AI Adoption in Sub-Saharan Africa: Choosing the Right Vehicles

Cédric Manouan[1,2], Miquilina Anagbah[4], N'guessan Yves-Roland Douha[2,3], and João Barros[2,3]
[1]Independent Researcher, Abidjan, Côte d'Ivoire
[2]Carnegie Mellon University Africa, Kigali, Rwanda
[3]Carnegie Mellon University, Pittsburgh, PA, USA
[4]Department of Electrical & Computer Engineering, University of Washington, Seattle, USA

https://tinyurl.com/compute-tracker-for-africa

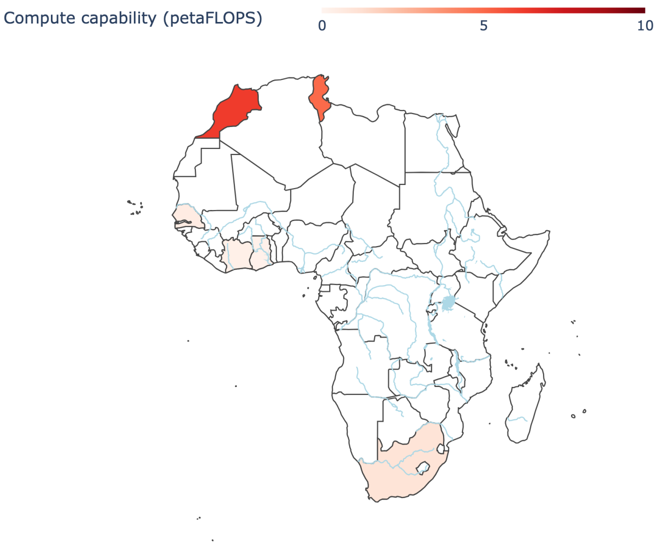


Fig. 1: Africa Compute Tracker (ACT). *Status by September 2025*

***Abstract*—Africa's participation in modern AI development is constrained by severe infrastructural and policy gaps. Important barriers include limited access to high-performance computing (HPC), restricted cloud access due to payment system mismatches, volatile exchange rates, and strict data sovereignty laws that fragment regional collaboration between African countries. Although recent public-private initiatives signal the growing interest in adopting AI in Africa, they remain insufficient and require better coordination between African countries. Drawing on Africa's declaration on AI adopted in early 2025 at the inaugural Global AI Summit on Africa, this paper primarily builds on the idea that sustainable AI capability requires robust digital foundations through balanced access to compute, data, and the energy that makes it possible. We refer to these foundations as the "right vehicles", considering them as crucial components for successful and sustainable AI adoption in Africa within the context of the global AI race. We also introduce the *Africa AI Compute Tracker (ACT)*, an interactive map that monitors the availability of AI-ready HPC systems throughout the continent. This tool represents the first open-source effort to consolidate data on Africa's evolving HPC landscape for tracking and monitoring, and aims to encourage more transparency from local AI stakeholders. By prioritizing infrastructure, accessibility, and affordability, this analysis underscores the urgency of tangible actions to close the AI divide and allow Africa to actively shape its AI future.**

***Index Terms*—African artificial intelligence; governance; data sharing; graphics processing units; compute; talent; policy.**

I. Introduction

Artificial intelligence (AI) is rapidly transforming industries, economies, and societies across the globe. From healthcare and agriculture to finance and education, AI systems offer unprecedented opportunities for automation, decision-making, and innovation [1]. However, access to these capabilities is unevenly distributed. While developed regions benefit from abundant computing clusters and AI-optimized cloud platforms equipped with the latest accelerator hardware, Africa seems to have no other choice but to contend with resources that often fall short of the demands of modern AI workloads and are several generations behind state-of-the-art technologies [2]. Although Africa boasts a youthful and burgeoning populace, dynamic

This work was not supported by any organization. Correspondence to: Cédric Manouan (email:cmanouan@alumni.cmu.edu).

tech sectors, and budding artificial intelligence (AI) communities, it is notably underrepresented in the higher spheres of AI advancement. This is primarily due to the fact that the continent contributes less than 5% to the global pool of research outputs [3, 4]. Despite these challenges, Africa's potential in the field of AI remains significant and ripe for exploration.

Africa's perceived lateness to the *global AI race* does not stem from a lack of talent or ambition, but from a structural digital divide rooted in limited *compute* infrastructure, restricted access to global cloud platforms, and systemic barriers to technological scalability. The implications are far-reaching: without the means—i.e., the right vehicles—to build and deploy AI systems, African countries risk deepening their dependency on foreign technologies, missing out on economic gains and limiting their ability to address uniquely regional problems with locally-informed solutions.

This paper argues that addressing Africa's AI divide requires a concerted effort to expand access to foundational resources encompassing (but not limited to) internet connectivity, AI-ready edge devices, and HPC systems. This article also discusses the key reasons that limit AI adoption in Sub-Saharan Africa (SSA). And, as such, it explores strategies that can enable the adoption of sovereign and inclusive AI in Africa under current conditions. Based on the concept of digital sovereignty, we also argue that effective AI participation for African states hinges on coordinated interventions across compute access, regulatory harmonization, financial systems, and energy infrastructure. Using a policy synthesis approach, we draw on publicly available second-order data on infrastructure development, emerging continental strategies, and analogies from high-performing global models to identify high-leverage interventions. Furthermore, our analysis draws on insights from *neural scaling laws* [5, 6] supporting the idea that access to adequate computing capacity and data is central to effective AI adoption.

The primary goal of this work is to help inform AI-related policies by presenting various cases from which we share recommendations. Another key contribution of this paper is the release of the *Africa Compute Tracker (ACT)*, a monitoring tool for HPC systems and AI-ready installations available on the continent (see Figure 1). This map relies on publicly available data on HPC system deployment and does not include systems with capabilities less than 100 teraFLOPS.[1] To make ACT a collaborative and evolving effort, a form[2] is open to submit details about HPC systems that are missing from the map.

The remainder of our analysis begins by presenting the general context of the study in section II, covering the geopolitics of AI chip shipments, foundational issues in SSA's AI landscape, and the enabling factors for successful AI adoption. In section III, we explore the policy measures that can drive the adoption of AI within the region, drawing lessons from various examples around the world. This section includes actionable pathways for building inclusive and forward-thinking AI systems on the continent. Section IV presents the limitations of this work, followed by the concluding remarks in section V.

II. Background and Related Work

Over the last decade, the global AI research and development landscape has been shaped by the widespread adoption of GPUs, which became essential with the rise of deep learning.[3] Their parallel processing capabilities made them particularly suitable for training large AI models, driving breakthroughs in fields such as computer vision [8], natural language processing (NLP) [9], speech recognition [6], or even scientific discoveries [10]. As a result, access to high-performance GPUs has become a key determinant of who can meaningfully participate in or shape cutting-edge AI development.

