

Harnessing AI and Data in African Libraries for Africa's Development

Dr Edmund Balnaves
ebalnaves@prosentient.com.au
Director, Prosentient Systems Pty Ltd
Chair, Information Technology Section, IFLA

Abstract

This paper explores the critical role African libraries can play in harnessing artificial intelligence and data to advance sustainable development across the continent. Positioned as trusted, equitable community institutions, libraries are uniquely equipped to mediate the adoption of AI technologies in ways that align with public good, ethical standards, and local needs. The analysis moves beyond a narrow focus on generative AI to consider a broader ecosystem, including machine learning for image recognition, big data analytics, and—crucially—AI-powered translation tools. This linguistic dimension is particularly vital in sub-Saharan Africa, where extreme linguistic diversity presents a significant barrier to information access. By leveraging open-source platforms and cloud-based services, libraries can deploy AI to translate and localize content, thereby dramatically widening access to educational, governmental, and cultural resources in local languages.

The successful integration of AI depends on African libraries actively building agency and local capacity. This involves engaging in foundational work to develop regionally contextualized large language models, applying retrieval-augmented generation to enrich global models with local knowledge, and utilizing AI for enhanced discovery and classification of cultural heritage materials. While challenges such as funding constraints, data silos, and nascent digitization efforts exist, a pathway forward is demonstrated through emerging initiatives and international collaborations, such as the IFLA symposia held across African universities. Ultimately, by adopting a framework of governance, education, and experimentation, libraries can ensure AI serves as a multiplier for the visibility and accessibility of Africa's rich research and cultural output, directly contributing to the continent's inclusive and equitable development.

Keywords: Artificial Intelligence, Big Data, Sub-Saharan Africa, Large Language models

AI has many dimensions

When we discuss Artificial Intelligence, it is easy to be focussed on the Generative systems used in text, audio and image applications. These are important, and valuable, but by no means the only parts of AI that are relevant to the library.

Harnessing AI and Data in African Libraries for Africa's Development

The United Nations has over the last 10 years been building on the 17 sustainability development goals that set a groundwork for planning around development in all regions of the world (UNDP (2017), UN (2025)) For libraries specifically the International Federation of Library Associations recognises the centrality of access to information as part of the UN Sustainability Goals (IFLA (2025)). This paper explores the ways in which Artificial Intelligence and Big Data can contribute to these goals in sub-Saharan Africa. The library is uniquely situated in this context with its historical centrality as a trusted resource for access to information. Artificial Intelligence has been evolving for over 20 years (Balnave (2025)). The use of AI for image recognition (including facial recognition), AI is becoming pervasive across many different applications, and as a result its accessibility has grown apace. Mobile device providers have moved rapidly to position products with embedded AI, leveraging the ability to describe and modify images. This has opened many areas of challenging ethical, cultural and legal issues, not the least around copyright and intellectual property rights.

The pace of development has been very rapid, and libraries face both a challenge and an opportunity in leveraging AI to fulfill their missions for accessibility and access to information.

Trust, the library and Artificial Intelligence

There are many contexts in which AI can enhance the work of libraries in supporting development in the African context. Libraries are a key source of trusted access to information. Numerous research studies point to the important place the library holds particularly in a world of complex information and mis-information. Fox(2017) Public libraries have a key role in providing equitable access to information. Libraries in educational information provide access to important and trusted research portals. Libraries in government can provide access to rich archives of historical reports by government instrumentalities.

This trust positions them uniquely to support users in adopting artificial intelligence technologies. Libraries are widely regarded as neutral, public-interest institutions that prioritise intellectual freedom, privacy, and equitable access to information. When libraries introduce AI tools or services, users are more likely to engage with them because they trust that these technologies are being selected, implemented, and explained in ways that serve user needs rather than commercial or surveillance interests. This credibility allows libraries to act as safe entry points for AI exploration, particularly for users who may be sceptical of large technology companies or unsure about the implications of AI systems.

