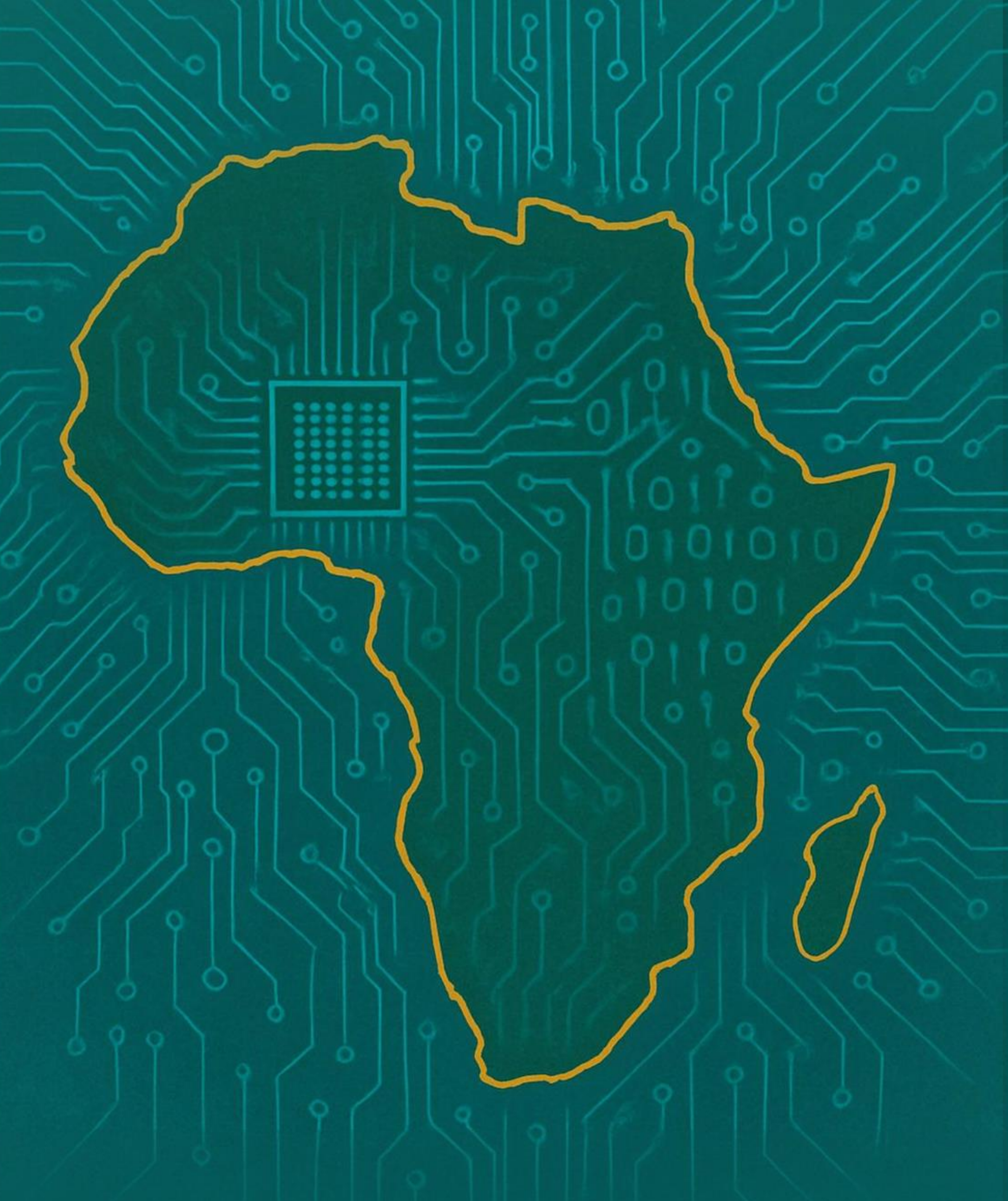

Interim Report | June 2025

Unlocking Compute in Africa

Africa Green Compute Coalition

This interm report is produced by the co-stewards of the Africa Green Compute Coalition (AGCC) and is based on interviews with stakeholders across Africa and around the world. In the spirit of openly sharing knowledge with the wider global ecosystem and in Africa, this interim report is intended to start a conversation and invite feedback to guide the mission of the Africa Green Compute Coalition.

Please write to us and share your thoughts: digital.support@undp.org info@axum.earth; admin@alliance4ai.org

Acknowledgements: Lead contributors: Robin Miller, Alim Ladha, Stephanie Mambo and Victor Kogo from Axum; Alex Tsado from Alliance4AI. Contributors from UNDP include Keyzom Ngodup Massally, Alexander Hradecky, Dwayne Carruthers, Maria Giulia Vitagliano, Romilly Golding, Yashica Yashica, Francesco Puggioni, Omar Jagne, Haley Iversen and Helin Su Aslan. Contribution from Kytabu by Tonee Ndungu.

They would like to thank United Nations Development Programme (UNDP), Cineca and Kytabu for their co-stewardship of the Africa Green Compute Coalition. We are grateful to the range of organizations, companies and initiatives whose work directly and indirectly contributed to this report: African Union NEPAD, AminiAI, AWS, BudgIT, CAIRD, Cassava Technologies, Cooper Union, DAIR, Deep Learning Indaba, Distributive, Equal Compute Network, Gates Foundation, Google, GSMA, IDRC, Office of the Technology Envoy – Government of Kenya, Lelapa, Masakhane, Microsoft, MinoHealth AI Labs, Mozilla Foundation, NVIDIA, Tunisia Novationcity, Udu Technologies, University of Cape Town AI Initiative.

Special thanks to all of the stakeholders across Africa and globally who participated in interviews, provided valuable perspectives and who continue to be invested in this topic.

AGCC co-stewards: AI Hub for Sustainable Development, Alliance4AI, Axum, Cineca and Kytabu.

Executive Summary

Africa's artificial intelligence (AI) revolution is not a distant possibility, but is happening now. The economic potential for Africa is transformative, with digital technologies projected to contribute $1.5 trillion to Africa's gross domestic product (GDP) by 2030.[1] A 2024 report by UNDP and the International Telecommunication Union (ITU) highlights the impact of AI innovations on 70 percent of SDG targets.[2]

Hundreds of AI solutions across the African continent are currently addressing critical challenges in agriculture, healthcare, finance and education, showcasing Africa's capacity to build contextually relevant solutions. However, unlocking this potential sustainably and at scale hinges on accessible, affordable and robust compute infrastructure. Despite hosting 18 percent of the world's population, Africa accounts for less than 1 percent of global data centre capacity.[3] This is creating a severe "compute gap" that risks widening the digital divide and relegating Africa to a consumer of AI rather than a producer of innovative solutions.

Preliminary analysis reveals that, among select groups of early-stage AI innovators, demand for AI model training alone reaches over 7 million GPU hours,[4] not including inference and other compute use cases. This represents only a fraction of the latent need that is constrained by systemic barriers and underscores the urgent need for strategic interventions to close the $2.5 billion compute gap identified in this interim report by 2030.[5]

This interim report outlines Africa's transformative AI potential, the critical compute infrastructure barriers that must be addressed, and targeted opportunities to accelerate early-stage innovation.

Section 1.0, Introduction, "Africa's Pivotal Moment for AI" highlights Africa's vibrant innovation ecosystem, where thousands of innovators are developing AI solutions for agriculture, healthcare, finance and education, supported by robust institutions like Co-Creation Hub, the African Institute for Mathematical Sciences (AIMS), and Masakhane. The section discusses over 100 AI solutions identified across key sectors,[6] points to national AI strategies in 13+ countries,[7] and discusses major investments such as Cassava Technologies' $720 million AI Factory[8] and Udu Technologies Africa GPU Hub.

Africa's demand for compute—which encompasses hardware, software and infrastructure—is surging, yet only 5 percent of AI talent has adequate access,[9] revealing a $2.5 billion

---

[1] PwC, _Sizing the prize What's the real value of AI for your business and how can you capitalise?_ Retrieved 2025
[2] UNDP, _Digital technologies directly benefit 70 percent of SDG targets, say ITU, UNDP and partners_, September 2023
[3] DFC.gov, _Tackling a critical need for data center infrastructure in Africa_, retrieved 2025
[4] AGCC analysis. See Annex for more detail on methodology and estimates.
[5] Sida, IDRC.CRDI, AI4D Africa, Genesis, _An estimate of $ 2.5bn is required by 2030 to close the compute gap to enable Africa to fully benefit from AI_, March 2024
[6] AGCC analysis
[7] TechAfrica News, _AI For Good: How Africa is Shaping the Responsible Use of Artificial Intelligence_, 2024
[8] CIO Africa, _Cassava Announces Plan to Build AI Factory_, March 2025

[9] UNDP, _Only five percent of Africa's AI talent has the compute power it needs_, November 2024

investment gap[10] and latent demand constrained by infrastructure, cost, skills and regulatory barriers. This positions compute as a critical enabler for the $1.5 trillion digital economy by 2030, according to a projection by consulting firm, PwC.[11]

Section 2.0, The Compute Bottleneck, details Africa's severe compute infrastructure deficit. Africa is home to only 1 percent of global data centre capacity despite being home to 18 percent of the world's population.[12] The continent has limited high-performance computing services, cloud service challenges, concentrated server infrastructure and high workstation costs. Nascent federated computing efforts face operational barriers such as unreliable power, slow connectivity, regulatory complexity and a critical skills gap (62 percent of firms lack the AI skills needed to leverage innovation, and 42 percent specifically struggle to hire AI-skilled employees).[13] These barriers create cost penalties and stifling innovation, despite growing investments from Microsoft, Google and others.

Section 3.0 Underlying Market Failures Preventing Self-Correction, reveals that Africa's compute infrastructure is stymied by interconnected market failures despite $ 4 trillion in domestic capital[14] and strong demand from AI innovators. Capital market inefficiencies such as risk mispricing and inappropriate evaluation frameworks create a $100 billion annual financing gap,[15] while coordination failures across power, connectivity and regulatory systems, alongside information asymmetries, skills shortages and network effects, perpetuate underinvestment, trapping the ecosystem in a low-equilibrium state that demands coordinated interventions to unlock innovation potential.

Section 4.0, The Investment Opportunity: Strategic Investment in Africa's Compute Future, reframes Africa's compute challenge as a $2.5 billion investment opportunity to catalyse digital transformation, highlighting a "flywheel effect" where targeted investments in model training for research, startups, developer platforms and Africa-based infrastructure activate latent demand, generate revenue, reduce risk perceptions and attract further funding, achieving self-sustaining growth, possibly supported by Africa's renewable energy potential.[16]

Section 5.0, Aligning for Action: The Africa Green Compute Coalition (AGCC), establishes the AGCC as a pivotal mechanism to translate this vision into action, fostering African agency and African infrastructure through five strategic pillars: coordinating GPU access; aggregating demand; shaping the compute market; leading advocacy, and developing talent pipelines. The AGCC will also leverage renewable energy for sustainable, cost-efficient infrastructure. By mobilizing hyperscale providers, development finance institutions, governments and innovation ecosystems, AGCC aims to create a resilient, equitable and green computing ecosystem that empowers African AI builders and aspires for self-sufficiency within 2–3 years.

---

[10] SIDA, IDRC.CRDI, AI4D Africa, Genesis, *An estimate of $2.5bn is required by 2030 to close the compute gap to enable Africa to fully benefit from AI*, March 2024

[11] PWC, *Sizing the Prize. Global Artificial Intelligence Study: Exploiting the AI Revolution*. 2024. https://www.pwc.com/gx/en/issues/analytics/assets/pwc-ai-analysis-sizing-the-prize-report.pdf

[12] DFC.gov, *Tackling a critical need for data center infrastructure in Africa*, retrieved 2025

[13] SoftwareOne Cloud, Mind the Gap! 62% of Companies Lack AI Skills to Leverage Rapid Innovation, June 2025

[14] State of Africa infrastructure Report, *Africa Finance Corporation*, 2025.

[15] African Development Bank, *Infrastructure Financing Gap*, 2023

[16] SIDA, IDRC.CRDI, AI4D Africa, Genesis, *An estimate of $2.5bn is required by 2030 to close the compute gap to enable Africa to fully benefit from AI*, March 2024

Introduction: Africa’s Pivotal Moment for AI

Innovation in action

Africa's technological landscape is distinguished by a vibrant and rapidly expanding innovation ecosystem, brimming with ingenuity and a profound commitment to solving local challenges. Hundreds, if not thousands, of African innovators are actively developing cutting-edge AI solutions tailored for social good and impactful development. These innovations span multiple sectors, demonstrating the continent's capacity to lead in AI development that is not only technologically advanced but also deeply relevant and beneficial to its communities.

