Online Article

Generative AI in Global Scientific Knowledge Production: Challenges and Prospects for Africa

This paper examines the transformative impact of large language models (LLMs) like GPT-4 on global scientific research, and highlights the potential benefits and the significant ethical challenges for Africa. It explores how generative AI can enhance Africa's scientific knowledge production by incorporating underrepresented sources like oral history and grey literature. It also addresses the risks of perpetuating Western-centric biases in AI models and emphasises the need for localised AI training and governance to protect African researchers' intellectual property rights. The paper advocates for international collaboration, open-source initiatives and grassroots engagement to ensure Africa's active participation in global scientific discourse, thereby enriching it with diverse perspectives.

Introduction

Generative artificial intelligence (AI) is revolutionising global scientific knowledge production through advancements in large language models (LLMs) like OpenAI's GPT-4, Google's BERT and Meta's Llama. These technologies not only transform the sourcing, production and dissemination of scientific knowledge but also raise crucial ethical questions about inclusivity and representation. Africa, representing 19 per cent of the world's population yet contributing only 1.2 per cent to global scientific output, faces a pivotal moment. This moment presents opportunities for innovation and significant ethical challenges, particularly in underrepresented regions of the world. How can African researchers use AI to incorporate oral history and grey literature, essential sources of information of-

Jacob Jaygbay*

Senior Content Developer, Cloud + AI, Microsoft Virginia, USA

ten overlooked by global databases? Al algorithms are increasingly involved in knowledge production in several disciplines, including healthcare, climate change or social science research. The paper explores how these tools can be used to augment research activities and enhance scientific knowledge production in Africa.

AI algorithms, some built on LLMs, rely heavily on data mined from existing literature and online sources, which cover little of Africa's scientific knowledge and even misrepresent it. This underrepresentation of African scientific research output in global datasets where LLMs mine their training

data poses significant challenges to how Africa's contribution to global scientific research is measured in the generative AI era. With African scholars producing only a fraction of what is published in globally recognised scientific journals, the training data for these AI algorithms risks perpetuating a Western-centric view of knowledge. The potential impact of this underrepresentation is broadly three-fold:

1. **Marginalised perspectives:** Important insights and Indigenous knowledge from Africa may be overshadowed.

2. **Bias:** The resulting AI models might inadvertently introduce or perpetuate biases that further marginalise African perspectives.

3. **Access to information:** Given that AI is increasingly used for decision-making, lack of representation could affect Africa's say in global policy-making processes that are AI driven.

Optimise scholarly knowledge production cost through AI

Incorporating generative AI into scientific research and writing presents significant opportunities and challenges for knowledge production in Africa, where research infrastructure is still developing. AI tools can enhance the efficiency and quality of science research outputs by automating tasks such as summarising, translating and generating content (Conde et al. 2024a, 2024b). Let us begin with how AI-powered machine translation can significantly improve the cost of production in research output across Africa's four official languages—Arabic, English, French and Portuguese—through three broad applications.

Reduce translation costs: Machine translation tools powered by AI can reduce the costs associated with human translation services. Researchers often need to translate their work to make it accessible to a broader audience, which can be expensive when using professional translators. AI tools like GPT-4 can quickly and efficiently translate large volumes of text, allowing researchers to allocate their budgets to other critical areas of their projects. This cost-saving measure is especially beneficial for researchers with limited funding, enabling them to produce and share more work without the financial burden of manual translation.

AI translation tools can also process multiple languages simultaneously, reducing the need for multiple translators proficient in different languages. This streamlining of the translation process cuts down on administrative and logistical expenses, further lowering overall production costs. By using generative AI for translation, researchers can ensure that their work is disseminated across diverse linguistic audiences without incurring prohibitive costs.

Enhance collaboration: AI-driven machine translation improves accessibility, allowing researchers to share their findings with colleagues across different linguistic backgrounds. This fosters greater collaboration and knowledge exchange among researchers who speak Arabic, English, French and Portuguese, for example. By breaking down language barriers, AI facilitates the pooling of resources, expertise and data, leading to more comprehensive and diverse research outcomes. Enhanced collaboration can lead to more innovative solutions and discoveries, benefiting the research community.