Trust also enables libraries to play a critical educational role in helping users understand how AI works and how it should be used responsibly. Because libraries are trusted sources of information literacy instruction, users are more receptive to guidance on topics such as algorithmic bias, data quality, transparency, and ethical use of AI-generated content. Librarians can frame AI not as an opaque or intimidating technology, but as a tool that requires critical evaluation—much like traditional information sources. This trusted instructional relationship empowers users to adopt AI thoughtfully, with an awareness of both its capabilities and its limitations.

Finally, trust in libraries supports inclusive and equitable AI adoption by ensuring that user interests remain central. Libraries' strong commitments to user privacy and confidentiality reassure patrons that their data will be protected when interacting with AI-enabled services. This is especially important for vulnerable or marginalised communities who may be disproportionately affected by misuse of AI or data exploitation. By embedding AI within trusted library services and values, libraries can reduce fear, build confidence, and foster informed participation in an AI-enabled information environment, helping users adopt these technologies in ways that are ethical, confident, and aligned with the public good.

Language and AI

AI-powered translation has significant value in the African context, where extraordinary linguistic diversity coexists with limited access to professional translation resources. With over two thousand languages spoken across the continent, many of which are under-resourced in digital environments, AI translation tools offer a scalable way to bridge communication gaps. These systems can support everyday access to information in health, education, agriculture, and government services by making content available in local and regional languages. This helps reduce reliance on colonial or global lingua franca alone and enables more people to engage with information in the languages they use most confidently.

In education and knowledge dissemination, AI translation can play a transformative role by expanding access to learning materials and research. Academic content, training resources, and open educational materials can be translated into African languages more rapidly and affordably than through traditional methods. This supports multilingual education initiatives and helps preserve and legitimise local languages in formal and digital settings. In research and cultural heritage contexts, AI-assisted translation also enables African knowledge, stories, and scholarship to be shared more widely, both within the continent and globally, fostering greater cross-cultural understanding.

AI translation also has strong economic and social implications across Africa. By lowering language barriers,

it can support regional trade, cross-border collaboration, and participation in the digital economy. Small businesses, NGOs, and community organisations can use AI tools to communicate across linguistic boundaries without prohibitive costs. While challenges remain—such as uneven accuracy for low-resource languages and the need for community involvement to ensure cultural and linguistic integrity.

Much of the work in language integration by major vendors such as Zoom, Wordly and Grammarly is driven by the commercial imperative of market size. Sadly, few have recognised the possibilities of the sub-Saharan market place.

Zoom for instance has now integrated live transcription captioning. Out of the 37 languages currently supported by Zoom, not one comes from an African context (excluding English and European languages). Wordly has a longer history in translation systems and includes Swahili, Zulu and Afrikaans. Grammarly does not include any sub-Saharan languages. Google translate however includes an impressive range of languages including Acholi, Alur, Baoulé, Bemba, Dinka, Domba, Dyula, Ga, Kanuri, Kiga, Kokongo, Kituba, Luo, Nuer, Nko, Rundi, Sango, Susa, Swati, Tiv, Tswana, Tumbuka, Venda, Wolof and others.

AI translation represents a powerful tool for inclusion, empowerment, and linguistic equity when developed and deployed in partnership with African language communities.

Open source systems are a key entry point for libraries. Systems such as Koha (for traditional library services) and DSpace (for digital libraries) provide an open platform for translation. Leveraging tools such as Google translate and the open translation platform provided by open source library systems represents a huge opportunity for African libraries to widen their accessibility. AI can play a key role to this transition by providing rapid addition of translated language-context sensitive systems.

Generative AI

The early stages of Generative AI have favoured easily accessible information sources in English language. While this is a point of criticism of AI, this early evolution of large language models has served as a proving ground for establishing the value of AI. It is important to remember that this aspect of AI is in its relative infancy.

Large language models (LLMs) have expanded into multiple languages primarily through the use of vast, multilingual training datasets that include text from many linguistic and cultural contexts. By training on diverse corpora—ranging from books and news articles to web content and public documents—LLMs learn shared patterns across languages, such as grammar, semantics, and discourse structures. This approach allows a single model to support dozens or even hundreds of languages, including low-resource languages that previously lacked sufficient data for standalone language technologies. As a result, multilingual LLMs can translate, summarise, and generate content across languages with increasing fluency and contextual awareness.