Recent research by GSMA, a non-profit organization, across Ethiopia, Kenya, Nigeria, and South Africa identified over 100 solutions focused on agriculture, healthcare, energy and climate action alone.[17] The Gates Foundation currently supports over 70 AI use cases in Africa (and even more across the Global South) through its Grand Challenges AI Community, which spans 15 countries and includes health, education, climate, agriculture, language, disability inclusion and gender equality.[18]

In agriculture, AI shows transformative potential through precision farming, water optimization, pest detection, and weather prediction. In Africa, 70 percent of smallholder farmers depend on rainfed farming systems, making them vulnerable to climate variability,[19] while more than half of Sub-Saharan Africa's employment and approximately 30 percent of GDP depends on agriculture.[20] Degas Ghana Limited uses AI to support over 45,000 smallholder farmers through two key applications. Their credit-scoring model analyses over 1.2 million data points, while their large language model (LLM)-based tools integrated into an agent app deliver accessible guidance for climate-smart, regenerative agriculture to boost sustainability, resilience, and yields.[21]

Healthcare applications demonstrate AI's capacity to address critical infrastructure gaps. African healthcare systems face severe specialist shortages; Ghana has only 3 radiologists per 1 million people, while South Africa has only 10 per million compared to Europe's 130 per million.[22] minoHealth AI Labs has fine-tuned multimodal LLMs for radiology to generate accurate medical image analysis and currently deployed in 30 health facilities across Ghana.[23] In Kenya, Penda Health’s AI-powered ”Chat na Penda” supports health information sharing and is delivering over 1,000 unique conversations per day across 19 clinics in Nairobi, enhancing access to quality care for patients.[24]

---

[17] GSMA, *AI for Africa: Use Cases Delivering Impact*, July 2024

[18] Gates Foundation, *Grand Challenges*, retrieved 2025

[19] Läderach, Peter; Martinez-Valle, Armando; Bourgoin, Clement; Parker, Louis (27 March 2019). *Vulnerability of the agricultural sector to climate change: The development of a pan-tropical Climate Risk Vulnerability Assessment to inform sub-national decision making*, March 2019

[20] World Bank, *Agriculture Remains Key to Sub-Saharan Africa’s Economic Future*, World Economic Forum, Oct. 2015

[21] Gates Foundation Global Grand Challenges, *Supporting Field Agents to Scale Climate Action*, retrieved 2025 / Degas, *Degas, Which Aims to Increase the Incomes of African Smallholder Farmers, Raises $ 7 Million to Further Strengthen Peripheral Businesses Using its Farmer Platform*, 2023

[22] Sarkodie, B.D., Ohene-Botwe, B., Mensah, Y.B. et al. *Density and regional distribution of radiologists in a low-income country: the Ghana situation*. Chin J Acad Radiol 6, 188–195, 2023

[23] Gates Foundation Global Grand Challenges, *Foundation Model for Radiology*, retrieved 2025

[24] Gates Foundation Global Grand Challenges, Penda Health

Financial inclusion represents another transformative application area. Across countries, fintech companies leverage AI for credit scoring, fraud detection, and mobile money services, extending banking access to millions previously excluded from formal financial systems.[25] Across Kenya, Nigeria and South Africa, AI-driven platforms are revolutionizing how underserved populations access financial services.[26]

The innovation ecosystem is further anchored by robust institutions that provide critical infrastructure for scaling AI solutions. Co-Creation Hub (CcHub) in Nigeria has supported numerous AI-driven ventures, including LifeBank's logistics optimization platform for medical supplies that ensures critical blood and oxygen reach health facilities efficiently.[27] Injini in South Africa focuses specifically on educational technology, backing AI startups like Siyavula that deliver personalized mathematics education through AI-adaptive learning platforms that adjust to individual student learning patterns.[28] In Kenya, iHub has mentored breakthrough innovations like Ubenwa, which uses AI to detect newborn health issues through cry analysis—demonstrating how African innovation addresses uniquely local challenges with global technological relevance.[29]

Academic research institutions provide the foundational infrastructure for advanced AI development across the continent. The African Institute for Mathematical Sciences (AIMS), with centres across Ghana, Rwanda and other locations, leads foundational AI research while training the next generation of African AI researchers. University-based laboratories, including those at Makerere University in Uganda, conduct cutting-edge research in natural language processing for local languages and agricultural optimization models. The Masakhane research community exemplifies grassroots academic collaboration, bringing together researchers from across Africa to develop natural language processing tools for African languages—demonstrating how African institutions can lead global AI development in areas where local expertise and cultural understanding provide competitive advantages.

The state of AI in Africa

Beyond the dynamism of the innovators themselves, the AI surge is supported by a relatively nascent but increasingly robust policy, research and investment environment. Policy frameworks are rapidly evolving to support AI development. More than a dozen African countries have developed national AI strategies and policies, including Benin, Egypt, Ethiopia, Ghana, Kenya, Mauritius, Nigeria, Rwanda, Senegal, Sierra Leone, South Africa, Tunisia, and Zambia.[30] These frameworks demonstrate shared objectives that focus on increased AI research and development, constructing robust data infrastructures, skills development, developing local AI solutions and creating AI funding and investment incentives. The African Union (AU) launched its Continental AI Strategy for Africa in 2024, building on the AI for Africa Blueprint developed in 2021 between the AU and the Smart Africa Alliance, a pan-African

[25] Wagdi, Nabil, and Tarek., *Who Gets the Money? A Qualitative Analysis of FinTech Lending and Credit Scoring through the Adoption of AI and Alternative Data,* ScienceDirect, 2024
[26] GSMA, AI for Africa: Use Cases Delivering Impact, July 2024
[27] Growth Capital Fund (CcHUB), *CcHUB Growth Capital and EchoVC invest in Lifebank*, January 2018
[28] Injini, *Improving education in Africa through innovation, entrepreneurship, research, and collaboration*, retrieved 2025
[29] Ubenwa.ai
[30] Global Centre on AI Governance. African Countries are Racing to Publish AI Strategies, 2 April 2025. Atlantic Council Emerging Technology Policies and Democracy in Africa. March 2025

initiative that seeks to accelerate Africa's digital transformation.[31]

Investment momentum reflects growing confidence in Africa's AI potential. Major infrastructure commitments include:

- **Cassava Technologies**: $720 million in partnership with NVIDIA to build Africa's first AI Factory, launching in South Africa by June 2025 with plans to expand to Egypt, Kenya, Morocco, and Nigeria;[32]
- **Microsoft and G42**: $1 billion investment to build a geothermal-powered data centre and establish an Azure cloud region in Kenya;[33]
- **Microsoft South Africa**: $300 million investment from 2025–2027 to scale cloud and AI infrastructure;[34]
- **Africa Data Centres**: Accessed $83 million of an approved $300 million from the U.S. International Development Finance Corporation (DFC) to extend data centres across major African capitals.[35]
- **Udu Technologies:** Went live with the first dedicated GPUs for rent and purchase by AI researchers in Africa, along with software integration services to lower barriers to entry and empower them to multiply their innovation output.

Global cloud providers are establishing significant regional presence. Amazon Web Services launched its Cape Town region in 2020 and plans Local Zones in Kenya by 2025.[36] Microsoft Azure operates data centres in South Africa and is expanding across the continent.[37] Google Cloud launched its first African region in South Africa in 2023,[38] complementing extensive network infrastructure, including the Equiano submarine cable.

Understanding compute and its criticality

At the foundation of this AI transformation is a critical infrastructure: compute—which is the computational backbone that determines the scope, scale and accessibility of AI technologies. Compute refers to the computational resources required for AI systems to perform tasks such as processing data, training machine learning models, and making predictions.[39] It encompasses three essential components: hardware (the physical processors, particularly Graphics Processing Units (GPUs), that perform calculations and execute instructions), software (the interfaces and platforms that allow users to develop applications utilizing hardware capabilities), and infrastructure (the physical facilities housing hardware, data

---

[31] African Union. Continental Artificial Intelligence Strategy. Endorsed by the African Union Executive Council, 45th Ordinary Session, Accra, Ghana, 18–19 July 2024. African Union, 9 Aug. 2024
[32] CIO Africa, Cassava Announces Plan to Build AI Factory, March 2025
[33] Reuters, *Microsoft ties up with UAE-based AI firm to invest $ 1 bln in Kenya data center*, 2024
[34] Datacenterdynamica, *Microsoft to invest $ 300m in cloud and AI infrastructure in South Africa – DCD*, 2025
[35] DFC, *Cassava Technologies Receives $ 90 Million Equity Investment for Expanding Digital Infrastructure Across Africa*, 2024
[36] Amazon Web Services, *Announcing the new AWS Africa (Cape Town) Region,* April 2020 / Amazon Web Services, *Amazon Web Services to Open Data Centers in South Africa,* Business Wire, October 2018
[37] Vanian, Jonathan, *Microsoft's Azure Cloud Data Centers Expands to South Africa,* Fortune, March 2019
[38] Patel, Niral, *Heita South Africa! The new Google Cloud region is now open in Johannesburg,* Google Cloud Blog, January 2024
[39] Financial Times AI Glossary, *Compute*, July 2023 / CLRN Team, *What Is Compute in AI?* California Learning Resource Network, January 2025

storage devices, and related network equipment).[40]

Compute serves multiple critical functions in AI development. It enables AI models to process vast amounts of data, perform complex calculations, and refine predictions such as for local climate change actions through iterative learning. The amount and quality of available compute directly impacts the performance, accuracy, and efficiency of AI systems. Organizations and nations controlling high-performance compute infrastructure gain significant advantages in AI development, creating competitive and strategic benefits that determine technological leadership.

Different phases of AI development place distinct demands on computational resources:

- **Model development and training**: As the most computationally intensive phase, this involves feeding vast datasets to algorithms to build and refine complex models. Training requires significant GPU hours and is typically performed on high-performance computing clusters or powerful cloud instances. This type of compute is primarily demanded by research institutions for foundational breakthroughs, by larger AI startups developing proprietary models, and by enterprises building sophisticated internal capabilities.[41]

- **Model deployment and inference**: The application of pre-trained models to new data for real-time predictions or actions. Inference is less computationally demanding than training and can often be executed on distributed, efficient platforms including local workstations, edge devices or smaller cloud instances. This is critical for AI startups deploying solutions to users, SMEs integrating AI into operations, and NGOs delivering AI-powered services in communities.[42]

- **Data operations and processing**: Encompasses data collection, cleaning, transformation, storage and management activities that prepare datasets for model consumption. Includes real-time data pipelines, batch processing workflows and data quality assurance processes that require consistent, scalable compute resources.[43]

- **Experimentation and research**: Covers iterative algorithm development, hyperparameter tuning, model comparison, and research simulations. This includes A/B testing of different approaches, sensitivity analyses and complex modelling scenarios in domains like climate science or drug discovery that require flexible, high-performance compute access.[44]

- **Production Operations & Monitoring**: Involves model monitoring, performance tracking, automated retraining workflows, and system maintenance. Requires reliable, always-on compute infrastructure to ensure deployed AI systems continue operating

[40] Ibid / AGCC Analysis

[41] Alzahabi, Ayman, *AI Explained: The Critical Phases of Training and Inference*, DZone

[42] Google Developers, *Machine Learning – Phases*, retrieved

[43] Ibid

[44] AI Architecture Handbook, *AI Model Lifecycle Management*, Retrieved 2025

effectively and adapt to changing data patterns.[45]

What's driving Africa' compute demand?

Africa's demand for compute resources is accelerating due to several converging technological, economic, and regulatory factors that are reshaping the continent's digital landscape.

1. **Digital transformation is happening across all sectors**: Both public and private organizations are rapidly adopting AI-powered solutions to enhance efficiency, improve service delivery and create new value propositions. The 2023 African data centre market was valued at approximately $3.33 billion, with projections rising to $6.46 billion by 2029, representing 11.7 percent annual growth.[46] Africa's cloud market is projected to grow 15 percent annually, reaching $18 billion by 2028, driven primarily by the finance, healthcare and telecommunications sectors.[47]

2. **Regional and national entities are prioritizing digital infrastructure:** Continental frameworks including the African Union's Digital Transformation Strategy for Africa (2020–2030)[48] and the Smart Africa Alliance (comprising 37 member states)[49] prioritize building digital infrastructure to support economic development. These initiatives create demand for local compute resources to support government digital services, smart city developments, and public sector AI applications.

3. **New regulations require local data processing:** Emerging regulatory frameworks across Africa require organizations to process and store data within national borders. Countries including Kenya, Nigeria, South Africa and Tunisia have implemented data protection laws with localization requirements,[50] creating demand for sovereign compute infrastructure that ensures compliance while maintaining performance.

4. **Digital adoption is exploding across the continent**: Internet penetration continues expanding across the continent, with increasing mobile connectivity and smartphone adoption driving data generation and real-time processing requirements. As more Africans access digital services, the volume of locally generated data requiring local processing capabilities also continues to grow.

5. **Several sectors show particularly strong demand patterns**:

---

[45] Data Science PM, *What Is the AI Lifecycle?* Retrieved 2025

[46] Arizton, *The Africa Data Center Market Investment to Reach $ 6.46 Billion by 2029*, Industry Today, March 2025

[47] Grant, Dee, *African Cloud Computing,* Black Economics, July 2024

[48] African Union Commission. The Digital Transformation Strategy for Africa (2020–2030). African Union Commission, May 2020

[49] Smart Africa, About us, retrieved 2025

[50] Adebisi, Abiodun, *Data Regulation in Africa: Free Flow of Data, Open Data Regimes and Cybersecurity,* Data Governance and Policy in Africa, Palgrave Macmillan, 2023

a. **Finance and fintech**: Mobile money adoption has driven banks and fintech companies toward cloud and AI adoption for services like fraud detection, credit scoring, and customer service automation.

b. **Healthcare and education**: Telemedicine platforms, e-learning systems, and AI diagnostic support solutions increasingly depend on scalable, secure compute infrastructure.

c. **Agriculture**: Image-based crop disease detection and predictive climate models require cloud-powered compute, including edge deployments for rural applications.

d. **E-commerce and manufacturing**: Platforms leverage cloud infrastructure for customer data management, logistics optimization, and IoT-driven analytics.

Quantifying the actual demand

Tackling an investment gap as large as $2.5 billion requires a clear focused first wave of investments that will not only deliver returns (social and commercial) but also begin to address some of the systemic barriers and enhance confidence in the AI market overall, ideally creating a flywheel for returns, impact and future investment.

Systematic analysis across six key demand segments reveals substantial and rapidly growing computational requirements. The methodology involved stakeholder interviews across demand pools, assessment of current and ideal GPU utilization, and segmentation of users into light, intermediate, and heavy compute categories.

Our analysis focuses specifically on model training demand—the most compute-intensive phase of AI development. Preliminary findings forecast immediate demand of 7 million GPU hours by researchers and early-stage innovators across key user groups over the next three years. This conservative estimate excludes inference and other compute uses, capturing only the most visible demand pools; the true underlying need is substantially larger as innovation accelerates continent-wide.

Latent Demand and Barriers to Realization

Beyond this quantified demand lies substantial latent demand; unmet need constrained by systemic barriers that prevent its full expression. According to an article by Zindi, UNDP and Alliance4AI, only 5 percent of Africa's AI talent currently has access to the computational power needed for complex tasks,[51] meaning African innovators may wait up to six days to iterate on models compared to 30 minutes for peers in G7 countries. This constraint means many potential AI applications never move beyond conceptual stages, with innovators unable to test, validate, or scale their solutions. The barriers to demand realisation include:

1. **Infrastructure accessibility**: The absence of accessible compute infrastructure means many potential users cannot experiment with AI applications or validate business models that would justify larger investments. Additionally, poor infrastructure drives high latency, making real-time AI inference applications unfeasible for many real-time use cases. Without the ability to test concepts affordably or deploy responsive solutions, many innovations remain theoretical.

