and consistency, this report uses the following definitions:

Language AI: AI systems designed to understand, generate, or process human language in text or speech form.

LLMs (Large Language Models): Deep-learning models trained on extensive text corpora to perform tasks such as translation, classification, generation, and reasoning.

SLMs (Small Language Models): Efficient, lower-parameter models optimized for on-device or low-resource environments.

ASR (Automatic Speech Recognition): Models that convert spoken audio into text.

TTS (Text-to-Speech): Systems that produce natural-sounding speech from text input.

NLU (Natural Language Understanding): AI capabilities that interpret meaning, intent, sentiment, or context in human language.

Multimodal AI: Models capable of processing multiple input types (e.g., text and audio).

Low-resource languages: Languages with limited digital text, speech data, or NLP tools available for model training.

2.3 Research Methodology and Data Sources

The methodology involves qualitative and quantitative analysis of various trusted sources, including:

- Academic publications and conference proceedings
- Government strategy documents and regulatory updates
- Press releases from organizations, corporations, and ministries
- Industry reports and sector analyses
- Dataset and model documentation from AI developers
- NGO and multilateral agency outputs
- Public technical benchmarks and open-source repositories

Data points from these sources were aggregated and compared to identify continental trends, sector patterns, and emerging gaps.

2.4 Verification Steps

To ensure accuracy and prevent misinterpretation, the following verification steps were applied:

- Cross-source validation: Claims from one source were corroborated by other different references wherever possible.
- Temporal consistency checks: Only advances that happened in 2025 were included in the main analysis.
- Terminology alignment: Definitions and model descriptions were standardized to avoid ambiguity.
- Bias mitigation: Care was taken not to overrepresent advances from countries with higher media coverage at the expense of smaller or less-documented markets.
- Factual consistency: Every factual point in the analysis is linked to a cited reference, dataset, or official announcement.

2.5 Scope and Coverage

This study covers all five African regions: North, West, East, Central, and Southern Africa. The study focuses on advances that occurred from January to December 2025. The scope includes:

- Dataset releases (speech, text, multimodal, image)
- Language models (LLMs, SLMs, ASR/TTS)
- Infrastructure growth (compute, cloud regions, telecommunications)
- Sector deployments (finance, healthcare, education, public services)
- Governance, regulation, and trust frameworks
- Partnerships and investments across government, academia, industry, and non-profit organizations (NGOs)

2.6 Limitations and Assumptions

To maintain methodological integrity, several limitations and assumptions are acknowledged:

- Data availability varies significantly across countries. Countries with stronger media or more active AI communities generated more publicly documented advances.
- Not all AI deployments are publicly disclosed, especially in private-sector environments, which means that some initiatives may be underrepresented.
- Rapidly evolving AI ecosystems may lead to new advances and updates shortly after publication; this report includes only information retrieved at the writing time.
- Language datasets and model descriptions were only included when linked to verified sources; informal claims or unverifiable announcements were excluded.
- Comparisons across regions assume similar definitions of "AI deployment" and "dataset release," though implementation maturity varies.
- These limitations do not diminish the report's value but clarify the boundaries within which the analysis should be interpreted.

equalyzAI

Africans want AI that get things done when they speak

At **EqualyzAI,** AI is not conversation.
AI is execution through voice.
We call it Afro-agentic!

www.equalyz.ai

Compute and Cloud Capacity

3.1.1 Local GPU and Compute Resources

In 2025, Africa's push toward building indigenous high-performance computing (HPC) and AI infrastructure transitioned from nascent ambition to measurable deployment, with major private and public-sector players catalyzing a new tier of localized compute capacity. Across the continent, three patterns defined the year's progress: massive data centre investments by global and regional technology firms, telecom-cloud expansions for AI workloads, and academic and ecosystem-oriented compute clusters for research and innovation. Together, these developments illustrate a continent moving from isolated pilot projects to a more coordinated compute landscape.

A defining moment in this shift was

US$700M

partnership between Nvidia and Cassava Technologies.

Nvidia's first direct infrastructure investment in Africa (1). The collaboration set a new benchmark for scale and ambition: Nvidia GPUs and AI compute stacks are being integrated into Cassava-built data centres in South Africa, with planned deployments in new markets such as Egypt, Nigeria, Kenya, and Morocco. According to the partnership announcement, the initiative is also tied to capacity-building efforts through Nvidia training programs, a recognition that hardware alone is insufficient without local skills development. Beyond its headline value, the partnership signals a deeper strategic shift: Africa is transitioning from simply consuming AI technologies to hosting the compute foundations needed to build and run its own enterprise-grade models. By placing advanced GPUs closer to African organizations, the initiative directly addresses the continent's longstanding reliance on offshore compute.

Source: **Techpoint Africa** *https://techpoint.africa/news/cassava-nvidia-ai-720-million/*

This continental push toward localized AI infrastructure was echoed in East Africa, where **Atlancis Technologies** launched the region's first GPU-powered "AI factory" under its Servernah Cloud brand (2). Hosted at the hyperscale, AI-ready iXAfrica Data Centres in Nairobi, the facility is engineered to deliver hyperscaler compute capacity for machine learning, deep learning, data analytics, and other intensive workloads. Reporting on the launch highlighted Atlancis's focus on energy-efficient hardware design and its framing of the factory as a national digital competitiveness asset. By leveraging Open Compute Project principles and Nvidia GPUs, the deployment is a catalyst for computational sovereignty: instead of routing AI workloads to distant providers, enterprises, and governments are equipped to train and deploy models from within the continent.

The telecom sector—historically one of Africa's most influential technology actors—also deepened its role in expanding AI-ready compute. **Vodacom's multi-year strategic collaboration with Google Clouds** reflects a wider trend in which telcos position themselves as gateways for enterprise AI adoption (3). Coverage of the agreement notes that Vodacom aims to improve network performance, analytics, and customer services by migrating core data platforms onto Google's infrastructure and integrating generative AI models such as Gemini. This move highlights the sector's strategic pivot from connectivity providers to enablers of digital intelligence at scale.

Source: **Techcabal.**

Simultaneously, **MTN Nigeria** undertook a major upgrade to its local data centre ecosystem with the launch of West Africa's largest Tier III data centre in Lagos (4). Initial reporting valued the investment at approximately US$235-255M (including cloud services), positioning it as part of MTN's broader US$1 billion cloud strategy aimed at strengthening Nigeria's data-hosting capabilities.

US$235-255M
(including cloud services),

The centre is designed to deliver scalable cloud and compute services that rival global hyperscalers like AWS, Google, and Azure. Its modular architecture and pay-as-you-use model—with services priced in local currency—aim to broaden access to enterprise-grade resources and anchor AI-driven innovation locally rather than offshore.
As these commercial deployments took shape, a new class of innovation-enabling environments also began emerging—most notably in Nigeria. **Itana's Full-Stack Growth Zone,** the country's first digital-only free zone, offers a regulatory and operational framework optimized for technology companies, including those building AI and compute-intensive products (5). By simplifying incorporation, compliance, and data-

handling processes, Itana serves as a bridge between the continent's expanding compute capacity and the startups, SMEs, and global firms seeking to utilize it. Its model demonstrates how regulatory innovation can complement infrastructure investment by accelerating the path from compute access to real-world deployment.

Beyond enterprise-focused investments, 2025 also witnessed investments in specialized computing within the research sector. In Uganda, **Makerere University's Infectious Diseases Institute** launched an upgraded HPC cluster to boost the institution capabilities for complex data analysis and computational research (6). Reporting emphasized its value for epidemiology, genomics, and climate-linked disease modelling, areas where African researchers have historically relied on international compute partnerships. Although smaller in scale compared to commercial data centres, these academic HPC deployments are critical drivers for local research communities tackling region-specific challenges.

Finally, the year's developments were underscored by a growing continental vision for more inclusive compute access. Initiatives led by **Udutech** and **Alliance4AI,** for example, endorse the development of an African "AI backbone" by reducing barriers to high-performance computing access for innovators and startups (7). Their model, focused on distributed access, affordability, and community empowerment, aligns with broader reporting that positions GPU availability as a potential economic differentiator for countries like Nigeria. If maintained, these ecosystem-oriented models could serve as the connective tissue linking national data centres, commercial platforms, and academic infrastructure, which ensures that Africa's emerging compute capacity extends well beyond major urban and economic hubs.

Source: **Makerere University**

3.1.1.1 GPU Access for Academic Research and Development (R&D)

Alongside commercial and industry-led compute investments, 2025 also marked a notable expansion in **research-focused GPU access** across Africa, which is an equally important pillar in the continent's AI readiness. While enterprise deployments received most of the attention, these research-oriented programs played a quieter but crucial role: they endow African academics, laboratories, and early-stage innovators with the computational foundation needed to contribute meaningfully to global AI research.
The most substantial of these initiatives came through the **AI Hub Compute Accelerator,** a joint effort that involves CINECA, AWS, and Microsoft. Through this initiative, African innovators received 1.5 million GPU hours and US$1 million in AWS credits.

1.5 million
GPU hours and
US$1 million
in AWS credits.

Positioning it at the forefront of coordinated compute-access interventions available to researchers on the continent (8). This support targeted a barrier highlighted repeatedly across Africa: the prohibitive cost of training and fine-tuning large-scale models. By offsetting these expenses, the Accelerator endowed research teams with the capacity to run experiments that would otherwise be infeasible in a purely local computing context. The initiative therefore served not only as a funding mechanism but as a catalyst for expanding Africa's computational research capacity.

At the same time, **Google Cloud expanded its academic research credit program** to faculty, PhD candidates, and postdoctoral researchers across **13 African countries**, including key research hubs such as Nigeria, Kenya, South Africa, and Egypt (9). By providing access to cloud GPUs and platforms like Vertex AI, Google's offering helped close the gap between African researchers and their peers in more compute-rich regions. The expanded eligibility also acknowledged a growing trend in African academia: the rise of multidisciplinary research teams that rely increasingly on model training and experimentation with generative AI systems.

Source: **Google Cloud**

Google's investment in academic capacity-building extended further through its sponsorship of the Data Science for Health Ideathon which is launched at the 2025 Deep Learning Indaba (10). Finalist teams were granted technical support and Vertex AI compute credits, which creates a practical experimentation environment for multidisciplinary research teams. Researchers across these programs could optimize limited computational resources through parameter-efficient techniques like LoRA (Low-Rank Adaptation) to fine-tune models, including Google's MedGemma, for localized health applications. This use of compute grants indicates that even if GPU access remains stringent, African researchers are increasingly adept at leveraging modern optimization strategies to reduce cost without compromising model quality.

Source: **TechCabal**

3.1.1.2 GPU Access for Innovation & Product Development

While academic communities made important gains through dedicated compute grants, 2025 also witnessed remarkable progress in **market-oriented GPU access** designed to speed innovation, product development, and enterprise adoption of AI across the continent. Unlike research-focused programs, these initiatives emphasize the creation of local ecosystems capable of building and deploying AI solutions at scale.

A major milestone is the launch of **Altron's AI Factory,** introduced as South Africa's first operational facility dedicated to enterprise-grade AI infrastructure and services (11). The platform provides organizations with direct access to HPC resources, model development environments, and technical support so as to reduce the barriers to building production-level AI applications. Altron positioned the AI Factory not only as a hardware resource, but as an end-to-end environment for accelerating business innovation in sectors such as finance, retail, logistics, and public services.

In parallel, **Cassava Technologies** expanded its role as a regional enabler of applied AI. Through its partnership with the **South African Artificial Intelligence Association (SAAIA),** Cassava engaged to widen access to compute services across Africa (12). The partnership is framed as part of a long-term effort to strengthen the continent's AI capacity by improving access not only to infrastructure but also to the networks and institutional support needed for experimentation and deployment. This collaboration reinforces a shift emphasized by many African digital policy leaders: reducing reliance on offshore cloud services and enabling more AI workloads, especially those involving sensitive or regulated data, to run locally. By aligning with an ecosystem-building organization like SAAIA, Cassava's approach emphasizes capacity development and market access alongside raw compute provision.

Cassava further deepened its ecosystem engagement through a separate collaboration with **Zindi,** Africa's largest community of data scientists (13). This partnership highlights African AI innovation by connecting community-driven problem solving with access so as to compute resources and visibility channels. By elevating solutions created by African practitioners, many of whom lack traditional access to enterprise-grade infrastructure, the collaboration demonstrates how compute access can serve as a multiplier for talent development. This model aligns with a broader continental trend: treating compute not only as a technical asset but as a catalyst for inclusive innovation pipelines that bring researchers, practitioners, and commercial actors into closer alignment.

3.1.1.3 "Last-Mile" GPU Access & Edge Computing

As continental-scale data centres and cloud partnerships expanded Africa's high-end compute capacity in 2025, a parallel movement emerged to solve a different but equally critical challenge: how to bring compute closer to the people and places that need it most. Stakeholders increasingly recognized that without last-mile access—affordable, local, and resilient—Africa's AI advancements would remain unevenly distributed. Against this backdrop, startups and research groups undertook the development of new pathways for providing GPU power and edge capabilities to underserved regions.

The **Frontier AI Accelerator** exemplified this shift. Launched in 2025, it provided African AI startups with compute credits, including GPU access, to help overcome one of the most stringent barriers facing emerging innovators: the cost and availability of model training infrastructure (14). Instead of centering only on mentorship or early-stage capital, the program placed compute itself at the heart of startup enablement. The approach acknowledged a key reality: African enterprises often have the ideas, but not the infrastructure to test and scale AI innovation.

Extending this logic beyond cohort-based acceleration, organizations such as **PAWA AI, Udu Technologies,** and the **Africa Compute Fund** advanced a complementary model focused on **affordable GPU networks deployed across multiple African countries** (15). By integrating mobile-money billing, they aligned compute provisioning with everyday financial practices on the continent. This strategic design choice strengthened accessibility for developers, researchers, and small businesses operating outside metropolitan centers. Whereas accelerators reduce up-front cost barriers, these GPU networks work to normalize continuous, pay-as-you-need access—effectively broadening the on-ramp for those seeking to prototype or deploy AI systems.

Source: **Christopher Sanchez | AI**

In parallel, research communities explored how **edge computing** could close the gap in regions where connectivity remains limited. Studies on **agentic educational content generation on low-power devices** such as Raspberry Pi and NVIDIA Jetson demonstrated that multilingual language models can run locally with low latency and strong performance (16). This research underscores a critical insight: even when cloud compute is unavailable, edge devices can maintain high-value AI use cases, particularly in schools, community hubs, and rural learning environments. These findings reinforce the idea that Africa's AI ecosystem must be architected with flexibility across connectivity tiers, from data-center-scale GPUs to self-contained local inference systems.

Moving from research prototypes to practical platforms, 2025 also saw the emergence of **next-generation edge AI delivery solutions** designed specifically for **African constraints.** One notable example is the "last-mile EDGE" initiative built around **Google's Coral NPU,** led by **African Google researcher, Abdoulaye Diack in Ghana** (17). This full-stack, open-source edge platform was engineered to address exactly the challenges facing low-power AI deployment on the continent, namely performance bottlenecks, device fragmentation, and the need for privacy-preserving, always-on inference across mobile and wearable devices.

At the same time, key industry players advanced cloud-based pathways that complement these distributed and edge-oriented efforts. **Cassava Technologies** continued to expand its **GPU-as-a-Service (GPUaaS)** footprint across the continent, which enables organizations to provision high-performance GPUs on demand without owning physical hardware (18). This model benefits especially enterprises and mid-scale teams that require reliable GPU capacity but cannot justify large capital expenditure. In contrast to the grassroots orientation of edge deployments or startup-focused accelerators, Cassava's GPUaaS targets organizations ready to scale, thereby filling a crucial middle layer in Africa's compute landscape.

Meanwhile, last-mile enablement also extended into the small-business economy. A collaboration between **Fastagger** and **Microsoft** brought **offline-capable edge AI tools** to merchants operating within Africa's **US$1.7 Trillion mobile-money market** (19).

Source: **STACK Cybersecurity**

These tools support routine tasks like product tagging and customer insights, operating directly on local devices without continuous internet access. Their offline-first design aligns with the operational realities of millions of small enterprises across the continent, where connectivity conditions vary from those presumed by most global AI solutions. This initiative illustrates that last-mile AI is not only a technical challenge but also a significant commercial opportunity grounded in Africa's unique economic patterns.

Across these varied initiatives, a coherent narrative reveals: Africa's AI progress depends more and more on the ability to decentralize compute, diversify how it is delivered, and align it with the lived realities of local users. From startup accelerators to cross-country GPU networks, from low-power edge devices to enterprise GPUaaS platforms, and from research labs to micro-merchants, last-mile compute access is becoming an essential counterpart to the continent's growing cloud and data-center capacity. By expanding who can run models, and where, these initiatives position Africa not only as a beneficiary of global AI innovation but as an active contributor shaping technologies that work at continental scale.

3.1.1.4 Innovative and Decentralized GPU Approaches

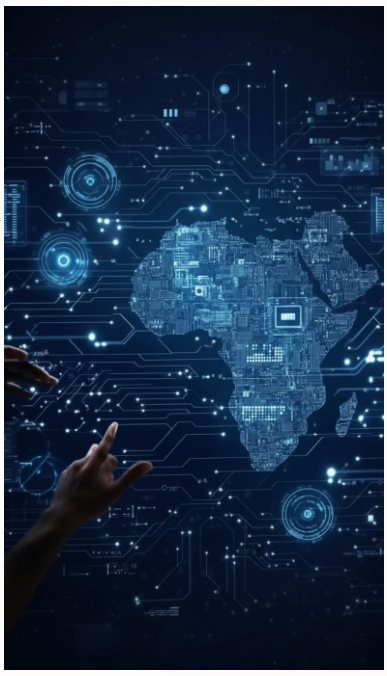
As Africa's AI ecosystem matures, 2025 also witnessed the emergence of **decentralized and distributed compute models** that complement traditional cloud and data-center infrastructure. While centralized approaches provide reliability and scale, decentralized approaches introduce redundancy, flexibility, and affordability, which represent the most important features especially in regions where connectivity or data-governance constraints restrict traditional GPU access. Across several initiatives, a shared ambition is becoming clear: to re-envision how compute is provisioned by distributing it closer to users.

Among the most forward-leaning examples is the **GPUAI Protocol,** which proposes a decentralized compute framework capable of transforming otherwise idle GPU resources into a unified mesh network (20). The protocol aggregates compute from a distributed pool of contributors instead of relying on hyperscale data centers, and coordinates AI training and inference workloads across this non-centralized hardware. Its value proposition lies in democratizing participation: anyone with spare GPU capacity can contribute and benefit from a shared compute economy. For regions where GPU availability is uneven or expensive, such decentralized pooling gives a potential pathway to more affordable and resilient AI development.

This logic extends naturally to **NodeGoAI,** which enables individuals and organizations to **monetize idle CPUs and GPUs** by contributing them to a peer-to-peer decentralized network designed for AI workloads, spatial computing, and high-performance tasks (21). While different in implementation, NodeGoAI shares key similarities with the GPUAI Protocol: both seek to unlock latent compute capacity across distributed actors and channel it toward productive AI use cases. The model blurs the line between compute providers and compute consumers, which positions users not only as end-points in the system but as active participants in a distributed computational marketplace. For Africa, access costs and hardware scarcity often restrict innovation; therefore, these models hold particular promise in lowering entry barriers.

A related but conceptually different strand of decentralization is embodied in **federated learning,** a technique championed by IBM and increasingly adopted in AI systems (22). Federated learning consists of distributing model training over multiple client devices, which enables each node to train locally while sharing only model updates and not raw data.

This approach is well-aligned with African contexts where data sensitivity, connectivity variability, and regulatory constraints can render centralized training obsolete. By keeping data on-device, federated learning improves data privacy and bandwidth efficiency by leveraging whatever compute is available at the edge. Where decentralized GPU networks expand who can provide compute, federated learning expands where learning can occur.



Source: **FurtherAfrica**

These approaches show how Africa's AI infrastructure ecosystem is expanding. The continent is starting to explore **hybrid models** that distribute computing across contributors and networks instead of depending only on centralized GPU clusters. These developments are still in their early stages of implementation; however, they show a forward-looking recognition: sustainable AI development in Africa will require both robust centralized infrastructure and decentralized systems that can adapt to local constraints and support broad participation. In this way, decentralized GPU protocols, peer-to-peer compute economies, and federated learning contribute all to the continent's toolkit for creating a resilient, inclusive AI future.

Source: **Forbes Africa**

3.1.1.5 Renewable Energy for Sustainable GPU Infrastructure

As Africa expands its AI and high-performance computing footprint, the question of **energy sustainability** has become more and more central. GPU clusters, data centres, and emerging decentralized compute networks are all highly energy-intensive, which raises concerns not only about cost but also about environmental impact. These challenges catalyzed many initiatives in 2025 that investigate how renewable energy and energy-efficient AI practices can support the next generation of compute infrastructure in Africa.

One notable contribution is the **AI Hub for Sustainable Development,** which positions renewable-powered compute as a foundation for Africa's AI ecosystem (23). Rather than treating energy as a downstream operational constraint, the initiative foregrounds sustainable power as a necessary condition for equitable digital transformation. The Hub's vision links climate resilience, energy access, and AI capacity-building. It seeks to expand access to AI infrastructure and lower carbon intensity by aligning compute resources with renewable energy solutions, especially solar in off-grid or unreliable-grid regions. This framing indicates that Africa cannot simply scale AI infrastructure by replicating energy-intensive models used in other regions. Instead, sustainability must be designed into the infrastructure from the outset.

At the same time, research communities have also begun interrogating the **energy profile of AI workloads themselves.** The study "Green Topics, Deep Roots: Energy-Aware Topic Modelling of Multilingual Nigerian Lyrics" exemplifies this shift (24). By analyzing the energy implications of multilingual topic-modelling tasks, the research highlights the costs associated with computational linguistics, especially in contexts where power is limited or expensive. Notably, the study does not argue for reduced ambition in African Natural Language Processing (NLP) research; rather, it demonstrates that by designing energy-aware approaches,

developers can achieve meaningful AI outcomes while reducing resource consumption. In regions with infrastructural constraints, such efficiency becomes a practical necessity to shape how AI is built, deployed, and maintained.

While these two efforts differ in scope, one focuses on renewable infrastructure and the other on energy-efficient AI methods, they share a common insight: **Africa's AI future must be sustainable not only in deployment, but in design.** Renewable-powered compute hubs expand the physical capacity to train and deploy models, while energy-aware algorithms reduce the burden placed on those systems. Together, they point toward an emerging paradigm in which the continent's AI growth is not limited by energy scarcity, but instead reinforced by innovative, climate-aligned approaches to computation.