NVIDIA GPUs have become the de facto hardware accelerator of choice for AI practitioners [11] due to their strong integration with the deep learning ecosystem from the outset [12, 13, 14]. These AI chips come in two main categories: 1) data center and workstation GPUs, which are optimized for professional applications, and 2) consumer GPUs designed primarily for gaming. Recent architectures namely Fermi (2010) [15], Kepler (2012) [16], Maxwell (2014) [17], Pascal (2016) [18], Volta (2017) [19], Turing (2018) [20], Ampere (2020) [21], Ada Lovelace and Hopper (2022) [22], and the latest Blackwell (2024) [23] have played a pivotal role in driving significant AI breakthroughs in the last two decades of rapid technological growth.

A. Geopolitics of Global AI Chip Shipments

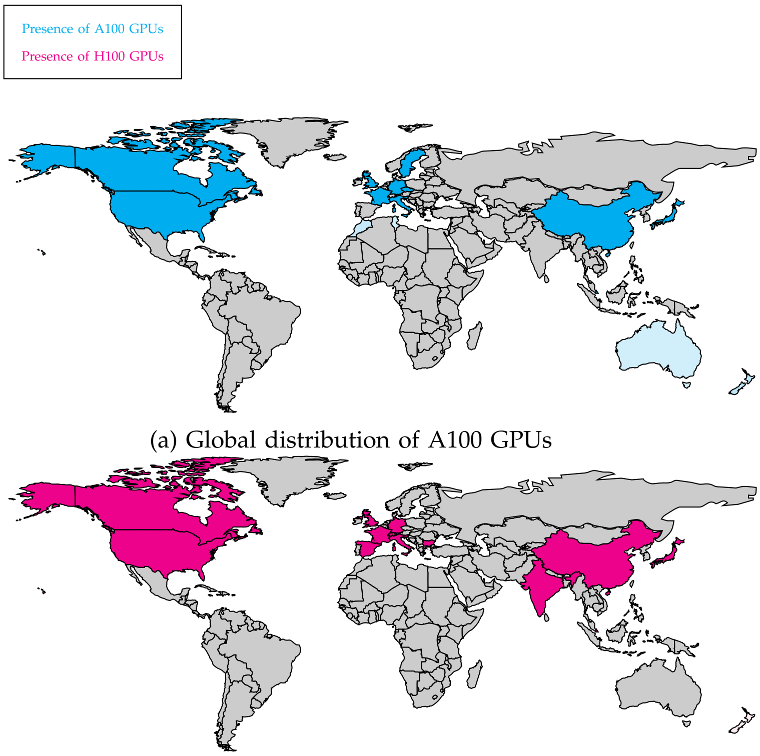
NVIDIA stands out as the main player in the AI accelerator field, securing a market share of about 94% in Q1 2025, while the remaining share belongs to AMD [24].

The state of AI report (compute index) 2024 [7] provides details about the evolution of AI compute platforms at the global scale. Figure 2a and Figure 2b (derived

[1] Considering systems with 100 teraFLOPS or higher is a design choice.

[2] https://forms.gle/yEJSvjLJ4WHDPZem7

[3] Deep learning is an approach to machine learning that uses (multi-layered) neural networks to automatically learn representations and patterns from large amounts of data.



(a) Global distribution of A100 GPUs

(b) Global distribution of H100 GPUs

Fig. 2: Advanced NVIDIA accelerators (Ampere/A100s and Hopper/H20s, H100s, H800s) across public, private, and national HPC infrastructure (2024). Diagram derived from [7] and based on public disclosure).

from the 2024 compute index report; see original data in Appendix VI-C) highlight the concentration of advanced NVIDIA GPUs in a handful of national and private infrastructure hubs, with no notable representation from sub-Saharan Africa. Organizations like Meta, Tesla, and DeepSeek control tens of thousands of these chips, while these resources seem far beyond the reach of many African institutions for reasons that are discussed in section II-B1.[4]

While notable initiatives are emerging on the continent, one typically has to look several NVIDIA generations back to see more of Africa's GPU assests on the world map. With clear disparities between sub-regions, Southern Africa is considered a part of the *compute South*—countries with presence of AI-ready compute—while most of the continent is part of the *compute desert*, i.e., countries with little to no presence of public cloud AI compute [25]. This trend underscores the need for pooled procurement and distributed HPC strategies in Africa, to facilitate access to modern AI hardware either through cloud platforms or on-premise installations.
The rapid adaptation of AI compute worldwide highlights the urgency for more equitable access to scalable, efficient, and sustainable digital infrastructure, especially in the global south. The presence of major cloud providers on the continent is predominantly concentrated in the southern region, with South Africa hosting data center facilities for companies such as Microsoft, Amazon, Google, Oracle, Huawei, and IBM. These data centers primarily serve local enterprises, offering services that include AI and machine learning capabilities. Nevertheless, the availability of cutting-edge GPUs in these data centers is limited, less documented, or arguably nonexistent [25]. As a result, this restricts the ability for local AI practitioners to build or deploy large-scale AI models unless they use overseas data centers, introducing latency and data sovereignty concerns (see section II-B1 for further details).

[4]The A100 is NVIDIA's advanced GPU equipped with its third generation tensor cores, offering exceptional parallel processing power, memory bandwidth, and scalability for modern AI workloads.

1) On the Power Dynamics

Starting in 2022, the United States (US) imposed restrictions on NVIDIA AI chip exports, with most countries—especially in Africa, the Middle East, and parts of Asia—falling into Tier 2 or Tier 3, the latter being the most restrictive category. These new policies are designed to limit the global proliferation of advanced AI hardware by imposing tighter controls on US-designed GPUs. The decsion was reportedly made to prevent China from accessing state-of-the-art AI compute that could enhance its internal capabilities, e.g., in the military domain. Nonetheless, one can argue that it has inadvertently motivated the innovation behind *DeepSeek*, one of the Chinese series of frontier AI models. Besides model architectures, DeepSeek introduced several innovations in hardware-aware optimization to achieve reduced memory usage, processing efficiency, and therefore cost reduction for training and serving large AI models.

For Africa, where access to advanced AI hardware is limited, those restrictions introduced by the US export controls further exacerbate the persistent digital divide [26]. Without carve-outs or special agreements, many African nations may have difficulties in acquiring the compute resources essential for modern AI research, development, and data center growth on their soil.

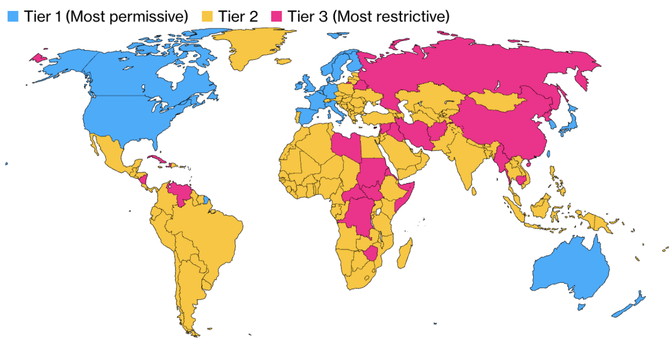


Fig. 3: US-imposed curbs on global AI chip shipments. Source: Bloomberg Reporting

B. Foundational Issues In Sub-Saharan Africa

As African policymakers become increasingly aware of the importance of computing power across the continent, it should be noted that the idea of AI compute encompasses more than technical access. This concept is tightly connected to those of affordability (section II-B1), reliability and environmental sustainability, (section II-B2), data security, and data sovereignty (section II-C2). In many developing regions, the rather limited digital infrastructure restricts access to cloud and HPC resources both locally and globally. In addition, the shortage of domestic professionals (science teachers and professors, data center engineers, network architects, cloud operations specialists, etc.) with expertise in AI-related domains also constrains the effective use of the digital infrastructure already accessible [27].