[51] UNDP, *Only five percent of Africa's AI talent has the compute power it needs*, November 2024

2. **Cost barriers**: Current compute costs create prohibitive entry barriers for smaller innovators. With factors such as taxes, import duties and currency fluctuation impact costs, many potential users are priced out of experimentation phases that would reveal larger demand.

3. **Complexity and skills barriers**: The perceived complexity of accessing and managing compute resources, combined with the critical skills gap (80 percent of companies report negative impacts from lack of AI skills according to an SAP survey),[52] prevents many organizations from articulating their true compute needs.

4. **Market readiness**: Limited access to patient capital, varying levels of market readiness for advanced AI solutions, and fragmented go-to-market strategies across diverse African economies create challenging environments for scaling AI applications, suppressing apparent demand below true need levels.

5. **Data and connectivity constraints**: Scarcity of high-quality, localized datasets essential for relevant AI development, combined with persistent connectivity challenges and high internet costs, directly impede widespread AI solution deployment and mask underlying demand.

6. **Risk aversion and investment gaps**: Perceived high risks by potential investors and customers, often related to market volatility, regulatory uncertainty, and infrastructure reliability, create funding gaps that prevent innovators from acquiring necessary compute resources, further suppressing visible demand signals.

This latent demand represents the difference between Africa's current compute utilization and its true potential if barriers were removed. Addressing infrastructure constraints would likely reveal higher demand than current projections, as improved accessibility would enable experimentation, validation and scaling of AI solutions across broader segments of Africa's innovation ecosystem.

The economic stakes

The economic implications of Africa's AI development trajectory are transformative for the continent's long-term prosperity. Stakeholders ranging from Google to Microsoft to the United Nations Economic Commission for Africa (UNECA) believe that Africa could capture $1.2–1.5 trillion in AI-driven economic value by 2030, representing nearly 10 percent of PwC's estimated $15.7 trillion global AI opportunity.[53] The continent's digital economy has already demonstrated remarkable growth, expanding from 1.1 percent of GDP in 2012 to 4.5 percent in 2020, with projections reaching 5.2 percent by 2025 and 8.5 percent by 2050,[54] representing a potential contribution of $712 billion by mid-century.

Sector-specific impact projections demonstrate AI's transformative potential. In agriculture, where over 70 percent of the population depends on farming activities,[55] AI applications for

---

[52] SAP News, *SAP Research Reveals Top Tech Skills Challenges for African Organisations*, SAP Africa News Center, March 2023.
[53] PwC, *Sizing the prize What's the real value of AI for your business and how can you capitalise?* Retrieved 2025
[54] Olatunji, Tolulope, *Africa's Digital Economy: Opportunities and Challenges in 2025*, Digital Economy Magazine, 2025
[55] Biteye, Mamadou, *70 % of Africans depend on agriculture for livelihoods*, World Economic Forum, 2016

precision agriculture, climate adaptation, and supply chain optimization could enhance productivity for millions of smallholder farmers. Healthcare applications addressing the severe shortage of medical professionals—Africa has 16 percent of the world's population but carries 25 percent of the world's disease burden and only has 3 percent of health workers[56]—could significantly improve access and quality of care across underserved communities.

The broader economic multiplier effects extend beyond direct AI applications. Local AI development creates high-value employment opportunities in technical fields, retains talent that might otherwise migrate to developed economies, and builds institutional capabilities that strengthen the continent's position in global value chains. Countries successfully developing local AI ecosystems can capture greater value from their data assets, increase domestic infrastructure independence, and potentially develop export capabilities in AI services and applications.

However, realizing this economic opportunity for people and sustainable development depends fundamentally on Africa's ability to develop, deploy, and scale AI solutions that address local challenges and leverage local opportunities. Success requires building AI applications, and ensuring they can operate at scale with reliable, affordable, and accessible computational resources that understand local contexts and comply with local regulations.

This economic potential faces a critical strategic bottleneck: the computational infrastructure needed to support AI development and deployment at scale. Computer power remains severely constrained across the continent, creating a fundamental dependency that could determine whether Africa shapes its AI future or remains relegated to consuming solutions developed elsewhere.

---

[56] McKinsey & Company. "Overcoming sub-Saharan Africa's Health Workforce Paradox." *McKinsey & Company*, 4 Nov. 2024.

The Compute Bottleneck

Despite Africa's vibrant AI innovation ecosystem and rapidly growing demand for computational resources, a critical constraint threatens to limit the continent's technological potential. The infrastructure necessary to support AI development and deployment at scale remains severely constrained, creating fundamental barriers that could determine whether Africa participates as an equal partner in the global AI economy or remains dependent on solutions developed elsewhere.

Africa's compute infrastructure reality

Africa's compute infrastructure remains nascent and severely constrained relative to both global standards and the continent's growing AI ambitions. While the Africa data centre market was valued at $3.49 billion in 2024 and is projected to reach $6.81 billion by 2030,[57] this growth originates from an extremely low baseline; Africa accounts for only 1 percent of global data centre capacity despite hosting 18 percent of the world's population.[58] This fundamental mismatch between population scale and infrastructure availability is just a proxy indicator of the compute deficit that constrains AI development and deployment across the continent.

For African innovators seeking computational resources, five primary pathways exist, each revealing distinct opportunities but also limitations that collectively illustrate why current supply remains inadequate despite rapidly growing demand. Understanding these pathways and their constraints offers insights into the infrastructure bottleneck that threatens to limit Africa's AI potential.

1.1.1 Figure: Types of compute[59]

Approach	Description	Examples
1 High-Performance Computing (HPC) credits	Groups of computers networked together to perform complex calculations and data processing. They typically consist of multiple nodes working in parallel, with specialized networking infrastructure and storage systems.	Toubkal Supercomputer; CHPC
2 Cloud credits	Promotional or subsidized units of cloud computing resources offered by cloud service providers to allow startups, researchers, non-profits or developers to access cloud compute infrastructure and services, allowing users to experiment, build, and scale resources without managing physical infrastructure.	Google Cloud; aws; Azure
3 Servers	Servers are specialized computing systems equipped with high-performance GPUs to handle the intensive computational demands of AI and ML workloads. These servers are designed to process large datasets, train complex models, and perform real-time inference tasks efficiently.	TERACO; Africa DataCentres
4 Workstations	Workstations are high-performance computers designed for technical or professional applications e.g., 3D design, video editing, engineering, AI/ML training. They typically have powerful processors, large memory capacity, and advanced graphics capabilities to handle demanding workloads.	DELL; hp
5 Federated Compute	A system allowing distributed computation while keeping data localized (a single problem is processed collaboratively in multiple computer systems, often at different locations, at the same time).	Deep Learning Indaba

[57] GlobeNewswire, *Africa Data Center Market Landscape Report 2025–2030: Africa's Data Center Market Set to Nearly Double, Valued at $ 6.81 Billion by 2030*, April 2025
[58] DFC.gov, *Tackling a critical need for data center infrastructure in Africa*, retrieved 2025
[59] AGCC analysis

1.1.2 High-Performance Computing: Limited Scale and Access

Africa's High-Performance Computing (HPC) landscape consists of approximately 35 HPC systems across 11 countries, primarily deployed through initiatives like the HPC Ecosystems Project.[60] However, the scale limitations are stark—only one system, the Toubkal supercomputer at Morocco's African Supercomputing Center (UM6P), ranks 246th among global supercomputers, with a LINPACK-rated performance of 3.15 petaflops.[61] South Africa's CHPC Lengau, while significant regionally with 1.03 petaflops performance,[62] demonstrates how even the continent's most advanced HPC infrastructure operates at scales far below global standards.

These systems offer access at costs ranging from $0.10–15/hour for CPU-based instances to $0.50-100+/hour for GPU-accelerated instances,[63] with complete clusters costing thousands to tens of thousands of dollars daily. However, accessibility remains constrained by multiple factors that limit effective utilization. Perceived complexity deters many potential users who lack familiarity with HPC environments and job scheduling systems. Inadequate internet infrastructure produces round-trip times exceeding six seconds in some countries, rendering real-time applications unfeasible and slowing data transfer for large-scale computations.[64] Most critically, limited technical expertise means many organizations cannot effectively utilize available HPC resources even when access is theoretically possible.

The geographic concentration of HPC resources exacerbates these limitations. Systems cluster in South Africa, Morocco, Egypt, and select East African countries, leaving vast regions without practical access to high-performance computational resources. This distribution pattern reflects broader infrastructure inequalities and constrains the development of distributed research and innovation capabilities across the continent.

Cloud: Promises Complicated by Market Constraints

Cloud computing credits provide valuable promotional or subsidized access to on-demand infrastructure, empowering African innovators to leverage world-class computational resources without the burden of managing physical hardware. Leading global providers, including Amazon Web Services (AWS), Microsoft Azure, and Google Cloud, offer significant credit programmes, with some initiatives, such as Google for Startups Accelerator Africa, providing up to $350,000 in Google Cloud credits to qualifying startups.[65] These programmes are designed to foster innovation, facilitate initial adoption of cloud services, and support the development of scalable, tech-driven solutions addressing local and global challenges.

---

[60] Johnston, Bryan J., et al., *Ten Years of the HPC Ecosystems Project: Transforming HPC in Africa for the Past Decade*, Practice and Experience in Advanced Research Computing (PEARC '24), July 2024

[61] Johnston, Bryan J., *African Supercomputing Center Inaugurates 'Toubkal,' Most Powerful Supercomputer on the Continent*, HPCwire, February 2021

[62] NICIS, *South African CHPC Unveils Lengau Supercomputer*, 2016

[63] Google Cloud, *Pricing – Compute Engine,* Google Cloud Platform, retrieved 2025 / DataCrunch.io, *GPU Cloud Pricing Comparison: AWS, GCP, Azure, Lambda, RunPod, Vast*, DataCrunch, May 2025 / High Performance Computing Center, University of Tartu, *Usage and Pricing*, UT HPC, retrieved 2025

[64] Barro, Pape A., Marco Zennaro, Jules Degila, and Ermanno Pietrosemoli. *A Smart Cities LoRaWAN Network Based on Autonomous Base Stations (BS) for Some Countries with Limited Internet Access*, 2019

[65] Aiyegbusi, Folarin, *Inviting Applications for the Google Startups Accelerator Africa*, Google for Startups Blog, April 2024

African users often come up against unique structural realities that can impact the efficacy of these programs and the overall cloud experience. A key challenge is the global scarcity and allocation dynamics of cutting-edge GPUs, such as NVIDIA H200, H100 or A100 systems. While cloud providers aim for equitable resource distribution, the intense global demand for these specialized accelerators, particularly for advanced AI model training, can lead to limited regional availability. African innovators, despite having substantial credit balances, may face delays or difficulty accessing the specific high-performance hardware essential for competitive AI development. Furthermore, the economic realities of exchange rate volatility in some African economies introduces further uncertainty. For instance, the Nigerian Naira devaluation in 2023 led to a reported ~234 percent increase in cloud service costs for some users when transitioning from credit-based usage to market rates,[66] underscoring the impact of currency fluctuations on operational budgets.

The economic landscape for cloud providers investing in Africa also presents complexities, contributing to pricing differences and service delivery models. Establishing and operating hyperscale data centres in Sub-Saharan Africa often requires higher infrastructure costs due to underdeveloped energy grids, requiring significant investment in backup power connectivity and systems. Higher bandwidth costs for international connectivity and the current lower economies of scale, compared to more mature markets, prevent providers from offering directly equivalent regional pricing. While specific figures can vary by service and provider, these factors contribute to cloud services costing more in Sub-Saharan Africa compared to equivalent usage in developed markets. Global cloud providers are actively investing in local infrastructure, exemplified by Microsoft's commitment to invest an additional ZAR 5.4 billion (approximately $254 million) by the end of 2027 to expand its cloud and AI infrastructure in South Africa, building on prior investments.[67] These ongoing investments aim to enhance local data centre presence, improve latency, and ultimately contribute to more localized and cost-efficient cloud services over time, reflecting a long-term commitment to fostering digital transformation on the continent.

Servers: Infrastructure concentration and quality caps

Server-based computing, where specialized systems equipped with high-performance GPUs are housed in data centres, is a pathway severely constrained by geographic concentration and infrastructure quality limitations. These servers are primarily located in South Africa, Nigeria, Egypt, Kenya, and Morocco, with access models including pay-as-you-go services, reserved instances, dedicated hosting and colocation arrangements.

The infrastructure quality presents some constraints for mission-critical AI applications Approximately 73 percent of African data centres are classified as Tier III,[68] which while offering good reliability, may pose limitations for applications requiring the highest levels of continuous availability and redundancy, for which Tier IV or specialized purpose-built facilities are often preferred. Infrastructure development costs reflect these quality and geographic challenges, ranging from $10–15 million per MW nationally, but escalating to $15 million per

[66] Jaiyeola, Temitayo, and Chinwe Michael, *Naira's Fall Drives Startups to Embrace Local Services*, BusinessDay, February 2024
[67] Reuters, *Microsoft to invest $ 300 million more in South Africa's AI infrastructure*, March 2025
[68] Mordor Intelligence, 2024

MW in challenging markets like Nigeria[69] where power, connectivity and security requirements drive additional complexity and expense.

The concentration of server infrastructure in major urban centres creates accessibility barriers for innovators and organizations outside these digital hubs. Countries across Central, West, and parts of East Africa lack adequate server infrastructure, forcing organizations to either accept higher latency and costs from distant servers or forgo server-based compute options entirely. This geographic inequality perpetuates innovation concentration patterns and limits the potential for distributed AI development across the continent.

Workstations: Local control limited by economic barriers

Workstations offer high-performance computing designed for technical applications including artificial intelligence/machine learning (AI/ML) training, providing dedicated compute power with advanced graphics capabilities that deliver predictable performance and greater control compared to cloud services.[70] This pathway proves particularly valuable in regions with limited connectivity, where local processing capabilities can overcome bandwidth constraints and provide consistent access to computational resources.