Source: **Hanwha Data Centers**

3.1.1.6 Free GPU Providers

While large-scale data centres and commercial GPU services continue to expand across the continent, an equally important development in 2025 was the emergence of **free or zero-cost GPU access** initiatives. These programs cover startup accelerators, pan-African compute networks, and global open-access platforms, and have an important role in preventing students, researchers, and innovators from being excluded from AI progress because of financial or infrastructural constraints.

A prominent example is the **EqualyzAI GPU4Good Initiative,** which provides free GPU access to researchers across Africa (25). GPU4Good was designed specifically to lower infrastructural inequality by enabling academic and socially oriented AI projects that often struggle to secure paid compute.

Source: **EqualyzAI**

By removing cost barriers, the initiative supports a wider range of research topics that might otherwise remain unexplored, including low-resource language technologies, climate modelling, and health analytics, among others. GPU4Good's mission reflects a growing recognition that democratizing access to compute ensures that African AI development is not confined to only well-funded institutions.

Complementing this philanthropic model, the **Africa Compute Fund** maintains a GPU compute network covering more than 20 African countries, with a mandate to support research, innovation, and capacity-building (26). Unlike one-off grants or isolated credit programs, the Fund provides sustained access to distributed GPU infrastructure so as to create continuity in how researchers and startups build and iterate on AI systems. Its focus on geographic reach addresses one of Africa's persistent challenges: the unequal distribution of digital infrastructure across regions. By connecting universities, labs, and innovation hubs to a shared compute backbone, the Africa Compute Fund underpins a more coordinated research environment across the continent.

Source: **Colegiolainmaculada.edu.co**

A more startup-focused pathway is provided through the **Frontier AI Accelerator,** which combines GPU credits with financial grants and mentorship (27). For early-stage African ventures, compute costs often constitute a major barrier to experimentation. By bundling compute access with structured programmatic support, the Accelerator helps teams move from concept to prototype to deployable product more quickly. In contrast to the purely research-oriented initiatives above, Frontier's model directly targets commercial innovation, linking compute access with ecosystem development goals such as market readiness and product scalability.

Finally, everyday academic experimentation continues to be supported by **global free GPU platforms,** most notably **Google Colab** and **GitHub**-integrated environments (28). Although these services give more limited computational capacity than specialized African initiatives, their accessibility makes them indispensable tools for self-taught learners, university students, and researchers. Since these platforms provide instant, browser-based access to GPUs without financial commitment, they represent the entry point for many African learners' first experiences with model training, experimentation, and other coding workflows.

While each initiative addresses a different segment of the ecosystem, they all converge on the same objective: ensuring that talent, not financial privilege, defines who can participate in the continent's AI future. As Africa is trying to reinforce its compute infrastructure, these free access to necessary resources will remain important to cultivate the next generation of entrepreneurs, innovators, and researchers.

3.1.2 Cloud Regions (Hyperscalers)

The increasing demand for AI and cloud-native services in Africa accelerated major developments among global hyperscalers in 2025, each seeking to extend regional presence while addressing critical gaps in cloud infrastructure across the continent. Despite their approaches differ, these initiatives mark a decisive shift: Africa is transitioning from a peripheral market served by distant data centres to an emerging cloud region with increasingly localized capacity.

A landmark milestone came from **Google Cloud launched its first African cloud region** in March 2025, marking a long-awaited turning point for latency-sensitive and data-residency–constrained workloads (29). The new region endows African businesses, governments, and developers with the ability to deploy applications closer to end users, improving many sectors like fintech, e-commerce, and digital public services. Broadly, the launch signals Google's recognition of Africa's expanding digital economy and its need for sovereign cloud capacity capable of supporting AI training, analytics, and enterprise-scale workloads.

Microsoft, meanwhile, advanced its continental strategy on multiple fronts. At **Microsoft Ignite 2025,** the company announced a set of new AI agents alongside updates specific to Africa, emphasizing deeper integration of generative AI tools into Azure's global ecosystem (30). These updates showed the opportunities and operational realities of using AI-driven cloud services in emerging markets. By featuring Africa at one of its flagship global events, Microsoft reinforced the continent's significance as a growth market for scalable AI solutions adapted to infrastructure variability.

Source: **CIO Africa**

However, the momentum was tempered by constraints. Reporting in late 2025 noted that **Microsoft Azure's global infrastructure push had begun to confront capacity limitations,** with Africa having slower expansion compared to other regions (31). While Microsoft continued to invest in regional connectivity and cloud availability zones, the pace of local buildout lagged behind the continent's rapidly increasing demand for compute and storage. Analysts emphasized that without significant acceleration in data-centre deployment, African organizations may continue to face higher latency, limited redundancy, and reliance on regions outside the continent for AI-intensive workloads.

These contrasting advances, Google's entry, Microsoft's innovation surge, and current capacity challenges, indicate a pivotal moment in Africa's cloud trajectory. On one hand, the increased attention of hyperscalers to the region shows the rising digital maturity of the continent, as well as its centrality to future market growth. On the other hand, the unequal rollout of infrastructure highlights the gap between demand for high-performance cloud services and the availability of local capacity to meet it.

Together, the events of 2025 demonstrate that way of Africa to cloud sovereignty will depend not only on hyperscaler expansion, but on sustained investment, diverse regional coverage, and alignment with local regulatory and economic priorities. As the adoption of AI accelerates, the strategic importance of locally anchored cloud regions will only intensify, forming how African organizations develop and use the next generation of AI-based systems.

Source: **Microsoft Azure**

Voice Agents That Serve, Sell, and Resolve in Local Dialects

Handle calls, resolve issues, and close deals through natural voice conversations.

www.equalyz.ai

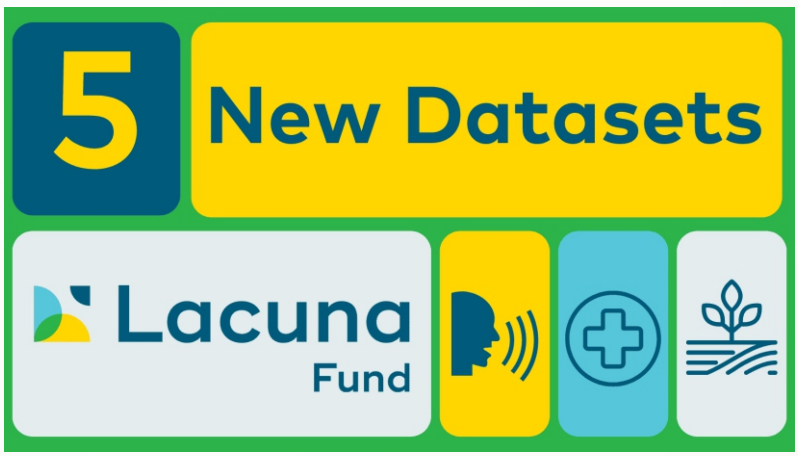
3.2. Data Ecosystem

3.2.1. New Datasets Released in 2025

The year 2025 marked a decisive acceleration in Africa's data ecosystem, reflected in the release of many datasets across regions, languages, and modalities. Despite different in goals, these initiatives share the same ambition: to close long-standing gaps in African language representation and to reinforce the continent's capacity to develop AI systems rooted in local language and cultural realities. Together, they illustrate a continent that is progressively shaping its own data landscape instead of relying on defined resources.

One notable development came through the **Lacuna PII Project,** which released new datasets in **Kanuri, Hausa, and Luganda** aimed specifically at **Personally Identifiable Information (PII) detection** (32). This work reflects a growing recognition that privacy-aligned NLP requires language-specific resources not only for global languages but also for under-resourced ones where PII characteristics may differ. The Lacuna project targets sensitive-data use cases; therefore, it tackles a fundamental requirement for responsible AI in African contexts.
Next to privacy and compliance, **2025 also saw a rise of reasoning-oriented dataset development.** Vambo AI introduced **Fikira 1.0,** a synthetic reasoning dataset spanning **more than 400 Million speakers and 10 African languages:** Amharic, ChiShona, Hausa, Igbo, Kinyarwanda, isiXhosa, isiZulu, Kiswahili, Tunisian Arabic, and Yorùbá (33). As an open-source resource for African language AI research, Fikira represents a major step toward multilingual reasoning evaluation and model alignment, filling a gap left by global corpora that rarely include African languages.

In parallel, **Nadhari AI Lab** introduced the **Swahili Thinking Dataset,** the first open-source **chain-of-thought (CoT) reasoning dataset** designed specifically for Swahili (34,35). Whereas Fikira provides larger multilingual coverage, Swahili Thinking focuses on native reasoning traces within a single high-resource African language, which enables LLMs to learn more authentic reasoning patterns rather than relying on translated or synthetic proxies. Together, these two initiatives, Fikira 1.0 and Swahili Thinking Dataset, advance the landscape of African reasoning datasets, one through linguistic breadth and the other through linguistic depth.
Mozilla continued its long-term investment in African language technologies through the **NaijaVoices project,** which released a new **Nupe dataset** to support speech and language research (36). As with earlier contributions in Hausa, Yoruba, and Igbo, this release underscores Mozilla's emphasis on community-led data creation and the cultivation of culturally grounded, high-quality resources. Nupe, spoken by millions but overlooked in NLP, now becomes more accessible for model training, evaluation, and fine-tuning.



Source: **Lacuna Fund**

Several 2025 releases expanded the continental footprint of speech datasets. For example, **IntronHealth's Afrispeech Dialog dataset** (37) introduced new multimodal speech resources to support real clinical and conversational applications, deepening the pool of healthcare-relevant linguistic data. Meanwhile, academic work in Ghana advanced the creation of **speech recognition resources for Ghanaian languages** (38), filling critical gaps for languages that have lacked sufficient annotated audio data for automatic speech recognition (ASR) model development.

As dataset creation accelerated, **discoverability itself became an essential enabler of ecosystem growth.** LanAfrica addressed this need by releasing **Version 2** of its continent-wide resource discovery platform (39). The platform consolidates Africa's expanding landscape of datasets, research papers, models, and tools into a single connected hub, equipped with an upgraded search system that makes finding resources faster and more comprehensive. In order to ensure that researchers and community contributors can navigate Africa's data assets in their native languages, LanAfrica introduced multilingual platform support that is currently available in English and French with African languages being developed.

Source: **International Development Research Centre**

Multi-language initiatives also gathered momentum. The **African Next Voices Project,** launched by the **Bill & Melinda Gates Foundation,** focused on enabling AI systems across African languages, reflecting a larger, donor-driven interest in accelerating the inclusion of Africa in global AI workflows (40). Similarly, the **African_UD project** expanded universal dependency treebank coverage for African languages (41), reinforcing the syntactic resources necessary for high-quality NLP modeling. These efforts complement each other: where African_UD targets structural language annotation, Next Voices emphasizes scalable data pipelines for low-resource languages.

In addition to these pipeline-oriented projects, 2025 also introduced Africa's first large-scale safety-alignment dataset: the **WildJailbreak Africa dataset,** created for jailbreak and instruction-safety tuning across **Acholi, Lugbara, Luganda, Swahili, and Ateso** (42). With roughly 300K translated samples, it fills a critical gap in safety-focused African-language resources. As the use of LLM increases across public and commercial domains, datasets like WildJailbreak Africa will shape more robust and culturally attuned guardrails.

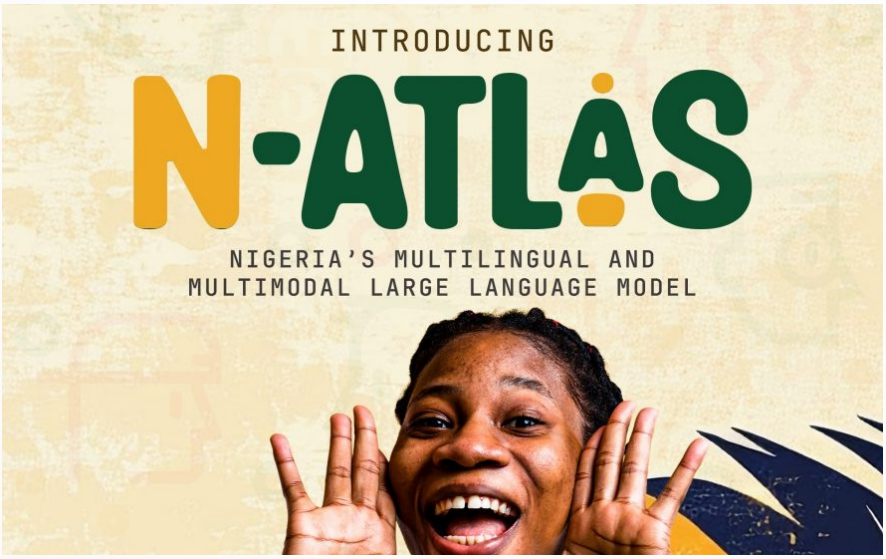
Local language inclusion also advanced in East Africa, with the **Lacuna Fund** supporting the creation of datasets for **low-resource Kenyan languages**, responding to long-standing calls to counter linguistic exclusion in computational systems (43). By focusing on languages with minimal digital representation, the project aligns with a growing movement to diversify datasets beyond the most widely spoken tongues.

In the same vein, **Translators without Borders (CLEAR Global)** released a Hausa TTS corpus sample (TWB-voice-TTS-Hausa-1.0-sampleset), which is part of its broader TWB Voice data-creation model, where community contributors record speech, transcribe, and rate outputs to generate paired voice–text datasets that can be used to build practical language technologies like ASR and TTS for marginalized languages, with an emphasis on reducing language barriers in low-resource settings.

Corporate and open-source collaborations also featured prominently. **Awarri,** in partnership with **N-ATLAS**, backed the development of African-language datasets to support AI systems deployed at scale across the continent (44). Later in 2025, Nigeria announced the official release of the **open-source N-ATLAS model,** presented at **UNGA80**, reinforcing national efforts to champion African languages in global AI research (45). These releases signal increasing corporate recognition that language technology is becoming a strategic asset in the digital economy of Africa.

The community-driven movement remained strong as well. The **NaijaVoices** community released a major new dataset focused on large-scale, culturally rich African speech data, created with extensive volunteer participation (46). Meanwhile, the **Deep Learning Indaba 2025** amplified this momentum by issuing a continent-wide call for African datasets, encouraging researchers, civil society groups, and language communities to contribute to the growing corpus of public resources (47). The Indaba's call reflects a shift from passive consumption to active creation of linguistic assets that mobilizes Africa's research community to impact global NLP development.

The dataset releases of 2025 show how Africa's data ecosystem is expanding and developing. Across low-resource languages, privacy-sensitive domains, speech and text modalities, and national-scale model initiatives, these efforts all reinforce the basis on which African AI capabilities will grow. Above all, they highlight an emergent principle: **Africa's AI future must be built on African data that is reflective of its people, its languages, and its cultural contexts.**



Source: **The Federal Ministry of Commmunications, Innovation, and Digital Economy**

3.2.2 Datasets Release by Countries

Following the surge of pan-African dataset releases in 2025, countries across the continent also accelerated their own contributions to the data ecosystem. These national releases were motivated by many factors, ranging from enhancing representation in global benchmarks to correcting racial and linguistic biases in AI systems.

NORTH AFRICA

Algeria

Algeria contributed datasets across both text and image modalities, each targeting long-standing gaps in AI representation.

The **AFRIHATE Algerian Arabic subset** introduced annotated tweets for hate-speech detection in local dialects (48). This dataset does not only reflect the increasing involvement of Algeria in global conversations around online safety, but also addresses the linguistic complexity of dialectal Arabic, which is an area historically underserved by major NLP resources.

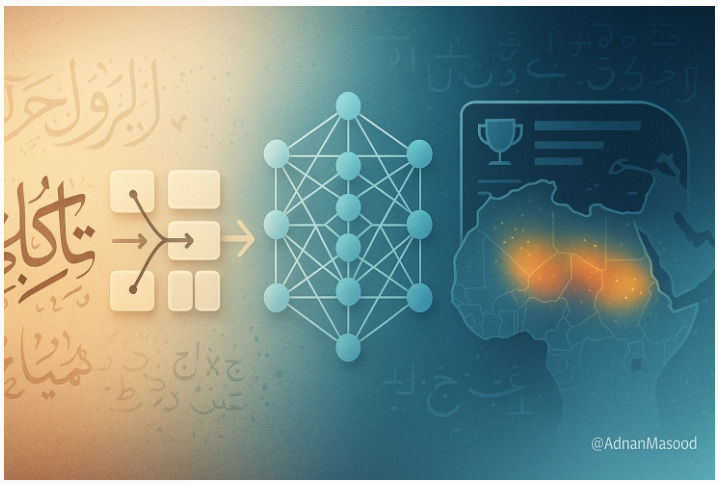
In parallel, Algeria prominently featured in the **FAGE_v2 facial dataset**, which includes **500 images of indigenous Algerian faces** to improve age-invariant facial recognition (49). By countering Western-centric bias in facial models, this dataset responds to a practical challenge: facial-recognition systems often underperform on North African populations due to sparse training data.

Egypt

Egypt likewise appeared in **FAGE_v2,** providing an additional regional subset that strengthens representation of North African phenotypes in facial datasets (49).
Beyond vision, Egypt contributed to regional NLP advancement through the **Arabic MMLU benchmark updates,** led by Intella and MENA NLP groups (50). These updates expanded evaluation datasets for Arabic, which is particularly important for a language spoken widely across North Africa but underrepresented in multi-task benchmarks.

Morocco

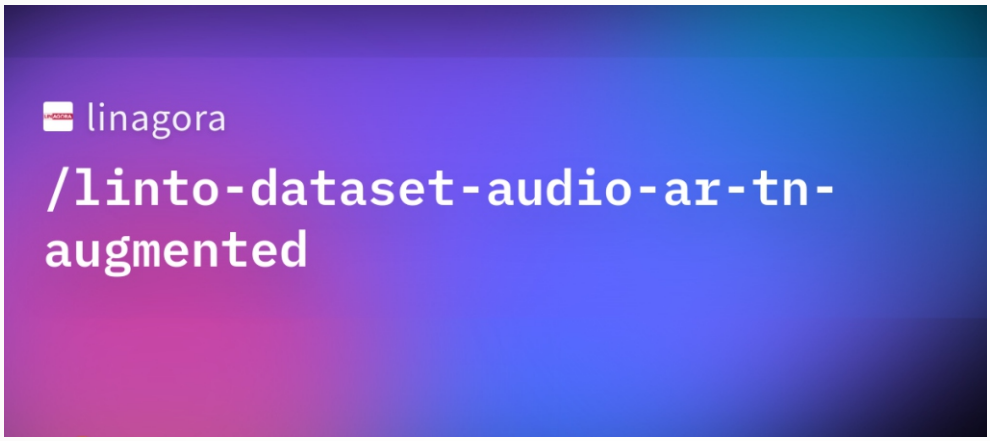
Morocco's primary 2025 contribution came through the **AFRIHATE Darija subset,** tailored for abusive-language detection in Moroccan Darija (48). Given Darija's distinct linguistic structure and divergence from Modern Standard Arabic, country-specific datasets such as this one play an essential role in enabling safe and culturally appropriate AI applications.



Source: **Medium**

Tunisia

Tunisia extended its speech resources by releasing **LinTO audio and textual datasets,** which notably increased available hours of Tunisian Arabic audio in 2025 (51). These additions mark important progress for ASR research in a country where speech recognition models historically relied on limited or non-representative



Source: **Hugging Face**

WEST AFRICA

Nigeria

Nigeria's data ecosystem rapidly grew across many domains like medical imaging, speech, and text, which reflects its position as one of Africa's most active AI development hubs.

In healthcare AI, the release of the **BraTS-Africa MRI dataset,** containing annotated brain tumour scans from 146 Nigerian patients, provided the first population-specific medical-imaging dataset of its kind (52). Its goal is explicit: retrain global models that demonstrate bias when applied to African scans.

Nigeria also saw major expansions in speech datasets.
The NaijaVoices dataset contributed **1,800 hours of speech collected from more than 5,000 speakers across Nupe, Yoruba, Pidgin, and Hausa.**

AfricanVoices.io released

3,000 hours

of high-quality Hausa, Igbo, Nigerian Pidgin, and Yoruba audio

Which further expanded the linguistic coverage for speech-enabled applications (54).
On the text side, Nigeria also features in AFRIHATE, which includes annotated hate-speech subsets in Hausa, Igbo, Pidgin, and Yoruba, important languages in online discourse (48).

Ghana

Ghana contributed datasets across text, image, and video modalities.

The Twi Emotions Corpus, including

400,000

sentiment-labeled sentences, is one of the largest emotion-classification datasets for a Ghanaian language

Ghana also appears in the **FAGE_v2 dataset,** adding local facial-image diversity critical for bias mitigation in vision models (49). Rounding out Ghana's 2025 releases, the **GSL Lexicon** expanded resources for **Ghanaian Sign Language,** supporting an often-overlooked modality in African dataset development (56).

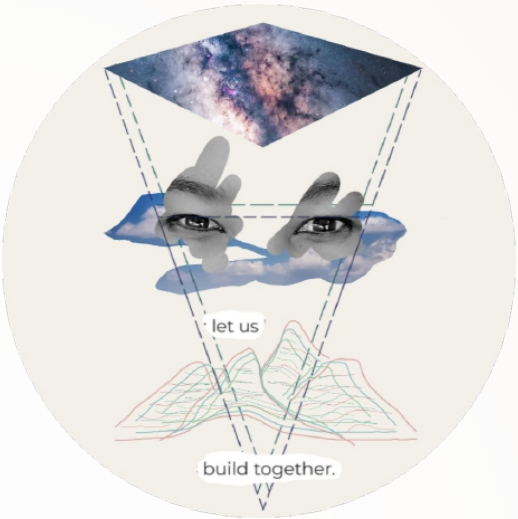
Source: **Kermal Erdem**

Senegal

Senegal released **WolBanking77,** a Wolof-language dataset for intent classification in banking applications (57). Its significance extends beyond the dataset itself: Wolof is the lingua franca for a large share of Senegal's population, yet it remains underrepresented in commercial NLP tools. As Senegal's digital-payments sector grows, resources like this support more accessible, local-language financial services. WolBanking77 therefore addresses a critical need: enabling financial-sector AI systems to operate in everyday language rather than relying on French-centric models that can exclude or misinterpret local users.

Benin & Togo

Both countries were represented in the 2025 update of **MasakhaPOS,** which added new part-of-speech (POS) tagging benchmarks for **Fon (Benin)** and **Ewe (Togo)** (58). These benchmarks are notable because they don't only expand linguistic coverage, but also because Fon and Ewe are widely spoken, socially fundamental languages in their respective countries, but remain almost entirely absent from global NLP datasets. These contributions are also consistent with the overarching mission of the Masakhane community: advancing high-quality NLP tools for African languages that remain underserved in commercial datasets.