AI translation tools can also improve researchers' access to work published in other languages, expanding their literature review and ensuring that they are up to date with the latest developments in their disciplines globally. This expanded access can accelerate the pace of research and reduce duplication of efforts, making the research process more efficient and cost-effective. African researchers can build on each other's work more effectively, driving progress and innovation.

Improve productivity: AI-powered machine translation significantly reduces the time required to translate research documents, articles and publications. Manual translation can be time-consuming, delaying the dissemination of research findings and hindering timely collaboration. AI tools can provide near-instant translations, allowing researchers to quickly share and publish their work. This time efficiency boosts productivity, enabling researchers to focus on their core tasks such as data analysis, experimentation and writing, rather than spending valuable time on translation.

The increased speed of translation also means that researchers can engage in real-time collaborations and discussions with their peers across different linguistic regions. This immediacy can enhance the quality and relevance of research, in that timely feedback and input can be incorporated into ongoing projects. By improving the speed and efficiency of communication, AI translation tools help researchers stay agile and responsive to added information and developments, enhancing the overall quality and impact of their work.

Thus, machine translation not only reduces production costs but also promotes a more inclusive and interconnected research community, driving innovation and progress in various disciplines of study.

AI governance: IP rights for African researchers

There are governance and intellectual property (IP) rights issues that are unresolved in this new generative AI era, which leave African scientific researchers vulnerable. Africa will need a united front to tackle these issues to protect its researchers.

The Malabo Convention is a legal framework for data protection and cybersecurity that was enacted in June 2023. It has been ratified by thirty-six of fifty-four African countries and was recently ratified by the African Union (AU). Although the convention is a commendable point of departure, it does not explicitly address IP rights for researchers within its provisions. The primary focus of the Convention is on establishing secure cyberspace, combatting cybercrime and protecting personal

data. It emphasises the establishment of national cybersecurity frameworks, protection of personal data and promotion of cybersecurity awareness and education but does not detail provisions regarding IP rights for researchers or other stakeholders. IP rights are crucial for researchers, especially in protecting their innovations and creations. Let us look at five of these IP issues that affect African researchers.

1. **Ownership and consent:** Data ownership and consent are foundational to the ethical use of AI in research. For African researchers, it is essential to have a clear understanding of who owns the data used to train and operate AI models like GPT- 4. Without clear ownership, there can be disputes over the rights to the data and the findings derived from it. In addition, researchers must ensure that the data they use is collected with explicit consent, respecting the privacy and rights of individuals and communities involved. This is especially pertinent in the African context, where data sovereignty and the rights of local populations need to be safeguarded against exploitation.

2. **Attribution and authorship:** When using generative AI tools, African researchers must navigate the complexities of determining how to appropriately credit contributions made by AI. Traditional notions of authorship may not fully account for the role of AI in generating content, which may lead to disputes over intellectual contributions. Furthermore, ensuring that human researchers receive proper recognition for their work, while also acknowledging the role of AI, is essential for maintaining the integrity of academic and scientific endeavours.

3. **Licensing and usage rights:** Understanding the licensing and usage rights associated with AI tools like GPT-4 is crucial for African researchers. These tools often come with specific terms of service and licensing agreements that dictate how the AI can be used, particularly regarding commercial applications. Researchers must be aware of these terms to avoid inadvertently violating licensing agreements that may undermine their research efforts. Clear comprehension of licensing terms ensures that researchers can use AI tools effectively while respecting IP laws.

4. **Bias and fair representation:** Bias in AI-generated outputs can significantly impact the credibility and fairness of research. African researchers must be vigilant about the potential for bias in AI tools that can perpetuate existing inequalities and misrepresent certain groups or perspectives. Ensuring fair representation in research output involves critically evaluating AI-generated content and supplementing it with diverse perspectives and rigorous analysis. Addressing bias is crucial for maintaining the integrity and ethical standards of research, especially in contexts where social and cultural nuances are important.