However, economic barriers limit workstation accessibility across Africa. High capital expenditure requirements range from $10,000 for basic AI-capable workstations to over $500,000 for enterprise-grade systems with multiple high-end GPUs.[71] This represents significant financial commitments for organizations operating in capital-constrained environments. Import duties can compound these costs, increasing total expenditure above global market prices due to various taxes, tariffs, and regulatory compliance requirements.

The economic barriers extend beyond initial capital costs to ongoing operational challenges. Customs clearance processes can delay hardware delivery by months, extending project timelines and increasing opportunity costs. Limited local technical support means organizations must either develop internal expertise or rely on expensive international service contracts for maintenance and troubleshooting. These factors combine to make workstation-based compute a challenging pathway despite its technical advantages for local AI development.

Federated compute: Nascent potential limited by infrastructure reality

Federated computing represents the most nascent pathway, offering a distributed model that processes data closer to where it's generated while keeping information localized—an approach particularly relevant for Africa's data sovereignty requirements and diverse regulatory landscape.[72] The conceptual appeal is significant: organizations could participate in collaborative AI development while maintaining local control over sensitive data and complying with national data protection requirements.

---

[69] Arizton, *Africa Data Center Construction Market Landscape Report 2025–2030: Africa Data Centers Cost ≈ $ 10M/MW (Rising to US$ 15M in Nigeria)*, via GlobeNewswire, February 2025.
[70] NVIDIA, *NVIDIA-Powered Workstations for Data Science*, Retrieved June 2025
[71] Wallner, Emil, *How I Built a €25K Machine Learning Rig*, 2025 / TRG Datacenters, *NVIDIA DGX Components, Pricing, and Other FAQs*, 2025
[72] FEDGEN, *Federated GENeral "Omics" (FEDGEN) Research*, retrieved 2025

Deployment of federated compute is currently limited due to infrastructure constraints and regulatory complexities across diverse African markets.[73] These systems require reliable, high-bandwidth connectivity between nodes, strong data management capabilities, and standardized technical protocols—requirements that exceed current infrastructure capabilities in most African markets.

Regulatory complexity further constrains federated compute development. Different data protection frameworks across African countries, from Nigeria's Data Protection Act requirements for local data storage to Kenya's consent requirements for cross-border transfers. These create compliance challenges that complicate federated system design and operations.[74] Harmonized regulatory approaches and improved infrastructure capabilities, could strengthen federated compute as a growing option for the continent.

The cost complexity

These infrastructure constraints translate into real cost complexities that compound barriers for African innovators. The cumulative effect of these cost disadvantages makes it harder for African AI developers to compete on an equal footing compared to their global counterparts.

Some studies suggest that African organizations relying on foreign compute services face premiums higher than equivalent usage in Europe or North America.[75] This differential seems to stem from limited local infrastructure, higher bandwidth costs for international connectivity, and exchange rate fluctuations when paying for services denominated in foreign currencies. For cloud computing specifically, GPU-enabled services may carry regional surcharges partly due to network infrastructure limitations and lower economies of scale that impact pricing.

Hardware acquisition also faces penalties due to import duties, taxes, and regulatory compliance requirements. Import duties and taxes increase GPU hardware costs by 20-30 percent above global market prices, with additional levies varying by country and specific hardware categories. High-performance GPUs required for AI training—such as NVIDIA H100 or A100 systems—face pressure, with total higher for imports under certain trade classifications. These costs reflect not just tariffs but the complex logistics of moving sensitive technology through African customs systems.

Energy costs create additional operational burdens despite Africa's average electricity price for businesses appearing competitive at $0.108 per kWh.[76] Significant variations exist across countries, with Sierra Leone facing costs of $0.342 per kWh, Côte d’Ivoire $0.227 per kWh, and Kenya $0.202 per kWh—among the highest globally for business electricity.[77] More critically, unreliable power supply necessitates expensive backup systems, with Nigeria alone spending an estimated $14 billion annually on generators for power backup.[78] Energy costs

---

[73] Adebayo, Shakir, et al. *Federated Learning and Privacy-Preserving Generative AI for African Contexts: A Research Framework*. ResearchGate, May 2024

[74] TechTrends, *Navigating the Challenges of Federated Learning: Privacy, Security, and Scalability in Distributed Data Environments*, December 2023

[75] AFRINIC, *Studying Performance Barriers to Cloud Services in Africa’s Public Sector*, AFRINIC Research, April 2020

[76] Globalen LLC, *Electricity Prices Around the World*, GlobalPetrolPrices.com, January 2025

[77] Globalen LLC, *Business Electricity Prices by Country – March 2024*, GlobalPetrolPrices.com, 2024

[78] Echewofun, Simon, *Businesses Spend $ 14 Billion Annually on Generators – Official*, Daily Trust, June 2024

typically represent 30-40 percent of data centre operations,[79] making these variations and reliability issues critically important for infrastructure viability.

Connectivity costs compound these challenges further. The average cost of 1GB of mobile data in Sub-Saharan Africa stands at approximately $3.31, compared to North Africa's $0.86.[80] For organizations requiring substantial data transfer to access cloud compute services, these costs accumulate rapidly. Fixed broadband access reaches only 22 percent of the African population,[81] limiting options for high-bandwidth compute applications and forcing reliance on more expensive mobile data alternatives.

Currency fluctuations add another layer of cost uncertainty for compute services. Nigeria's experience illustrates this volatility dramatically—the country experienced a ~234 percent increase in cloud service costs from early 2023 to late 2023 due to Naira devaluation, with services costing ₦458,000 rising to approximately ₦1.52 million. This exchange rate risk makes long-term planning and budgeting for compute resources challenging, forcing organizations to either accept currency exposure or pay additional premiums for hedging.

Operational barriers beyond cost

While cost represents a significant constraint, operational barriers create equally formidable challenges that prevent even well-funded organizations from accessing compute resources reliably. These systemic constraints reflect deeper infrastructure deficiencies that compound access difficulties.

Power infrastructure challenges represent perhaps the most fundamental barrier. Approximately 600 million people, or 43 percent of Africa's population, lack access to reliable electricity, with most concentrated in Sub-Saharan Africa.[82] For businesses, frequent power outages and unstable electricity supply pose persistent challenges: 80 percent of African businesses experience regular outages averaging 8 hours duration.[83] Data centres and compute facilities require 99.9 percent+ uptime, necessitating expensive backup power systems that significantly increase both capital and operational costs. The reliability gap forces organizations to invest heavily in redundant systems, effectively doubling power infrastructure requirements and associated expenses.

Connectivity infrastructure limitations persist despite progress in submarine cable deployment and fibre network expansion. Average internet download speeds in Sub-Saharan Africa reach approximately 12.11 Mbps, compared to over 100 Mbps in developed countries.[84] Round-trip times can exceed 6 seconds in some countries,[85] rendering real-time AI applications unfeasible and significantly slowing data-intensive model training processes. According to

---

[79] Majavu, Wabo, *Navigating the Energy Transition: The Cost Implications for Africa's Data Centers*, Africa Hyperscalers News, August 2024

[80] Statista, *Average price for 1 GB of mobile data in Africa as of 2023, by country (in U.S. dollars)*, , September 2023

[81] Africa Broadband Forum. *Africa Broadband Outlook 2023*. Omdia for Huawei & World Broadband Association, October 2023

[82] International Energy Agency, *Africa Energy Outlook 2022*

[83] Oseni, Musiliu Olalekan, *Costs of Unreliable Electricity to African Firms*, Energy for Growth Hub, August 2019

[84] Statista, *Average internet download speed in Sub-Saharan Africa as of 2023, by country (in megabits per second)*, Statista, August 2023

[85] Barro, Pape A., Marco Zennaro, Jules Degila, and Ermanno Pietrosemoli. *A Smart Cities LoRaWAN Network Based on Autonomous Base Stations (BS) for Some Countries with Limited Internet Access*, 2019

ITU, only 38 percent of Africa's population was online in 2024,[86] with rural areas particularly underserved, limiting the addressable market for compute services and constraining the potential user base that could justify infrastructure investments.

Regulatory and compliance complexity creates additional operational friction across diverse African markets. Data sovereignty requirements vary significantly; Nigeria's Data Protection Act (2023)[87] mandates local storage of personal data, while Kenya's Data Protection Act requires explicit consent for cross-border transfers.[88] While these requirements serve important sovereignty purposes, they add complexity to cross-border compute services and infrastructure development. Organizations must navigate multiple regulatory frameworks simultaneously, increasing compliance costs and legal risks.

Supply chain and logistics constraints further complicate infrastructure deployment. Import restrictions and export licensing requirements, particularly for high-performance GPUs, can delay access to advanced hardware by months. Customs clearance processes add additional delays and costs, while logistics infrastructure limitations complicate delivery and installation of sensitive equipment.[89] These factors extend infrastructure deployment timelines from months to years and increase project risks substantially, deterring potential investors and operators.

Security concerns, both physical and cyber, create additional operational burdens. Data centre security markets project significant growth, reflecting increased awareness of security risks but also indicating current vulnerabilities. Cybersecurity threats target African infrastructure disproportionately, while physical security concerns in some regions necessitate additional protective measures that ultimately increase costs beyond levels typical in developed markets.

The critical skills gap

Underlying all these infrastructure and operational challenges is a fundamental human capital constraint that both explains and perpetuates Africa's compute limitations. The shortage of specialized technical talent required to design, deploy, operate and innovate within compute infrastructure creates a self-reinforcing cycle that impedes progress across the entire ecosystem.

Recent research by SAP across Kenya, Nigeria, and South Africa revealed the magnitude of this challenge. Ninety-nine percent of companies view AI skills as essential for corporate success, yet 90 percent report negative impacts due to skills shortages, including project delays, failed innovation initiatives, and inability to undertake new AI-related work.[90] This gap extends far beyond AI development to encompass the full spectrum of compute infrastructure management, from data centre engineers and network architects to cloud operations specialists and AI/ML operations experts.

The compute ecosystem requires diverse technical expertise. Data centre engineers capable

---

[86] International Telecommunication Union. *Facts and Figures 2024 – Internet Use*. ITU, November 2024
[87] Strathmore University—CIPIT, *Navigating the Crossroads: The Challenges of Cross-Border Data Flows under Domestic Laws in Africa*, CIPIT, 2024
[88] Ibid
[89] AGCC stakeholder interviews
[90] SAP News, *SAP Research Reveals Top Tech Skills Challenges for African Organisations*, SAP Africa News Center, March 2023

of designing and managing complex facilities are limited, while network architects who can optimize high-bandwidth connectivity for AI workloads are difficult to source. Cloud operations specialists who can manage scalable platforms effectively, and AI/ML operations experts who can deploy and monitor AI models efficiently, represent particularly critical shortages. These skills are essential not just for utilizing existing infrastructure, but for the capacity building necessary to increase local expertise and develop sustainable, African-owned compute ecosystems.

A paradox emerges in skills investment patterns that reveals underlying market failures. While 94 percent of African organizations offer monthly training and skills development programmes, budget allocation for skills development has surprisingly declined since 2023.[91] This decline occurs precisely when demand for AI skills is expected to increase dramatically in 2025, with 85 percent of organizations prioritizing AI development skills and 83 percent focusing on Generative AI capabilities.[92] The gap between recognized need and actual investment reflects the challenge of individual firms bearing training costs while competitors can potentially recruit trained employees, leading to collective underinvestment in human capital development.

Available technical expertise tends to concentrate in major urban centres and established institutions, leaving vast regions underserved and compounding geographic infrastructure imbalances. Countries like Nigeria face challenges, with over 70 percent of engineering graduates lacking practical skills required by industry employers,[93] exacerbating gaps in critical areas like renewable energy integration and cybersecurity for compute infrastructure. This unequal geographic distribution of skills reinforces infrastructure concentration patterns, creating barriers for distributed compute development that could serve underserved regions.

The skills shortage directly constrains both infrastructure deployment and utilization effectiveness. Organizations may gain access to compute resources but lack the expertise to optimize configurations, manage security protocols, or develop innovative applications that maximize infrastructure value. This limitation reduces the return on infrastructure investments and slows the development of local expertise necessary for sustainable compute ecosystem development. The result is a cycle where limited skills constrain infrastructure development, which in turn limits opportunities for skills development, perpetuating dependence on external expertise and solutions.

This comprehensive skills challenge represents more than a simple human resource constraint—it constitutes a fundamental barrier to achieving the digital sovereignty and local innovation capacity that sustainable compute infrastructure development requires. Without addressing this gap through coordinated investment in education, training, and institutional capacity building that parallels infrastructure development efforts, Africa's compute ambitions will remain constrained by human capital limitations regardless of available financial resources or political commitment.

Growing investment and initiative momentum

Despite the infrastructure constraints documented above, Africa's compute landscape is

[91] Pillay, Nazia, *AI Skills Development in Africa – New Report Findings Revealed,* SAP Africa News Center, June 2025
[92] Bannier, Jason, *AI Skills Essential, Say 99% in SAP Survey*, SAP Africa News Center, May 2025
[93] Aluko, Olusola, *Addressing the Skills Gap in Nigeria's Engineering Sector: A Call to Action*, BusinessDay NG, September 2024

experiencing unprecedented momentum, driven by major investment commitments, policy initiatives and collaborative efforts that validate the continent's strategic importance for global AI development. These initiatives provide a crucial foundation for expanded development, while highlighting the coordination opportunities that could amplify their collective impact.

Major infrastructure investments demonstrate growing private sector confidence in African compute markets. Cassava Technologies' $720 million partnership with NVIDIA to build Africa's first AI Factory represents the largest single commitment to African AI infrastructure, with deployment of 3,000 NVIDIA GPUs and planned expansion to 12,000 units across Egypt, Kenya, Morocco, Nigeria and South Africa.[94] Additionally, Udu Technologies' launch of the Africa GPU Hub provides the software-platform complement that will bring GPUs to AI developers in Africa, reducing costs, improving availability and lowering support barriers for innovation.[95] Microsoft's combined investments—including the $1 billion geothermal-powered data centre partnership with G42 in Kenya and $300 million cloud infrastructure expansion in South Africa[96]—supports its planned Local Zones expansion.