Notable efforts to accelerate the continent's digital transformation are underway, although core barriers still prevent Africa from fully harnessing its potential in the AI value chains.

1) Barriers to Cloud Access

Payment Limitations

Cloud service providers inherently require payment methods such as internationally accepted bank cards. However, in many African countries, mobile money[5] is the dominant—and widely preferred—medium for personal and business transactions due to its accessibility and deep integration into the daily life of millions of people [28, 29]. But despite its widespread adoption across the continent, mobile money is barely supported by cloud platforms, except for some local ones like ST Digital.[6] This mismatch is a significant constraint, especially for local startups, researchers, and small businesses that lack access to traditional banking services or international credit lines. In fact, this situation is common in Africa given the prevalence of informal businesses with low digital use [30]. As a result, a large segment of potential cloud users across the continent remains excluded from full participation in the digital economy, while already contributing to the overall economy with its about 85% of informal workers [31].

This is arguably an untapped opportunity for enhancing digital services and fostering economic growth on the African continent.

Currency and Pricing Issues

Globally accessible cloud services are predominantly priced in US dollars, particularly exposing users in low- and middle-income countries (LMICs) to volatile exchange rates and currency conversion fees. In countries with weak or unstable currencies, this can make the effective cost of cloud access prohibitively high. For example, in Zimbabwe and Sudan, currency instability and inflation have severely eroded purchasing power.[7] In Nigeria and Ghana, frequent devaluations of the Naira and Cedi, respectively, have increased the local cost of dollar-denominated digital services. Naira[8] devaluation has negatively affected small and medium-sized enterprises (SMEs) in the short term by increasing costs for imported inputs, making it harder to access foreign exchange and reducing consumer purchasing power. And even though local SMEs may

[5]Mobile money is a digital financial service that allows users to store, send, and receive money using a mobile phone without needing a traditional bank account.

[6]ST Digital is a pan-African company offering digital transformation and cloud services, with presence in more than five countries across Africa.

[7]The root causes of the currency instability differ for both countries. In Zimbabwe, these issues stem from a history of macroeconomic instability and frequent currency crises, while Sudan's economic collapse is largely due to the ongoing war, which has paralyzed production and increased reliance on mineral revenues.

[8]The naira is the currency of Nigeria.

become more competitive exporters due to the same situation, this devaluation can potentially have a negative impact on Nigeria's overall economic growth in the long run [32, 33]. Thus, when converting the prices for cloud-based compute into African currencies, users from the region may face significantly increased costs. In the end, these economic conditions make it difficult for local AI practitioners—a population mostly comprising students and young innovators (see section II-C3)—to consistently access cloud platforms, thereby widening the digital divide.

Challenges in Latency

The limited presence of major cloud providers with data centers in the continent implies that data often has to travel long distances, leading to increased latency and adversely affecting the performance of latency-sensitive applications [34]. In addition, reliable internet is crucial for efficient use of cloud services as it enables seamless data transfer, real-time processing, and smooth operation of applications hosted on the cloud. Hence a stable and fast internet connection ensures minimal latency, reducing the potential for delays in service delivery. On the other hand, excess dependence on data centers located outside the continent increases Africa's exposure to latency issues and increases its vulnerability to service interruptions and external network failures.

Most of the time, African users are routed to cloud endpoints located in North America or Europe, bypassing the continent entirely and increasing network inefficiencies. Study by Babasanmi and Chavula [35] found that accessing servers from Amazon Web Services (AWS) or Azure cloud regions from Africa incurs median round-trip times (RTTs) of 74 to 84 milliseconds, compared to 12 to 15 ms typically experienced in Europe and North America. Moreover, intra-African routing remains inefficient as more than 38% of first-hop internet connections from African probes exit the continent, and nearly 50% of inter-African traffic is routed through non-African exchange points. This creates arguably avoidable bottlenecks and exposes the region to greater risks from global network outages [36]. Furthermore in the context of speech/audio-based use cases, which are gaining traction in African AI, long distances between users and servers can lead to slower response times or delayed content loading [25].
In contrast, using content delivery networks (CDNs) with local points of presence (PoPs) in Africa reduces latencies to 29–65 ms, indicating an overall improvement of 25-87% [35].
The aforementioned measurable gaps underscore the urgency of building locally-based cloud infrastructure to ensure resilience, low-latency access, and service continuity.

2) Energy Requirements

Reliability and scale of electricity supply constitute a fundamental bottleneck in deploying HPC systems in SSA. AI-ready data centers and supercomputers are not only expensive capital investments, but also extremely power-hungry. A standard data center's yearly electrical consumption can be equivalent to that of anywhere between 25,000 to 100,000 households [37, 38]. And remarkably, some of those currently under construction are projected to use energy levels twenty times greater than this average consumption, requiring up to several megawatts [38].
Unfortunately, national grids in many SSA countries are currently unable to consistently provide the required level of electricity needed to power up the continent's AI ambitions. Although this problem is not uniquely African, SSA has particularly acute constraints. According to Afrobarometer,[9] less than half of Africans use electricity supplied by a reliable grid connection [39] and rural areas are especially disadvantaged. Almost a quarter of all African households use off-grid solutions, primarily solar panels and diesel generators [40].

Load-shedding, old transmission infrastructure, and a high cost of off-grid back-up diesel power make it unlikely that significant AI infrastructure will be sustainable.
Data centers globally accounted for roughly 1.5% of electricity usage (415 terawatt-hours) in 2024, which is growing at more than four times the overall growth of electricity demand. In particular, AI data centers are becoming similar to industrial operations such as aluminum smelters, often concentrated in specific areas [38]. For example, nearly half of all data center capacity in the US is located in only five regional clusters which facilitates the planning and delivery of energy. In comparison, there are no such regional clusters in SSA, but state-owned facilities like those presented in section II-B3.

The International Energy Agency (IEA) is projecting that data centers will be responsible for roughly 10% of the upcoming demand growth for electricity across the world by 2030, with regional variances. In developing economies like SSA, it is projected that up to 5% of the new electricity demand will come from data centers, while the overall electricity demand for developing economies continues to grow rapidly as they electrify. This situation adds another layer of pressure to energy systems that are already strained for reliability and affordability.
The AI-energy nexus in Africa, therefore, is nuanced. On the one hand, building competitively impactful

[9] Afrobarometer is a pan-African, research network with regional and national partners across Africa.