Advances in model architecture and training techniques have further accelerated multilingual expansion. Techniques such as cross-lingual transfer learning enable knowledge gained in high-resource languages to improve performance in related or less-resourced languages. Subword tokenisation methods allow models to handle diverse writing systems and morphological structures efficiently, reducing the need for language-specific engineering. More recently, instruction tuning and reinforcement learning across multilingual prompts have improved the models' ability to follow instructions consistently across languages, making them more practical for real-world use in education, government, and research.

As we have seen, the expansion of LLMs into multiple languages has also been driven by growing demand for inclusive and globally accessible AI systems. Organisations developing LLMs increasingly prioritise linguistic diversity to reduce digital inequality and support local knowledge creation. This can lead to partnerships with cultural institutions, universities, and open-data initiatives to incorporate non-English and regionally specific content. While challenges remain—such as uneven performance quality and cultural bias—the multilingual growth of LLMs represents a significant step toward more equitable access to AI-powered tools and information worldwide.

Data depth

All AI depends on Big Data to train models that can then be used to answer questions or undertake tasks (identify this image, answer this question, construct this image). The larger the data source, the more effective the potential for training large language models and machine learning data sets. Generative AI systems have leveraged crawling systems to assembly content across the vast resources of the internet.

Big data in libraries remains under-explored territory (Garoufallou, E., & Gaitanou, P. (2021)). Ironically, the broad crawling systems of generative AI systems have made better usage of library data than libraries themselves. Data sources in research data systems includes analytical usage data as well as a great depth of research results and publication information. Big data relevant to libraries is also situated in the wider context of demographic and government information relevant to decision making for libraries (Ong (2013), Lun et al (2017), Lee (2023)). Libraries should not forget the opportunities of big data in open government demographic and population open government resource.

The silo-ed nature of repositories in Africa, the nascent digitisation of assets, combined with the great diversity of languages may appear to the evolution of these resources as Big Data. It is a hurdle bounded by time and effort. African institutions are actively engaged in exploring the potential for AI by leveraging the benefits of the communities of effort already being undertaken. African institutions are actively exploring Big Data and AI. Since 2020 the IFLA Information Technology Section has been promoting a series of Symposia exploring various dimensions of AI and Robots in libraries. This has included three symposia specifically in the African context. The Section has collaborated with South African Universities in a series of three Symposia on Robotics and AI. Their second IT section symposium was held at University of Pretoria in University of Pretoria, Department of Library Services. (2021). The 3rd IFLA Symposium on Artificial Intelligence was held in North West University in 2023 (IFLA (2023)). The 5th IFLA Symposium on AI was held on 16–20 June, 2025 hosted by the University of KwaZulu-Natal (UKZN) Library in Pietermaritzburg, South Africa ([IFLA](#) (2025)). A further symposium is planned for September 2026. This sort of international collaborative engagement exemplifies the ways in which African libraries are engaging in the complex area of AI adoption.

A great example of leveraging multiple facets of bug data is the Zamani Project. Using laser scanning, photogrammetry, satellite imaging, GIS mapping and panoramic photography 3d models are built. This includes terabytes of 3d point-cloud and spatial data that can be purposed in many ways. Uses of this big data include:

- conservation modelling
- monitoring environmental damage
- digital tourism and virtual access
- research datasets for archaeology and architecture



3D Model of the Nelson Mandela statue created by Xhanti Mpakama. Zamani Project (2026)

African reports and research

Educational institutions and Governments in Africa are a rich source of information. Stellenbosch University in Cape Town was an early adopter of open source for their journal management and publishing frameworks. The SunScholar repository now represents an excellent academic repository with over 57thousand publications. Parliamentary libraries in have also been active in adoption repositories for legislative output. Health libraries are leveraging evidence-based research to enhance their discovery systems. Nevertheless funding remains an impediment to AI adoption (Lee(2023)). This can be ameliorated by the use of cloud-based and open source systems (Balnaves (2024)).

Training AI in African-context information

To engage in the generative context for enrichment of African resources, Africa institution swill need to engage in Large language models (LLMs) across sub-Saharan Africa as tools - to expand access to information, education, and public services. Vulavula is an example of a large language model built to the African context by Lelapa AI (Ravindran, Sandeep (2023) Lelapa(2025)). Universities and research institutions are already leveraging the major LLMs such as GPT, LLAMA to support teaching and learning through automated tutoring, research assistance, and academic writing support, particularly in contexts where staff-to-student ratios are high.