Source: **GitHub**

EAST AFRICA

Kenya

Kenya continued to demonstrate leadership in multilingual dataset creation and socially grounded data resources. Under the **African Next Voices project,** new speech datasets for **Kikuyu and Dholuo** were released, extending coverage to two of the country's major indigenous languages (59). This expansion is critical in a country where linguistic diversity plays a central role in education, media, and civic life; voice-enabled systems, for instance, cannot be equitable without coverage of commonly spoken local languages.

Kenya is also represented within **FAGE_v2**, which includes a dedicated **Kenya facial-image subset** (49). As AI vision models have repeatedly been shown to underperform on African phenotypes, the addition of Kenyan faces contributes to correcting long-standing demographic imbalances in global facial recognition datasets.

Beyond language and vision, Kenya also saw contributions to **humanitarian data infrastructure,** with updated datasets on refugees and asylum seekers hosted through UNHCR (60). These resources support operational planning and social-protection policies in a country that hosts large displaced populations. They also underscore an important dimension of East Africa's data landscape: the growing use of AI-ready datasets to inform humanitarian and development decision-making, not only commercial or academic applications.

Rwanda

Rwanda made significant strides in speech and text datasets. The **Kinyarwanda ASR Track A** initiative released more than 500 hours of transcribed Kinyarwanda speech, which provides one of the largest ASR corpora for any African language to date (61). Such depth is essential to develop reliable voice interfaces and education tools in Rwanda, where Kinyarwanda is the national lingua franca.

Rwanda also contributed to the **AFRIHATE dataset,** with a dedicated Kinyarwanda subset for harmful-content detection (48). As digital participation expands across the region, the availability of culturally and linguistically grounded moderation datasets becomes increasingly important for online safety and platform trust.

Ethiopia

Ethiopia's dataset releases captured both linguistic and visual domains. The country contributed many text subsets to **AFRIHATE,** covering **Amharic, Oromo, and Tigrinya,** three of its most widely spoken languages (48). These additions are noteworthy for their linguistic coverage and also their applicability to Ethiopia's multilingual digital sphere, where the lack of reliable local-language datasets limited NLP research.

Ethiopia is also featured in **FAGE_v2,** which includes an Ethiopia facial-image subset (49). Representation in facial datasets is important in Ethiopia, one of the most ethnically diverse countries on the continent, since it helps reduce model error rates in vision-based identity verification or health-screening tools.

Source: **Hugging Face**

Uganda

Uganda saw advancements in text resources through expansions to **MakerereNLP translation datasets** in **Luganda** and **Runyankore** (62). These updates reinforce Luganda–English and other parallel corpora and thereby support machine translation, digital content creation, and language-learning tools. Since Luganda is one of the most commonly spoken languages in Uganda, better translation datasets make digital public services and community applications more accessible.

Source: **Waywithwords.net**

Tanzania

Tanzania contributed to the AFRIHATE project by releasing a **Swahili hate-speech detection subset** (48). Since Swahili is both a national language and a lingua franca across East Africa, high-quality text resources are essential for content moderation, public-sector communication, and cross-border digital applications.

Somalia

Somalia added two different dataset types that reflect both digital and socio-economic priorities.

First, its **Somali AFRIHATE** subset supports harmful-content detection in Somali, a language with limited prior NLP resources. This helps address the region's growing need for language-appropriate online governance tools.

Second, Somalia contributed to development and policy analytics with the **Ultra-Poor Graduation Midline Survey dataset,** which captures socio-economic indicators for vulnerable populations (63). These datasets inform poverty-alleviation strategies, program design, and humanitarian evaluations, showing that East African data ecosystems are increasingly incorporating linguistic and socio-economic dimensions of AI relevance.

SOUTHERN AFRICA

South Africa

South Africa continued to play a leading role in African dataset development by contributing both linguistic and visual resources.

The **Mafoko Multilingual Dataset** introduced a rich terminology corpus by covering all **11 official South African languages** (64). This dataset addresses a critical challenge in multilingual NLP: terminological inconsistency across domains like healthcare, law, and public administration. By standardizing terminology across the country's diverse linguistic landscape, Mafoko promotes more accurate translation systems and language technologies used in education, government, and digital public services.

In addition to terminology resources, South Africa contributed to code-switching research through the **AfroCS-xs dataset,** which includes English–Zulu and English–Sesotho code-switched text (65). Code-switching is deeply embedded in everyday South African communication but underrepresented in commercial NLP datasets. AfroCS-xs therefore fills a crucial gap and supports ASR, sentiment analysis, and conversational AI systems that must operate in multilingual and mixed-language environments.

South Africa is also represented within the **FAGE_v2 facial-image dataset,** which includes a specific **South Africa subset** (49). As global face-recognition models often underperform on African populations, this dataset provides essential demographic diversity to help reduce bias in identity verification, biometrics, and vision-based diagnostic systems.

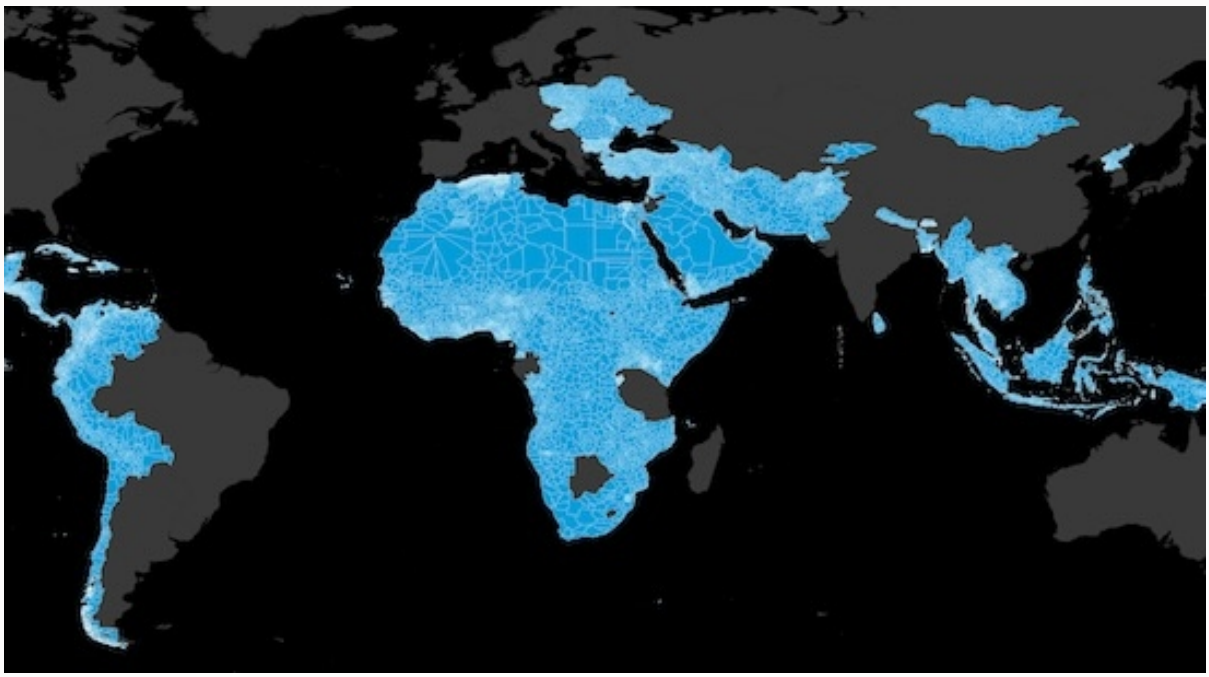
Source: **ImageFX**

Namibia

Namibia also appears in **FAGE_v2** with its own country-specific facial-image subset (49). Despite having a smaller population than South Africa, Namibia's representation contributes to much-needed diversity in facial datasets, especially for Southern African ethnic groups that have been omitted from global vision benchmarks. This representation is essential to boost AI systems used in digital-ID programs, border control, and civic registration.

Zimbabwe, Botswana & Mozambique

In these countries, contributions in 2025 were primarily motivated by humanitarian and development needs. Updated datasets on population density and displacement, hosted through the Humanitarian Data Exchange (HDX), provide key inputs for early-warning systems, crisis modeling, and resource allocation (66). As droughts, climate shocks, and economic disruptions continue to impact mobility patterns in Southern Africa, the availability of humanitarian data becomes important not only for aid delivery but also for AI models used in resilience planning and forecasting.



Source: **Humanitarian Data Exchange**

CENTRAL AFRICA

Democratic Republic of the Congo (DRC)

The DRC appeared in both visual and humanitarian datasets. A **DRC-specific facial-image subset** within **FAGE_v2** (49) improves representation for one of Africa's most diverse populations, supporting more accurate vision models in a region where biometric systems are increasingly used in public-service delivery. In addition, updated **UNHCR displacement data** for eastern DRC (December 2025) (67) provides high-granularity insights into population movements resulting from conflict. These datasets are essential for crisis modeling and reflect the growing role of AI in peacekeeping support, humanitarian logistics, and risk assessment.

Angola

Angola contributed its own **FAGE_v2 facial-image subset** (49), adding visual diversity from Central–Southern Africa. In contexts where identity verification and biometric systems are expanding, such demographic inclusion is essential to reduce misclassification rates and ensure that downstream applications are equitable across cultural groups.

Source: **ImageFX**

Cameroon

While Cameroon did not have a dedicated language or speech dataset release in 2025, it was included in the **West Africa Land Use/Land Cover (LULC)** geospatial dataset (68). Although its name emphasizes West Africa, this time-series dataset extends into Cameroon's territory to provide key environmental inputs for climate modeling, agriculture, and land-use planning. Such geospatial datasets often represent the foundational layers for AI-based climate resilience and natural-resource management tools across Central Africa.

Central African Republic (CAR) & Chad

Both countries contributed by updating **UNHCR Regional Response datasets** released in November 2025 (60). These datasets contain information on protection, displacement, and humanitarian assistance for communities affected by instability and conflict. In settings where timely data is often scarce, they provide valuable real-world signals that support AI systems used in early-warning analysis, crisis mapping, and humanitarian planning.

equalyzAI

Content That Speaks Every Dialect, Not Just Language.

Generate authentic local voiceovers for adverts, media contents and learning videos

www.equalyz.ai

3.2.3 African Languages Hackathons

Following the surge of new datasets across the continent, 2025 also saw a parallel rise in **hackathons and community-driven challenges** devoted to African language technologies. While different in sponsor, format, and target audience, these initiatives share a common goal: to mobilize Africa's expanding pool of students, researchers, engineers, and community linguists to develop data resources and tools that reflect their own linguistic and cultural contexts. augmentation.

Source: **Zindi**

Following the surge of new datasets across the continent, 2025 also saw a parallel rise in **hackathons and community-driven challenges** devoted to African language technologies. While different in sponsor, format, and target audience, these initiatives share a common goal: to mobilize Africa's expanding pool of students, researchers, engineers, and community linguists to develop data resources and tools that reflect their own linguistic and cultural contexts.

Zindi, which is one of Africa's most prominent data science platforms, played a very notable role. Its **Speech Dataset Catalogue** (69) operates as a curated starting point for learners and practitioners looking to be involved in African-language speech datasets. By highlighting locally relevant corpora and guiding users toward suitable resources, the catalogue helped improve participation in speech technology beyond formal research labs. In this context, Zindi's **Your Voice Challenge** invited participants to create speech datasets using their own devices and their own languages (70). This challenge underscored an important insight: in multilingual societies where many languages lack formal digital corpora, ordinary citizens can help create datasets.

Google also intensified its involvement through the **Introduction to ASR: African Challenge,** hosted on the Zindi platform (71). This beginner-friendly competition aimed to demystify ASR for African learners and expand language-specific data resources. By lowering the entry barrier for ASR experimentation and placing African languages at the heart of the challenge, the initiative played a dual role: capacity building and dataset augmentation.

Source: **Zindi**

Beyond competitions, 2025 saw increased institutional support for African speech data creation through Mozilla's **Open Multilingual Speech Fund** (72). Implemented through in-country programs tied to Common Voice, this initiative endorses community-led speech collection across African languages. The Fund prioritizes languages that have lacked representation in mainstream ASR datasets, which ensures that the voices of local communities shape the development of speech technology.

At the same time, the **Masakhane African Languages Hub** launched a call for proposals to support African language data projects (73). As one of the continent's leading grassroots NLP research communities, Masakhane's RFP encouraged open-source and collaborative dataset creation aligned with its mission: "for Africans, by Africans." This move reflects a shift from sporadic dataset releases to coordinated, community-driven research pipelines.

Multilingual data creation also extended into the open knowledge domain through the **WikiIndaba 2025 Hackathon** (74). The event was hosted by the African Wikimedia community, and it succeeded in mobilizing contributors to improve language resources, translations, and data across Wikipedia and other platforms. For many African languages, Wikimedia projects are among the largest publicly available text sources; improving them has direct benefits for NLP training and evaluation.

Finally, the **Nigeria National AI Hackathon 2025** (75) demonstrated the increasing national-level interest in African language AI. The hackathon included tracks that emphasized language technologies, encouraging teams to refine or create datasets that capture Nigeria's linguistic diversity. By anchoring the event in a national strategy, Nigeria demonstrated its intention to develop domestic talent pipelines capable of producing language AI tailored to local needs.

3.3 Large Language Models

The year 2025 marked a turning point in the **training and fine-tuning of large language models (LLMs)** for African languages. Across African startups, telecom operators, and research institutions, a shared urgency emerged: to make sure that the next generation of AI systems can understand and interact in the languages spoken across the continent.

One of the year's most significant global developments was the release of **Meta's Omnilingual ASR,** a model capable of capturing 1600 Languages including more than 500 low-resource languages for which no transcripts previously existed (76). At this scale, the model represents a shift in speech technology: languages once entirely excluded from digital ecosystems can now benefit from baseline ASR support.

1600 Languages

Source: **Simplilearn**

Omnilingual ASR provides a foundation for researchers and developers, on which regionally fine-tuned models can be built, even if the model itself is not yet optimized for African dialectal variation or code-switching, which are common characteristics in African communication patterns.

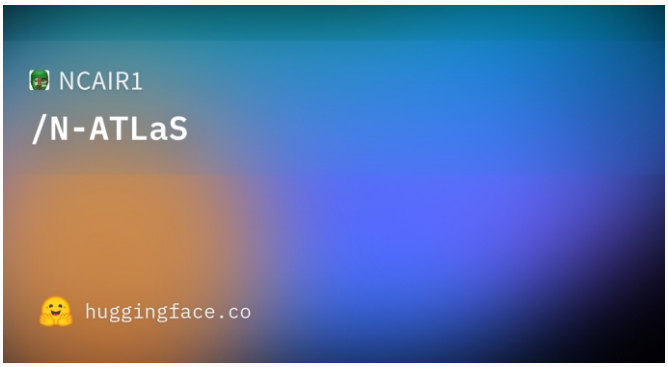
Building on this movement toward large-scale multilingual inclusion, the AfriqueLLM suite introduced a complementary advance: a set of open LLMs adapted to **20 African languages** through continued pre-training (CPT) on **26B tokens,** using base models such as Llama 3.1 (8B), Gemma 3 (4B & 12B), and Qwen 3 (8B & 14B) (77). AfriqueLLM's empirical study revealed an important insight for African LLM development: **base model capability often outweighs pre-existing multilingual coverage,** with AfriqueQwen-14B outperforming larger models such as Gemma 3 27B despite being less than half the size.

In parallel to these multilingual efforts, several African-led companies improved **speech-first models** rooted in local linguistic realities. **Khaya AI,** for example, enhanced its speech-recognition technology so as to support **32 African languages** (78,79), addressing tonal languages, dialectal shifts, and multilingual code-switching. The model's focus on "accurate language understanding and efficient digital communication" makes it an important player in sectors where voice interfaces can enhance accessibility, including education, customer support, fintech, and public services. Similarly, **Spitch** continued to develop **speech-to-text and text-to-speech solutions** tuned specifically to African languages and accents (80). Their work shows a common theme across the continent: AI speech systems must bridge not only linguistic diversity but also acoustic diversity, from rural soundscapes to urban call-center environments.

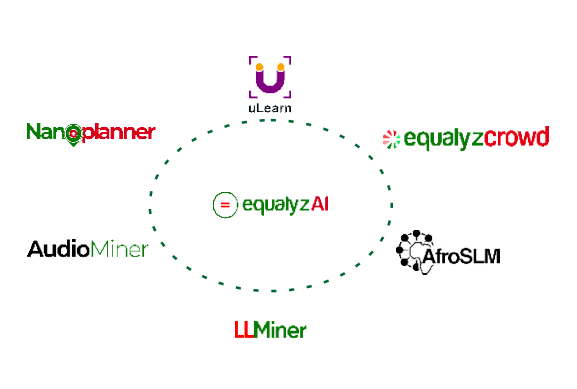
Yet speech technology alone cannot serve Africa's most common communication styles, especially code-switching and informal text. The study in (81) demonstrated how multilingual ASR models can be fine-tuned to understand Yoruba-English mixtures, which is an important breakthrough for aligning speech technology with daily communication patterns across West Africa.

In parallel, **AfroXLMR-Social** addressed a critical gap in African NLP by adapting pretrained models to **social-media text** using domain-adaptive pre-training (DAPT) and task-adaptive pre-training (TAPT) (82). By introducing the **AfriSocial** dataset and training models across **19 African languages,** researchers demonstrated substantial performance gains in sentiment analysis, emotion detection, and hate-speech identification—tasks where traditional models fail due to over-reliance on religious or formal text sources. This work highlights a broader trend: African NLP must reflect where Africans actually communicate, not just where training data has historically existed.

This momentum also translated into national and enterprise-driven language models. In Nigeria, **Awarri,** in collaboration with the **National Centre for Artificial Intelligence and Robotics (NCAIR),** released N-ATLAS, an African, open-source language model unveiled at UNGA80 in September 2025 (45). **N-ATLAS** is a push toward linguistic sovereignty that encourages African nations to form the development and governance of their own language models rather than relying only on foreign technologies.



Source: **Hugging Face**



Source: **EqualyzAI**

In the fintech sector, **EqualyzAI launched AfroSLM 1.0,** a multimodal AI model supporting **three African languages within financial domains** (83). AfroSLM was developed for applications such as compliance automation, customer support, and fraud-prevention workflows, and it reflects how African fintech ecosystems need domain-specific language tools that understand both linguistic nuance and financial terminology.

Telecoms and enterprise AI companies also contributed their own specialization layers. **Orange** announced a collaboration with **OpenAI and Meta** to **adapt and fine-tune LLMs for African regional languages** that existing generative models fail to support (84). This initiative is particularly significant given the presence of Orange across Francophone Africa; many regional dialects such as Central African French, Senegalese Wolof-inflected French, or North African Darija-influenced variants are poorly served by mainstream LLMs. Orange's involvement demonstrates that language inclusion is becoming a priority for telecom operators to expand digital access, customer engagement, and AI-enabled services.

Within the MENA and North African region, **WideBot** continued developing **AQL Mind,** an Arabic-centric LLM tuned for dialect-rich enterprise contexts (85). While primarily targeting Arabic-speaking markets, AQL Mind's approach resonates with challenges seen across Africa: a single language label (e.g., "Arabic") often masks variation that can profoundly affect AI performance. Approaches like WideBot's highlight the importance of dialect-sensitive fine-tuning for both enterprise and consumer use cases.

Meanwhile, domain-specific fine-tuning expanded further with Crane AI Labs that released Swahili Gemma 1B, a fine-tuned Gemma 3 instruction model specialized for **English-to-Swahili translation and Swahili conversational AI,** outputting in Swahili (86). This model sits within a growing class of compact, high-utility LLMs tailored to linguistic communities that are underserved by mainstream multilingual LLMs.

Community technologists and individual developers continued to broaden the ecosystem from below. **CDIAL's Indigenius** model represents one of the more ambitious grassroots contributions, a multilingual, speech-first model fine-tuned on a **proprietary 3-billion-parameter base** to support low-resource languages (87). Indigenius reflects a shift toward models that prioritize speech as a primary interface, indicating that many African languages are readily more spoken than written.

At the individual level, a University of Lagos graduate launched **YarnGPT** that supports **Nigerian English, Yoruba, Igbo, and Hausa,** which illustrates the expanding accessibility of LLM training and fine-tuning (88). The model shows how students and independent developers can now contribute models designed for specific linguistic and cultural contexts. Similarly, **SabiYarn 125M** expanded lightweight generative modeling for Nigerian languages through a compact architecture that excels at text generation and downstream fine-tuning while remaining resource-efficient (89).

All these efforts reinforce a central theme of this white paper: **Africa is no longer simply adapting global AI models, but it is actively shaping them** by developing language models that capture its voices, dialects, and lived experiences.

3.4 African NLP/LLM Publications

Source: **EqualyzAI**

Following major advances in dataset creation and model adaptation across the continent, 2025 also delivered a remarkable body of **peer-reviewed research, technical whitepapers, and open-source model releases** that together signal the maturation of Africa's NLP and LLM research ecosystem.

A key conceptual contribution came from **EqualyzAI,** whose whitepaper on Small Language Models (SLMs) for the Global South (90) articulated a strategic case for resource-efficient architectures designed for low-compute environments. In regions where bandwidth, hardware, and energy constraints limit adoption of frontier LLMs, SLMs provide a more practical pathway for deploying AI in education, finance, health, and government services. The paper reflects a growing recognition across African AI communities: model size must be balanced with accessibility, cultural alignment, and operational constraints.

Another important contribution, **Masakhane's publication on African NLP datasets** (91), provided a comprehensive mapping of available resources for building African AI. While the report is descriptive, its influence is strategic: by illuminating the fragmented state of existing datasets, Masakhane highlights the need for continued investment in open, shared, and community-maintained data infrastructure. For researchers and policymakers, the publication also serves as a diagnostic tool—revealing coverage gaps by language, modality, and region, and reinforcing the principle that sustainable African NLP must be grounded in African-created and African-owned data.

CLEAR Global (Translators without Borders) released a playbook for low-resource language data collection through its TWB Voice work, centered on community-led, ethical, and use-case-driven data creation. The approach emphasizes defining practical goals upfront, recruiting native speakers as contributors, and combining recording, transcription, and peer review to ensure quality and represent real speech diversity such as accents, dialects, and informal usage. Strong attention is given to informed consent, contributor rights, and responsible data sharing, with guidance on packaging and releasing datasets, often via platforms like Hugging Face, to support scalable ASR and TTS development for marginalized languages.