AI infrastructure requires continuous and substantial energy, which would further strain numerous (already fragile) grids and increase energy demand. On the other hand, the push for AI offers an opportunity to rethink power infrastructure altogether. Countries like Kenya, South Africa, Rwanda, and Senegal currently striving for nationwide electrification, are already leading renewable energy transitions [41] and could be considered as pilot sites for integrated and clean energy-powered AI hubs or enclaves.
With specific SSA regions strategically developed, HPC infrastructure could serve as both a digital enabler as well an anchor load to entice investment into reliable and renewable energy systems. Meeting these broader needs requires collaborative development on infrastructure and policy, and coordinated alignment with national energy plans and digital transformation aspirations. Without doing so, the region risks being excluded from the computational backbone of the AI era, not due to lack of innovation, but because, quite literally, the lights will not stay on.

3) Access to High-performance Computing

Africa hosts up to 35 supercomputing systems across 11 countries [42] aimed at bolstering research and development (see table I for a summary of recent installations). Meanwhile, when comparing Africa's HPC landscape to NVIDIA's state-of-the-art GPU offerings, the gap in compute capability becomes clear and concerning. These systems, though limited in scale, represent important progress in establishing AI-capable platforms in the region. Their presence signals a growing interest in HPC, even as most of the continent's compute capacity remains CPU-centric and largely focused on traditional scientific workloads. Let's explore these installations in a broader context.

Morocco's *Toubkal*[10] supercomputer delivers 3.15 petaFLOPS.[11] This performance is relatively equivalent to an NVIDIA Deep GPU Xceleration (DGX) A100 node, capable of delivering about 5 petaFLOPS (up to 10 petaOPS in INT8) of performance. *Toubkal* ranked 316th globally, before it was upgraded with about 20 Ampere A100 GPUs in December 2021.
South Africa's *Lengau* supercomputer is powered by 36 Tesla V100s and provides up to 1.029 petaFLOPS. It was ranked 121st on the world's TOP500 list of supercomputers at its launch by the center for high performance computing (CHPC). Though, in the current HPC landscape, this is below the performance of a single DGX-2 V100[12] node (2 petaFLOPS) released two years later.
Other notable GPU-accelerated supercomputers in SSA include Senegal's *Taouey* and that of Côte d'Ivoire. *Taouey*, powered by 48 V100s with an overall compute capability of 537.6 teraFLOPS, has reportedly supported applications in climate modeling, agriculture research, and AI through NLP for local languages.
Côte d'Ivoire's supercomputer is powered by 24 Kepler K80 GPUs, each offering up to 5.6 teraFLOPS of single-precision performance for a total compute capability of 322.56 teraFLOPS.

In contrast, NVIDIA's DGX Blackwell B200 launched in 2024 delivers between 72 and 144 petaFLOPS per node, figures that eclipse the aforementioned African HPC installations by multiple orders of magnitude. This disparity reveals both a performance deficit and a missed opportunity for Africans, resulting in domestic innovators being compelled to rely on smaller, and sometimes less competitive AI models. In reality, it is common for African AI practitioners to access advanced AI compute resources through cloud platforms as exemplified by the recent development of an open source text-to-speech (TTS) model for Senegal's wolof [43]. Over the long term, this situation could further impair SSA's involvement in the AI value chains. In particular, it may place SSA's AI workforce at a disadvantage by necessitating reliance on AI technologies developed and mainly accessible in *GPU rich* regions [2].

A focus is made on the age and type of the accelerator hardware powering HPC installations because for Africa to contribute to AI development, it is crucial to prioritize recent hardware with cutting-edge features. For example, techniques like *flash attention* and its variants [44, 45, 46] rely on specialized hardware such as Ampere- or Hopper-based GPUs, which boast advanced processing units for mixed-precision computations, warp-level parallelism primitives, and featuring asynchronous tensor cores or tensor memory accelerator (TMA). Many of these features available on recent GPUs are essential for performance optimizations. Older accelerators often lack these capabilities, restricting their ability to fully exploit the algorithms' potential or even run them at all. This is a major concern, considering the growing interest for large language model (LLM)-based applications in Africa.
Unlike *flash attention*, DeepSeek's *multi-head latent attention* (MHLA or MLA) technique [47] can run on an NVIDIA V100, but i significantly slower due to the

[10]Toubkal is the only African supercomputer listed in the TOP500 dataset of November 2024; see top500.org/lists/top500/list/2024/11/
[11]This is the performance reported by the host institution before upgrade was made in December 2021. We believe the overall capacity might have improved to over 9 petaFLOPS.
[12]The V100 was rapidly adopted because it introduced NVIDIA tensor cores, units specifically designed to accelerate matrix operations particularly for deep learning and HPC.

V100's limitations in VRAM[13] and memory bandwidth.

TABLE I: GPU-powered HPC installations in Africa (based on published data)

Country	Installation	Performance (petaFLOPS)	GPU Acc.
Tunisia	DGX A100 (2020)	5.0	A100 (x8)
Morocco	Toubkal (2020)†	3.15	None
South Africa	Lengau (2016)‡	1.029	V100 (x36)
Senegal	Taouey (2019)¶	0.5376	V100 (x48)
Côte d'Ivoire	Unnamed (2018)§	0.3226	K80 (x24)

†https://cc.um6p.ma/toubkal-super-computer
‡https://wiki.chpc.ac.za/chpc:lengau
¶https://cineri.sn/carasteriqtiques/
§https://cncci.edu.ci/cncci/

Figure 4 reinforces the centrality of hardware access by tracking the increase in citations of specific NVIDIA chips in AI-related publications. We see that starting from its release in 2020, the A100 has gained more popularity to rapidly establish itself as the most frequently cited accelerator by 2024. Following a similar trend, the Hopper H100 is quickly rising in prominence. It is also worth noting the decrease in citations of the Tesla V100 as the AI community gradually shifts toward the latest chips. Meanwhile, the growing presence of the RTX 3090 and 4090[14] suggests that many researchers, particularly in constrained environments, are relying on consumer-grade accelerators for their tasks. This is mostly the case in academic and small business settings where budget limitations necessitate the use of lower-grade hardware for R&D activities.

Implications Local AI Talents

In an article co-authored with Zindi and Alliance4AI, the UNDP reveals that only 5% of Africa's AI talent has access to the computational resources required to support complex AI tasks [2]. This conclusion was drawn from an analysis of compute usage by up to 11,000 data scientists from the Zindi community, which represents both Africa's largest network of AI practitioners and the continent's reference platform for AI-related competitions. Furthermore, the remaining 95% of AI builders on the platform rely on accelerators provided by Google's hosted cloud service *Colab*,[15] offering default access to GPUs like the NVIDIA Kepler K80, Turing T4, or Pascal P100, free of charge for a couple of hours weekly or A100s on paid subscriptions. However, it's important to highlight that the A100 significantly outperforms *Colab*'s free GPUs for development workloads. Its larger memory (40 or 80 GB as opposed to the 16 or 24 GB) eliminates the need for complex model parallelism techniques [48]. Consequently, the limited access experienced by African innovators implies they may have to wait way longer to iterate on experiments, compared to a few minutes for their peers in developed countries [2].