The next stage for African development is to engage in the work of language-specific enhanced use of AI. (Hussen, Sewunetie et al. 2025) highlight the need for foundational linguistic work to enhance the development of African models.

By enabling interaction in widely spoken regional and local languages alongside English, French, and Portuguese, LLMs help reduce linguistic barriers and improve inclusivity for students and researchers who have historically been underserved by global digital platforms. In government agencies and public institutions, LLMs have the potential to improve service delivery, knowledge management, and citizen engagement. Chatbots and virtual assistants can help citizens navigate administrative processes, access health or agricultural information, and understand legal or regulatory requirements. For institutions with limited human resources, LLMs offer the potential to scale expertise cost-effectively, while also supporting internal workflows such as document summarisation, policy drafting, and data analysis. This can only be achieved by strong investment in this area.

Agentic AI

The term Agentic AI refers to the process of giving active agency to AI in workflow and task processing. The potential for Agentic AI is only now emerging. It has implications for the library in a number of workflow tasks, including cataloguing, research services support and various library service delivery functions. The degree to which Agentic AI represents an opportunity for enhanced service delivery or a risk to library staff is not clear at this time. The ethical elemetns surrounding generative AI have even greater relevance to Agentic AI. In particular, the unsupervised use of Agentic AI with autonomous operation with little or no human interaction should be carefully scrutinised.

Harnessing AI and Data in African Libraries for Africa's Development.docx

Cultural Heritage issues

Institutions are increasingly aware of the need to align these uses with local governance frameworks, data protection laws, and ethical guidelines. Cultural heritage institutions, libraries, and NGOs in sub-Saharan Africa are also leveraging LLMs to support digitisation, translation, and preservation of indigenous knowledge. By assisting with transcription, metadata creation, and multilingual translation, LLMs can help make local archives and research outputs more visible both regionally and globally. At the same time, institutions are actively engaging with questions of bias, data sovereignty, and sustainability, recognising that effec-

tive use of LLMs requires local capacity building, contextual training data, and partnerships that place African institutions in control of how these technologies are developed and deployed. IFLA has established the IFLA Digital Heritage Network. This is a further area where the intersection of AI and Digital Cultural Heritage can be explored.

Conclusion

There are many pathways to harnessing AI in a productive way. Many of the AI toolkits are open source, including those for machine learning, translation and model generation. Others have cloud-based services that can come at quite low cost. These enable early experimentation with AI even when the institution is not ready for full AI adoption. University libraries can partner productively with their IT departments to explore and discover AI services.

Some of the steps in the adoption of AI include:

Governance: establishing the principles around use of AI in your institution

Education: digital literacy education that encompasses AI

Engagement: exploring collaborative development of systems for your institution.

Experimentation: undertaking engagement with AI systems by leveraging open source and accessible cloud-based trial systems.

Evaluation: continually review the AI deployment to ensure right-sizing of the investment for the institution.

The sub-Saharan African region has over 1500 spoken languages across many linguistic groups. This paper has explored the opportunities for this region to actively engage in AI to support the goals of development in the region. Critical in this process is the development of regional tools for better language and cultural awareness, and the leveraging of existing tools to enhance the language translation of resources. This will be an enabler of the progressive expansion of the accessibility and visibility of the rich diversity of resources available in this region.

Leveraging AI offers the potential provide a significant multiplier in the reach of the legislative, report and research output of the region. There are many entry points to exploring this opportunity. They involve engaging in AI to explore its possibilities, as many institutions in Africa already are. This engagement can be done at many levels:

- By the use of existing large language models and using Retrieval Augmented Generation (RAG) to enhance these models with institutional and regional content
- By building local large language models and training models with African content
- By deploying AI language translation tools to extend the reach of African content
- To deploy machine learning AI tools for image and text classification and discovery
- Through the use of AI in chat services to extend the reach of the institution

The best way to manage risks around AI is to have agency in its use and adoption by understanding its functionality and utility toward enhancing discoverability of African research within and outside Africa, broadening the accessibility of the rich cultural resources of the region for the benefit of African development. Many aspects of AI, including generative AI are in their early stages still, and the ethno-centricity of the current AI models will be ameliorated by the active agency that African institutions take in the adoption of AI.