5. **Compliance with local and international IP laws:** Navigating the complexities of local and international intellectual property (IP) laws is a significant concern for African researchers using generative AI tools. Understanding these laws is essential for safeguarding intellectual property rights and avoiding legal disputes that could hinder the dissemination and commercialisation of research findings. Compliance with IP laws ensures that the AI-augmented work of African researchers is legally protected and recognised on a global scale.

A way forward

The good news is that generative AI's impact on academic research in Africa is nascent. By responding collectively in collaboration with IT professionals, research institutions, grassroots organisations and African scientific researchers, Africa can ride the tide of this new AI technology to increase the quota of Africa's contribution to global scientific knowledge production. Here are three recommendations to consider.

1. **Train LLMs locally with African scientific research content:** The first step towards resolution is to train AI models on more localised and diversified datasets, incorporating literature and oral traditions specific to Africa. Projects that digitise African languages, for instance, could make local insights more accessible, thus making them available for training AI systems. Any of the major players in the LLM space would benefit from such collaboration if done at the pan-African institutional or consortia levels. A few of the following pan-African research institutions could organise an African response: The Council for the Development of Social Science Research in Africa (CODESRIA), in Dakar, Senegal; the African Academy of Sciences (AAS), in Nairobi, Kenya; and the Council for Scientific and Industrial Research (CSIR), in Pretoria, South Africa.

2. **Tap into international open-source AI projects:** The purpose of international collaboration in research is to foster cooperation and the exchange of knowledge and resources between researchers from different countries or regions. It allows researchers to use their diverse abilities, perspectives and resources to address com-

plex scientific questions and challenges that may require multidisciplinary approaches. International collaboration can lead to the sharing of data, methodologies and best practices, accelerating the pace of scientific discovery and innovation. It also promotes cultural understanding, builds scientific networks and enhances the global impact and relevance of research findings. Additionally, international collaboration can help address global issues that require collective efforts, such as climate change, infectious diseases and sustainable development. A collective global effort towards open-source AI can also help in diluting the centralised control of knowledge. By making algorithms and training data publicly accessible, open-source projects can ease the inclusion of diverse perspectives, including those from African researchers.

3. **Engage grassroots stakeholders:** In addition to institutional collaboration, encouraging joint ventures among AI developers and technology companies investing across the continent, local African science researchers and research institutions can produce models that are fine-tuned to the cultural and contextual sensitivities of the continent. Such partnerships could help prevent ethical oversights while respecting and incorporating Indigenous knowledge systems and grey literature.

Although the current infrastructure and data biases risk further marginalising Africa, concerted efforts among African stakeholders can reverse this trend. Localising AI training, fostering grassroots collaboration, enabling policy interventions and supporting open-source movements can prevent the erosion of African perspectives and significantly enrich global scientific discourse. In an interconnected world, the production of knowledge is a collective endeavour that is most fruitful when it is inclusive, diverse and fair.

To succeed, African governments, research institutions and researchers must collectively invest in addressing the resource-intensive requirements of these LLMs. These include new computers equipped with not only Central Processing Units (CPUs) and Graphic Processing Units (GPUs) but also Neural Processing Units (NPUs) for processing AI inquiries. There is a lot to learn, and one must resist the urge to kowtow to the brazen promise of generative AI for advancing scientific research in Africa. However, Africa cannot afford to be a spectator. Africa must forge new alliances outside the traditional scientific research community and learn together with the rest of the world as generative AI transforms the way we produce, retrieve, and consume scientific knowledge.

*Jacob Jaygbay is a software engineer and currently a Senior Content Developer at Microsoft. He is a former Editor at the Council for the Development of Social Research in Africa (CODESRIA) in Dakar, Senegal. He studied software engineering at George Mason University and holds an MPhil degree in Digital Publishing from the University of Stirling, Scotland, UK. He has published multiple articles on scholarly knowledge production and self-censorship in Africa.

Bibliography

Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., and Amodei, D., 2020, 'Language models are few-shot learners', *Advances in Neural Information Processing Systems,* Vol. 33, pp. 1877–1901.