Emerging African Compute Initiatives Cassava Technologies, in collaboration with NVIDIA, is building a network of AI factories across Africa. This project is estimated at about $720 million and plans to deploy up to 12,000 GPUs[16] across multiple countries, including South Africa, Egypt, Nigeria, Kenya, and Morocco [49]. A few months following Cassava AI's announcement, Synectics Technologies and Schneider Electric have also unveiled a partnership to establish Uganda's first AI factory within the 600MW Karuma hydropower station.[17] Referred to as the *Aeonian Project*, this 100MW hybrid Tier-4 Plus (4+) off-grid facility[18] will be backed by global partners including NVIDIA, the GIZ, HAUS, the EU Development Fund, and MDCS.AI among others. The project comprises multiple stages, with its initial phase—which encompasses a 15MW AI module and a 10MW sovereign supercomputer (USIO)—planned for launch in the second half of 2026 with full capacity expected by H2 2027. The USIO supercomputer is expected to incorporate NVIDIA's GB300 Blackwell GPU, along with a complete NVIDIA enterprise AI ecosystem, designed to support diverse application areas like healthcare and life sciences among others.[19] Notably, the *Aeonian Project* will operate solely on renewable energy, making it one of the world's greenest AI facilities.

Another key initiative led by Microsoft in partnership with UAE-based AI firm G42 is the announcement of a $1 billion investment to build a data center in Kenya, aiming to expand Microsoft's Azure services to East Africa [50].

The Africa Green Compute Coalition (AGCC), a collaborative initiative led by the UNDP, has been created to coordinate sustainable AI computing across Africa. As part of the AI Hub for Sustainable Development,[20] AGCC advocates for a robust, resilient, and sustainable computing ecosystem by establishing a pan-African governance and financing framework. It seeks to unlock the potential of both cloud and

[13]Video Random Access Memory (VRAM) is a dedicated type of memory on a GPU, used to store and quickly access graphics-related data.

[14]NVIDIA RTX 3090s and 4090s (24GB VRAM each) are originally gaming GPUs that meet the requirements for relatively low-scale AI R&D.

[15]https://colab.research.google.com/signup

[16]Details about the GPUs powering these AI factories have not been disclosed, but it is expected to be a mix of high-end data center units.

[17]www.uegcl.com/power-plants/karuma-hydropower-project

[18]Typically, the installation will be a hybrid hyperscale data & high-performance computing centre (DHPC).

[19]https://renewableenergynews.co.ke/synectics-technologies-and-schneider-electric-announce-africas-first-sovereign-supercomputer-and-ai-factory-ecosystem-the-aeonian-project/

[20]The AI hub for sustainable development is an initiative co-led by the United Nations Development Programme (UNDP) and the Italian G7 presidency, see www.aihubfordevelopment.org.

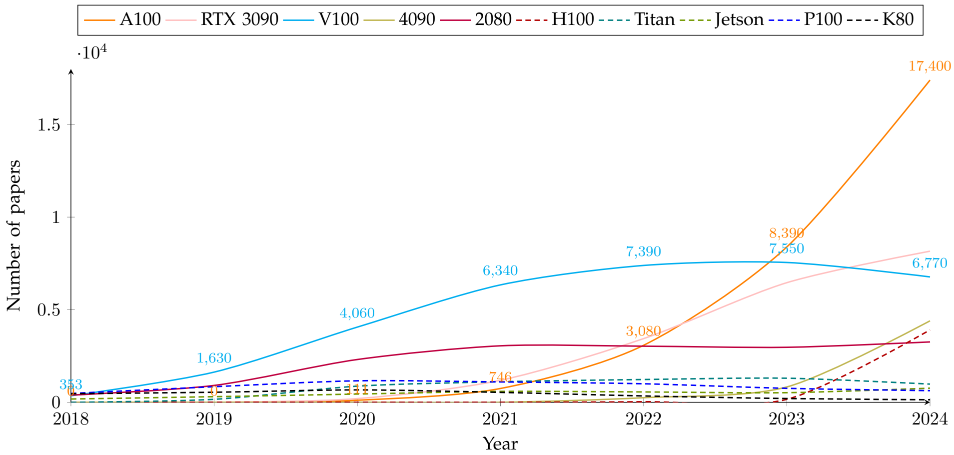


Fig. 4: Cited NVIDIA GPU usage in open source AI papers from 2018–2024 (estimates from the state of AI report/compute index and Zeta Alpha).

on-premises GPU-as-a-service options. The interim findings from the AGCC report [27] present an immediate investment opportunity worth $150 million, addressing a documented demand for **7 million GPU hours** solely for AI model training. These initiatives collectively mark a significant step toward embedding Africa in the global AI value chains by addressing a key prerequisite: democratized access to AI-ready compute infrastructure and cloud-based platforms. Situating HPC resources within African borders will stimulate regional innovation ecosystems, support the development of domestic AI startups, and foster public-private partnerships (PPPs) to deploy AI applications in priority sectors.

It is essential to acknowledge that African countries are not equally prepared for AI adoption, with varying levels of prerequisites before full implementation can occur. However, by observing how AI compute projects are being rolled out, it seems there may be competing efforts emerging across different parts of the continent. Having said that, we should also recognize that regional coordination should complement rather than compete with each country's pursuit of sovereignty in AI.
All in all, this situation underscores the need for close collaboration between African nations to ensure that common interests are satisfied. Transparent cooperation can help address regional disparities in AI readiness, while respecting national autonomy in decision-making processes.

C. On the Enabling Environment

1) The African AI Governance and Policy Landscape

Leaders around the world have recognized the need for global governance of AI as well as an enhancement of global cooperation [51]. Along the same lines, the AI governance landscape in Africa is evolving to address unique regional challenges while leveraging local opportunities. It involves establishing comprehensive frameworks that include laws, strategies, and policies tailored to African needs [52]. Key priorities revolve around 1) ensuring equitable benefits from AI advancements, 2) protecting human rights, and 3) safeguarding the environment amid rapid technological growth. Governance efforts also focus on creating robust ecosystems for AI development, preventing exploitation of resources, and fostering accountability in AI deployment.

Considering that approximately 30% of AU member states have published validated or draft AI strategies [53]—with over 25 more actively developing theirs—one cannot neglect the growing recognition of AI's strategic importance to African leaders. This somewhat proactive approach demonstrates the commitment of numerous governments to harness the potential of AI for economic growth, innovation, and social advancement. Yet, the existence of these strategies alone is not a guarantee of success. While the prioritization of stand-alone strategies is promising, effective implementation definitely hinges on a comprehensive ecosystem. Such an ecosystem should be grounded in essential foundational activities like capacity building, fostering collaboration among stakeholders, providing access to cutting-edge technologies, ensuring sustainable funding models, and establishing robust digital infrastructure. These elements are crucial for fully realizing the potential of these strategies across various domains. Without substantial investment in these areas, AI strategies risk remaining aspirational documents rather than catalysts for tangible progress.
Rapid development of these strategies also highlights the potential risk of outpacing available resources and

expertise, emphasizing the critical need for a balanced approach that integrates policy with practical execution. Attempting to build the foundations by establishing supportive digital infrastructure and demonstrating immediate impact simultaneously often leads to prioritization of short-term (but more visible) outcomes at the expense of long-term (but more impactful) actions. Moreover, going down that route risks creating a cycle of fragmented efforts, where initial AI projects might produce immediate benefits but ultimately fail to scale due to an underdeveloped basis encompassing an appropriately skilled workforce, or AI-specific regulations. The pressure to showcase rapid progress can inadvertently undermine the very foundations needed for lasting success.