Similarly, **AfroCS-xs,** a benchmark for code-switched English–Zulu and English–Sesotho text (65), pushed forward a crucial frontier in African NLP: modeling multilingual environments where speakers blend languages dynamically. As code-switching remains deeply embedded in cultural and social communication across Southern Africa, this work clarifies the limitations of monolingual-only models and highlights the need for linguistically grounded evaluation benchmarks.

Lelapa AI's **InkubaLM compression breakthrough** (92) further emphasized the importance of efficiency, demonstrating how model compression techniques can drastically reduce computational requirements while retaining competitive performance, a pivotal advancement for African

deployment contexts. Viewed against the backdrop of limited GPU availability and high cloud costs, InkubaLM's work underscores a broader design philosophy gaining traction across the continent: African AI requires models that are powerful enough to matter, but lean enough to run where they are needed most.

A significant share of 2025 publications focused on **reinforcing support for individual low-resource languages** to address the acute scarcity of digitally aligned linguistic resources. **Yankari,** a monolingual Yoruba dataset introduced by Maro Akpobi (93), expanded high-quality training material for one of West Africa's most widely spoken languages, supporting improved downstream tasks such as MT, QA, and sentiment analysis.

The work on **Amharic text complexity classification** (94) similarly tackled a practical gap, not simply how to process Amharic text, but how to evaluate reading difficulty, which has implications for education, government communication, and public information systems in Ethiopia. By introducing a supervised tool for complexity estimation, the study highlights a key insight: NLP for African languages must often solve pedagogical and civic communication challenges, not just standard machine-learning tasks.

Two publications explored languages with even fewer digital resources. **Kpelle machine translation corpus** (95) introduced one of the first functional MT datasets for Liberia's Kpelle language and **Medumba resource-gathering challenges** (96) documented the socio-technical barriers, such as speaker dispersion, annotation scarcity, and lack of standardized orthography, that continue to impede corpus creation for languages spoken by smaller groups in Cameroon. In parallel, **YodiV3** was introduced to support Togolese languages Ewe and Mina through the Eyaa-Tom dataset (97), which reinforces NLP tooling in a region where linguistic diversity is rich but underrepresented in computational contexts. By incorporating the Lom metric, the work pushes beyond data collection toward more contextually grounded evaluation frameworks, signaling how research is beginning to develop language-appropriate methods rather than applying generic metrics developed for high-resource languages.

Another major cluster of publications addressed **speech recognition in low-resource settings,** a domain critical for expanding AI access across Africa's predominantly oral linguistic cultures. The study on tolerance of repetition in Swahili ASR (98) explored how speech-recognition accuracy degrades under repeated phrasing, an important real-world phenomenon in conversational Swahili. In addition, the paper on **Fine-tuning Whisper Tiny for Swahili ASR** (98) provided pragmatic recommendations for adapting lightweight models in resource-constrained scenarios. The authors highlighted challenges such as inconsistent orthography and limited annotated audio but demonstrated that careful fine-tuning can result in meaningful performance gains.

In Ethiopia, **exploring transliteration-based zero-shot transfer for Amharic ASR** (99) evaluated methods that bypass data scarcity by leveraging related scripts. Since Amharic shares structural and phonological parallels with other Ethiopic-script languages, transliteration opens pathways for transferring knowledge from better-resourced languages or from text corpora into ASR models. At a broader systems level, the meta-analysis on **Automatic Speech Recognition for African Low-Resource Languages** (100) synthesized challenges and

Source: **Cohere**

emerging opportunities across multiple languages. The paper highlighted pervasive structural barriers—including severe underrepresentation in global datasets, inconsistent writing systems, dialectal variation, and the absence of standardized orthographies for many languages.

Shifting from core modeling to domain applications, the applied NLP papers of 2025 demonstrated the growing alignment between African AI research and sectoral priorities—especially in health, agriculture, and financial safety. The work on **efficient RAG agents for Kenyan farming advisory** (101) explored how quantized language models can power agricultural decision support, addressing localized challenges such as limited connectivity and device constraints. Similarly, research on **drug-information access through local-language RAG systems** (102) focused on comprehension and health literacy, showing how multilingual RAG tools can reduce misinterpretation of medication guidelines. Finally, the study on **multilingual NLP for African healthcare** (103) documented bias, translation errors, and explainability issues in clinical NLP systems, reinforcing the need for careful model evaluation before deployment in sensitive domains.

In parallel, a substantial group of papers engaged with **evaluation, safety, and robustness,** themes increasingly important as LLMs begin serving African users. The study (104) investigated the detectability of machine-generated Hausa text, which represents a key issue for combating misinformation in one of West Africa's most commonly spoken languages. **Culturally nuanced LLM evaluation** (105) challenged standard benchmarks by emphasizing scenarios where cultural cues and contextual meaning heavily influence model correctness. Further extending this line of inquiry, research on the **Robustness of Yorùbá QA models to typographical noise** (106) examined error tolerance in environments where orthographic variation is common. In the education domain, the paper on **document VQA for African standardized exams** (107) evaluated multimodal LLMs in an educational context, which revealed where current global models fail to understand region-specific exam formats. Finally, the review on **HausaNLP status** (108) provided a landscape review highlighting systemic barriers in Hausa NLP development.

Alongside these evaluation-centered studies, several publications explored machine translation (MT) and cross-lingual transfer, emphasizing resource scarcity and linguistic diversity. Research on **pretraining strategies for low-resource MT** (109) compared monolingual vs. parallel data approaches, identifying trade-offs relevant to many African languages. Expanding beyond translation into reasoning, **English–Yorùbá open-book comprehension (Y-NQ)** (110) introduced an evaluation dataset targeting high-level reasoning, not just translation accuracy. At the model-optimization level, work on **multilingual translation using multi-armed bandits** (111) explored adaptive optimization strategies for LLM-based MT systems across many languages.

Extending beyond Sub-Saharan Africa, the MENA region contributed notable Arabic-centric LLM research. **Command R7B Arabic** (112) introduced a small, enterprise-focused multilingual model tailored to culturally rich and dialectally complex Arabic contexts. This work reflects a broader regional trend: prioritizing models that can navigate dialect variation while meeting enterprise performance and deployment constraints.

These contributions demonstrate an intellectual shift: African researchers are not only filling resource gaps, but they are shaping methodological innovation, evaluation standards, and model-design principles that better reflect the realities of African languages and societies.

3.5 Leading Research Institutions-driven Initiatives

As Africa's AI ecosystem expanded in 2025, research institutions across the continent increasingly stepped into leadership roles, which not only shape talent but also the strategic direction of locally relevant AI research. These initiatives share a common purpose: to build institutional capacity that is strong enough to sustain Africa's long-term participation in global AI innovation.

A landmark example came from the **University of Lagos (UNILAG),** which in October 2025 hosted **Africa's first OpenAI Academy** (113). Designed as an intensive training program, the academy brought together students and researchers to work directly with tools and curricula shaped by OpenAI. Beyond technical upskilling, the academy represented something more significant: a recognition by global AI leaders that African universities are not merely training sites but strategic partners in shaping the continent's AI trajectory. For Nigeria's academic sector, the collaboration reinforced UNILAG's growing position as a hub for AI learning, experimentation, and innovation.

Reinforcing the emphasis on talent development, **Google.org recently announced a N3 Billion grant to support** Nigerian AI talent (114) This investment focuses on building sustainable digital and AI skills with a special interest in Natural Language Processing (NLP) so as to build local capacity to support the localized innovations and create new jobs. The grant aligns with Nigeria's growing prominence as one of Africa's largest digital economies, which reflects a clear trend: building talents who can translate local datasets into homegrown innovations at scale.

Meanwhile, the **Masakhane Research Foundation** continued advancing grassroots NLP research through its **Language Datasets RFP** (115). By funding community-led dataset creation, Masakhane directly addresses one of Africa NLP's most persistent bottlenecks: the scarcity of high-quality, openly available linguistic resources across the continent's thousand-plus languages. Unlike corporate or university-led programs, Masakhane's model prioritizes *community ownership* and participatory research, a defining approach that has helped to position it as one of Africa's most influential NLP collectives. The RFP's emphasis on dataset diversity and linguistic inclusivity reflects Masakhane's broader mission: building African AI "with Africans, by Africans, for Africans."

Source: **TechCity**

3.6 Benchmark and Evaluation Progress

Source: **NLP for Africa**

While talent development and institutional capacity accelerated in 2025, another critical pillar also advanced: **benchmarking and evaluation.** For African AI systems to be trustworthy and comparable they need rigorous benchmarks to capture linguistic, socio-cultural, and domain-specific realities unique to the continent. Three major benchmark initiatives were released in 2025, signaling a leap toward that goal. Three major benchmark initiatives released in 2025 signal a notable leap toward that goal.

The most expansive of these is **Sahara,** introduced as a comprehensive benchmark for African NLP and showcased at ACL 2025 (116). The benchmark covers a large number of African languages and NLP tasks and it moves beyond piecemeal evaluation by providing a multilingual benchmark suite. This enables researchers to compare model performance across tasks and languages while identifying systematic gaps. Sahara's breadth reflects a growing recognition: Africa's linguistic landscape cannot be represented by only a handful of major languages, but benchmarks must scale to reflect real-world diversity.

In parallel, **IrokoBench** was introduced at NAACL 2025 (117) so as to evaluate African languages through more fine-grained linguistic phenomena. Unlike Sahara that focuses on breadth, IrokoBench provides depth through the examination of linguistic features, reasoning patterns, and task-specific challenges that are often overlooked in major LLM benchmarks. This benchmark highlights an important shift in African NLP research: evaluation is no longer about if a model can operate on African languages, but about **how well** it copes with the nuances embedded within them.

AfroBench (118) extended the ecosystem by introducing a set of evaluations designed to measure model performance across multiple modalities and tasks in the African context. AfroBench is notable for its structured approach to culturally grounded evaluation that captures domain knowledge, sociolinguistic cues, and region-specific reasoning tasks that traditional benchmarks fail to reflect the shift toward a **layered evaluation framework** that moves beyond basic NLP tasks to assess higher-order reasoning in culturally situated settings.
Extending this trajectory toward even more localized evaluation, **Crane AI Labs** introduced the **Ugandan Cultural Context Benchmark (UCCB),** the first comprehensive QA suite designed to assess a model's cultural reasoning and contextual understanding specific to Uganda (119). UCCB includes 1,039 question–answer pairs across 24 cultural domains, from history and social norms to linguistic nuances and artistic traditions. The benchmark fills a crucial gap in global evaluation suites, which often lack the specificity needed to define if an AI system can reason correctly within culturally grounded African contexts. By focusing on Uganda's diverse cultural setting, including Ugandan English and local language elements, UCCB adds a new dimension to African AI evaluation: country-level cultural competence, not only linguistic correctness or multimodal ability.

equalyzAI

The next AI stack for africa is not app-first. It is voice-first, offline-first, and culturally fluent

At **EqualyzAI,** we are already building the AI that works for African's realities

www.equalyz.ai

Major Model Developments & Releases in 2025

4.1 African Language LLM Releases

The year 2025 marked a defining moment for African language technologies; several countries released their own large language models (LLMs) developed for local languages. Although these models vary in scale and purpose, they all reveal an increasing continental confidence: African institutions are no longer waiting for global models to adapt to them; instead, they are **developing their own AI systems** that are specific to their languages and cultures. Four LLM releases stood out during the year:

Source: **Recollect**

Nigeria: N-ATLAS v1, the First National Multilingual LLM

Nigeria's release of **N-ATLAS v1** (120) represented a landmark in African AI development. Launched in partnership with Awarri and the Federal Ministry of Communications, Innovation & Digital Economy, N-ATLAS stands as **Nigeria's first multilingual LLM,** created to advance language technologies at scale.

Unlike generic multilingual models that are trained on global datasets, N-ATLAS is specifically designed to support Nigerian linguistic diversity and local AI innovation. Its launch makes Nigeria one of the first African nations to articulate a national strategy around LLM development, moving beyond dataset creation to **sovereign AI capability** with public-sector relevance.

Uganda: Sunflower, a Local-Language "ChatGPT" for National Use

Uganda introduced **Sunflower,** which is described in government communications as a "ChatGPT for local languages" (121). The model was developed to support Ugandan languages through conversational interfaces so as to enhance access to information, public services, and educational resources.

The significance of Sunflower is in its orientation: it is a **government-backed LLM designed for domestic needs,** instead of an academic or private-sector prototype. For a multilingual country with important rural populations, the development of an AI system that can handle local languages is a key enabler for digital inclusion and public-sector modernization.

South Africa: InkubaLM, a Small Language Model for Edge Deployment

South Africa's **InkubaLM** (122) provided a different innovation pathway: efficiency. InkubaLM is a **Small Language Model (SLM)** that was developed by Lelapa AI to optimize for operation on **edge devices,**

which is a critical design choice for regions where bandwidth, compute, and energy constraints limit the use of larger LLMs.

Rather than competing on scale, InkubaLM advances computational accessibility. Its design demonstrates a growing African emphasis on resource-aware architectures that can operate in low-connectivity environments, power offline applications, and reduce dependency on cloud-based GPU infrastructure.

Kenya, Tanzania, Rwanda: UlizaLlama, Swahili LLM for Healthcare

In East Africa, **UlizaLlama** emerged as one of the first open-source **Swahili-language LLMs focused on healthcare** (123). The model was developed in partnership with Jacaranda Health and global health research collaborators so as to support clinical communication tasks for maternal and reproductive health, where language accessibility can directly impact health outcomes.

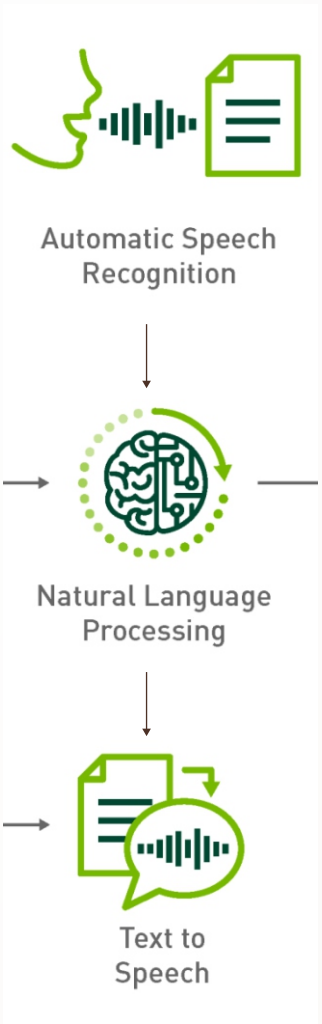
The focus of UlizaLlama on healthcare indicates an important trend: African LLMs are not only being developed for general-purpose use, but increasingly for **sector-specific tasks** where local-language competence is crucial. For health workers and patients who use Swahili as their primary language, UlizaLlama enables more accurate and culturally aligned interactions than English-centric models can provide.

4.2 ASR/TTS Developments

Alongside the progress in African language LLMs, **Automatic Speech Recognition (ASR)** and **Text-to-Speech (TTS)** technologies also saw a notable progress in 2025 across the continent. This indicates a clear shift: as African populations interact increasingly with digital services through voice interfaces, speech data has become as important as text datasets.

One of the most ambitious initiatives is the **African Voices Project,** which aims to create one of the **largest open speech datasets** ever assembled for African languages (124) and coordinated by University of Pretoria (South Africa), Data Science Nigeria (Nigeria) and Maseno University (Kenya) with the funding support of Gates Foundation. In order to capture dialects, accents, and pronunciation patterns that are still underrepresented in global ASR datasets, the project focuses on gathering large amounts of native African speech. Its scope is unusually broad: instead of targeting a single language group, the project is curating resources across multiple regions, each characterized by linguistic complexity and rich oral traditions.

The significance of this effort is twofold. First, the African Voices Project prioritizes open datasets, which addresses one of the most persistent barriers to African speech technology development: the scarcity of publicly accessible, large-scale corpora that can be used for training and evaluation. Second, the project indicates that African ASR/TTS systems must be built on authentic, community-driven audio data, not approximations derived from non-African accents or text-to-speech simulations.



Source: **Rahul LinkedIN**

4.3 Deployment Platforms & Developer Tools

As African organizations are transitioning from prototyping to large-scale AI adoption in 2025, a parallel evolution occurred in the **platforms and tools** used to implement these systems. The emergence of infrastructure partnerships and developer-focused ecosystems indicates a shift toward more accessible and production-ready environments for African innovators and enterprises.

A notable progress in this space was the launch of **Liquid G,** a program led by Liquid C2 in partnership with **Google Cloud** (125). Liquid G is positioned as an AI-powered cloud distribution platform that expands Google Cloud's reach across Africa by endowing enterprises with access to AI tooling, compute, and managed services via a unified channel. The initiative is notable because it focuses on in-region support, thereby addressing a long-standing challenge: African companies often rely on offshore cloud regions, resulting in higher latency, limited resilience, and persistent data-sovereignty concerns. By combining Liquid C2's continental network footprint with Google Cloud's AI stack, Liquid G positions itself as an integrated deployment layer for organizations seeking to operationalize AI at scale.

Safaricom's $500M AI strategy, announced in parallel, further illustrated how major African enterprises are deepening their technical infrastructure while cultivating developer ecosystems (126). A key component of the strategy is a **developer portal** that is designed to expose AI services, APIs, and sandbox environments to Kenya's expanding tech community. This approach recognizes that in order to translate AI capabilities into real-world impact, enterprises must encourage not only internal teams but also startups and third-party developers who develop customer-facing applications. Safaricom's investment indicates a shift in the role of telecoms in East Africa, from network providers to AI-based digital platforms able to support financial services and small-business productivity tools.

Source: **Liquid C2**

4.4 Industry Product Launches

Alongside these platform-level shifts, 2025 saw a wave of **AI-powered products** that integrate the African market in sectors with deep public importance: healthcare, telecommunications, and financial services. All these launches have one common theme: AI become more and more embedded into frontline service delivery, not just back-end automation.

In Rwanda, the Ministry of Health introduced two AI-based applications: **e-Buzima** and **e-Banguka** (127). In keeping with Rwanda's long-standing leadership in the adoption of health technology, these tools aim to optimize diagnostic processes, patient engagement, and community health workflows. The deployments emphasize the value of AI in health systems, where information sharing, early detection, and effective triage continue to be major obstacles. These applications are specific to Rwanda's care-delivery ecosystem rather than being generic AI assistants, show how public-sector innovation can drive AI integration across the continent.

In the telecommunications sector, **Vodacom's collaboration with Google Cloud** enabled the rollout of new AI-powered customer care systems across its African markets (128). By embedding generative AI into frontline support, Vodacom seeks to lower response times, standardize service quality, and quickly manage growing customer bases. This rollout also positions Vodacom as a main channel for delivering AI utility to consumers who may not directly interact with AI tools but nonetheless benefit from improved service reliability and customized support. It indicates the broader telecom trend: moving from connectivity providers to AI-enabled digital service platforms.

In financial services, **Qore** introduced **Recova,** which is an AI solution designed to support credit recovery in banking (129). With credit access expanding across African markets, non-performing loan management has become more and more important. Recova uses AI to spot repayment trends, customize outreach strategies, and improve operational efficiency, which supports banks as they navigate the trade-off between risk management and customer engagement. As with health and telecom, Recova shows how AI is being directly woven into main business processes in high-impact sectors.

Source: **Qore**


equalyzAI
Ask by Voice.
Get Answers.
Take Action.
No Dashboard.
Speak naturally to data and receive clear, actionable insights without dashboards.
www.equalyz.ai

Investments, Partnerships & Policy

5.1 Investments

5.1.1 Startup Funding Rounds

Africa's startup ecosystem had another year of significant financial momentum, with $140M raised in September alone and the total 2025 funding reaching $2.2B (130).

$140M
raised in September alone and the total 2025 funding reaching
$2.2B

This sustained investment tempo reflects an ecosystem that has matured beyond sporadic mega-deals into a more steady and diversified capital environment. Notably, the composition of funded startups shows a growing appetite among investors for companies that integrate AI into financial services, logistics, and consumer technology, sectors where data availability and digital adoption have increased remarkably over the past five years.

Within this broader funding landscape, **Kredete's $22 million Series A round** affirms the role of AI in Africa's financial sector (131). Kredete develops AI-based credit-scoring solutions to serve populations that were excluded from formal lending because of limited credit histories. By leveraging machine learning models, the company tackles a persistent constraint across African markets: insufficient financial identity for millions of individuals and small businesses. Kredete's funding shows that investors recognize that AI can play a transformative role, especially in areas with both social impact potential and strong commercial viability like credit risk modeling, financial inclusion, and digital-lending ecosystems.

Source: **TechCabal**

5.1.2 NLP/LLM Grants

While startup funding accelerated AI commercialization in 2025, a parallel stream of investment emerged from organizations seeking to empower the foundation on which African AI must be built. These grants focus on capacity building, linguistic inclusion, and community-driven R&D. Two notable commitments shaped this landscape.

A major milestone came when the Masakhane Research Foundation received a
$3M grant
to expand its African Languages Hub, led by Gates Foundation.

Masakhane will use this investment to support dataset creation, model development, and researcher training across dozens of African languages. The grant reinforces Masakhane's distinctive contribution to the global AI landscape: demonstrating that high-quality, open, multilingual NLP research can be driven from and by Africa, even for languages with minimal digital presence.

In parallel, Google.org committed

N3 Billion

to strengthen Nigeria's AI ecosystem (114)

While not exclusively an NLP grant, a significant portion of the investment supports digital skills training, talent development programs, and AI literacy efforts, which are necessary for advancing multilingual model development in one of Africa's largest tech markets. This investment reflects the long-term strategic interest of Google in fostering technical capacity in Nigeria.

5.1.3 Corporate and Innovation Investments

Along with startup financing and research-focused grants, 2025 also saw major **corporate investments** that seek to accelerate the AI innovation ecosystem of Africa. These commitments reveal that global technology companies increasingly recognize that Africa is not simply an emerging market, but it is a strategic frontier for AI research and talent development.

A standout example came from Google, which announced a

$37Million

"Africa AI" investment package

designed to support the future of AI research both within Africa and in collaboration with global institutions (132). The package includes funding for academic partners, research fellowships, and infrastructure support. What distinguishes this initiative is its dual focus:

- **Strengthening Africa's research ecosystem** by ensuring that local researchers can contribute at the frontier of global AI development; and
- **Boosting responsible AI globally** by incorporating African linguistic, cultural, and socio-technical perspectives.

According to Google's announcement, a considerable share of the investment is directed toward **AI research and innovation hubs, education and training programs,** and support for work on African languages. This reflects a clear shift in corporate strategy: African researchers and language contexts are increasingly seen as essential contributors to the diversity and robustness needed in global AI systems.