Conde, J., Martínez, G., Reviriego, P., Salvachúa, J. and Hernández, J. A., 2024a, 'Designing Metadata for the Use of Artificial Intelligence in Academia', Paper delivered at the 39th ACM/SIGAPP Symposium on Applied Computing (SAC '24), 8–12 April 2024, Avila, Spain. New York: ACM.

Conde, J., Reviriego, P., Salvachúa, J., Martínez, G., Hernández, J. A. and Lombardi, F., 2024b, 'Understanding the Impact of Artificial Intelligence in Academic Writing: Metadata to the Rescue', *IEEE Computer,* Vol. 1, No. 1.

Curtis, N. and ChatGPT, 2023, 'To ChatGPT or not to ChatGPT? The Impact of Artificial Intelligence on Academic Publishing', *Pediatric Infectious Disease Journal*, Vol. 42, No. 4 (April), p. 275.

D'Arcus, B. and Giasson, F., 2009, *Bibliographic Ontology Specification. Technical Report.* Structured Dynamics LLC.

Devlin, J., Chang, M. W., Lee, K. and Toutanova, K., 2019, 'BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding'. *arXiv preprint arXiv:*1810.04805.

Haslhofer, B. and Klas, W., 2010, 'A Survey of Techniques for Achieving Metadata Interoperability', *ACM Computing Survey*, Vol. 42, No. 2, Article 7 (Mar).

Hassan Y., 2023, 'Governing algorithms from the South: A case study of AI development in Africa', *AI and Society*, Vol. 38, pp. 1429–1442.

Kaplan, J., McCandlish, S., Henighan, T., Brown, T. B., Chess, B., Child, R., ... and Amodei, D., 2020, 'Scaling Laws for Neural Language Models'. arXiv:2001.08361.

Kojima, T., Gu, S. S., Reid, M., Matsuo, Y. and Iwasawa, Y., 2023, 'Large Language Models are Zero-Shot Reasoners'. arXiv:2205.11916

Messeri, L. and Crockett, M. J., 2024, 'Doing More but Learning Less: The Risks of AI in Research', *Nature*, 7 March.

Myklebust, J. P., 2024, 'Research explores how AI might change knowledge production, *University World News*, 13 April.

National Science Foundation, National Science Board, National Center for Science and Engineering Statistics, 2024, *Publications Output: U.S. Trends and International Comparisons* (NSB-2023-33).

Nyokong, T., Ngoy, B. P. and Amuhaya, E. K., 2021, 'Overcoming hurdles facing researchers in Africa', *Nature Materials*, Vol. 20, p. 570. DOI:10.1038/s41563-021-00961-0.

Ray, P. P., 2023, 'ChatGPT: A comprehensive review on background, applications, key challenges, bias, ethics, limitations, and future scope, *Internet of Things and Cyber-Physical Systems*, Vol. 3, pp. 121–154.

Schmidt, E., 2023, 'The Download: How AI Will Transform the Way Science Gets Done', *MIT Technology Review*, 31 July, pp. 9–11.

Smith, J., 2023, 'Frank Herbert of Dune on Artificial Intelligence', *Medium*, 15 January.

UNESCO, 2022, 'Transforming knowledge for Africa's future: Seven horizons for the African Union Decade of Accelerated Action for the Transformation of Education and Skills Development in Africa (2025–2034)', Programme and meeting document ED-2025/ED-FL1/2.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł. and Polosukhin, I., 2017, 'Attention is all you need', *Advances in Neural Information Processing Systems*, 30. arXiv:1706.03762

Wei, J., Wang, X., Schuurmans, D., Bosma, M., Ichter, B., Xia, F., ... and Zhou, C., 2022', 'Chain of Thought Prompting Elicits Reasoning in Large Language Models'. arXiv:2201.11903:

Woolston, C., 2019, 'Meeting the challenges of research across Africa', *Nature*, 29 July.

World Economic Forum (WEF), 2024, 'There's a science research gap in Africa. Here's how to fill it', WEF 12 February.