African nations encounter considerable obstacles as they establish AI policy frameworks grounded on unfulfilled digitalization ambitions. A critical concern is the excessive focus on integrating AI across various economic sectors without due consideration of how other digital technologies could bolster the intended socioeconomic objectives [54]. Moreover, while some countries have adopted multi-stakeholder engagement, the limited representation of diverse groups in the formation of their AI policy frameworks may exacerbate digital disparities within the continent. Lastly, often perceived as mostly consumers of AI technologies [55, 56] and with many local AI policy frameworks lacking clarity concerning the continent's active role within global AI governance frameworks [57], the region may struggle to effectively contribute to shaping the future direction of AI on a global scale.

2) *Data Sovereignty Concerns*

Policies on data sovereignty, although essential for protecting national interests and citizen privacy, can impose significant constraints on local AI development, particularly in emerging markets like those across Africa. These policies often mandate that data generated within a country must be stored, processed, and sometimes even analyzed within national borders. In essence, this protects against foreign exploitation and ensures regulatory control, but can inadvertently result in limited access to global-scale cloud platforms that offer the compute and tooling necessary for AI research and deployment.

On the one hand, the African Union (AU) is promoting cross-border data flows (CBDFs) with its data policy framework [58]. This framework emphasizes a comprehensive approach to data governance that extends beyond just personal data protection, which has been a primary focus globally and on the continent. More specifically, it is deeply connected with the development of regional and continental data infrastructure to support digital integration, research, development, and innovation in areas including AI and big data analytics. On the other hand, there seems to be a perceived disconnect between continental initiatives and local priorities within member states [59] as exemplified by the low rate of ratification of the *Malabo Convention*,[21] which only entered into force in 2023 after being ratified by Mauritania almost ten years after its adoption. This reluctance can be attributed to several factors. Primarily, data policy is often seen as a matter deeply tied to national sovereignty, with countries prioritizing their own regulatory frameworks over supranational agreements. This perspective is evident in the varied approaches to data protection and localization across African countries like Senegal, Nigeria, Zambia, Côte d'Ivoire, Benin, and Morocco, where some enforce strict mandates that require substantial policy adjustments and investment to align with AU's goal of harmonized cross-border data flows. This situation is also the consequence of the differences in levels of digital readiness, which can influence a country's willingness or ability to commit to such conventions. Secondly, political considerations and the complexities involved in coordinating multiple stakeholders across different jurisdictions contribute to the slow ratification process. In fact, intergovernmental collaboration involves navigating bureaucratic processes and institutional capacities that may differ significantly across the continent. Ensuring effective communication and alignment among diverse legal systems and regulatory bodies requires resources that should be strategically allocated.

Collectively, these factors highlight an obstacle that the AU is trying to overcome: reconciling diverse local priorities with overarching regional goals for data governance across the continent [60].

Considering that many African countries are still building their foundational digital infrastructure, e.g., top tier data center or high-speed fiber networks, there should be a balanced approach to regulation in the data domain. As a result of restrictive policies, African researchers and innovators may be forced to work with limited datasets or underpowered systems, which consequently can slow down innovation. Additionally, compliance burdens can deter international cloud and AI companies from offering services in countries with strict or ambiguous sovereignty laws, further isolating local ecosystems.

Addressing these challenges requires a concerted effort from all stakeholders in order to implement policies that promote data accessibility without hindering innovation that can potentially benefit many African communities.

[21] Article 36 of the AU Convention on Cyber Security and Personal Data Protection (commonly referred to as the Malabo Convention) states that the treaty would come into effect once 15 ratifications were achieved.

3) Digital Skills and Training Initiatives

The AI talent readiness index for Africa [61] highlights distinct regional patterns in the continent's capacity to effectively participate in AI value chains as an actual contributor. The first pillar of this index assesses nations' digital skills capabilities. It assesses the availability and quality of AI-related education, the volume of science, technology, engineering, and mathematics (STEM) graduates, and the overall AI literacy within the workforce. Key indicators include secondary school completion rates, the proportion of the labor force holding higher education degrees, the representation of women in STEM fields, the number of institutions offering AI/Machine Learning education, and the density of developers per million people.

Overall, North Africa leads with Tunisia, Egypt, Algeria, and Morocco among the top 10 countries, driven by advanced educational environments, high developer density (over 4,000 per million people in Tunisia), and key digital transformation initiatives led by local governments. With a regional average of 38.2 out of 100, North Africa comprises up to 85 institutions with dedicated AI training.

East Africa demonstrates good progress, as evidenced by Kenya, Mauritius, and Rwanda collectively nurturing distinct talent ecosystems, despite pronounced disparities within the region. This relatively positive trajectory is reflected in a regional average score of 32.7, illustrating both advancements at individual levels and ongoing challenges in cultivating a skilled workforce regionally.

West Africa—with an average score of 27.6—holds is in the middle position. In this region, Ghana stands out for its well-rounded growth, while Nigeria is noted for its considerable but yet to be fully realized potential, highlighted by its large economy and dynamic startup scene.

Central Africa presents a challenging scenario with an average score of 19.4 out of 100, indicating substantial hurdles that need immediate attention. To develop sustainable AI talent ecosystems, the region requires critical interventions aimed at improving digital infrastructure, expanding electricity and internet access, and enhancing governance frameworks. Cameroon leads the region with a score of 42.35 and is 11th overall.

South Africa leads the Southern region with its cutting-edge digital infrastructure and leadership prowess. Nevertheless, the region encounters difficulties in extending digital skills beyond its urban centers. And although South Africa secures first place continent-wide with an impressive score of 52.15, the regional average remains at 35.3, indicating a gap that needs to be bridged for more widespread digital literacy.

Many academic institutions across Africa are contributing to cultivating technical talent, while community-driven efforts like the Masakhane project[22] demonstrate the impact of grassroots collaboration in developing AI tools for African languages.

Academic Institutions

The Mohammed VI Polytechnic University (UM6P) in Morocco, which has established an International Center for AI to foster Moroccan expertise in AI and Data Sciences.

CMU-Africa, a partnership between Carnegie Mellon University (CMU) and the government of Rwanda, offers a special program in engineering artificial intelligence that equips learners with advanced skills covering key sectors such as transportation, energy, and healthcare.

The University of Pretoria (UP) in South Africa hosts the data science for social impact group,[23] a research group that has been a very active contributor to Africa's AI advancement, particularly in African language NLP and ML research.

The African Institute for Mathematical Sciences (AIMS)—with centers in South Africa, Rwanda, Ghana, Cameroon, and Senegal—offers advanced training in AI and machine learning, including programs like the AI for Science Master's in partnership with Google DeepMind.