Government recognition of compute infrastructure's strategic importance is seen in national AI strategies across nearly a dozen countries, as well as the African Union's Continental AI Strategy for Africa. These frameworks demonstrate a sophisticated understanding of the foundational role of compute infrastructure for AI development, while creating policy environments that support expanded investment and development.

Academic and research momentum continues to expand through networks like AI for Development (AI4D), which supports 22 projects across 18 African countries;[97], the Grand Challenges AI Community with nearly 70 Africa-focused initiatives across 15 countries (and more than100 innovators across 20 countries in the Global South)[98] and research institutions [99]from Makerere University's Centre for Artificial Intelligence to Strathmore University's iLab Africa Center[100] Other institutions of note include the African Institute for Mathematical Sciences, Deep Learning Indaba and the Masakhane research community. Innovation ecosystems—such as the Co-Creation Hub, iHub and AfriLabs' network of nearly 500 innovation hubs in 5[101] countries—demonstrate the continent’s substantial entrepreneurial capacity. Platforms like Zindi, with its growing network of innovators across Africa,[102] also indicate a need for compute resources to scale AI applications.

International partnerships between African institutions, global technology companies and development organizations are demonstrating successful coordination and momentum. Examples include Google's AI Lab in Accra, IBM Research collaborations and growing development engagement from finance institutions such as DFC, Finnfund and the African

---

[94] CIO Africa, *Cassava Announces Plan to Build AI Factory*, March 2025

[95] UduTech Africa GPU Hub Launch https://udutech.co/udutech-bringing-africa-ai-dreams-to-reality/

[96] Reuters, *Microsoft ties up with UAE-based AI firm to invest $ 1 bln in Kenya data center*, 2024 / Datacenterdynamica, Microsoft to invest $ 300m in cloud and AI infrastructure in South Africa – DCD, 2025

[97] AI4D, *Projects*, retrieved 2025

[98] Gates Foundation, *Grand Challenges*, retrieved 2025

[99] Makerere University, *Artificial Intelligence*, retrieved 2025

[100] Strathmore University, *iLabAfrica Center*, retrieved 2025

[101] Africa World Initiative, *Connecting Africa’s Innovation Dots: AfriLabs Welcomes 18 New Hubs to Its Pan-Afrikan Network,* December 2024

[102] Zindi.Africa

Development Bank, together with various multilateral initiatives. Alliance4AI's work building university AI clubs, big tech mentorship programmes for Africa AI start-ups and the collaborative work of ecosystem actors—including the Equal Compute Network, the Rockefeller Foundation, Mozilla, the Gates Foundation and a range of impact investors exploring AI as a complement to their portfolios—further illustrate significant activity across the ecosystem.

These initiatives collectively represent significant investments and demonstrate clear recognition of Africa's compute potential across the private sector, government, academic and international stakeholder communities. They provide a substantial foundation for coordinated approaches that could achieve greater scale and systematic impact through aligned implementation and shared infrastructure development.

Underlying Market Failures Preventing Self-Correction

Why markets can't self-correct

While the infrastructure investments documented above demonstrate a growing recognition of Africa's compute potential, these promising initiatives will face the same constraints in compute development across the continent. The constraints documented in the previous section persist not due to a lack of demand or available capital, but because of fundamental market failures that prevent efficient resource allocation toward compute infrastructure development. These systemic barriers create conditions in which rational individual decision-making by market participants produces collectively suboptimal outcomes, trapping Africa's compute ecosystem in a persistent state of underinvestment despite compelling underlying economics.

The puzzle is particularly acute given Africa's substantial financial resources. The continent possesses an estimated $4 trillion in domestic capital across commercial banking assets, pension funds, insurance companies, sovereign wealth funds and central bank reserves.[103] Simultaneously, the documented demand from thousands of AI innovators demonstrates clear market opportunities with significant potential returns. Yet these two elements—available capital and proven demand—fail to connect naturally through market mechanisms and are creating a 100 billion annual infrastructure financing shortfall that persists despite obvious economic complementarity.[104]

Understanding why markets can't bridge this gap reveals a need for coordinated intervention that addresses multiple, interconnected failures simultaneously. These failures span the investment ecosystem and reinforce one another, keeping infrastructure development far below its potential.

Capital allocation inefficiencies and risk pricing failures

The failure of capital markets to efficiently price and allocate resources toward African compute infrastructure reflects systematic biases and structural inefficiencies that distort investment flows, despite underlying project viability. These distortions create a persistent misallocation, leaving potentially profitable infrastructure investments unfunded, while capital continues to flow toward lower-return alternatives.

Risk assessment mechanisms systematically overstate infrastructure risks through methodologies designed for different market and historical contexts. African countries face borrowing cost premiums averaging 2.9 percentage points above comparable developed market rates, translating to $24 billion in additional annual interest payments and $46 billion in foregone lending capacity.[105] However, these premiums often substantially exceed actual historical default rates. Infrastructure projects in Sub-Saharan Africa show default rates as low

---

[103] State of Africa infrastructure Report, *Africa Finance Corporation*, 2025
[104] African Development Bank, *Infrastructure Financing Gap*, 2023
[105] International Development Institute, *The African Debt Dilemma: Unpacking the Three Unfavourable Factors*, ODI Think Change, 2024

as 2.1 percent in some analyses: below Western Europe (5.9 percent) and well under Asia (8.8 percent),[106] and well below the risk levels implied by current pricing.

This systematic risk mispricing creates self-fulfilling prophecies, where inflated financing costs reduce project viability, thereby validating risk concerns and maintaining high risk premiums. The cycle becomes particularly vicious for innovative infrastructure like compute facilities, where credit rating agencies lack historical performance data and default to conservative assumptions that may bear little relationship to actual project economics under proper management and appropriate risk mitigation.

Capital market structures compound pricing failures through inappropriate evaluation frameworks that favour short-term, liquid investments over long-term infrastructure development. Much of Africa's $777 billion in institutional investor assets remains allocated to government securities or foreign investments rather than domestic infrastructure. This is not due to a lack of opportunity, rather to regulatory frameworks and investment mandates designed for developed market conditions.[107] These constraints force even patient capital toward suboptimal allocations that generate lower returns while failing to address infrastructure needs.

The result is a systematic misallocation where high-return infrastructure projects remain unfunded while capital flows toward lower-yield alternatives. This inefficiency persists because individual financial institutions cannot unilaterally change risk assessment methodologies or regulatory frameworks. Coordinated action is required to correct market-wide pricing distortions that no single actor can resolve alone.

Coordination failures across interdependent investments

Compute infrastructure development requires synchronized investment across multiple domains—power systems, connectivity networks, regulatory frameworks, research and innovation ecosystems and human capital—that individual market participants cannot coordinate. Each investment domain depends on others for viability, yet private actors lack mechanisms to ensure coordinated development, creating systematic underinvestment in complementary infrastructure.

The coordination problem can be clearly seen in the circular dependencies between infrastructure components. Data centre operators require reliable power and connectivity before committing to facility development, while utility and telecommunications providers need committed anchor demand before investing in capacity upgrades. Similarly, equipment suppliers require sufficient market scale to justify local operations and support infrastructure, while potential customers need reliable local supply chains before committing to infrastructure deployment.

These coordination challenges become more complex when multiple jurisdictions are involved. Cross-border compute initiatives that could achieve economies of scale and risk diversification face coordination problems across different regulatory frameworks, technical standards and economic policies. While regional integration could create larger, more viable

---

[106] Adesina, Akinwumi, *Africa's Infrastructure Finance Default Rate at 2.1% Is Lowest in the World*, Nairametrics, November 2023

[107] Africa Finance Corporation, *Briefing*, June 2025

markets for infrastructure development, the coordination mechanisms necessary to achieve such integration do not emerge naturally from individual market participant decisions.

The coordination failure extends to timing mismatches between different investment cycles. Infrastructure development requires long-term capital commitments with extended payback periods, while potential users often need immediate access to computational resources for time-sensitive innovation cycles. This temporal mismatch prevents the alignment of supply and demand that would justify infrastructure investment, even when underlying economics are favourable.

Successful coordination requires either dominant market players with sufficient scale to internalize coordination benefits or institutional mechanisms that align incentives across multiple participants. In most African markets, no single private actor has the scale to coordinate infrastructure development across sectors. At the same time, weak institutional coordination keeps progress fragmented, with coordination failures preventing otherwise viable infrastructure development.

Information asymmetries and market intelligence deficits

Efficient market functioning requires symmetric access to accurate information about costs, demand and technical requirements. Africa's compute infrastructure market suffers from systematic information asymmetries that prevent potential investors and users from identifying mutually beneficial opportunities, leading to systematic misallocation despite underlying economic complementarity.

Critical market intelligence remains fragmented across multiple stakeholders without mechanisms for aggregation or standardization. Potential infrastructure providers lack reliable data about actual demand patterns, user requirements and willingness to pay for different service levels.[108] Simultaneously, potential users lack sufficient information about available options, comparative costs and technical capabilities of different infrastructure approaches. This bilateral information deficit prevents efficient matching between supply and demand capabilities.

The information problem extends to technical knowledge about optimal infrastructure configurations for African operating environments. Global infrastructure providers often apply standardized solutions developed for different contexts without adequate understanding of local power reliability, connectivity constraints or regulatory requirements. This knowledge gap leads to suboptimal infrastructure design that increases costs and reduces effectiveness, while local market participants lack access to global best practices that could improve project viability.

Investment evaluation suffers from insufficient historical data about infrastructure performance, user adoption patterns, and revenue generation potential in African markets. This data scarcity forces conservative assumptions that may systematically underestimate project viability while preventing the development of appropriate risk management and financing mechanisms. The absence of standardized performance metrics and reporting frameworks compounds this problem by preventing accumulation of market intelligence that

[108] AGCC stakeholder interviews

could inform future investment decisions.

Regional information-sharing mechanisms remain underdeveloped, forcing each market to independently develop knowledge that could be shared across similar contexts. While continental organizations like Smart Africa aim to facilitate information-sharing, practical mechanisms for aggregating and disseminating market intelligence about infrastructure development remain limited, constraining the development of regional expertise and best practices that could improve investment outcomes across multiple markets.

Skills market failures and underinvestment in human capital

The systematic underinvestment in specialized technical skills necessary for compute infrastructure development represents a classic market failure, where the incentives of individual firms diverge from collectively optimal human capital development. This failure both constrains current infrastructure utilization and perpetuates the shortage of local expertise necessary for sustainable ecosystem development.

The underinvestment paradox emerges clearly from recent research showing that while 99 percent of companies recognize AI skills as essential,[109] and 94 percent offer training programmes, budget allocation for skills development has declined since 2023.[110] This reduction occurs despite 85 percent of organizations prioritizing AI development skills,[111] revealing a systematic divergence between stated priorities and actual investment behaviour —reflecting underlying market failure rather than a lack of recognition of the importance of skills.

The market failure stems from individual firms' reluctance to bear training costs when competitors can recruit trained employees, creating free-rider dynamics where each organization prefers others to invest in training while benefiting from the resulting skilled workforce. This problem becomes particularly acute in mobile African labour markets, where skilled professionals frequently migrate between organizations or countries, reducing individual firm returns on training investments and incentivizing collective underinvestment despite clear aggregate benefits.

A geographic concentration of skills investment compounds the market failure by creating regional imbalances that limit infrastructure development potential in underserved areas. Training programmes cluster in major urban centres with established educational infrastructure, leaving vast regions without adequate human capital to support distributed infrastructure development. This concentration reinforces infrastructure inequalities and constrains the distributed development patterns that could serve Africa's diverse geographic and economic landscape more effectively.

The skills shortage extends beyond technical capabilities to encompass business and operational expertise specific to African infrastructure development contexts. Project management, regulatory compliance and business development capabilities remain scarce, limiting successful project execution even when technical solutions and financing are available. Traditional educational systems often fail to produce workforce capabilities aligned

[109] Bannier, Jason, *AI Skills Essential Say 99% in SAP Survey*, SAP Africa News Center, May 2025
[110] Pillay, Nazia, *AI Skills Development in Africa – New Report Findings Revealed,* SAP Africa News Center, June 2025
[111] Waters, John K., *Report: 85% of Organizations Are Evaluating or Using Some Form of AI,* THE Journal, February 2025

with infrastructure development needs, while private sector training programmes remain narrowly focused on immediate operational requirements rather than broader ecosystem development priorities.

Network effects and critical mass coordination problems

Compute infrastructure exhibits strong network effects, where value increases exponentially with user adoption and application diversity. Yet achieving critical mass requires coordinated initial investment that individual market actors cannot sustain independently. This creates self-reinforcing cycles, where insufficient adoption constrains infrastructure investment, while limited infrastructure constrains adoption—trapping the ecosystem below viability thresholds.

Current utilization patterns illustrate the critical mass challenge: with only 5 percent of Africa's AI talent having access to adequate computational power,[112] insufficient demand density prevents infrastructure providers from achieving the economies of scale necessary for competitive pricing.[113] Low utilization rates increase per-unit costs, which constrains adoption and perpetuates low utilization in a self-reinforcing cycle that individual market participants cannot break independently.

The network effects problem becomes particularly acute for specialized infrastructure like AI-optimized compute clusters, where fixed costs for cooling, power management and technical support can only be amortized across substantial user bases. Without sufficient demand concentration, providers cannot achieve utilization rates necessary for competitive pricing—limiting accessibility and constraining market development below self-sustaining levels.

Platform coordination challenges compound network effects constraints by requiring simultaneous development of infrastructure capabilities and application ecosystems. Compute platforms become valuable as more developers create applications and more users adopt them, but early-stage platforms cannot attract developers without users or users without applications. This coordination challenge requires either substantial patient capital to subsidize early adoption or market-making interventions that facilitate the coordination necessary to achieve critical mass.

Regional fragmentation multiplies coordination challenges by preventing African markets from achieving continental scale that could support viable infrastructure development. Regulatory differences, technical standards variations and limited cross-border connectivity constrain development of pan-African platforms that could aggregate demand across multiple countries to achieve critical mass. This forces individual markets to independently surmount coordination challenges that could be addressed more efficiently at a regional scale.