References

- Abba, T. (2024). Use of artificial intelligence technologies in rendering library services: An empirical evidence from university libraries in Africa. *African Journal of Library, Archives and Information Science*, 34(1), 23-35.
- Balnaves, E (2024). "New Horizons in Artificial Intelligence in Libraries" De Gruyter Saur. <https://repository.ifla.org/items/0695f7ce-d945-468b-8768-737a8321df11>
- Balnaves, E (2025).
- Fox, Katie (2017). Libraries are the Most Highly Trusted Resource for Information on Recent Pew Survey. *Library Research Service* September 14, 2017. <https://www.lrs.org/2017/09/14/libraries-are-most-highly-trusted-resource-for-information-on-recent-pew-survey/>
- Garoufallou, E., & Gaitanou, P. (2021). Big Data: Opportunities and Challenges in Libraries, a Systematic Literature Review. *College & Research Libraries*, 82(3), 410. doi:<https://doi.org/10.5860/crl.82.3.410>
- Hussen, K. Y., W. T. Sewunetie, A. A. Ayele, S. H. Imam, S. H. Muhammad and S. M. Yimam (2025). "The State of Large Language Models for African Languages: Progress and Challenges." <https://doi.org/10.48550/arXiv.2506.02280>
- IFLA (2023) *3rd IFLA Symposium on Artificial Intelligence*. IFLA. <https://www.ifla.org/events/3rd-ifla-symposium-on-artificial-intelligence> IFLA
- IFLA & University of KwaZulu-Natal Library Services. (2025). *5th IFLA Artificial Intelligence Symposium*. <https://ailib.ukzn.ac.za/> (Ailib)
- IFLA. (2025). *Powering sustainable development*. IFLA. Retrieved January 4, 2026, from <https://www.ifla.org/units/sustainable-development/>
- Lee, Jungyeoun, and Nahyun Kwon. "The development of a public library data collection and analysis platform and its uses: a case study of the National Library Big Data Platform in Korea." *Malaysian Journal of Library and Information Science* 28.2 (2023): 89-102.
- Lelapa AI. (2025). *Lelapa AI – Resource-efficient AI solutions from Africa to the world*. <https://lelapa.ai/Lelapa>
- Note:
- Lee, J., & Kwon, N. (2023). The development of a public library data collection and analysis platform and its uses: a case study of the National Library Big Data Platform in Korea. *Malaysian Journal of Library and Information Science*, 28(2), 89-102.
- Lee, Jungyeoun, and Nahyun Kwon. "The development of a public library data collection and analysis platform and its uses: a case study of the National Library Big Data Platform in Korea." *Malaysian Journal of Library and Information Science* 28.2 (2023): 89-102.
- Lu, N., Song, R., Heng, D. L. G., Gottipati, S., & Tay, A. (2017). Using data analytics for discovering library resource insights: Case from Singapore Management University.
- Ong, G. (2014). A smarter library: using data analytics to improve resource management and services at NLB. *Singapore: Civil Service College*.
- Ravindran, Sandeep (2023) Frustrated that AI tools rarely understand their native languages, thousands of African volunteers are taking action By Sandeep Ravindran. *Science* 381 (6655), pp 262-265 <https://www.science.org/doi/pdf/10.1126/science.adj8519>
- UNDP. (2017). *The sustainable development goals*. Stylus Publishing, LLC.

UN (2025). *The 17 goals*. <https://sdgs.un.org/goals>

University of Pretoria, Department of Library Services. (2021). *International Symposium on Robotics and AI in Libraries and the Education Sector*. University of Pretoria. <https://www.up.ac.za/departments-of-library-services/news/call-paperspresentations-international-symposium-robotics-and-ai-libraries-and-education>

Zamani Project (2026). “3D model of Nelson Mandela statue”. <https://www.zamaniproject.org/site-south-africa-cape-town-city-hall.html>