Grassroot AI Communities

Besides efforts by academic institutions, *Deep Learning Indaba (DLI)*[24] serves as a continental movement for strengthening AI capacity by organizing regular gatherings across Africa, promoting collaboration and knowledge sharing among local researchers, AI practitioners, and especially students.

Local AI communities like Galsen AI in Senegal,[25] AI Kenya,[26] Mbaza NLP in Rwanda, Data Science Nigeria,[27], and numerous others are working diligently to promote widespread access to AI throughout the region. These communities primarily consist of young developers, researchers, and students.

Examination of recent data from DLI [62, 63] reveals that Africa's most active AI practitioners are predominantly students at both undergraduate and postgraduate levels. This contrasts with traditional African research habits, which have typically relied on PhD students and senior researchers to drive academic endeavor. The shift reflected in DLI reports highlights a growing trend

[22] www.masakhane.io
[23] https://www.dsfsi.co.za/
[24] Deep Learning Indaba is the largest AI-focused conference on the African continent.
[25] https://galsen.ai
[26] https://kenya.ai
[27] https://datasciencenigeria.org

toward empowering early-career researchers[28] (ECRs). Enhancing opportunities for young researchers to rise through the ranks and take on prominent roles can further strengthen Africa's burgeoning AI ecosystem.

In addition to AI-specific initiatives, valuable lessons can be drawn from the continent's growing robotics ecosystem, which is evolving despite similar constraints. Programs like the African Robotics Network (AFRON) and its successor, AfRob, alongside regional robotics competitions and training efforts by institutions such as Ashesi University, Fundi Bots in Uganda, and the Rwanda National Robotics Program, have demonstrated the power of community-driven innovation in the face of infrastructure scarcity [64]. These robotics initiatives emphasize the importance of hands-on education, low-cost but efficient hardware, and regional collaboration, all of which are equally critical to fostering local talent and encouraging homegrown solutions. Integrating robotics, AI, and embedded systems training across African educational systems, from basic to advanced levels, could help build a more robust pipeline of practitioners equipped to navigate and close the digital divide.

4) *Scaling Laws and Implications for Africa*

In the context of AI, scaling laws are empirical relationships that describe how the performance of AI models (particularly LLMs) improves predictably as we increase their size, the amount of data they are trained on, or the available computing power. Nowadays, these laws are crucial to understanding the limitations and potential of AI models, often prior to running experiments to help properly scope them.

In their seminal work [5], OpenAI argued that the training objective[29] of a language model scales as a power-law with respect to model parameters, dataset size, and compute capabilities. However, larger models or more data do not always result in better outcomes unless they are balanced correctly with the available computing resources. They typically argue for a prioritization of model scaling over data. For example, research indicates that there is an optimal point where increasing a model's size and data input can lead to diminishing returns without corresponding increases in other areas.

Building upon [5], the *Chinchilla* scaling laws—proposed by Google DeepMind researchers—suggest that for a given compute budget, the optimal strategy for training large language models is to increase both the model size (number of parameters) and the training data size (number of tokens) at equal rates. Specifically, their findings show that a more balanced scaling of parameters and data leads to more compute-efficient and better-performing models. This latter formulation of the scaling laws emphasizes that beyond a certain point, increasing model size without proportionally increasing data does not yield significant performance gains.

The development and even the deployment of state-of-the-art AI models like OpenAI's GPT-4, DeepSeek, or Meta's LLama underscores the critical role of substantial compute, extensive and quality datasets, or simply put: significant investment. These models have been trained on vast datasets using thousands of advanced GPUs, which therefore results in significant costs. Developing them reportedly implied spending between a few million and several hundreds of million US dollars. As an example, training LLama 3 405B [65], considering the 16,000 H100 GPUs used, can be roughly estimated at over $60 million, based on GPU rental costs of approximately $1,500 per GPU per month over a three-month period.[30]

There have also been debates regarding the disclosed budget for the development of DeepSeek-V3 [66], estimated at roughly $5.6 million for the official training run for up to 2 months. Usually, these estimates primarily account for GPU rental/usage during the final training phases, omitting substantial investments in hardware acquisition, data center infrastructure maintenance, and data procurement. Moreover, the development process of new deep learning models has continuously involved extensive R&D activities, including multiple experiments for hyperparameter tuning and ablation studies, which contribute significantly to the overall cost. Therefore, when considering the comprehensive infrastructure-related expenses encompassing hardware, energy, data acquisition/pre-processing, and iterative experiments, the actual cost for developing DeepSeek-V3 likely far exceeds the publicly stated figures.[31]

Reflecting about scaling laws in the African context can help calibrate ongoing efforts toward AI adoption. Although limited access to compute and large-scale data constrains the ability for local AI practitioners to develop frontier models exclusively relying on domestic resources, understanding these laws allows for strategic

[28]Early career researchers are scientists or scholars in the initial stages of their research careers who may be pursuing advanced degrees such as Master's or PhDs; recent graduates who have started working on research projects independently for the first time.

[29]In the context of machine learning, an objective refers to quantifying the goal the model is built to achieve during its training process using mathematical functions, optimization algorithms, and evaluation metrics to measure performance.

[30]See https://lambda.ai/pricing for an estimate of the pricing. Renting H100 GPUs typically costs over $2.2 per GPU-hour, for reserved instances.

[31]In their analysis, the authors pointed out that the estimated expenses solely cover the official training of DeepSeek-V3, while omitting costs related to prior research and ablation experiments on architecture, algorithms, or data.

allocation of the available resources.

For Applied Research

Scaling laws do not discourage African nations from pursuing frontier AI development. Rather, they highlight the strategic choices that should be made for the continent to fully participate in this revolution. Training large models entirely from scratch may be financially or technically out of reach for many countries on the continent, but leveraging existing open-source models and adapting them to local languages and contexts offers a practical and impactful alternative. In particular, the high-demand for local language integration, cultural nuances, and sector-specific needs especially in agriculture, health, education, and public services creates opportunities where fine-tuning could result in more relevance and lasting impact.

For Basic Research

AI breakthrough have historically been fueled by the concurrent availability of relevant compute, datasets, and adequate skills. Even modest investments in compute capacity can have outsized effects. A notable example is NVIDIA's 2016 donation of one of the firsts DGX-1 supercomputer to OpenAI, which drastically accelerated the firm's early breakthroughs. Similarly, African AI practitioners don't necessarily need on-premise hyperscale infrastructure before starting building for their communities. Strategic partnerships around cloud compute platforms can unlock local development, while concurrent efforts are made for deployment of AI factories within the region.

Building on Africa's AI declaration [67], each African country should further articulate its aspirations for engaging in both basic and applied research to ensure that investments are appropriately directed.[32] For those with existing AI strategies, it involves clarifying their implementation roadmaps to facilitate collaboration on shared goals.