The reinforcing cycle of market failure

These market failures operate synergistically to create self-reinforcing cycles, where each constraint amplifies others, making it increasingly difficult for market mechanisms alone to achieve optimal infrastructure development. The interplay between capital allocation inefficiencies, coordination failures, information asymmetries, skills shortages and network effects creates a persistent equilibrium far below socially optimal levels— one that becomes

---

[112] UNDP, *Only five percent of Africa's AI talent has the compute power it needs*, November 2024

[93] Alex Tsado, Celina Lee, *Only five percent of Africa's AI talent has the compute power it needs*, UNDP, November 2024

more entrenched over time.

Capital mispricing constrains infrastructure investment, which limits market development and validates conservative risk assessments in self-fulfilling cycles. Coordination failures prevent the synchronized investment necessary for viable infrastructure development, while information asymmetries prevent potential participants from identifying coordination opportunities. Skills shortages limit both infrastructure development capabilities and utilization effectiveness, reducing returns on infrastructure investment and constraining the market development that could support expanded skills investment.

Network effects amplify all these constraints by creating winner-take-all dynamics where early disadvantages become increasingly difficult to overcome. Without critical mass, African compute markets cannot achieve the economies of scale economies needed to compete with established global alternatives—while those alternatives become increasingly attractive as African markets remain subscale and expensive.

The persistence of these market failures—despite substantial investment momentum, policy recognition and demonstrated innovation capacity—reveals the coordination challenge at the heart of Africa's compute gap. Major infrastructure commitments and government AI strategies across multiple countries demonstrate clear recognition of the opportunity and substantial resource mobilization. However, these initiatives remain geographically concentrated and technically fragmented, limiting their ability to achieve the cumulative impact necessary to overcome systematic market failures.

Individual initiatives are optimized for their own specific technical and commercial requirements, but lack mechanisms to ensure compatibility with other infrastructure development efforts. Geographic concentration patterns persist across different initiative types, with investments clustering in major markets while leaving vast regions underserved. Technical fragmentation limits interoperability and network effects that could amplify the impact of individual investments, while timing mismatches between initiatives prevent complementary effects that could overcome market failures more effectively than isolated efforts.

The cumulative effect traps Africa's compute ecosystem in a low-equilibrium state where limited infrastructure constrains innovation potential, while constrained innovation reduces demand for infrastructure investment. Breaking this cycle requires coordinated intervention that addresses multiple market failures simultaneously, creating positive reinforcement in the opposite direction:a flywheel effect where initial infrastructure investment catalyses demand growth, which supports additional infrastructure development, skills investment and ecosystem development in self-reinforcing cycles that market mechanisms alone cannot initiate.

This analysis demonstrates why the promising innovations documented across African AI development remain constrained despite the clear potential for value creation and growing investment recognition. The market failures preventing compute infrastructure development are systematic and interconnected. They require coordinated solutions that address multiple constraints simultaneously to create the positive reinforcement necessary for sustainable ecosystem development.

The Investment Opportunity: Strategic Investment in Africa's Compute Future

Africa's compute infrastructure challenge represents more than a gap to be filled; it presents a strategic investment opportunity to catalyse the continent's digital transformation while generating sustainable returns. The convergence of documented demand, proven deployment models and clear pathways to self-sustaining operations creates an investment thesis that transcends traditional development finance to offer compelling commercial prospects.

The true opportunity lies in recognizing that Africa's demand for compute forms an interconnected innovation ecosystem where investment in one area amplifies returns in others. Research institutions create talent and fuel startup growth; developer platforms create skilled workforces;, and successful startups validate markets that attract further investment. This interconnectedness favours a portfolio approach to compute investment, deploying infrastructure across multiple segments simultaneously to capture aggregated demand, network effects and create self-reinforcing cycles that no single-segment investment could achieve alone.

Understanding the investment landscape

The estimated $2.5 billion required to unlock Africa's AI potential reflects not a monolithic infrastructure build, but a sophisticated portfolio of targeted investments across diverse market segments.[114] This figure, drawn from AI4D's assessment of Africa's compute ecosystem needs, emerges from comprehensive analysis of existing demand pools, infrastructure requirements and the economic dynamics that will drive sustainable growth in Africa's compute ecosystem. AGCC analysis identifies an immediate investment opportunity of $150 million to address 7 million GPU hours of documented demand for model training alone—representing the most visible and quantifiable portion of this broader infrastructure requirement.

Strategic deployment of this investment creates a "flywheel effect":—initial infrastructure enables innovation, which generates revenue, which validates market viability, which attracts additional investment, creating a self-reinforcing cycle of growth. Unlike traditional aid-dependent models, this approach builds toward commercial sustainability from the outset.

The 7 million GPU hours of investment opportunity to support training and early-stage innovation represent a strategic entry point for catalysing a flywheel effect in Africa's AI sector. This investment thesis targets coordinated deployment across research labs, early-stage startups, government facilities and SMEs rather than isolated funding approaches. Portfolio clustering—whether geographic, sectoral or focused on green compute opportunities—creates the innovation ecosystems necessary to initiate self-reinforcing cycles of growth and

---

[114] Sida, IDRC.CRDI, AI4D Africa, Genesis, *An estimate of $ 2.5bn is required by 2030 to close the compute gap to enable Africa to fully benefit from AI*, March 2024

development.

Philanthropic and catalytic funding play a crucial role in this early-stage investment approach, which mirrors successful blended finance strategies that have mobilized over $200 billion[115] globally by demonstrating market viability and de-risking subsequent commercial capital deployment. The key differentiator lies in coordinated investment timing and robust monitoring frameworks that track ecosystem development metrics—including talent pipeline growth, innovation velocity and infrastructure utilization rather than immediate financial returns.

Success requires patient capital from foundations and impact investors that understand the ecosystem development timeline while building toward commercial sustainability. This initial investment creates the foundational infrastructure and talent base that sets the flywheel in motion, positioning Africa's AI sector for growth-stage opportunities when market conditions mature for larger commercial investment.

The investment opportunity divides into complementary streams that serve different segments of Africa's innovation ecosystem. Research infrastructure supports the academic and foundational AI research community, whichis already generating immediate demand. Innovation hubs and startup accelerators require compute resources that can scale rapidly with entrepreneur needs. Developer platforms must serve Africa's growing community of more than 70,000 data scientists distributed across 47 countries.[116] Enterprise and government deployments demand sovereign, secure infrastructure for sensitive applications. Each stream represents distinct investment characteristics, risk profiles and return mechanisms.

Strategic investment allocation

Africa's compute ecosystem exhibits strong network effects, where different user segments reinforce each other, creating opportunities for portfolio investments that generate multiplicative rather than additive returns.

The most effective approach combines three complementary strategies. Geographic clustering concentrates investments in innovation hubs where research institutions, startup ecosystems, developer communities and enterprise clients operate within the same metro areas, achieving economies of scale while creating dense innovation clusters. Sectoral alignment focuses compute infrastructure on specific domains like healthcare or agriculture, serving academic research centres, industry startups and government applications within integrated ecosystems to build domain expertise and competitive advantages. Green compute integration leverages Africa's abundant renewable energy potential, balancing urban accessibility with sustainable power sources through peri-urban facilities or distributed models.

This coordinated approach creates powerful self-reinforcing cycles, where each segment strengthens the others. Research institutions provide stable utilization and talent development that feeds startup growth. Startups validate market potential and attract enterprise adoption. Developer platforms create skilled workforces that enterprises need. Enterprise clients

[115] Convergence. State of Blended Finance 2024: https://www.convergence.finance/resource/state-of-blended-finance-2024/view
[116] The Borgen Project, *How AI Upskilling Is Creating Jobs in Africa: The Case of Zindi*, retrieved June 2025

generate premium revenue streams that subsidize access for research and innovation segments. Rather than isolated investments, this transforms the estimated 7 million hours of quantified demand into coordinated ecosystem investment, generating multiplicative returns.[117]

Research institutions across Africa demonstrate immediate, quantifiable demand that provides a stable foundation for infrastructure investment. Academic labs and universities currently supported by networks like AI4D represent 48 initiatives (22 of these in Africa), currently generating over 80,000 GPU hours for training alone over the next three years.[118] This translates to more than $1.6 million in compute requirements necessary to enhance Africa's research community.[119] Universities offer particularly attractive investment prospects because their compute needs are predictable, their funding sources are diversified, and their research outputs create positive externalities that benefit the broader innovation ecosystem.

The startup and innovation ecosystem presents a different but equally compelling investment profile. Venture builders like UNDP's Timbuktoo initiative support 10,000 startups across Africa, with current cohorts already generating more than 220,000 GPU hours to train 5 million people over three years.[120] Innovation hubs supporting approximately 250 startups across the continent project more than 700,000 GPU hours over the same period.[121] These investments require more flexible, scalable infrastructure but offer higher growth potential as successful startups scale their compute requirements exponentially.

Developer platforms represent the largest single demand segment, with Zindi's community projecting more than 6 million GPU hours for model training over three years in compute requirements.[122] This significant scale reflects the latent demand across Africa's technical community—professionals who possess the skills to build AI solutions but lack access to the computational resources needed to realize their potential Platform investments offer the highest leverage, serving thousands of users through shared infrastructure while building the human capital foundation for Africa's AI economy.

Enterprise and government applications require sovereign infrastructure that ensures data compliance and security. These investments command premium pricing due to their security and sovereignty requirements, while offering stable, long-term revenue streams from large organizations with substantial budgets.[123] The regulatory environment increasingly favours local infrastructure, with countries like Nigeria and Kenya implementing data localization requirements that create captive demand for domestic compute resources.

Rather than pursuing these segments in isolation, portfolio allocation creates multiplicative value through geographic clustering in innovation hubs, sectoral alignment around domains like healthcare or agriculture, and integration of Africa's renewable energy potential. This coordinated approach transforms the estimated $150 million in quantified demand into

---

[117] AGCC analysis
[118] Ibid
[119] Ibid
[120] Ibid
[121] Ibid
[122] Ibid
[123] Ibid

ecosystem investment, generating exponential returns.

Challenging investment misconceptions

Several persistent myths constrain capital allocation toward African compute infrastructure, despite compelling evidence of market viability and attractive returns. Addressing these misconceptions is essential for mobilizing the investment necessary to realize Africa's compute potential.

Myth 1: Africa lacks sufficient demand for advanced compute infrastructure

Reality: This fundamentally misunderstands the continent's innovation dynamics and the scale of untapped demand. Preliminary demand analysis reveals 7 million GPU hours of immediate, quantifiable need for model training alone—worth approximately $150 million—yet this represents only the most visible demand from a subset of African innovators and excludes inference and other compute uses.

The true scale of latent demand becomes apparent when examining sector-specific applications where AI addresses critical challenges that traditional solutions cannot solve effectively. Agriculture exemplifies this dynamic: 70 percent of smallholder farmers depend on rainfed systems,[124] vulnerable to climate variability, while the sector employs over half of Sub-Saharan Africa's workforce and contributes approximately 30 percent of GDP.[125] AI applications for precision farming, pest management and climate adaptation are not luxury technologies, but essential tools for food security and economic stability.

Healthcare applications address fundamental infrastructure gaps—Ghana's ratio of 3 radiologists per million people compared to Europe's 13 per 100,000[126] creates compelling demand for AI-powered diagnostic systems that can operate at scale. Major infrastructure investments by Microsoft ($1 billion in Kenya), Cassava Technologies ($720 million AI Factory), and similar commitments by Google and AWS provide market validation from sophisticated investors.

Myth 2: Workstation capital expenditure is inherently too high for African contexts

Reality: This perception reflects often ignores total cost of ownership dynamics. The analysis becomes more nuanced when considering project scale and duration. For initiatives with budgets exceeding $50,000, infrastructure ownership can prove more cost-effective, while projects in the $20,000–$50,000 range benefit from hybrid approaches that combine cloud flexibility with owned infrastructure for core workloads.[127]

African organizations often face a price distortion when compared to equivalent usage in developed markets, stemming from currency fluctuation, regional pricing structures, and limited local infrastructure. When combined with import duties on GPU hardware— often as

---

[124] Läderach, Peter; Martinez-Valle, Armando; Bourgoin, Clement; Parker, Louis (27 March 2019). *Vulnerability of the agricultural sector to climate change: The development of a pan-tropical Climate Risk Vulnerability Assessment to inform sub-national decision making*, March 2019

[125] World Bank, *Agriculture Remains Key to Sub-Saharan Africa's Economic Future*, World Economic Forum, October 2015

[126] Sarkodie, B.D., Ohene-Botwe, B., Mensah, Y.B. et al. *Density and regional distribution of radiologists in a low-income country: the Ghana situation*. Chin J Acad Radiol 6, 188–195, 2023

[127] AGCC analysis

high as can often be up to 20-30 percent[128]—the economics favour strategic capital investment that eliminates ongoing foreign exchange exposure while providing predictable operational costs.

Myth 3: GPU location is irrelevant for performance and compliance

Reality: This assumption fundamentally misunderstands Africa's unique digital landscape and regulatory environment. The physical location of GPUs critically impacts performance, cost-efficiency and legal compliance across multiple dimensions.

Latency creates immediate performance barriers and impacts in Africa's variable internet infrastructure. Cross-continental connections often introduce significant delays that directly impact application effectiveness. For real-time applications—such as fraud detection in financial services, remote surgery and telemedicine, chatbots and virtual assistants, voice-based AI applications, early warning systems and smart city management—even milliseconds of delay can render solutions ineffective or create poor user experiences that undermine adoption.[129]

Data sovereignty requirements increasingly demand local processing. A growing number of African countries are enacting data protection regulations with provisions stipulating that sensitive or personal data must be processed and stored within national borders,[130] mirroring GDPR stipulations.[131] Deploying AI applications on hardware outside the country of data origin creates material compliance risks, including legal challenges, substantial fines and reputational damage.

Cost and resource efficiency seem to favour local infrastructure. International data transfers for AI training or inference incur substantial bandwidth[132] costs that are often hidden in cloud pricing. Data centres consume massive amounts of energy—the International Energy Agency reports global data centres used 240-340 TWh in 2022 (equivalent to 1-1.3 percent of global electricity), with some facilities matching the power consumption of medium towns.[133]

Beyond the technical and legal aspects, local infrastructure development fosters innovation and digital sovereignty by enabling African businesses, researchers and developers to build solutions tailored to local challenges while developing domestic expertise, creating jobs, and reducing foreign dependence. Additionally, localized data processing reduces exposure to security vulnerabilities that affect data traveling across international networks—significantly reducing the attack surface,enhancing overall data security and simplifying compliance with data protection best practices.