Source: **Nigeria Communications Week**

5.2 Partnerships

5.2.1 Government–Industry Collaborations

Government–industry collaborations grew considerably in 2025 as African looked for partners to speed up digital transformation. These collaborations demonstrate a shared priority: ensuring that AI systems implemented in Africa are in line with national objectives, regulatory environments, and the continent's digital-trust agenda.

One example comes from **Benin,** which signed a landmark agreement with the **International Institute of Digital Innovation in Africa (IIDiA)** to expand digital trust services and endorse the country's digital-sovereignty posture (133). According to reporting, the agreement reinforces trusted digital transactions, secure digital identities, and national digital infrastructure. For a country scaling its e-government and digital-finance services, such investments are crucial: AI-driven services cannot achieve public trust or regulatory compliance without trusted digital foundations. Therefore, this collaboration indicates that African governments increasingly view digital trust infrastructure as the bedrock upon which future AI systems will operate.

Source: **Freepik**

5.2.2 University–Industry Collaborations

Parallel to government collaborations, **university–industry collaborations** were also increased in 2025 so as to support the research capacity of Africa and accelerate talent development. These collaborations show that universities also play an important role in shaping the long-term AI trajectory of Africa, not only corporations and governments.

A prominent example is the collaboration between **OpenAI and the University of Lagos (UNILAG),** that launched the first **OpenAI Academy** in Africa as part of UNILAG's 2025 International Week (113). The Academy exposed students and researchers to AI development frameworks and model-building practices. Notably, it showed a shift in how global AI leaders engage with African institutions: not as peripheral training hubs, but as strategic partners whose expertise is important to shape a globally representative AI ecosystem. For UNILAG, this collaboration affirmed its role as a leading center for AI education in West Africa.

In a different but complementary direction, the **Matsuo Lab at the University of Tokyo** partnered with **UNDP Africa** to explore how AI adoption can support development priorities across the continent (135). As discussed in UNDP's Afri-Converse forum, the collaboration focuses on building AI capacity and assessing the practical way through which African countries can use AI for social impact, especially in economic planning, governance, and service delivery. While OpenAI's collaboration emphasizes technical upskilling, the Matsuo Lab partnership emphasizes development-oriented research, which provides African policymakers and institutions with analytical frameworks to navigate AI's opportunities and risks.

5.2.3 Donor-Funded and NGO Initiatives

Source: **Nigeria Communications Week**

Next to government- and industry-led collaborations, **donor-funded programs and NGO collaborations** continued to play a pivotal role in forming Africa's AI-ready digital ecosystem in 2025. These initiatives function at a different layer of the ecosystem: while governments focus on national strategy and companies invest in infrastructure and applications, donor programs often target the **social foundations** needed for inclusive AI adoption.

A noteworthy example is the expanded collaboration between **TECNO and UNICEF** to scale the **Learning Passport** program across Nigeria (136). The Learning Passport is designed as a digital learning platform for children in underserved regions so as to provide curriculum-aligned educational content that is accessible online and offline. The 2025 expansion, enabled through TECNO's mobile technology and UNICEF's educational programming, focuses on reaching even more learners who still cannot receive a high-quality education due to poor connectivity, resource constraints, and remote location.

What makes this initiative pertinent to the AI trajectory of Africa is not the direct use of AI, but the enabling conditions it reinforces. By extending access to digital learning materials, improving device availability, and supporting localized educational content, the Learning Passport helps cultivate the future AI workforce long before formal AI training begins. Furthermore, by reaching marginalized communities, the program counters one of the most persistent risks in Africa's digital transition: that AI benefits increase disproportionately to already-connected populations.

This collaboration also shows the unique value proposition of donor-led initiatives within Africa's technological landscape. Unlike commercial programs, UNICEF and TECNO's efforts prioritize educational equity and child-centered development outcomes. And in contrast to purely government-led models, they engage external expertise and private-sector collaborations to speed up implementation at scale.

5.2.4 Regional Cross-border Partnerships

Regional, cross-border partnerships also contribute to forming the AI and digital transformation of Africa in 2025. These alliances reflect an increasing recognition that no single country can fully capture the benefits of AI alone. Instead, regional coordination is essential to expand markets and pool investment into shared digital infrastructure.

A clear illustration of this trend came during the **Nigeria–South Africa Investment Dialogue,** held ahead of the G20 Summit (137). The dialogue was hosted with the participation of senior government representatives and major African enterprises including MTN, and it focused on improving economic cooperation across the continent's two largest and most influential economies. While the discussion covered many investment themes, its importance to AI lies in the mutual priority expressed by both countries: improving regional digital infrastructure, expanding digital-skills pipelines, and making Africa a unified economic bloc that can impact global innovation agendas.

The dialogue emphasized that **AI progress in Africa is increasingly formed by regional economic diplomacy** and not only national policy. Nigeria and South Africa are uniquely positioned to support continent-wide initiatives, from data-governance harmonization to cross-border digital trade. Their collaboration sends an important signal to investors, development partners, and neighboring governments: The digital transformation of Africa is progressing not through isolated national efforts but through coordinated frameworks that amplify the scale and impact.

5.3 Policy and Regulation

As AI adoption accelerated across the continent in 2025, policymakers increasingly turned their attention to the frameworks needed to guide responsible and rights-preserving implementation. The dominant fact of the year is a transition from dispersed and exploratory discussions to multi-level coordinated governance efforts that cover regional commitments, national strategies, sector-specific regulation, and global standard-setting.

Momentum at the regional level was strengthened by the **Global AI** Summit on Africa, where leaders supported the **Kigali Declaration on AI** (138). The declaration highlights common priorities like fair access, reducing algorithmic harms, and the significance of inclusive governance mechanisms. Discussions at the summit made clear that the digital future of Africa cannot rely only on national strategies, but coordinated, continent-wide approaches are necessary so as to address challenges in data governance, model evaluation, and capacity building. In this way, the Kigali Declaration serves as a statement of Africa's intention to actively contribute to global AI governance norms, as well as a policy compass.

Within this continental environment, individual countries continued to advance their domestic AI frameworks. The **National AI Strategy** of Nigeria was presented as a roadmap to convert the country's AI potential into measurable socio-economic gains (139). Commentary on the strategy highlights its focus on operational execution, emphasizing applications in public service delivery, digital identity, innovation ecosystems, and trusted data governance. Regulatory evolution also extended into adjacent fields such as data protection.

Early 2025 saw **new privacy laws were passed in Botswana and Malawi** (140), which reflects a growing continental movement toward strengthening data rights as a foundation of reliable AI. These legal updates are essential for enabling safe implementation of AI systems that depend more and more on sensitive data. Moreover, harmonized privacy rules help create the regulatory clarity required for regional research collaboration and responsible cross-border data flows.

Africa's policy discourse also intersected with global initiatives to address the cultural dimensions of AI. UNESCO, working alongside the G20, advanced efforts to promote **linguistic diversity in the digital age** (141). With thousands of languages spoken across the continent, this advocacy is particularly important: without deliberate policy interventions, AI systems risk reinforcing global linguistic hierarchies and overlooking local languages. UNESCO's involvement indicates that AI governance must extend beyond technical ethics to include cultural equity and linguistic representation.

Source: **Tech Labari**

equalyzAI

Voice Creation That Sounds Local, Trusted and Human.

Create localized audio content quickly across multiple African dialects

www.equalyz.ai

Sector-by-Sector Deployments

6.1 Financial Services

AI adoption in Africa's financial sector accelerated markedly in 2025, formed by two simultaneous forces:

1. escalating cyber and fraud threats to target high-growth fintech markets, and
2. increasing demand for data-driven decision-making in credit and consumer finance.

These forces, both destabilizing and transformative, created a landscape where AI became not optional, but essential infrastructure.

Late 2025 saw what industry observers described as an **"AI fraud crisis,"** as cybercriminals weaponized deepfake audio, synthetic identities, and sophisticated phishing techniques to exploit Africa's fast-growing digital payments ecosystem (142). Fintechs and banks faced more and more incidents of impersonation and fraudulent authorization attempts that are exacerbated by the rapid uptake of mobile money and digital banking of the region. In response, financial institutions began improving their investments in biometric verification, behavioral analytics, and real-time anomaly detection. These tools are a defensive evolution: AI is not only powering innovation in payments and lending, but also becoming the main shield against AI-driven fraud.

Against this backdrop, the

$22M

Series A funding raised by Kredete (131) highlighted the other side of the fintech transformation:

the push to expand credit access using AI-driven assessments. Kredete's platform uses machine learning so as to generate credit profiles for consumers who lack formal credit histories, which represents a longstanding barrier to financial inclusion in Africa. By analyzing nontraditional signals like mobile behavior or transaction patterns, AI credit scoring allows lenders to evaluate risk in ways impossible under traditional systems. Kredete's funding reflects investor confidence not only in one company, but in the broader thesis that AI can close the credit-access gap in Africa without compromising risk management.

Amid both opportunity and risk, regulators also stepped forward. The **South African Financial Sector Conduct Authority (FSCA)** published a report on AI's adoption in banking that describes how institutions are employing AI for risk modeling, fraud prevention, customer service, and compliance (143). The report shows an increasing awareness that AI is now deeply integrated in financial pipelines. Its publication marks a shift toward risk-aligned AI governance in Africa's most mature financial market, emphasizing transparency and consumer protection.

Source: **fscasouthafrica**

6.2 Healthcare

Source: **Techpoint Africa**

Healthcare emerged as another pillar of AI transformation of Africa in 2025, driven by the continent's urgent needs for multilingual patient engagement, scalable clinical support, health-worker augmentation. In contrast to financial services, where AI mainly optimizes transactions and risk, in healthcare AI aims to increase the reach and efficacy of human providers, especially in settings with limited resources.

Intron Health continued to demonstrate the value of clinical speech recognition in African hospitals (144). Its tools lessen the administrative burden and support environments where there is a severe shortage of clinicians by assisting clinicians in more effectively documenting patient encounters. Speech-to-text systems tailored to African accents and clinical vocabulary represent an important step to close operational gaps in overstretched health systems.

At the patient-engagement level, **Jacaranda Health** reported impact data from PROMPTS and **UlizaLlama,** the Swahili LLM developed for healthcare communication (123). These tools provide maternal health guidance, triage support, and information in local languages so as to bridge gaps in understanding, compliance, and early detection. The integration of linguistically aligned AI into maternal health workflows shows how **language technology is becoming a determinant of care quality** in multilingual populations.

A similar shift toward accessible, multilingual patient engagement was demonstrated by **Data Science Nigeria's family planning chatbot,** developed in partnership with **DKT** and **Tulane University New Orleans, United States** (145). Designed to improve the access to reliable reproductive-health information, the chatbot provides personalized counselling in **Pidgin, Hausa, and Yoruba languages** so as to enable millions of speakers to obtain guidance in their preferred language.

Beyond frontline care, **HelloCareAI** advanced virtual ward technology, supported by new funding (146). Virtual wards use AI-based monitoring to assist patients outside traditional hospital settings; this strategy is especially pertinent in African regions where hospital capacity is limited or geographically uneven. A different approach to care's continuity is provided by AI-enabled remote monitoring that reduces strain on facilities while improving patient outcomes.

These advances were reinforced by new dataset initiatives like the **African Next Voices medical ASR dataset** (124), which improves speech resources for clinical contexts. With many African patients more comfortable communicating orally than textually, high-quality medical speech datasets are crucial for building safe, reliable clinical decision-support systems and patient-facing applications.

6.3 Education

Source: **EduTimes Africa**

Education is one of the most active and strategically important domains for integrating AI in Africa in 2025. Despite that efforts varied across countries and institutions, they all pointed toward a shared continental priority: preparing a generation of learners and educators not only to use AI, but to form its development responsibly. The year saw the convergence of digital literacy campaigns, university–industry training programs, national AI tools for local languages, and cross-country assessments of workforce readiness, together they form a layered picture of the evolving education landscape of Africa.

Digital Literacy and Responsible AI Awareness

The year's most visible public-facing effort came from the **InnTech Summit 2025**, which placed a strong spotlight on the **"Digital Literacy for All"** initiative (147). Discussions at the summit emphasized that as AI systems is diffused through public services, finance, healthcare, and daily communication, basic digital literacy is no longer optional. Calls for responsible AI adoption highlighted the necessity of providing young people, teachers, and civil servants with the basic knowledge needed to navigate an increasingly algorithmic world. The summit's framing made clear that African countries must prioritize not only access to AI tools, but also public understanding of risks, ethics, and digital rights, a fundamental component of equitable AI-enabled education systems.

Local-Language AI Tools for Education and Public Engagement

In East Africa, AI adoption in education intersected closely with language accessibility. The **Uganda Ministry of ICT's launch of the Sunflower local-language AI model** (121) demonstrated how language technologies can support national education and public-information goals. Sunflower was designed to understand and interact in Ugandan languages so as to create opportunities for **inclusive learning tools,** local-language digital assistants, and enriched educational platforms that reach communities often left behind by English-heavy digital content. This initiative reflects a regional vision in which AI becomes a **linguistic bridge** that reduces educational inequities related to language mismatch and enables more culturally relevant digital learning environments.

Continental Readiness: Skills, AI Integration, and Gaps

These national and institutional initiatives are in line with broader findings from **Africa Education Watch's 2025 Skills & Tech Report** (148). The report highlights a dual dynamic: even though AI and technology skills are becoming essential to Africa's economic competitiveness, many education systems still struggle with technology access, teacher readiness, and curriculum integration. The analysis indicates that AI in education is not only a technological issue but also a question of **systemic preparedness** that requires investment in teacher training and support for digital infrastructure in schools. In this sense, the report situates the Uganda and UNILAG initiatives within a wider continental narrative: Africa is moving quickly to integrate AI into learning systems, but uneven progress risks increasing skill disparities across regions.

Gaps, Risks & Critical Challenges

7.0 Technical Challenges

Despite significant momentum in AI infrastructure, dataset creation, and model development throughout 2025, the African AI ecosystem continues to face many structural and technical challenges. These constraints which stem from historical underrepresentation, linguistic diversity, and resource limitations, form not only the feasibility of AI development but also the quality, inclusivity, and long-term sustainability of implemented systems. The following subsections outline the most persistent obstacles identified across research publications, industry deployments, and community-led initiatives.

7.1 Data Scarcity in Low-Resource Languages

Many African languages are still lacking the large, annotated datasets necessary for effectively training advanced machine learning and NLP systems. Even though recent dataset releases have started to close these gaps, most languages across the continent, including those widely spoken, **still lack sufficient resources,** which hinders advancements in ASR, translation, and LLM fine-tuning. Sentiment analysis, information retrieval, and conversational agents are among the downstream tasks that are hampered by models' inability to capture grammar, semantics, and cultural nuance in the absence of robust corpora.

7.2 Dialect Diversity and Code-Switching

Even when datasets exist, **dialect variation** poses additional modeling challenges. Languages such as Swahili, Arabic, Yoruba, and Hausa exhibit considerable geographic and social variation, and these dialects can differ at the levels of vocabulary, pronunciation, and orthography. In daily communication, code-switching, combining multiple languages within a sentence or conversation, is widespread, especially in urban centers and multilingual communities. Because most global NLP models assume monolingual input, African language systems often fail when confronted with these authentic patterns of speech and writing.

7.3 Benchmarking and Evaluation Gaps

Although new benchmarks such as Sahara, IrokoBench, and AfroBench have expanded evaluation coverage, many African languages still lack **task-specific and culturally grounded evaluation frameworks.** This makes it difficult to diagnose model limitations, compare model performance across languages, or determine if improvements stem from learning or from artifacts in training data. Inadequate benchmarks may result in models that seem performant but fail in real-world applications, especially for high-stakes domains like financial services, healthcare, and public information delivery.

Source: **Owlcation**

7.4 Bias and Representativeness

A recurring issue across both private and academic models is the **underrepresentation** of African speakers, dialects, and cultural contexts in training data. Bias manifests in several ways:

- Heavier weighting of Western or global-dominant languages leads to poorer performance on African languages.
- Vision models trained on non-African facial datasets show higher error rates for African phenotypes.
- Linguistic bias can emerge when datasets fail to capture regional idioms, politeness structures, or discourse norms.

Such gaps risk reinforcing existing inequities, excluding users from effective AI experiences, or propagating harmful stereotypes.

7.5 Compute Limitations

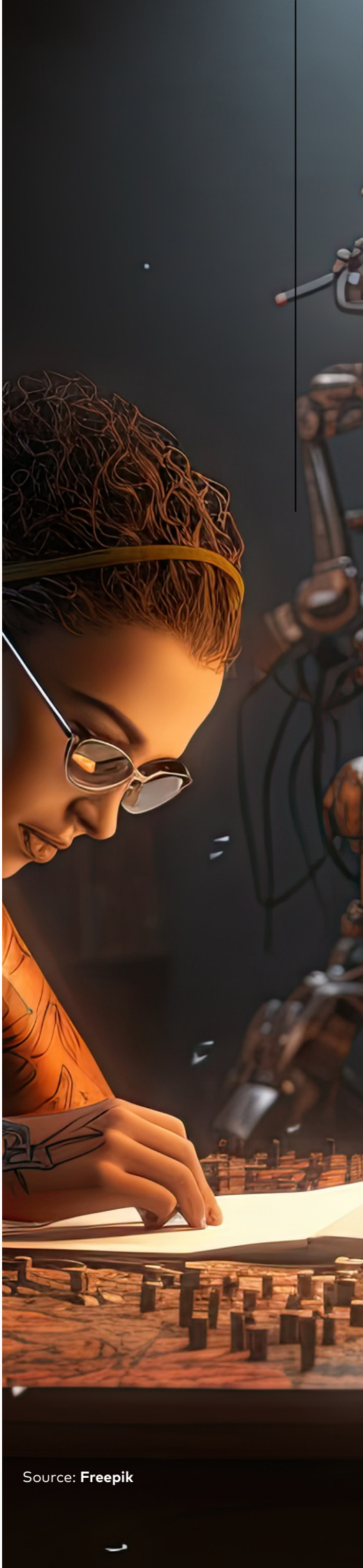
Even though regional cloud availability and localized GPU access are increasing, **compute scarcity** remains a challenge, especially for startups, research groups, and smaller institutions. Training large-scale models or even mid-scale LLMs needs resources that are still expensive or inaccessible across much of the continent. This constraint forms technical strategy: African researchers often depend on smaller models, parameter-efficient fine-tuning, or low-compute architectures, which may limit experimentation with more advanced model families.

7.6 Latency and On-Device Performance Issues

Infrastructure limitations like low bandwidth, inconsistent connectivity, and unstable power supply create real constraints in the implementation of AI. Latency issues can hinder interactive applications (e.g., real-time ASR, conversational assistants, and customer-care agents), making the cloud paradigm unreliable in many rural or peri-urban settings. This necessitates **edge-optimized or hybrid models** that can run offline or with minimal network dependence, which is a design pattern reflected more and more in African LLM and ASR research.

7.7 Speech Data Quality Challenges

Speech technologies face unique difficulties beyond simple data scarcity. Many African languages lack standardized orthographies, and speech recordings often exhibit **variable accents, background noise, code-switching, and low-quality audio capture,** especially when sourced from mobile devices or community environments. These factors reduce the reliability of ASR systems and introduce noise into downstream NLP systems. These limitations can directly impact information accuracy and service delivery, especially in health and emergency settings, where ASR has significant potential.



Source: **Freepik**

8.0

Public Trust, Safety & Responsible AI

As AI systems became more ubiquitous across Africa in 2025, the question of **public trust** moved to the center of national and industry agendas. While technical progress accelerated, so did concerns about data exploitation, misinformation, and unequal regulatory oversight. The year's developments made clear that responsible AI in Africa is not simply a technical challenge but a complex issue with roots in governance, social norms, user protection, and the unique linguistic and cultural fabric of the continent.

The following subsections present the main themes that formed the evolving responsible-AI landscape of Africa.

Source: **PLMR**

8.1 Misinformation and Deepfake Risks

African journalists, policymakers, and digital platforms faced increased threats from **misinformation and deepfakes,** especially during elections, public-health campaigns, and high-profile political events. Long-standing weaknesses like inadequate fact-checking infrastructure and quick mobile-first information flows were exacerbated by the proliferation of AI-generated audio and video. As deepfakes became easier to create using low-cost tools, the line between real and manipulated content blurred, eroding public trust in institutions and media. Countries recognized more and more the need for coordinated responses that combine media literacy initiatives with detection technologies and regulatory oversight.

These dynamics highlight a crucial realization: **misinformation in Africa is not only about content volume but also about contextual fragility,** where linguistic diversity, trust networks, and variable digital literacy provide fertile ground for synthetic manipulation.

8.2 Voice Cloning Safety Issues

Alongside visual deepfakes, **voice cloning** emerged as a pressing safety issue. With the continent's rapid adoption of voice-driven services, from mobile money authentication to customer support, malicious actors began exploiting cloned voices for fraud and impersonation. In markets where oral communication is culturally prominent and literacy barriers persist, voice interfaces carry disproportionate influence. The rise of cloned audio therefore elevated risks in sectors such as finance, telecoms, and emergency response.

Source: **Consumer Reports**

The challenge is compounded by Africa's diverse accents and tonal languages, which many global safety tools are not yet designed to detect. Building safeguards for African voice profiles requires not only technical innovation but also **region- and language-specific datasets** for detecting irregularities.

8.3 Transparent Data Sourcing

With increased model deployment came renewed scrutiny on **data sourcing practices,** especially where African language datasets are involved. Developers, researchers, and civil society groups raised concerns about the lack of transparency around how training data is collected, annotated, stored, and governed. As more institutions began building their own LLMs and ASR systems, questions of ownership, consent, compensation, and community involvement gained urgency.

A growing movement across African AI communities called for **open, auditable documentation,** similar to data cards or model cards, to ensure that datasets reflect ethical collection processes. In multilingual societies, where linguistic marginalization has historical roots, transparent sourcing is not simply a procedural step but a matter of protecting cultural identity and ensuring equitable participation.

8.4 Governance and National Regulation

Governance frameworks gained momentum as countries worked to define national AI strategies, update data-protection laws, and adopt continental commitments such as the Kigali Declaration. Governments increasingly recognized that effective AI regulation cannot rely solely on importation of global norms; it must address African-specific priorities—such as linguistic inclusion, access disparities, cyber fraud, and public-sector capacity constraints.

Regulatory discussions centered on areas like algorithmic transparency, accountability in high-risk domains (e.g., credit scoring and healthcare), and cross-border data-sharing. These efforts reflect an important trend: **African governments are shifting from reactive regulation toward proactive governance,** aligning policy with long-term digital-development goals.