Myth 4: Sole reliance on cloud and HPC credits provides sustainable long-term access

Reality: While foreign cloud and HPC credits offer valuable initial access, relying solely on them may not be sustainable for a continent's long-term digital sovereignty and strategic

---

[128] AGG stakeholder interviews / Cruz, Marcio, Paulo Bastos, and Lucio Castro, *Boosting Digital Development in Africa: The Role of Trade Policy*, World Bank Blogs: Development Talk, December 2024

[129] GeeksforGeeks, *What Is Latency in Computer Network?* GeeksforGeeks, September 2023

[130] Monyae, David, *Data Sovereignty and African Integration: The Time Is Now*, *The African*, August 2022

[131] DataProtection.Africa, *Mapping the progress (and delays) for data protection in Africa*, November 2023

[132] Shrivastava, Anand Babu, *AI Model Training Rekindles Interest in OnPremises Infrastructure*, SiliconANGLE, October 2023

[133] International Energy Agency. *Electricity 2023: Analysis and Forecast to 2025*. IEA, 2023

positioning.

Current credit models create dependency cycles requiring continuous external funding, with innovation-stage companies spending 30–50 percent of grant funding on cloud GPU services.[134] When credits expire, only approximately 5 percent of recipient organizations deliver competitive AI solutions, meaning 95 percent require renewed funding to continue development.[135]

Cost volatility adds uncertainty—Nigeria experienced an estimated 234 percent cloud service cost increases during 2023 due to currency devaluation.[136] Limited customization means platforms cannot always accommodate African-specific requirements or data sovereignty mandates.

Local infrastructure achieves self-sustainability through different dynamics. Workstation and server components generate revenue streams that cover operational costs beyond year three, while providing data sovereignty, reduced latency and consistent access that cloud credits cannot guarantee.

Myth 5: Investing in on-premises GPUs is not sustainable

Reality: The workstation and rack server components of compute programs achieve self-sustainability after three year, contrary to perceptions that on-premises infrastructure requires perpetual external funding. Leveraging Africa's abundant renewable resources could offer a sustainable and environmentally responsible path to digital infrastructure development.

Direct ownership could provide cost savings over a 3– to 5- year period, with the right conditions in place. Beyond cost savings, local compute can further foster ecosystems of ancillary services, creating jobs in hardware maintenance, technical support and energy management, while building local capacity essential for long-term digital independence.

On-premises infrastructure generates sustainable revenue streams through multiple channels. Training programs, technical consulting and specialized services create ongoing income that covers operational costs. Local expertise development ensures systems remain optimized and current as technology evolves, while shared infrastructure models distribute costs across multiple users to improve economic viability.

The sustainability model differs fundamentally from cloud credits, which require continuous external funding. Once established, workstation and server infrastructure can operate indefinitely with modest operational expenditure covered by user fees and service revenues, creating genuine independence from donor dependency cycles.

Myth 6: African Compute Investment Carries Prohibitive Risks

Reality: The African data centre market's projection—from $3.3 billion in 2023 to $6.46 billion by 2029—represents compound annual growth of 11.7 percent,[137] demonstrating robust

---

[134] AGCC analysis
[135] Ibid
[136] Jaiyeola, Temitayo, and Chinwe Michael, *Naira's Fall Drives Startups to Embrace Local Services*, BusinessDay, February 2024
[137] Arizton, *The Africa Data Center Market Investment to Reach $ 6.46 Billion by 2029*, Industry Today, March 2025

commercial viability.

In addition, risk perception often exceeds reality in African infrastructure projects. While borrowing cost premiums of 2.9 percentage points increase capital costs, actual default rates for infrastructure projects in Sub-Saharan Africa measure 2.1 percent in certain analyses[138]—substantially lower than commonly assumed. The continent's $100 billion annual infrastructure financing shortfall[139] reflects capital allocation challenges and risk perception rather than fundamental project viability.

The flywheel effect: Creating self-reinforcing sustainable growth

Investing in local African AI ecosystem suppliers and African-owned compute infrastructure can generate a "flywheel effect", a self-reinforcing cycle that transforms initial capital deployment into sustained ecosystem growth. This dynamic distinguishes strategic compute investment from traditional infrastructure development by creating compounding returns that extend far beyond the original investment scope.

The flywheel begins when accessible compute infrastructure activates latent demand among African innovators. Researchers who previously waited weeks for cloud access can iterate on models in real time. Startups that could not afford enterprise cloud pricing can now develop competitive AI solutions. This activation of previously constrained talent creates immediate value while demonstrating market viability.

As these newly enabled innovators develop solutions and generate revenue, they validate the commercial potential of African AI applications. Success stories reduce risk perceptions among investors, lowering capital costs and attracting additional funding. Each successful AI startup or breakthrough research project becomes a case study that facilitates further investment.

Scale effects amplify this growth as infrastructure reaches critical mass. Shared resources become more efficient while local expertise develops organically. The flywheel effect can generate value beyond direct financial returns through local job creation, skills development and ecosystem services that keep value creation within Africa.[140]

Africa's abundant renewable energy resources add a crucial dimension to this flywheel effect. Green compute infrastructure reduces operational costs while attracting climate-conscious investment, creating sustainable competitive advantages. Initiatives such as Teraco's 120-megawatt solar installation[141] demonstrate how environmental sustainability enhances commercial viability, positioning Africa's compute sector for long-term growth aligned with global sustainability trends.

These dynamic transforms compute infrastructure from a cost centre into a growth engine that

---

[138] Adesina, Akinwumi, *Africa's Infrastructure Finance Default Rate at 2.1% Is Lowest in the World*, Nairametrics, November 2023

[139] African Development Bank, *Infrastructure Financing Gap*, 2023

[140] The Times, *Is the tech sector losing its lustre?* March 13, 2025 (In the UK, beyond traditional tech centres like London, regional cities including Manchester, Edinburgh, Bristol, Leeds and Birmingham are experiencing notable growth in their tech sectors, particularly in AI, fintech, cybersecurity, and more—creating high-salary job opportunities (£70k–£150k) and forming a distributed innovation ecosystem that fuels local economic expansion)

[141] Teraco, a Digital Realty Company, *Teraco Commences Construction on 120 MW Utility-Scale Solar Power Plant*, Teraco, November 2024

generates sustained competitive advantages for African innovation. The flywheel effect ensures that well-targeted initial investments create lasting value that compounds over time, generating both attractive financial returns and transformational development impact.

Realizing this potential requires coordinated action across multiple stakeholders. The Africa Green Compute Coalition emerges as the mechanism to orchestrate this coordination, translating the flywheel concept from theoretical framework into practical implementation across the continent's diverse innovation landscape.

Aligning for Action: The Africa Green Compute Coalition (AGCC)

This assessment reveals a pivotal moment for Africa. The burgeoning demand for AI, coupled with persistent infrastructure challenges and market failures, underscores the urgent need for strategic, coordinated action. This critical juncture demands fostering African agency and ownership of compute infrastructure, not merely bridging a technological gap, but cultivating digital sovereignty and enabling locally driven innovation essential for Africa's future prosperity.

The Africa Green Compute Coalition is a community that aims to translate this vision into tangible action. As a cornerstone of the broader action plan, AGCC is committed to spearheading the development of a resilient, dynamic and green computing ecosystem designed to empower African AI builders and ensure the continent's equitable participation in the global AI revolution. The AGCC model is designed to transition to self-sufficiency after 2 to –3 years,[142] ensuring the coalition's long-term sustainability and independence from continuous donor dependency.

AGCC's strategic foundation

AGCC's mission is guided by four foundational principles:

1. **Empowering African AI builders** by facilitating accessible and equitable infrastructure and resources, ensuring both public and private sectors can actively contribute to the growth of AI communities across Africa.
2. **Building a resilient, equitable and sustainable market** that increases access to GPU hours while recognizing the delicate balance between expanding access, serving demand, and supporting a sustainable market that drives long-term innovation.
3. **Building upon existing compute infrastructure** by prioritizing the use and enhancement of existing technological systems and frameworks to avoid unnecessary duplication and ensure efficient utilization of available resources.
4. **Facilitating research-driven, evidence-based decision making** that provides rigorous research and data-driven insights to guide investment, strategy, funding and policy in line with the evolving needs of the African context.

AGCC recognizes and seeks to complement and build upon the valuable efforts of numerous existing initiatives, innovators and organizations already shaping Africa's digital and AI landscape. This includes major infrastructure projects, burgeoning venture capital investment, the critical role of philanthropy, major industry actors and governments. AGCC aims to align with and support these efforts, acting as a catalyst for synergy, accelerating investment pathways, and actively identifying and de-risking opportunities for new, impactful investments

[142] AGCC analysis

The AGCC's strategic pillars

AGCC operates through five interconnected strategic pillars designed to address the multifaceted challenges and leverage the unique opportunities identified in this assessment.

1. **Coordinating GPU access:** The AGCC works to optimize the utilization of existing GPU resources across the continent and facilitate equitable access to new, high-performance compute capacity. It will consider distributed 5-30 MW data centres and modular compute clusters that align with Africa's current infrastructure capacity and complement larger hyperscale investments. This approach prioritizes transferring ownership to local grantees to build sustainable local capacity. The coalition collaborates with local and global compute providers to ensure affordable and reliable access to compute while building GPU clusters to offer affordable and accessible GPU-as-a-service, prioritizing access for Africa's startups and innovators.
2. **Aggregating demand:** By consolidating and articulating the diverse compute needs across various sectors—from AI startups and academic researchers to government initiatives and large enterprises—AGCC will aggregate demand in a unified voice. This enables more effective engagement with hardware manufacturers and service providers, potentially unlocking better pricing, customized solutions and targeted infrastructure investments in Africa. By connecting AI builders with new infrastructure and organizing purchasing power, AGCC anchors strategic investments while supporting innovators with go-to-market strategies to secure early anchor clients, driving utilization and return on investment (ROI) for large-scale deployments.
3. **Shaping the compute market:** AGCC aims to actively shape the African compute market by orchestrating strategic partnerships, promoting awareness, and mobilizing capital at different stages. This includes advocating for transparent market information, stimulating competition to ensure accessible, affordable and high-quality compute services. Market shaping efforts focus on fostering local entrepreneurship in compute provision and related services, while identifying complementary investments to unlock value of compute. AGCC prioritizes partnerships with local suppliers who maintain stronger relationships with the continent and understand regional contexts. Africa's 5-30 MW infrastructure requirements create significant opportunities for local suppliers specializing in modular deployments and solutions that serve diverse African markets effectively. This includes de-risking adoption via shared procurement models and subsidy pilots in support of Africa's vibrant startup and innovation research ecosystem.
4. **Leading policy and advocacy support:** AGCC aligns closey with national and regional policy frameworks and contributes to shape them. A key focus is coherence with the African Union's Data Policy Framework and other emerging regulations concerning data governance, digital sovereignty and cross-border data flows. By working with policymakers, AGCC will help create an enabling regulatory environment that supports the secure, ethical and sustainable growth of Africa's compute infrastructure.
5. **Talent pipeline development:** Recognizing that infrastructure is only as effective as the human capital that builds and utilizes it, AGCC places strong emphasis on talent pipeline and venture building development. This includes fostering skilled professionals through targeted training programmes, academic partnerships and initiatives focused on data centre operations, AI/ML ops, green compute technologies, advanced AI development, and venture building to ensure viability of sustainable business models.

Stakeholder mobilization and partnership strategy

AGCC's effectiveness depends on its ability to convene and mobilize a diverse coalition of critical stakeholders whose cooperation is essential for systemic change. The coalition brings together hyperscale providers, development finance institutions, philanthropic organizations, sovereign wealth funds, government agencies, academic institutions and private sector actors around shared objectives that benefit all participants.

The coalition mobilizes hyperscale providers including AWS, Microsoft Azure and Google Cloud, which contribute technical expertise, global connectivity and substantial financial resources. AGCC engages these actors as partners rather than competitors, identifying opportunities for collaboration that serve both commercial objectives and development goals. Development Finance Institutions provide patient capital, risk mitigation instruments and policy expertise essential for large-scale infrastructure development, with organizations like the African Development Bank, DFC and Finnfund already demonstrating engagement in African compute infrastructure.

Academic and research networks—including AI4D, which supports 22 initiatives across 18 African countries, the Gates Foundation AI Community of Practice with nearly 70 Africa-focused initiatives and institutions like the African Institute for Mathematical Sciences—provide crucial research capabilities and talent development. Innovation ecosystems, through organizations like Co-Creation Hub, iHub and AfriLabs' network of approximately 500 innovation centres, demonstrate substantial entrepreneurial capacity seeking compute resources to scale AI applications.

Government engagement across countries developing national AI strategies—including Egypt, Kenya, Mauritius, Rwanda, Senegal, Sierra Leone, South Africa and , Tunisia, —creates policy environments that support infrastructure development. The African Union's Continental AI Strategy for Africa offers continental frameworks aligned with AGCC's mission, while initiatives like UNDP's Timbuktoo, aiming to mobilize $1 billion over 10 years to support 10,000 startups across Africa,[143] demonstrate the scale of coordination opportunities available.

Green compute leadership

AGCC's commitment to green computing positions Africa to leverage its abundant renewable energy resources while building sustainable infrastructure that attracts climate-conscious investment. Africa's unique renewable energy potential creates opportunities to develop compute infrastructure that reduces operational costs while addressing power reliability challenges through innovative hybrid approaches.

The coalition explores comprehensive off-grid and hybrid energy solutions that minimize dependence on existing grids while ensuring reliable operations. These range from solar kits paired with battery systems for AI workstations to modular compute infrastructure deployed directly at renewable energy generation sites. Successful implementations across the continent demonstrate the viability of this approach— organizations like CoLab in Nigeria

---

[143] UNDP, *UNDP And African Leaders Launch Timbuktoo Initiative to Unleash Africa's Startup Revolution*, United Nations Development Programme, January 2024

operate battery-powered systems with 3–4 day off-grid capacity,[144] while facilities such as Africa Data Centre's 7.5MW site combine 1 MW solar arrays with 96-hour generator backup for hybrid uptime and cost-efficient energy use.[145]

Large-scale initiatives—Teraco's 120 MW solar PV plant in South Africa and Moroccan data centres sourcing 100 percent electricity from wind farms[146]—exemplify the transformative potential for sustainable compute infrastructure development. The coalition prioritizes partnerships with providers already utilizing green energy while supporting those in fossil fuel-dependent areas to develop roadmaps to net-zero emissions, creating a comprehensive approach to sustainable compute across diverse energy contexts.