8.5 Safety Frameworks and Industry Practices

Industry actors also advanced responsible-AI practices, responding to consumer expectations and emerging regulatory requirements. Telecoms, banks, and health-tech companies began integrating internal **AI safety frameworks,** including risk assessment protocols, bias audits, and ethical review processes. Many enterprises adopted global standards but adapted them to African contexts, where infrastructure constraints, linguistic diversity, and socio-economic sensitivities demand tailored approaches.

These practices illustrate the maturation of Africa's AI industry: companies are no longer focused only on performance metrics but increasingly on **trustworthiness, explainability, and alignment with local norms.**

Source: **Usercentrics**

8.6 Ethical Concerns in Data Collection

Finally, data-collection ethics remained a foundational challenge. In many African regions, datasets are collected in environments where power imbalances, limited digital literacy, and socio-economic vulnerabilities raise questions about **meaningful consent.** Additional concerns involve cultural sensitivity, misuse of community-generated data, and inadequate protections for sensitive information, especially in sectors like health, biometrics, and mobile communication.

The year highlighted the need for ethical guidelines that go beyond legal compliance to encompass community engagement, culturally appropriate explanations, and mechanisms for redress. As AI systems expand their footprint, ethical data practices will be essential to maintaining public legitimacy and protecting historically marginalized groups.

equalyzAI

Turn Everyday Conversation into Clear Business Intelligence

Deliver insights by voice, grounded in local context and dialect

www.equalyz.ai

2026 Outlook and Predictions

9. Strategic Trends & Predictions for 2026

Africa's rapid AI expansion in 2025 prepared the stage for a new phase of adoption that is defined not only by technical innovation, but also by how these technologies permeate daily life, reshape service delivery, and impact sector-wide strategies. Looking ahead to 2026, several interconnected trends point to where the continent's AI ecosystem is headed.

Source: **7Hills Computers**

The following subsections outline the key developments likely to shape Africa's AI trajectory in 2026, highlighting how technical constraints, linguistic realities, market forces, and governance priorities converge to define the continent's emerging AI landscape.

9.1 Mass Adoption of Voice AI

Voice is becoming Africa's most scalable interface for AI. Since millions of users are relying on oral communication, mobile phones, and voice-driven services, 2026 will likely see **voice AI integrated into customer service, public information systems, and healthcare triage tools.** The increasing availability of African speech datasets and models, especially for Swahili, Yoruba, Amharic, and Kinyarwanda, will speed up real-world deployments, making voice a primary access layer for digital experiences.

9.2 Rise of Offline, On-device LLMs

As compute constraints persist and connectivity remains uneven, **small, efficient language models designed for edge devices** are expected to gain significant traction. Models such as InkubaLM and AfroSLM demonstrated that compressed architectures can operate effectively on low-power hardware, foreshadowing a broader adoption of **offline-capable AI assistants, mobile apps, and service tools** that do not rely on cloud latency or expensive GPUs.

9.3 Local-Language AI Assistants

Building on the emergence of models like Sunflower and UlizaLlama, 2026 will likely bring a proliferation of **AI assistants tailored to African languages and cultural contexts.** These assistants will support sectors such as education, agriculture, and public health—providing accessible digital guidance in users' native languages and bridging gaps where literacy or English proficiency remain barriers.

9.4 Standardization of African LLM Benchmarks

With Sahara, IrokoBench, and AfroBench establishing momentum, **African-language evaluation frameworks are expected to become more standardized and widely used in** 2026. These benchmarks will help researchers and companies compare LLM performance more reliably across languages and tasks, reducing their reliance on global evaluation techniques that fail to capture African linguistic realities. Benchmark convergence will also strengthen policy-making and procurement standards for AI systems deployed in public-sector environments.

9.5 Telco-led Language AI Distribution

Telecom operators are expected to evolve into major distribution channels for AI services, leveraging their vast user bases, established billing rails (including mobile money), and network infrastructure. Initiatives such as Vodacom–Google Cloud partnerships offer early signals of a trend where telcos provide embedded AI solutions—including language services, customer-care automation, and voice APIs—to enterprises and SMEs, mirroring the role telcos played in Africa's mobile-money revolution. The recent efforts by **GSMA** to mobilize a pan-African network of telecommunications operators and local startups represent an important step toward advancing African language development as a foundation for inclusive, locally driven innovation. By convening mobile network operators, ecosystem builders, and indigenous technology providers, GSMA is helping to unlock telco assets to support the creation of language datasets, speech technologies, and locally relevant digital services.

Source: **MTN Group**

9.6 AI-first Customer Service Expansion

Customer experience functions across banking, telecoms, e-commerce, and public services are shifting toward **AI-first operating models.** By 2026, generative AI, NLP, and voice-bot systems will handle a majority of frontline interactions, reserving human agents for high-complexity or high-emotion cases. This shift will be driven by improvements in voice recognition, local-language comprehension, and the availability of AI developer platforms tailored to African markets.

9.7 Multimodal (Speech + Text) Solutions for Low-Literacy Users

With large portions of the population still facing literacy challenges, multimodal AI solutions that combine **speech, text, and simple interfaces** will be essential in expanding digital inclusion. In 2026, expect broader deployment of multimodal systems in: agriculture advisories, maternal health communication, financial education, and e-government services.

Looking ahead, multimodal tools in 2026 are likely to evolve in several directions:

- **Conversational audio agents** that guide users through complex tasks such as loan applications, medical instructions, or government-service enrollment.
- **Hybrid interfaces** where speech handles explanation and interaction, while minimal text summaries, icons, or structured prompts reinforce clarity.
- **Localized voice models** tuned to regional accents and dialects, allowing users to interact naturally without shifting to formal variants of their languages.

These solutions will help users with limited text-literacy navigate more and more digital contexts while ensuring that AI adoption remains equitable.

9.8 Emergence of Regional Fine-tuning Hubs

The rapid expansion of GPU access, cloud regions, and local LLM initiatives sets the stage for **regional fine-tuning hubs** across Nigeria, Kenya, South Africa, Rwanda, and Egypt. These hubs will specialize in adapting foundation models to local linguistic, regulatory, and sector-specific needs. Rather than training frontier models, Africa's advantage will lie in **contextual fine-tuning,** enabling models to perform effectively in multilingual environments, code-switched interactions, and domain-specific workflows such as healthcare triage or credit scoring.

Source: **TechCabal**

9.9 Rationale, Impact and Sector Alignment

These strategic trends reflect a deeper shift in Africa's AI landscape:

Rationale:

Africa's linguistic diversity, connectivity constraints, and mobile-first usage patterns shape AI innovation around **local context, efficiency,** and **inclusivity,** rather than scale alone.

Impact:

These shifts will influence everyday experiences:

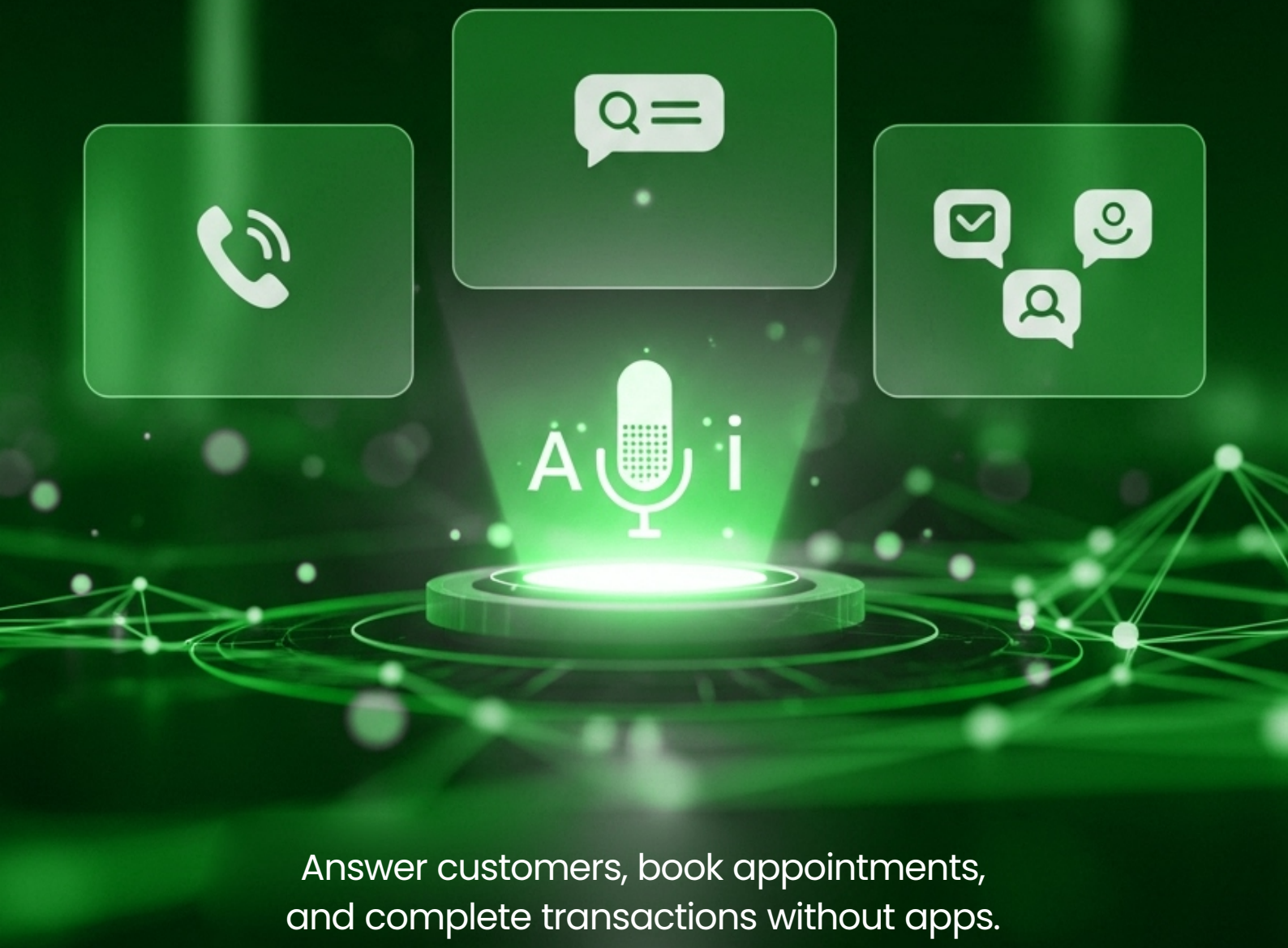
- customer service will be more automated and language-aware;
- health and education tools will become more accessible through voice and multimodal channels;
- enterprises will gain AI support without needing deep technical infrastructure;
- regulators will have clearer benchmarks and safety frameworks for oversight.

Sector Alignment:

- Financial services will benefit from telco-led AI distribution and fraud-resistant voice biometrics.
- Healthcare will expand multimodal interfaces and offline decision-support tools.
- Education will adopt local-language assistants and AI-driven learning platforms.
- Government services will deploy AI-first engagement systems and standardized evaluation frameworks.

equalyzAI

From Calls to Closures, Powered by Local Voice AI



Answer customers, book appointments,
and complete transactions without apps.

www.equalyz.ai

10.0

EqualyzAI Perspective & Call to Action

10.1 EqualyzAI Contributions in 2025

Throughout 2025, EqualyzAI contributed to Africa's AI landscape through initiatives focused on **democratizing access, supporting low-resource languages,** and **promoting responsible AI practices** tailored to Global South realities. The organization's whitepaper on **Small Language Models (SLMs)** articulated a clear framework for deploying AI in environments marked by bandwidth, compute, and energy constraints that are common across African institutions and communities.

EqualyzAI also improved the engagement with researchers and ecosystem partners by emphasizing the significance of African-first dataset strategies and advocating for inclusive, contextually grounded evaluation frameworks. These contributions align with the growing consensus in academia, industry, and civil society: AI models need to be adapted to African languages, African infrastructures, and African use cases, rather than being retrofitted as an afterthought.

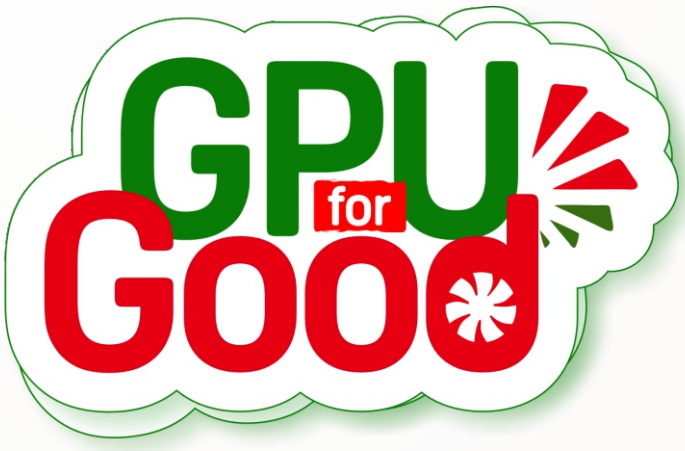
In this way, the role of EqualyzAI in 2025 was both technical and conceptual, developing useful tools while forming a comprehensive view of what responsible AI deployment should look like throughout the continent.

10.2 Platforms, Tools, and Deployments

EqualyzAI's commitment to accessibility took more concrete form through a suite of platforms and tools aimed at lowering barriers to AI participation:

GPU4Good

The **GPU4Good initiative** expanded access to compute resources for African researchers, students, and innovators who often struggle to afford cloud GPUs or high-end hardware. By providing free or subsidized access to training workloads, GPU4Good directly addressed one of the most noticeable disparities in Africa: the high cost of creating or fine-tuning AI models. This initiative complemented regional GPU hubs and cloud expansions discussed earlier and helped ensure that compute does not become a bottleneck to research creativity or startup innovation.



uLearn

Through **uLearn,** EqualyzAI targeted another foundational need: accessible learning resources for AI and data science. The platform provided curated modules, hands-on exercises, and regional examples designed to build practical AI literacy. In a landscape where digital-skills readiness remains uneven, uLearn helped bridge the learning gap between curiosity and capability—particularly for early-career professionals and students in low-resource settings.

Itoju – African TradoMedLLM

EqualyzAI's **Itoju** deployment model further highlighted the organization's focus on responsible and context-aware applications. Itoju is designed to support healthcare workflows by providing localized AI models that assist decision-making without replacing health workers. Its emphasis on patient safety, cultural appropriateness, and transparent model behavior mirrored broader conversations across Africa's health-tech ecosystem—where AI must complement strained health systems, not add complexity or risk.

Together, these platforms illustrate EqualyzAI's operational philosophy: **build tools that work with the constraints of the Global South, not against them; and ensure those tools strengthen local capacity rather than deepen dependency.**

10.3 Identified Gaps and Ecosystem Needs

While 2025 brought remarkable progress, EqualyzAI's engagement with communities, researchers, and institutions faced several persistent challenges that must be addressed to sustain inclusive AI growth.

- **Language and Data Gaps Remain Significant.** Despite major dataset releases, many African languages still lack the depth, diversity, and annotation quality required for robust NLP and ASR systems. Dialect variation and code-switching further complicate model training and evaluation. Therefore, more community-driven, ethically sourced datasets are essential.
- **Compute Access Is Expanding, But Not Equitably.** Although GPU availability has increased, startups, universities, and local research groups still face unequal access.. Public compute infrastructure, especially for experimentation and fine-tuning, needs stronger investment, governance, and accessibility frameworks.

- **Responsible AI Practices Must Be Mainstreamed.** Many organizations still use AI without clear guidelines on consent or data governance. As AI enters critical domains like finance and healthcare, Africa must standardize safety frameworks in accordance with regional realities.
- **Skills Gaps Persist Across Multiple Levels.** Even with the emergence of advanced training programs, basic digital and AI literacy remains uneven across African regions. Therefore, scalable, multilingual training pathways are crucial to guarantee that AI adoption does not exacerbate existing inequalities.
- **Fragmented Regulatory Ecosystems.** National AI strategies and privacy laws advanced in 2025, but regional harmonization remains limited. To support cross-border data flows, benchmark convergence, and shared safety standards, Africa needs more coordinated regulatory frameworks.

APPENDICES

A. Directory of African Language Datasets (2025)

A.1 Pan-African and Multi-Country Datasets

Dataset	Language(s)	Modality	Primary Use Case
Lacuna PII Project (32)	Kanuri, Hausa, Luganda	Text	PII detection, privacy-aligned NLP
Fikira 1.0 (Vambo AI) (33)	Amharic, ChiShona, Hausa, Igbo, Kinyarwanda, isiXhosa, isiZulu, Kiswahili, Tunisian Arabic, Yorùbá	Text	Multilingual reasoning evaluation & LLM alignment
Swahili Thinking (Nadhari AI Lab) (35)	Swahili	Text	Native reasoning traces for Swahili LLMs
NaijaVoices – Nupe (36)	Nupe	Speech	ASR, speech research
Afrispeech Dialog (37)	Multi-language	Speech + Text	Clinical and conversational AI
Ghanaian ASR Resources (38)	Ghanaian languages	Speech	ASR development
African Next Voices (Gates Foundation) (40)	Multiple African languages	Speech	Low-resource language enablement
African_UD (41)	Multiple African languages	Text)	Dependency parsing, NLP modeling
WildJailbreak Africa (42)	Acholi, Lugbara, Luganda, Swahili, Ateso	Text	Safety alignment & instruction tuning for African languages
Lacuna Fund (Kenya) (43)	Kenyan low-resource languages	Speech + Text	Linguistic inclusion
Awarri × N-ATLAS datasets (44)	African languages	Speech	Large-scale AI deployment
N-ATLAS (Open-source) (45)	African languages	Text	National language modeling
NaijaVoices (Community Speech) (46)	Multiple Nigerian languages	Speech	Volunteer-driven speech corpora
Deep Learning Indaba Call (47)	Continental	Speech + Text	Open dataset creation

A.2 North Africa

Country	Dataset	Modality	Focus
Algeria	AFRIHATE (Algerian Arabic)(48)	Text	Hate-speech detection
	FAGE_v2 (Algeria subset)(49)	Images	Facial recognition bias mitigatior
Egypt	FAGE_v2 (Egypt subset)(49)	Images	Demographic representation
	Arabic MMLU updates(50)	Text	Arabic NLP evaluation
Morocco	AFRIHATE (Darija)(48)	Text	Abusive-language detection
Tunisia	LinTO speech & text(51)	Speech + Text	Tunisian Arabic ASR

A.3 West Africa

Country	Dataset	Modality	Focus
Nigeria	BraTS-Africa MRI(52)	Medical imaging	Brain tumor analysis
	NaijaVoices (1,800 hrs)(53)	Speech	Multilingual ASR
	AfricanVoices.io (3,000 hrs)(54)	Speech	Speech-enabled AI
	AFRIHATE (Hausa, Igbo, Pidgin, Yoruba)(48)	Text	Online safety
Ghana	Twi Emotions Corpus(55)	Text	Sentiment analysis
	GSL Lexicon(56)	Sign language	Accessibility
Senegal	WolBanking77(57)	Text	Banking intent classification
Benin	MasakhaPOS– Fon (58)	Text	POS tagging
Togo	MasakhaPOS– Ewe (58)	Text	POS tagging

A.4 East Africa

Country	Dataset	Modality	Focus
Kenya	Next Voices (Kikuyu, Dholuo)(59)	Speech	ASR
	FAGE_v2 (Kenya subset)(49)	Images	Vision bias
	UNHCR Refugee Data(60)	Tabular	Humanitarian analytics
Rwanda	Kinyarwanda ASR (500+ hrs)(61)	Speech	Voice interfaces
	AFRIHATE (Kinyarwanda)(48)	Text	Harmful content
Ethiopia	AFRIHATE (Amharic, Oromo, Tigrinya)(48)	Text	Multilingual moderation
	FAGE_v2 (Ethiopia subset)(49)	Images	Demographic inclusion

Uganda	MakerereNLP MT datasets (62)	Text	Translation
Tanzania	AFRIHATE (Swahili) (48)	Text	Content moderation
Somalia	AFRIHATE (Somali) (48)	Text	Online safety
	Ultra-Poor Graduation Survey (63)	Socio-economic	Development analytics

A.5 Southern Africa

Country	Dataset	Modality	Focus
South Africa	Mafoko Multilingual Terminology (64)	Text	Terminology standardization
	AfroCS-xs (65)	Text	Code-switching NLP
	FAGE_v2 (SA subset) (49)	Images	Vision fairness
Namibia	FAGE_v2 (Namibia subset) (49)	Images	Representation
Zimbabwe / Botswana / Mozambique	HDX population & displacement data (66)	Geospatial	Crisis modeling

A.6 Central Africa

Country	Dataset	Modality	Focus
DRC	FAGE_v2 (DRC subset) (49)	Images	Vision models
	UNHCR displacement data (67)	Tabular	Conflict analytics
Angola	FAGE_v2 (Angola subset) (49)	Images	Biometric fairness
Cameroon	West Africa LULC dataset (68)	Geospatial	Climate & land use
CAR & Chad	UNHCR Regional Response datasets (60)	Tabular	Humanitarian planning

A.7 Hackathons and Community Dataset Initiatives

Initiative	Organizer	Focus
Zindi Speech Dataset Catalogue (69)	Zindi	Dataset discovery
Your Voice Challenge (70)	Zindi	Community speech data
Intro to ASR: African Challenge (71)	Google × Zindi	ASR capacity building
Open Multilingual Speech Fund (72)	Mozilla	Community speech corpora
Masakhane Language Data RFP (73)	Masakhane	Open NLP datasets
WikiIndaba Hackathon (74)	Wikimedia Africa	Open text resources
Nigeria National AI Hackathon (75)	Government	Language AI

B. Directory of African Language AI Models

B.1 Pan-African and Global Models with African Language Coverage

Model	Organization	Language Coverage	Modality	Primary Focus
Omnilingual ASR (76)	Meta	500+ African & low-resource languages	Speech	Universal ASR baseline
AfriqueLLM (77)	McGill NLP	20 African + EN/FR/AR/PT	Text	CPT-adapted open LLMs
Whisper (fine-tuned variants) (98)	Open-source / Research groups	Swahili, Amharic, others	Speech	Lightweight ASR adaptation
Khaya AI Speech Models (78,79)	Khaya AI	32 African languages	Speech	ASR for public services & enterprise
Spitch African Speech Stack (80)	Spitch	Multiple African languages & accents	Speech	Speech-to-text & TTS

B.2 National and Government-Backed Language Models

Model	Country	Organization	Modality	Key Contribution
N-ATLAS (45)	Nigeria	Awarri × NCAIR	Text	Open-source national LLM