This green compute strategy creates sustainable competitive advantages that extend beyond environmental benefits. Reduced operational costs and climate-conscious investment attraction position Africa's compute sector for long-term growth aligned with global sustainability trends, while Africa's abundant renewable energy resources add a crucial dimension to the flywheel effect that enhances commercial viability.

AGCC in Action

The strategic framework outlined above is already being operationalised through the **AI Hub for Sustainable Development's** flagship programmes launching in 2025. AGCC has been instrumental in bringing together the critical partners and strategic thinking that are shaping these initiatives.

The **Compute Accelerator Programme** directly addresses immediate demand by providing GPU access and technical support to African AI builders and innovators, targeting organizations at different stages of compute readiness across Africa's innovation ecosystem.

The **Infrastructure Builder Programme** tackles supply-side challenges by supporting African infrastructure builders to develop foundational compute infrastructure—such as data centres, connectivity solutions, energy infrastructure and hardware innovations—while facilitating critical partnerships, capital access and regulatory guidance.

Together, these programmes activate the flywheel effect described in this report—connecting African innovators with infrastructure, resources, partnerships and technical expertise needed for sustainable ecosystem development.

Catalysing Africa's compute future

The evidence is clear: Africa has substantial documented demand for compute across research institutions, innovation hubs and developer platforms; proven pathways to significant cost savings through local infrastructure ownership; and compelling investment opportunities that can activate the flywheel effects necessary for sustainable growth.

AGCC provides the coordination mechanism to transform fragmented individual efforts into

---

[144] AGCC stakeholder interviews

[145] Africa Data Centres, *EADC goes solar for power back-up*, Africa Data Centres, April 2018

[146] Teraco, a Digital Reality company, *Iozera AI Data Center to Be Powered by Noor Solar Complex and Koudia Al Baida Wind Farm in Morocco*, Absolute Fusion, May 2024

systematic impact. Through strategic pillars—such as coordinating GPU access, aggregating demand, shaping the market, and leading advocacy—AGCC mobilizes a diverse coalition of hyperscale providers, development finance institutions, philanthropic organizations and sovereign wealth funds.

The question is not whether Africa needs compute infrastructure, it is whether stakeholders will act decisively to capture the opportunities this analysis has documented. AGCC's success will be measured by Africa's emergence as a confident participant in shaping the global AI future while serving African priorities and advancing continental development objectives.

ANNEX - Glossary of Terms

Term	Definition
Access model	The framework or set of rules that governs how users, devices, or systems can access computing resources—such as data, applications, storage or hardware (e.g., GPUs, servers). It defines who can access what, when, how, and under what conditions.
Affordability	The cost-effectiveness and financial accessibility of accessing and using computing resources.
Air cooling	Managing the temperature of electronic components by utilizing air as the cooling medium. This process involves directing air over hot components or surfaces to transfer heat and then expelling it away from the system.
Artificial Intelligence (AI)	Computer systems capable of performing tasks that typically require human intelligence, such as reasoning, learning and decision-making.
AI Hub for Sustainable Development	A collaborative initiative co-led by the Italian G7 Presidency and the United Nations Development Programme (UNDP), aimed at harnessing AI to advance sustainable development, particularly across Africa.
AI Laboratories (Labs)	A dedicated research and development facility focused on the creation, testing and implementation of AI technologies.
Bandwidth	The maximum amount of data that can be transmitted over a network or communication channel in a given amount of time. It is typically measured in bits per second (bps)
Capital Expenditure (CapEx)	Funds used to acquire, upgrade or maintain physical assets such as property, buildings, technology, or equipment.
Colocation	The practice of housing privately owned servers and networking equipment in a third-party data center facility. Instead of building and maintaining their own data centers, organizations rent space in colocation centers, which provide power, cooling, physical security, and high-speed internet connectivity.
Cloud credits	Promotional or subsidized units of cloud computing resources offered by cloud service providers to allow startups, researchers, non-profits or developers to access cloud compute infrastructure and services, allowing users to experiment, build, and scale resources without managing physical infrastructure.
Compute	Refers to the computational resources required for AI systems to perform tasks, such as processing data, training machine learning models, and making predictions. Specifically, it refers to the i) hardware—the chips (e.g., GPUs that perform calculations, process data and execute instructions), ii) the software—the interface that allows users to develop applications utilizing the raw capability of the hardware and iii) infrastructure—the physical locations that store the hardware, data storage devices and related network equipment.
Content Delivery	Geographically distributed servers that store and deliver cached content,

Network Points of Presence (CDN PoPs)	e.g., web pages or images to end-users based on their location. They are designed to reduce latency and enhance the performance of web applications by bringing content closer to users.
Central Processing Unit (CPU)	The primary component of a computer that performs most of the processing. It interprets and executes instructions from software and hardware, acting as the "brain" of the computer.
Data centres	Facilities that house computer and networking systems, with redundant power supplies, cooling systems, and security measures to store, process, and distribute data.
Data localization	The practice of storing and processing data within the geographic boundaries of the country or region where it was collected.

Data security	The practice of protecting digital information from unauthorized access, corruption, or loss throughout its lifecycle. This encompasses the implementation of technologies, policies, and procedures designed to ensure the confidentiality, integrity, and availability of data.
Data sovereignty	Refers to the principle that data is subject to the laws and regulations of the country or region where it is generated or processed.
Decentralized	Refers to distributing computing resources—such as processing power, data storage, and network functions—across multiple locations or nodes rather than relying on a single centralized system.
Demand	The need or requirement for computational resources—such as processing power, memory, storage, and networking capacity—by users, applications or services. In the context of this analysis, demand refers to requirements for compute resources by users.
Demand Pool	A demand pool refers to a defined group or category of compute users characterized by their need for computational resources.
Edge computing	A distributed computing framework that brings enterprise applications closer to data sources such as Internet of Things (IoT) devices or local edge servers. More broadly, it refers to any design that pushes computation physically closer to a user, to reduce latency.
Ethernet	The technology provides rules that allow network-connected devices to talk to one another without packet collisions (a packet collision occurs when two or more connected devices on a shared network attempt to transmit data packets at the same time).
Federated compute	A system allowing distributed computation while keeping data localized (a single problem is processed collaboratively in multiple computer systems, often at different locations, at the same time).

Gigabit	A unit of digital information equal to one billion (10^9) bits, commonly used to measure data transfer speeds in networks.

Graphics Processing Unit (GPU)	A specialized electronic circuit designed to accelerate images and visual data processing.
Green computing	Refers to the environmentally responsible design, manufacture, use, and disposal of computers and related technologies. In the context of this report, it refers to the leveraging of green energy solutions to power and cool compute resources/hardware.
Green energy	Electricity produced from natural, renewable sources that have minimal environmental impact. These sources include solar, wind, geothermal, biomass and hydroelectric power.
GPU hours	Refers to the total amount of time that a Graphics Processing Unit (GPU) is utilized for processing tasks. This metric is commonly employed in cloud computing and high-performance computing environments to quantify GPU usage and associated costs.
High Performance Compute (HPC)	Groups of computers networked together to perform complex calculations and data processing. They typically consist of multiple nodes working in parallel, with specialized networking infrastructure and storage systems.
HPC credits	Virtual currency or allocation systems used by cloud service providers and research institutions to manage and track the usage of computational resources. They allow users to access and utilize computing resources without the need for upfront payment. They are often provided to researchers, students, or institutions to facilitate computational tasks that require significant processing power.
HPC supercomputer	A type of HPC system that operates at or near the highest operational rate for computers, processing massive sets of data and complex calculations at rapid speeds. It achieves this by utilizing parallel processing techniques, where multiple processors work simultaneously on different parts of a task, significantly reducing computation time.
Hyperscale	The capability of computing architectures and data centers to scale rapidly and efficiently in response to increasing demand. This involves the seamless addition of computing resources—such as servers, storage, and networking components—to accommodate large-scale workloads.

Internet Exchange Point (IXP)	A physical infrastructure through which Internet service providers (ISPs), content delivery networks (CDNs) and other network operators exchange internet traffic between their networks.
Internet of Things (IoT)	A network of physical objects—devices, vehicles, appliances and other items—embedded with sensors, software and other technologies that enable them to connect and exchange data with other devices and systems over the internet.
Latency	The time it takes for data to travel between two points. Low latency means quick response times, ideal for real-time activities.
Load scheduling	The process of efficiently assigning computational tasks (or jobs) to various resources (such as processors, cores, or servers) to optimize performance metrics like processing time, throughput, and resource utilization.
Mattei Plan	The Mattei Plan, as it was presented in January 2024, is a project that brings together cooperation projects managed by Italian public and private companies in nine African countries, under the supervision of a steering committee,[1] headed by Ambassador Fabrizio Saggio and reporting directly to the Prime Minister's Office, for a planned duration of four years.[2] In its first formulation, the Plan had five "pillars": education and training, health, agriculture, water, and energy. Later in 2024, a sixth pillar was added, that of physical and digital infrastructures.
Multi-Factor Authentication (MFA)	A security mechanism that requires users to provide two or more independent credentials to verify their identity before gaining access to a system, application or data.
On-premise	Software, hardware or IT infrastructure that is physically located within the premises of an organization, rather than hosted on external servers or in the cloud.
Operating Expenditure (OpEx)	ongoing costs a business incurs through its normal operations. These are expenses necessary to maintain day-to-day activities and generate revenue.

Operating system	System software that manages computer hardware, software resources, and provides common services for computer programmes. It acts as an intermediary between users and computer hardware, enabling users to interact with the system and run applications effectively.
Optimization	The process of improving the efficiency, performance or resource usage of a computer programme or system. This can involve enhancing algorithmic performance, reducing computation time, minimizing

	memory usage, or balancing load across systems.
Port	A virtual point where network connections start and end. Ports are used by operating systems to manage multiple simultaneous network connections and direct data to the correct application or service.
Quota	A predefined limit set by a system administrator that restricts how much of a resource a user or application can consume, such as disk space or number of files. In the context of this report, it refers to the number of GPUs AI builders are allocated.
Queue	A linear data structure or abstract data type in which the elements are arranged in a sequential order and follow the FIFO (First In First Out) principle. Queues are widely used in scheduling, task management, buffering and communication between processes or systems.
Rack	A server rack is a structure used to house technical equipment including servers, networking devices, cables and other data center equipment. These racks are designed to optimize space and improve efficiency in data centres.
Reliability	The probability that a system will perform without failure under given conditions for a specified period.
Seed funding	The initial capital provided to a startup or early-stage company to help develop its product, conduct market research and cover initial operating expenses.
Servers	Specialized computing systems equipped with high-performance GPUs to handle the intensive computational demands of AI and Machine Learning (ML) workloads. These servers are designed to process large datasets, train complex models, and perform real-time inference tasks efficiently.
Software	A set of instructions, data or programmes used to operate computers and execute specific tasks. It encompasses everything from operating systems and applications to utilities and middleware, enabling hardware to execute desired functions.
Tensor Processing Units (TPU)	Custom-developed application-specific integrated circuits (ASICs) designed to accelerate machine learning workloads, particularly Google's TensorFlow framework.
Tiers	A standardized level that describes the infrastructure performance, reliability, and redundancy of a data center. The Uptime Institute’s Tier Classification System is the most widely used standard and defines four tiers: Tier I, Tier II, Tier III, and Tier IV.
Workstations	High-performance computers designed for technical or professional applications, e.g., 3D design, video editing, engineering, AI/ML training.

	They typically have powerful processors, large memory capacity, and advanced graphics capabilities to handle demanding workloads.

Annex: Methodology – Understanding demand for GPU hours for training amongst select groups of early-stage innovators

The intention of the preliminary estimate is to define a quantifiable level of compute use and demand, primarily for model training, necessary to accelerate early-stage AI innovation across selected user groups in Africa. The estimates are not meant to be exact and, as with other aspects of the report, are intended to solicit engagement and input to shape investment opportunities that advance the AI innovation ecosystem.

Approach: Using a bottom-up approach, engaged with five proximate groups of early-stage innovators (i.e. networks or entities actively operating across Africa) in order to deeply understand specific compute needs). These groups included a research network; an AI innovator network; venture builders; a large-scale developer platform; SMEs and government labs. Two to three stakeholders were selected from each group to more deeply understand their specific needs for early-stage AI innovation. Specific questions and areas of discussion included:

Overview of core use cases, research focus and intended impact.

Current and optimal compute utilization:

- *Current GPU usage (GPU hours)* — the current number of hours used for model training in either cloud or on-premise setups.
- *Ideal GPU usage for model training (if no constraints).* i.e. number of GPU hours they would use if neither cost, access nor technical capacity was an issue)

Analysis of compute use relative to peers — considered the representative nature of the interviewee relative to others. Defined: heavy, intermediate and light:

- **Heavy** – Users that rely on multiple high-performance GPUs to run continuous, large-scale computations. These users often operate dedicated GPU clusters, on-premise AI servers, or cloud-based GPU instances. Activities include large scale AI model training.
- **Intermediate** – Users requiring computational power for frequent AI/ML workloads, moderate 3D rendering and data-intensive applications. These users often need mid-range GPUs with more VRAM and processing cores. Activities include DL training and inference, cloud-based AI development, etc.
- **Light** – Users who run occasional, low-intensity tasks that do not require sustained GPU performance. These tasks can be handled by consumer grade or cloud-based free tier GPUs. Activities include data visualization, basic ML and inference, software development and Edge AI & IoT applications.

Composition of each "pocket of demand". Once categorized, engaged with interviewees and other stakeholders to consider overall composition of each group, *e.g., 60 percent light users, 30 percent intermediate users, 10 percent heavy users for a research network.*

Extrapolated the total number of users within each demand pool.

Consideration for 3-year growth rate.