B.3 Startup-Led and Private-Sector African LLMs

Model	Organization	Language(s)	Modality	Application Domain
AfroSLM 1.0 (83)	EqualyzAI	African languages (finance-focused)	Text + Multimodal	Fintech & compliance
Swahili Gemma 1B (86)	Crane AI Labs	Swahili (EN↔SW input)	Text	Translation & conversational AI
InkubaLM (compressed) (92)	Lelapa AI	African languages	Text	Efficient deployment
AQL Mind (85)	WideBot	Arabic dialects (MENA & North Africa)	Text	Enterprise LLM
Orange LLM Adaptations (84)	Orange × OpenAI × Meta	Francophone African languages	Text	Telecom & customer services

B.4 Community, Academic, and Grassroots Models

Model	Origin	Language(s)	Modality	Distinguishing Feature
AfroXLMR-Social (82)	Academic research	19 African languages	Text	Domain-adapted social media NLP (DAPT + TAPT)
SabiYarn 125M (89)	Academic research (Nigeria)	EN, YO, HA, IG, Pidgin	Text	Lightweight multilingual Nigerian LLM

Model / System	Sector	Language(s)	Contribution
Indigenius (87)	CDIAL	Low-resource African languages	Speech-first Multilingual grassroots model
YarnGPT (88)	Independent (Nigeria)	Nigerian English, Yoruba, Igbo, Hausa	Text Student-led LLM
Masakhane Experimental Models (91)	Masakhane Community	Multiple African languages	Text Research & benchmarking

B.5 Domain-Specific and Applied Language Models

Model / System	Sector	Language(s)	Contribution
Agricultural RAG Agents (101)	Agriculture	Kenyan languages	Advisory systems
Health Information RAG Systems (102)	Healthcare	Local African languages	Drug & medical guidance
Clinical NLP Systems (103)	Healthcare	Multilingual	Bias-aware deployment

B.6 Evaluation, Safety, and Robustness Models

Model / Study	Language(s)	Focus
Hausa Text Detection (104)	Hausa	Misinformation detection
Culturally Grounded LLM Evaluation (105)	Multilingual	Context-sensitive benchmarking
Yorùbá QA Models Robustness (106)	Yorùbá	Typographical noise tolerance
Exam VQA Evaluation (107)	Multilingual	Education assessment
Ugandan Cultural Context Benchmark (UCCB) (119)	Ugandan English + local language elements	Country-level cultural reasoning and contextual LLM evaluation

C. Research Methodology Tables

Table D1. Data Collection and Source Overview

Methodological Componen	Description
Data Sources	Peer-reviewed publications, technical reports, official websites, open datasets, institutional and organizational releases (2025)
Modalities Covered	Text, speech, vision, multimodal, tabular
Geographic Scope	Pan-African, with country-level granularity where applicable
Language Coverage	African local languages, lingua francas, and multilingual benchmarks
Inclusion Criteria	Public availability, relevance to African AI ecosystems, documented methodology

Table D2. Literature Review Methodology

Dimension	Approach
Review Type	Structured narrative review
Time Frame	January 2025 – December 2025
Databases Consulted	arXiv, ACL Anthology, IEEE Xplore, institutional repositories
Screening Criteria	Relevance, methodological rigor, reproducibility, regional grounding
Output	Thematic synthesis across datasets, models, infrastructure, and policy

D. Glossary of Key Terms

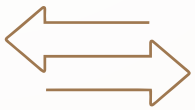
Artificial Intelligence (AI)
Computer systems that are designed to perform tasks that typically need human intelligence, such as reasoning, perception, learning, and language understanding.

Natural Language Processing (NLP)
A subfield of AI focused on enabling computers to process, understand, and generate human language.

Large Language Model (LLM)
A neural network that is trained on large text corpora to generate, summarize, translate, or reason over natural language.

Automatic Speech Recognition (ASR)
Technology that converts spoken language into written text.

Small Language Model (SLM)
A compact language model optimized for efficiency, enabling deployment in low-resource or constrained compute environments.



Code-Switching
The practice of switching between two or more languages or dialects within a single conversation.

Low-Resource Language
A language with limited digitally available data, tools, or annotated resources for AI training.

Model Compression
Techniques that reduce model size and computational requirements while maintaining its performance.

Multimodal AI
AI systems that integrate multiple data modalities, such as text, images, speech, or video.

Dataset Curation
The process of collecting, cleaning, annotating, and validating data for AI training and evaluation.

Federated Learning
A decentralized machine learning approach where models are trained across multiple devices or organizations without the need to centralize raw data.

Digital Inclusion
Efforts to ensure equitable access to digital technologies, skills, and services across populations.

GPU (Graphics Processing Unit)
A specialized processor optimized for parallel computation, commonly used for training and deploying AI models.

Responsible AI
The design and deployment of AI systems that are ethical, transparent, and aligned with societal values.

Compute Infrastructure
The hardware, software, and energy resources required to train, run, and scale AI systems.

Downstream Tasks
Applications or evaluations performed using a trained model, such as translation, sentiment analysis, or question answering.

Digital Sovereignty
A state's ability to control its digital infrastructure, data, and technologies in alignment with national laws, values, and strategic interests.

Data Governance
The frameworks and rules that define how data is collected, stored, shared, and used.

Public–Private Partnership (PPP)
A collaboration between government entities and private-sector organizations so as to sponsor, design or provide digital and AI-enabled services.

Biometric Verification
The use of biological or behavioral characteristics like voice, fingerprints, or facial features for identity verification.

Digital Literacy
The ability to understand, use, and interact with digital technologies, including AI-based systems as well as awareness of data privacy.

Virtual Wards
AI-based healthcare models that enable remote monitoring and care of patients outside hospital settings.

Endnotes

1. Ashiru G. Nvidia Partners with Cassava to Build AI-Ready Data Centres Across Africa in US$700 Million Deal [Internet]. Tech In Africa. 2025. Available from: https://www.techinafrica.com/nvidia-partners-with-cassava-to-build-ai-ready-data-centres-across-africa-in-us700-million-deal/
2. Agency E. Ecofin Agency. Atlancis Technologies Deploys GPU-Powered AI Factory in Kenya. Available from: https://www.ecofinagency.com/news-digital/1011-50320-atlancis-technologies-deploys-gpu-powered-ai-factory-in-kenya
3. As WTDC-25 ends, report highlights ongoing digital development gaps - Developing Telecoms [Internet]. Available from: https://www.developingtelecoms.com/telecom-business/telecom-regulation/19431-as-wtdc-25-ends-report-highlights-ongoing-digital-development-gaps.html
4. Oni A. MTN to launch data centre to boost Nigeria's tech ecosystem [Internet]. TheCable. 2025. Available from: https://www.thecable.ng/mtn-to-launch-data-centre-to-boost-nigerias-tech-ecosystem/
5. Eleanya F. Inside Itana, Nigeria's first digital free zone [Internet]. TechCabal. 2025. Available from: https://techcabal.com/2025/06/10/inside-itana-nigerias-first-digital-free-zone/
6. Kansiime T. Nilepost News. Uganda's Health Research Leaps Forward with New US-Backed Supercomputing Facility at Makerere. Available from: https://nilepost.co.ug/index.php/health/261504/ugandas-health-research-leaps-forward-with-new-us-backed-supercomputing-facility-at-makerere
7. Eleanya F. Nigeria's next big export isn't oil, it's AI computing power [Internet]. TechCabal. 2025. Available from: https://techcabal.com/2025/10/23/nigerias-next-big-export-isnt-oil-its-ai-computing-power/
8. Co-creating for Africa's AI Ecosystem: Infrastructure for Tomorrow, Resources for Today [Internet]. Available from: https://www.aihubfordevelopment.org/co-creating-for-africas-ai-ecosystem-infrastructure-for-tomorrow-resources-for-today
9. Google Cloud [Internet]. Google Cloud pour les chercheurs. Available from: https://cloud.google.com/edu/researchers
10. Spotlight on innovation: Google-sponsored Data Science for Health Ideathon across Africa [Internet]. Available from: https://research.google/blog/spotlight-on-innovation-google-sponsored-data-science-for-health-ideathon-across-africa/
11. Altron launches SA's first operational AI Factory delivering enterprise-grade AI Infrastructure, platform and services [Internet]. Available from: https://www.altron.com/press-releases/altron-launches-south-africas-first-operational-ai-factory-delivering-enterprise-grade-ai-infrastructure-platform-and-services
12. Saaia-admin. Cassava Technologies partners with the SAAIA to grow AI access [Internet]. SAAIA. 2025. Available from: https://saaiassociation.co.za/cassava-technologies-partners-with-the-south-african-artificial-intelligence-association-to-boost-local-access-to-ai-compute-services/
13. Jack G. Cassava Tech - Collaborates with Zindi to showcase African AI innovation [Internet]. Cassava Tech. 2025. Available from: https://www.cassavatechnologies.com/cassava-technologies-collaborates-with-zindi-to-showcase-african-ai-innovation/

14. Writer S. ITWeb Africa. 2025. Accelerator unlocks compute to boost Africa's AI innovators. Available from: https://itweb.africa/article/accelerator-unlocks-compute-to-boost-africas-ai-innovators/lwrKxq3Ya1J7mg1o
15. AI Hub [Internet]. Available from: https://www.aihubfordevelopment.org/programmes/compute-accelerator-programme
16. Gupta R, Bhatia G. Agentic Educational Content Generation for African Languages on Edge Devices. In: 2025 IEEE Global Humanitarian Technology Conference (GHTC) [Internet]. 2025 [cited 2026 Jan 2]. p. 191–2. Available from: http://arxiv.org/abs/2511.07437
17. (5) Post | LinkedIn [Internet]. Available from: https://www.linkedin.com/posts/adiack_big-news-in-the-edge-ai-space-my-journey-activity-7384367034521767936-f_O8/
18. Cassava AI - forges ahead with its GPU-as-a-Service offering [Internet]. 2025. Available from: https://www.cassava.ai/2025/11/11/forges-ahead-with-gpu-as-a-service/
19. Musoba J. Fastagger, Microsoft AI to Help Africa's $1.7T Mobile Money Market [Internet]. TechMoran. 2025. Available from: https://techmoran.com/2025/11/06/fastagger-and-microsoft-bring-edge-ai-to-africas-1-7-trillion-mobile-money-market/
20. Solution: The GPUAI Protocol | GPU AI [Internet]. 2025. Available from: https://docs.gpuai.space/solution-the-gpuai-protocol
21. OverView | NodeGo Documentation [Internet]. 2024. Available from: https://docs.nodego.ai/nodego-ecosystem/overview
22. What Is Distributed Machine Learning? | IBM [Internet]. 2025. Available from: https://www.ibm.com/think/topics/distributed-machine-learning
23. VC4A [Internet]. The AI Hub for Sustainable Development, Accelerator on VC4A. Available from: https://vc4a.com/the-ai-hub-for-sustainable-development/
24. Folorunso SO, Sina AT, Oladipo FO, Rukayat GO. Green Topics, Deep Roots: Energy-Aware Topic Modelling of Multilingual Nigerian Lyrics. In 2025 [cited 2026 Jan 2]. Available from: https://openreview.net/forum?id=tYC19w4d1D
25. GPU4Good - EqualyzAI [Internet]. 2025. Available from: https://equalyz.ai/gpu4good/
26. Africa Compute Fund [Internet]. Africa Compute Fund. Available from: https://www.africacomputefund.com
27. Acalytica [Internet]. [cited 2026 Jan 2]. frontier-ai-accelerator. Available from: https://acalytica.com/frontier-ai-accelerator
28. Bisong E. Google Colaboratory. In: Bisong E, editor. Building Machine Learning and Deep Learning Models on Google Cloud Platform: A Comprehensive Guide for Beginners [Internet]. Berkeley, CA: Apress; 2019 [cited 2025 May 12]. p. 59–64. Available from: https://doi.org/10.1007/978-1-4842-4470-8_7
29. Writer S. Google Cloud Celebrates First Cloud Region In Africa [Internet]. CIO Africa. 2025. Available from: https://cioafrica.co/google-cloud-celebrates-first-cloud-region-in-africa/
30. Writer S. Microsoft Ignite 2025 Unveils AI Innovations [Internet]. CIO Africa. 2025. Available from: https://cioafrica.co/microsoft-ignite-2025-unveils-ai-innovations/
31. Africa D. Dawan Africa. 2025. Microsoft's Cloud Surge Redefines Global Tech, But Africa's Data Future Still Lags Behind. Available from: https://dawan.africa/news/microsofts-cloud-surge-redefines-global-tech-but-africas-data-future-still-lags-behind

32. mschink. Lacuna Fund Releases 18 New AI Datasets Empowering Local Communities to Tackle Challenges in Agriculture, Climate, Health and Language [Internet]. Lacuna Fund. 2025. Available from: https://lacunafund.org/lacuna-fund-releases-18-new-ai-datasets-empowering-local-communities-to-tackle-challenges-in-agriculture-climate-health-and-language/
33. vamboai/fikira · Datasets at Hugging Face [Internet]. Available from: https://huggingface.co/datasets/vamboai/fikira
34. Nadhari/Swahili-Thinking · Datasets at Hugging Face [Internet]. [cited 2026 Jan 9]. Available from: https://huggingface.co/datasets/Nadhari/Swahili-Thinking
35. Introducing The Swahili Thinking Dataset – Nadhari AI Lab [Internet]. Available from: https://www.nadhari.ai/swahili-thinking.html
36. Mozilla Data Collective [Internet]. Speech Data Collection for The Nupe Language. Available from: https://datacollective.mozillafoundation.org/datasets/cmihkoth10246md07dtojxehg
37. intronhealth/afrispeech-dialog · Datasets at Hugging Face [Internet]. 2025. Available from: https://huggingface.co/datasets/intronhealth/afrispeech-dialog
38. Wiafe I, Abdulai JD, Ekpezu AO, Helegah RD, Atsakpo ED, Nutrokpor C, et al. Advancing automatic speech recognition for low-resource ghanaian languages: Audio datasets for Akan, Ewe, Dagbani, Dagaare, and Ikposo. Data Brief. 2025 Aug 1;61:111880.
39. Lanfrica Labs [Internet]. Available from: https://lanfrica.com/en
40. Ouitona S. L'intelligence artificielle en langues africaines : quand le continent décide de rattraper son retard [Internet]. Afrik. 2025. Available from: https://www.afrik.com/l-intelligence-artificielle-en-langues-africaines-quand-le-continent-decide-de-rattraper-son-retard
41. The Center for Digital Humanities at Princeton [Internet]. African_UD: Universal Dependencies Treebank for African Languages. Available from: https://cdh.princeton.edu/projects/african_ud-universal-dependencies-treebank-for-african-languages/
42. CraneAILabs/wildjailbreak-africa · Datasets at Hugging Face [Internet]. Available from: https://huggingface.co/datasets/CraneAILabs/wildjailbreak-africa
43. Stasimioti M. New Dataset for Low-Resource Kenyan Language Tackles AI Exclusion [Internet]. Slator. 2025. Available from: https://slator.com/new-dataset-for-low-resource-kenyan-language-tackles-ai-exclusion/
44. Koigi MR| B. MAfrica Business Communities. MTN backs African language datasets for AI development. Available from: https://africabusinesscommunities.com/artificial-intelligence/mtn-backs-african-language-datasets-for-ai-development/
45. Sen A. Nigeria Unveils Open Source N-ATLAS AI Model To Champion African Languages At UNGA80 - Open Source For You [Internet]. 2025. Available from: https://www.opensourceforu.com/2025/09/nigeria-unveils-open-source-n-atlas-ai-model-to-champion-african-languages-at-unga80/
46. Emezue C, Community N, Awobade B, Owodunni A, Emezue H, Emezue GMT, et al. The NaijaVoices Dataset: Cultivating Large-Scale, High-Quality, Culturally-Rich Speech Data for African Languages [Internet]. arXiv; 2025 [cited 2026 Jan 2]. Available from: http://arxiv.org/abs/2505.20564
47. African Datasets [Internet]. Deep Learning Indaba 2025. Available from: https://deeplearningindaba.com/2025/african-datasets/

48. Muhammad SH, Abdulmumin I, Ayele AA, Adelani DI, Ahmad IS, Aliyu SM, et al. AfriHate: A Multilingual Collection of Hate Speech and Abusive Language Datasets for African Languages [Internet]. arXiv; 2025 [cited 2026 Jan 2]. Available from: http://arxiv.org/abs/2501.08284
49. FAGE_v2 Dataset [Internet]. Available from: https://www.kaggle.com/datasets/ajewoleolaitan/fage-dataset
50. Koto F, Li H, Shatnawi S, Doughman J, Sadallah AB, Alraeesi A, et al. ArabicMMLU: Assessing Massive Multitask Language Understanding in Arabic [Internet]. arXiv; 2024 [cited 2026 Jan 2]. Available from: http://arxiv.org/abs/2402.12840
51. Naouara H, Lorré JP, Louradour J. LinTO Audio and Textual Datasets to Train and Evaluate Automatic Speech Recognition in Tunisian Arabic Dialect [Internet]. arXiv; 2025 [cited 2026 Jan 5]. Available from: http://arxiv.org/abs/2504.02604
52. Adewole M, Rudie JD, Gbadamosi A, Zhang D, Raymond C, Ajigbotoshso J, et al. The BraTS-Africa Dataset: Expanding the Brain Tumor Segmentation Data to Capture African Populations. Radiol Artif Intell. 2025 Jul;7(4):e240528.
53. naijavoices/naijavoices-dataset-compressed · Datasets at Hugging Face [Internet]. Available from: https://huggingface.co/datasets/naijavoices/naijavoices-dataset-compressed
54. Nigeria DS. African Voices. African Voices - Multilingual Speech Dataset. Available from: https://africanvoices.io
55. ghananlpcommunity/twi-emotions-corpus-400k · Datasets at Hugging Face [Internet]. [cited 2026 Jan 2]. Available from: https://huggingface.co/datasets/ghananlpcommunity/twi-emotions-corpus-400k
56. Data sets [Internet]. Available from: https://www.universiteitleiden.nl/sign-languages/datasets/data
57. Kandji AK, Precioso F, Ba C, Ndiaye S, Ndione A. WolBanking77: Wolof Banking Speech Intent Classification Dataset. In 2025 [cited 2026 Jan 2]. Available from: https://openreview.net/forum?id=7k0JBDeHAv
58. masakhane/masakhapos · Datasets at Hugging Face [Internet]. 2025. Available from: https://huggingface.co/datasets/masakhane/masakhapos
59. TRT Afrika - AI is finally trying to speak African languages. Will this end a historic neglect? [Internet]. Available from: https://www.trtafrika.com/english/article/359e1362af39
60. Power BI Report [Internet]. Available from: https://app.powerbi.com/view?r=eyJrIjoiZGYyNWVhYTEtNzkwNS00MDRhLWJjYmMtM2M4MjQ2NWUzMDBiIiwidCI6ImU1YzM3OTgxLTY2NjQtNDEzNC04YTBjLTY1NDNkMmFmODBiZSIsImMiOjh9
61. Kinyarwanda Automatic Speech Recognition Track A [Internet]. Available from: https://kaggle.com/kinyarwanda-automatic-speech-recognition-track-a
62. Masakhane - MakerereNLP: Text & Speech for East Africa [Internet]. Available from: https://www.masakhane.io/ongoing-projects/makererenlp-text-speech-for-east-africa
63. IFPRI Datasets [Internet]. IFPRI. Available from: https://www.ifpri.org/publications/datasets/
64. Marivate V, Dzingirai I, Banda FE, Richard L, SINDANE T, Madumo K, et al. Mafoko Companion Dataset for Mafoko Open Multilingual Terminologies Paper [Internet]. Zenodo; 2025 [cited 2026 Jan 3]. Available from:

https://zenodo.org/records/17484992
65. Olaleye K, Oncevay A, Sibue M, Zondi N, Terblanche M, Mapikitla S, et al. AfroCS-xs: Creating a Compact, High-Quality, Human-Validated Code-Switched Dataset for African Languages. In: Che W, Nabende J, Shutova E, Pilehvar MT, editors. Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers) [Internet]. Vienna, Austria: Association for Computational Linguistics; 2025 [cited 2026 Jan 3]. p. 33391–410. Available from: https://aclanthology.org/2025.acl-long.1601/
66. Humanitarian Datasets | Find Crisis Data | HDX [Internet]. Available from: https://data.humdata.org/dataset
67. Region Regional Bureau for West and Central Africa [Internet]. Available from: https://data.unhcr.org/en/situations/rbwca
68. West Africa Land Use Land Cover Time Series [Internet]. U.S. Geological Survey; 2020 [cited 2026 Jan 3]. Available from: https://catalog.data.gov/dataset/west-africa-land-use-land-cover-time-series
69. Zindi. 10 best African language datasets for data science projects [Internet]. Available from: https://zindi.africa/learn/10-best-african-language-datasets-for-data-science-projects
70. Zindi. Your Voice, Your Device, Your Language Challenge ☎ - Win 1 000 CHF [Internet]. Available from: https://zindi.africa/competitions/your-voice-your-device-your-language-challenge
71. Zindi. Google NLP Hack Series: Intro to ASR Africa Challenge - Win $3 000 USD [Internet]. Available from: https://zindi.africa/competitions/google-asr-hack-series-africa-asr-data-challenge
72. Mozilla Foundation [Internet]. Programmatic work. Available from: https://www.mozillafoundation.org/en/common-voice/in-country-programmes/
73. Masakhane - Announcements [Internet]. Available from: https://www.masakhane.io/masakhane-african-languages-hub/announcements
74. Event:WikiIndaba 2025 Hackathon Pre-Event - Meta-Wiki [Internet]. Available from: https://meta.wikimedia.org/wiki/Event:WikiIndaba_2025_Hackathon_Pre-Event
75. Nigeria National AI Hackathon 2025 – Community [Internet]. Available from: https://deep-communities.ai/hackathon/nigeria-national-ai-hackathon-2025/
76. Bhati D. India Today. 2025. Meta claims its new AI can understand more than 1600 languages, is superintelligence next? Available from: https://www.indiatoday.in/technology/news/story/meta-claims-its-new-open-source-ai-can-understand-more-than-1600-languages-is-superintelligence-next-2817258-2025-11-11
77. AfriqueLLM - a McGill-NLP Collection [Internet]. 2026. Available from: https://huggingface.co/collections/McGill-NLP/afriquellm
78. Ghana NLP. Khaya African Language Translation and Speech Recognition AI Demonstrates Major Improvements [Internet]. Medium. 2022. Available from: https://nlpghana.medium.com/khaya-african-language-translation-and-speech-recognition-ai-demonstrates-major-improvements-931711300497
79. (10) Post | LinkedIn [Internet]. Available from: https://www.linkedin.com/posts/paulazunre_announcing-speech-

recognition-and-generation-activity-7392686968364736512-7kD3/?utm_source=share&utm_medium=member_desktop&rcm=ACoAABze08EB_u4jYuf7rrm5BehtVFRAiYNRsRE
80. Omoniyi F. AI is learning to speak African languages, thanks to these startups [Internet]. TechCabal. 2025. Available from: https://techcabal.com/2025/07/17/african-ai-startups/
81. Babatunde OB, Olufemi VT, Bolarinwa E, Moshood KY, Emezue CC. Beyond Monolingual Limits: Fine-Tuning Monolingual ASR for Yoruba-English Code-Switching. In: Winata GI, Kar S, Zhukova M, Solorio T, Ai X, Hamed I, et al., editors. Proceedings of the 7th Workshop on Computational Approaches to Linguistic Code-Switching [Internet]. Albuquerque, New Mexico, USA: Association for Computational Linguistics; 2025 [cited 2026 Jan 9]. p. 18–25. Available from: https://aclanthology.org/2025.calcs-1.3/
82. Belay TD, Azime IA, Ahmad IS, Adelani DI, Abdulmumin I, Ayele AA, et al. AfroXLMR-Social: Adapting Pre-trained Language Models for African Languages Social Media Text [Internet]. arXiv; 2025 [cited 2026 Jan 9]. Available from: http://arxiv.org/abs/2503.18247
83. admin. EqualyzAI Unveils AI Solutions to Break Language Barriers and Drive Inclusion in Africa - [Internet]. 2025. Available from: https://fintechmagazine.africa/2025/02/23/equalyzai-unveils-ai-solutions-to-break-language-barriers-and-drive-inclusion-in-africa/
84. Rousseau R. Orange to expand open-source AI models to African regional languages for digital inclusion [Internet]. Newsroom Groupe Orange. 2024. Available from: https://newsroom.orange.com/?p=54149
85. Dabafinance [Internet]. Dabafinance - Egypt's Widebot Raises $3M to Build Arabic Large Language Model. Available from: https://dabafinance.com/en/news/egypts-widebot-raises-3m-to-build-arabic-large-language-model
86. Google DeepMind [Internet]. Swahili Gemma 1B from Crane AI Labs bridges the gap with powerful on-device performance. Available from: https://deepmind.google/models/gemma/gemmaverse/craneai/
87. Aimuengheuwa J. CDIAL AI Launches Indigenius AI Agent Pro [Internet]. Techeconomy. 2025. Available from: https://techeconomy.ng/cdial-ai-launches-indigenius-ai-agent-pro/
88. Emmanuel V. Inside YarnGPT: The text-to-speech model built for Nigerian languages [Internet]. Condia. 2025 [cited 2026 Jan 3]. Available from: https://thecondia.com/yarn-gpt-nigerian-text-to-speech/
89. Otoibhi J, Damilola O, David O. SabiYarn: Advancing Low Resource Languages with Multitask NLP Pretraining. In: Lignos C, Abdulmumin I, Adelani D, editors. Proceedings of the Sixth Workshop on African Natural Language Processing (AfricaNLP 2025) [Internet]. Vienna, Austria: Association for Computational Linguistics; 2025 [cited 2026 Jan 9]. p. 95–107. Available from: https://aclanthology.org/2025.africanlp-1.14/
90. White Papers - EqualyzAI [Internet]. 2025. Available from: https://equalyz.ai/white-papers/
91. Some African NLP Datasets That You Can Use To Build African AI | Lanfrica Docs [Internet]. 2025. Available from: https://lanfrica.com/docs/some-african-nlp-datasets-that-you-can-use-to-build-african-ai
92. user. African Innovation: Major Breakthrough In AI For Africa LLM [Internet]. Broadcast Media Africa. 2025. Available from: https://news.broadcastmediaafrica.com/2025/06/10/african-innovation-major-breakthrough-in-ai-for-africa-llm/
93. Akpobi M. Yankari: Monolingual Yoruba Dataset. In: Lignos C, Abdulmumin I, Adelani D, editors. Proceedings of the Sixth Workshop on African Natural

Language Processing (AfricaNLP 2025) [Internet]. Vienna, Austria: Association for Computational Linguistics; 2025 [cited 2026 Jan 3]. p. 1–6. Available from: https://aclanthology.org/2025.africanlp-1.1/
94. Nigusie G. Supervised Machine Learning based Amharic Text Complexity Classification Using Automatic Annotator Tool. In: Lignos C, Abdulmumin I, Adelani D, editors. Proceedings of the Sixth Workshop on African Natural Language Processing (AfricaNLP 2025) [Internet]. Vienna, Austria: Association for Computational Linguistics; 2025 [cited 2026 Jan 3]. p. 7–14. Available from: https://aclanthology.org/2025.africanlp-1.2/
95. Yamoah KA, Weako J, Dorley E. Building a Functional Machine Translation Corpus for Kpelle. In: Lignos C, Abdulmumin I, Adelani D, editors. Proceedings of the Sixth Workshop on African Natural Language Processing (AfricaNLP 2025) [Internet]. Vienna, Austria: Association for Computational Linguistics; 2025 [cited 2026 Jan 3]. p. 52–63. Available from: https://aclanthology.org/2025.africanlp-1.8/
96. Moteu Ngoli T, Christabel M, Yopa N. Challenges and Limitations in Gathering Resources for Low-Resource Languages: The Case of Medumba. In: Lignos C, Abdulmumin I, Adelani D, editors. Proceedings of the Sixth Workshop on African Natural Language Processing (AfricaNLP 2025) [Internet]. Vienna, Austria: Association for Computational Linguistics; 2025 [cited 2026 Jan 3]. p. 136–42. Available from: https://aclanthology.org/2025.africanlp-1.18/
97. Justin BE, Xegbe KF, Essuman CNN, Samuel AKM. YodiV3: NLP for Togolese Languages with Eyaa-Tom Dataset and the Lom Metric. In: Lignos C, Abdulmumin I, Adelani D, editors. Proceedings of the Sixth Workshop on African Natural Language Processing (AfricaNLP 2025) [Internet]. Vienna, Austria: Association for Computational Linguistics; 2025 [cited 2026 Jan 3]. p. 143–9. Available from: https://aclanthology.org/2025.africanlp-1.20/
98. Sharma AK, Pandya M, Shukla A. Fine-tuning Whisper Tiny for Swahili ASR: Challenges and Recommendations for Low-Resource Speech Recognition. In: Lignos C, Abdulmumin I, Adelani D, editors. Proceedings of the Sixth Workshop on African Natural Language Processing (AfricaNLP 2025) [Internet]. Vienna, Austria: Association for Computational Linguistics; 2025 [cited 2026 Jan 3]. p. 74–81. Available from: https://aclanthology.org/2025.africanlp-1.11/
99. Nigatu HH, Aldarmaki H. Exploring Transliteration-Based Zero-Shot Transfer for Amharic ASR. In: Lignos C, Abdulmumin I, Adelani D, editors. Proceedings of the Sixth Workshop on African Natural Language Processing (AfricaNLP 2025) [Internet]. Vienna, Austria: Association for Computational Linguistics; 2025 [cited 2026 Jan 3]. p. 64–73. Available from: https://aclanthology.org/2025.africanlp-1.10/
100. Imam SH, Sani B, Gete DK, Ahmed BY, Ahmad IS, Abdulmumin I, et al. Automatic Speech Recognition for African Low-Resource Languages: Challenges and Future Directions. In: Lignos C, Abdulmumin I, Adelani D, editors. Proceedings of the Sixth Workshop on African Natural Language Processing (AfricaNLP 2025) [Internet]. Vienna, Austria: Association for Computational Linguistics; 2025 [cited 2026 Jan 3]. p. 89–94. Available from: https://aclanthology.org/2025.africanlp-1.13/
101. Owiti TL, Kipkebut AK. Enhancing AI-Driven Farming Advisory in Kenya with Efficient RAG Agents via Quantized Fine-Tuned Language Models. In: Lignos C, Abdulmumin I, Adelani D, editors. Proceedings of the Sixth Workshop on African Natural Language Processing (AfricaNLP 2025) [Internet]. Vienna, Austria: Association for Computational Linguistics; 2025 [cited 2026 Jan 3]. p. 24–30. Available from: https://aclanthology.org/2025.africanlp-1.5/

102. Ismail AI, Ibrahim BO, Adekanmbi O, Adebara I. Retrieval-Augmented Generation Meets Local Languages for Improved Drug Information Access and Comprehension. In: Lignos C, Abdulmumin I, Adelani D, editors. Proceedings of the Sixth Workshop on African Natural Language Processing (AfricaNLP 2025) [Internet]. Vienna, Austria: Association for Computational Linguistics; 2025 [cited 2026 Jan 3]. p. 108–14. Available from: https://aclanthology.org/2025.africanlp-1.15/
103. Okafor U. Multilingual NLP for African Healthcare: Bias, Translation, and Explainability Challenges. In: Lignos C, Abdulmumin I, Adelani D, editors. Proceedings of the Sixth Workshop on African Natural Language Processing (AfricaNLP 2025) [Internet]. Vienna, Austria: Association for Computational Linguistics; 2025 [cited 2026 Jan 3]. p. 221–9. Available from: https://aclanthology.org/2025.africanlp-1.32/
104. Sani B, Soy A, Imam SH, Mustapha A, Aliyu LJ, Abdulmumin I, et al. Who Wrote This? Identifying Machine vs Human-Generated Text in Hausa. In: Lignos C, Abdulmumin I, Adelani D, editors. Proceedings of the Sixth Workshop on African Natural Language Processing (AfricaNLP 2025) [Internet]. Vienna, Austria: Association for Computational Linguistics; 2025 [cited 2026 Jan 3]. p. 82–8. Available from: https://aclanthology.org/2025.africanlp-1.12/
105. Ochieng M, Gumma V, Sitaram S, Wang J, Chaudhary V, Ronen K, et al. Beyond Metrics: Evaluating LLMs Effectiveness in Culturally Nuanced, Low-Resource Real-World Scenarios. In: Lignos C, Abdulmumin I, Adelani D, editors. Proceedings of the Sixth Workshop on African Natural Language Processing (AfricaNLP 2025) [Internet]. Vienna, Austria: Association for Computational Linguistics; 2025 [cited 2026 Jan 3]. p. 230–47. Available from: https://aclanthology.org/2025.africanlp-1.33/
106. Okewunmi P, James F, Fajemila O. Evaluating Robustness of LLMs to Typographical Noise in Yorùbá QA. In: Lignos C, Abdulmumin I, Adelani D, editors. Proceedings of the Sixth Workshop on African Natural Language Processing (AfricaNLP 2025) [Internet]. Vienna, Austria: Association for Computational Linguistics; 2025 [cited 2026 Jan 3]. p. 195–202. Available from: https://aclanthology.org/2025.africanlp-1.29/
107. Olufemi VT, Babatunde OB, Bolarinwa E, Moshood KY. Challenging Multimodal LLMs with African Standardized Exams: A Document VQA Evaluation. In: Lignos C, Abdulmumin I, Adelani D, editors. Proceedings of the Sixth Workshop on African Natural Language Processing (AfricaNLP 2025) [Internet]. Vienna, Austria: Association for Computational Linguistics; 2025 [cited 2026 Jan 3]. p. 150–7. Available from: https://aclanthology.org/2025.africanlp-1.22/
108. Muhammad SH, Ahmad IS, Abdulmumin I, Lawan FI, Imam SH, Aliyu Y, et al. HausaNLP: Current Status, Challenges and Future Directions for Hausa Natural Language Processing. In: Lignos C, Abdulmumin I, Adelani D, editors. Proceedings of the Sixth Workshop on African Natural Language Processing (AfricaNLP 2025) [Internet]. Vienna, Austria: Association for Computational Linguistics; 2025 [cited 2026 Jan 3]. p. 176–91. Available from: https://aclanthology.org/2025.africanlp-1.27/
109. Nguefack IN, Finkelstein M, Sakayo TS. Pretraining Strategies using Monolingual and Parallel Data for Low-Resource Machine Translation. In: Lignos C, Abdulmumin I, Adelani D, editors. Proceedings of the Sixth Workshop on African Natural Language Processing (AfricaNLP 2025) [Internet]. Vienna, Austria: Association for Computational Linguistics; 2025 [cited 2026 Jan 3]. p. 31–8. Available from: https://aclanthology.org/2025.africanlp-1.6/

110. Costa-jussà MR, Chen J, Adebara I, Chuang J, Ropers C, Sánchez E. Y-NQ: English-Yorùbá Evaluation dataset for Open-Book Reading Comprehension with Open-Ended Questions. In: Lignos C, Abdulmumin I, Adelani D, editors. Proceedings of the Sixth Workshop on African Natural Language Processing (AfricaNLP 2025) [Internet]. Vienna, Austria: Association for Computational Linguistics; 2025 [cited 2026 Jan 3]. p. 248–54. Available from: https://aclanthology.org/2025.africanlp-1.34/
111. Singh PR, Prasad K, Zaki M, Wasnik P. In-Domain African Languages Translation Using LLMs and Multi-armed Bandits [Internet]. arXiv; 2025 [cited 2026 Jan 3]. Available from: http://arxiv.org/abs/2505.15069
112. Alnumay Y, Barbet A, Bialas A, Darling W, Desai S, Devassy J, et al. Command R7B Arabic: A Small, Enterprise Focused, Multilingual, and Culturally Aware Arabic LLM [Internet]. arXiv; 2025 [cited 2026 Jan 3]. Available from: http://arxiv.org/abs/2503.14603
113. Ogunshakin N. Scripting Africa's AI Future: UNILAG and OpenAI Host Africa's First OpenAI Academy [Internet]. University Of Lagos. 2025. Available from: https://unilag.edu.ng/scripting-africas-ai-future-unilag-and-openai-host-africas-first-openai-academy/
114. Onyeagoro J. Nigeria Receives N3 Billion Investment from Google.org to Advance Digital Transformation [Internet]. TechAfrica News. 2025. Available from: https://techafricanews.com/2025/12/01/nigeria-receives-n3-billion-investment-from-google-org-to-advance-digital-transformation/
115. Masakhane - Ongoing Projects [Internet]. Available from: https://www.masakhane.io/ongoing-projects
116. Sahara Benchmark [Internet]. Available from: https://africa.dlnlp.ai/sahara/
117. Adelani DI, Ojo J, Azime IA, Zhuang JY, Alabi JO, He X, et al. IrokoBench: A New Benchmark for African Languages in the Age of Large Language Models. In: Chiruzzo L, Ritter A, Wang L, editors. Proceedings of the 2025 Conference of the Nations of the Americas Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 1: Long Papers) [Internet]. Albuquerque, New Mexico: Association for Computational Linguistics; 2025 [cited 2026 Jan 5]. p. 2732–57. Available from: https://aclanthology.org/2025.naacl-long.139/
118. AfroBench [Internet]. Available from: https://mcgill-nlp.github.io/AfroBench/index.html
119. CraneAILabs/UCCB · Datasets at Hugging Face [Internet]. Available from: https://huggingface.co/datasets/CraneAILabs/UCCB
120. fmcwp. Nigeria Launches Landmark AI Model Powered by Awarri to Advance Languages and AI at Scale [Internet]. The Federal Ministry of Communications, Innovation and Digital Economy. 2025. Available from: https://fmcide.gov.ng/nigeria-launches-landmark-ai-model-powered-by-awarri-to-advance-languages-and-ai-at-scale/
121. Uganda launches an Artificial Intelligence (AI) language mod... | Ministry of ICT and National Guidance [Internet]. Available from: https://ict.go.ug/media/news/uganda-launches-an-artificial-intelligence-ai-language-model
122. Shikwambane N, https://elixirdsgn.com ED Developer: Ronnie Mbugua, Website: Africa's First Multilingual Small Language Model (SLM) Gets Even Smaller - Thanks to Top African Innovators [Internet]. Lelapa. 2025. Available from: https://lelapa.ai/africas-first-multilingual-small-language-model-slm-gets-even-smaller-thanks-to-top-african-innovators/
123. Jacaranda Health • AI for Global Health Research [Internet]. [cited 2026 Jan 5]. Available from: https://ai-globalhealthresearch.tghn.org/partners/jacaranda-health/

124. Standard B. African languages for AI: The project that's gathering a huge new dataset [Internet]. 2025. Available from: https://www.business-standard.com/technology/tech-news/african-languages-for-ai-the-project-that-s-gathering-a-huge-new-dataset-125101900137_1.html
125. Onyeagoro J. Liquid C2 Launches AI-Powered Google Cloud Distribution Programme "Liquid G" for Africa [Internet]. TechAfrica News. 2025. Available from: https://techafricanews.com/2025/12/01/liquid-c2-launches-ai-powered-google-cloud-distribution-programme-liquid-g-for-africa/
126. Mbego S. Safaricom Unveils $500M AI Strategy [Internet]. CIO Africa. 2025. Available from: https://cioafrica.co/safaricom-unveils-500m-ai-strategy/
127. Militzer J. Africa is Done Waiting: 10 Key Takeaways from Africa HealthTech Summit 2025 [Internet]. NextBillion. 2025 [cited 2026 Jan 5]. Available from: https://nextbillion.net/africa-is-done-waiting-10-key-takeaways-from-africa-healthtech-summit-2025/
128. Google Cloud Press Corner [Internet]. Vodacom Announces Multi-Year Strategic Collaboration with Google Cloud to Boost Africa's AI Advancement. Available from: https://www.googlecloudpresscorner.com/2025-11-25-Vodacom-Announces-Multi-Year-Strategic-Collaboration-with-Google-Cloud-to-Boost-Africas-AI-Advancement
129. Partner. Nexus 2025: Qore's flagship conference sets a new standard for Africa's credit infrastructure [Internet]. TechCabal. 2025. Available from: https://techcabal.com/2025/11/26/qore-flagship-conference/
130. Kennedy F. African Startups Raise $140 Million in September, Bringing 2025 Total Funding to $2.2 Billion [Internet]. Empower Africa. 2025. Available from: https://empowerafrica.com/african-startups-raise-140-million-in-september-bringing-2025-total-funding-to-2-2-billion/
131. Founders A. Fintech and AI Lead: African Startups Raise $140 Million in September, Pushing 2025 Total to $2.2 Billion [Internet]. ArabFounders. 2025. Available from: https://arabfounders.net/en/african-startups-funding-september-2025/
132. Supporting the future of AI Research in Africa and globally [Internet]. Available from: https://blog.google/intl/en-africa/company-news/outreach-and-initiatives/supporting-the-future-of-ai-research-in-africa-and-globally/
133. Onyeagoro J. Benin Signs Landmark Agreements to Expand Trust Services and Accelerate Digital Transformation [Internet]. TechAfrica News. 2025. Available from: https://techafricanews.com/2025/07/17/benin-signs-landmark-agreements-to-expand-trust-services-and-accelerate-digital-transformation/
134. KingFadray. Google commits N3bn to advance Nigeria's national AI strategy [Internet]. STAR 101.5 FM. 2025. Available from: https://starfm.ng/2025/11/29/google-commits-n3bn-to-advance-nigerias-national-ai-strategy/
135. UNDP [Internet]. AFRI CONVERSE2025#4 - Africa Taking off with AI: When the Fastest-growing Continent Adopts the Potential of AI. Available from: https://www.undp.org/africa/events/afri-converse20254-africa-taking-ai-when-fastest-growing-continent-adopts-potential-ai
136. TECNO and UNICEF Deepen Partnership to Expand 'Learning Passport' Programme for More Children Across Africa [Internet]. Available from: https://www.unicef.org/nigeria/press-releases/tecno-and-unicef-deepen-partnership-expand-learning-passport-programme-more-children
137. amogelangmaluleka. Nigeria–South Africa Investment Dialogue Advances Regional Cooperation | [Internet]. 2025. Available from: https://www.mtn.com/nigeria-south-africa-investment-dialogue-advances-

regional-cooperation/
138. UNDP [Internet]. [cited 2026 Jan 6]. Key takeaways from the Global AI Summit on Africa. Available from: https://www.undp.org/rwanda/blog/key-takeaways-global-ai-summit-africa
139. Amos. Nigeria's AI moment: From national strategy to real impact [Internet]. The Guardian Nigeria News - Nigeria and World News. 2025 [cited 2026 Jan 6]. Available from: https://guardian.ng/opinion/columnists/nigerias-ai-moment-from-national-strategy-to-real-impact/
140. Bimonthly Update on Privacy in Africa (January - February 2025) [Internet]. Available from: https://www.techhiveadvisory.africa/insights/bimonthly-update-on-privacy-in-africa-january---february-2025
141. UNESCO and G20 promote Linguistic Diversity in Digital Age | UNESCO [Internet]. Available from: https://www.unesco.org/en/articles/unesco-and-g20-promote-linguistic-diversity-digital-age
142. Cent Capital [Internet]. [cited 2026 Jan 6]. Cent Capital. Available from: https://www.cent.capital/news/financial-fraud-consumer-protection/phishing-cybercrime-finance/ai-fraud-crisis-hits-africas-fintech-20251201
143. Artificial Intelligence in the South African Financial Sector. [cited 2026 Jan 6]; Available from: https://www.resbank.co.za/en/home/publications/publication-detail-pages/media-releases/2025/artificial-intelligence-in-the-south-african-financial-sector
144. Aimuengheuwa J. Top 10 AI Startups to Watch in 2025 [Internet]. Techeconomy. 2025. Available from: https://techeconomy.ng/top-10-ai-startups-to-watch-in-2025/
145. Bassey C, Oluigbo T. Powering a Better Future [Internet]. Available from: https://deepdive.nigeriahealthwatch.com/powering-a-better-future/
146. AI In Healthcare Market Size & Share | Industry Report, 2033 [Internet]. Available from: https://www.grandviewresearch.com/industry-analysis/artificial-intelligence-ai-healthcare-market
147. Adepoju L. InnTech Summit 2025 Concludes with Strong Calls for Responsible AI, Digital Literacy, and Sustainable Tech Adoption - EduTimes Africa [Internet]. 2025. Available from: https://edutimesafrica.com/inntech-summit-2025-concludes-with-strong-calls-for-responsible-ai-digital-literacy-and-sustainable-tech-adoption/
148. Africa P. Africa's Education Revolution: Skills, Tech, and AI in 2025 [Internet]. Powering Africa. 2025 [cited 2026 Jan 6]. Available from: https://powering.africa/africas-education-revolution-skills-tech-and-ai-in-2025/

equalyzAI

EqualyzAI Inc.
USA: 2001 L Street N.W. Washington DC 20036, USA
Nigeria: 33, Queens Street, Yaba, Lagos, Nigeria.
www.equalyz.ai