The increasing focus on cost-effective solutions should not be misconstrued as a one-size-fits-all approach for all use cases. While there are indeed applications that can benefit from frugal AI models, some like written and spoken language processing inherently thrive when using large, sophisticated models. In low-resource automatic speech recognition (ASR), scaling up model size tends to bring proportionally larger gains than in high-resource settings. The study by Chen et al. on scaling laws for multilingual speech recognition and translation models [68] provides an excellent demonstration of the point in question. This research reveals that transitioning from relatively compact models (around 1B parameters) to larger speech models (approximately 9B parameters) significantly reduces word error rates for languages with limited training data (less than 35 hours). Thus, when data availability is restricted but can be improved in some dimensions (quality, diversity, or distribution), investing in a larger shared multilingual model and expanding data where feasible yields substantial returns, rather than solely relying on smaller models to extract marginal improvements.

It is therefore essential to clearly delineate the needs of each use case and match these with the most appropriate models accordingly.

III. Africa's Path Forward

A. Accelerating and Adapting Capacity Building Efforts

Empowering Africans with advanced AI skills is essential to unlock the continent's potential and foster a vibrant ecosystem of talented professionals who can contribute significantly to both local and global innovation. To this end, efforts should be redoubled to better equip early-career researchers with the necessary expertise to transform the youthful population into skilled contributors in AI and related fields. Achieving this goal requires not only providing access to education and training but also creating an environment that encourages and retains local talents. Incentives such as valuing research through grants allocations, decent salaries, and opportunities for career advancement can go a long way in attracting and retaining top talent in the region.

B. Breaking the Cycles of Short-term Efforts

Despite deep potential and ambitious experiments, uneven capacity and lack of strategic scaling frameworks hinder the transition from pilot to widespread adoption [69, 70]. Looking at some of the factors influencing post-hackathon project continuation in the region, Ratsoga and Primus [71] argue that discontinuity is largely due to insufficient funding, skills gaps, weak infrastructure, limited policy support, and poor integration of projects into existing systems. This suggests existing local creativity but also highlights a persistent gap between ideation and operational maturity in Sub-Saharan Africa's digital ecosystem. These short-term innovation cycles, often driven by visibility goals or donor incentives, seldom transition into scalable, production-ready systems.

AI-focused entrepreneurship in Africa is growing but remains structurally constrained. While more than 200 deep-tech startups have emerged across the continent—79% founded post-2014—the majority operate in early stages with limited access to funding,

[32] The declaration was adopted at the inaugural Global AI Summit on Africa in April 2025.

infrastructure, or specialized support. AI and machine learning represent the most active tech cluster within deep-tech, with applications spanning agriculture, healthtech, fintech, and logistics. The gap between university research and market-ready talent also hampers scalability [72].

Funding to AI startups in Africa is increasingly becoming more concentrated, as countries like Kenya, Tunisia, Egypt, South Africa, and Nigeria reportedly account for over 90% of total investments in AI on the continent.[33] While this trend highlights the growing interest from global investors in the region's burgeoning AI landscape and offers promising opportunities for early-stage AI companies, there is still a need to address the cyclical nature of short-term projects that can create a sense of recurring restart rather than sustained growth. To foster long-lasting development in AI, it's essential to focus on establishing consistent and reliable funding mechanisms for startups across the continent. This could involve creating an environment for venture capital (VC) funds dedicated exclusively to AI investments—as also suggested by Africa's declaration on AI with a continental AI fund [67]—and offering government grants for promising projects, or implementing tax incentives for companies investing in AI R&D.

Several African countries have already made strides by establishing startup acts or dedicated legal frameworks aimed at fostering the activities of domestic startups and SMEs.[34] However, the impact on the ground remains limited. Bridging this gap necessitates addressing potential obstacles such as bureaucratic red tape, lack of resources for enforcement, and insufficient coordination between government agencies and private sector stakeholders.

Without sustained institutional support and robust mechanisms for long-term development, promising ideas by African innovators risk being abandoned before achieving the expected real-world impact.

C. *Effectively Realizing Africa's AI Ambitions*

Africa's declaration on AI is a key instrument that aims to facilitate mutualization of efforts among African countries in AI adoption. By aligning with existing frameworks such as the AU continental AI strategy [73], Smart Africa's blueprint on AI for Africa [74], the AU data policy framework [58], the "Malabo Convention" on data protection [53], and the UN global digital compact [75], it ensures consistency in AI-related policies across the continent.

To support implementation, the declaration proposes a $60B AI fund drawing participation from public, private, and philanthropic sources. Additionally, it emphasizes the need for national policies and frameworks to be aligned with the AU continental AI strategy, as well as establishing an African AI council co-chaired by the AU Commission and the International Telecommunications Union (ITU), and led by the Smart Africa Steering Committee. These measures aim to create a united front for AI adoption in Africa, although they do not necessarily prevent countries from undertaking unilateral initiatives as it has been the case for several years.

The realization of the continent's ambitions for AI requires a deliberate approach, as adoption will unfold gradually rather than through top-down implementation. Although AI strategies have been developed since 2018,[35] translating these into tangible outcomes demands sustained effort to address the aforementioned systemic challenges. Gradual adoption allows for iterative testing, learning from early successes and failures, and adapting frameworks to local contexts. This approach also ensures that AI integration is both sustainable and inclusive, prioritizing sectors with immediate impact—such as healthcare, agriculture, and education—while building foundational capabilities, i.e., data infrastructure, talent pipelines, and regulatory frameworks.

Once implementation is initiated across the continent, accelerating progress will then depend on coordinated actions among all stakeholders in order to scale pilot projects, allocate resources efficiently, and foster a culture of innovation that aligns with the African context.

D. *Adapt Successful Models in other Regions*

Scaling AI infrastructure in Africa faces significant hurdles due to the region's fragmented and often unreliable energy landscape [38], gaps in mobile network coverage, and limited fixed broadband penetration [76]. Given the high power demands of AI development workloads, African countries should tread carefully while concurrently addressing the pressing need for household electrification (see sections II-B2 and II-B3). This delicate balance calls for innovative solutions that ensure a sustainable future for both technology growth and equitable access to energy resources, and can potentially be adapted from successful models applied in other regions of the world.

[33] https://blog.startuplist.africa/articles/ai-revolution-in-africa-2025

[34] Countries with such laws include Tunisia, Senegal, Nigeria, Côte d'Ivoire, Ethiopia, and Ghana.

[35] In November 2018, Mauritius became the first African country to publish an AI strategy.

1) The EU Approach to Regional HPC Capabilities

The European HPC Joint Undertaking (EuroHPC)[36] represents an interesting model of regional cooperation. It strategically locates its compute centers near stable and renewable energy sources across the EU, prioritizes energy-efficient hardware, and builds infrastructure through cooperative continental efforts. This partnership involves 36 European countries, for a budget of around EUR 8 billion from 2021 to 2027.
JUPITER, Europe's fastest (exascale) and most energy-efficient supercomputer, takes center stage in the EuroHPC JU network.[37] Powered by an astounding 24,000 of NVIDIA's latest Grace Harper (GH200) chips, this powerhouse serves as a testament to the region's unwavering dedication to providing cutting-edge resources and support for local innovators. This commitment reinforces Europe's position at the forefront of AI research, fostering innovation and progress in a variety of fields, contrasting the growing narrative about the continent's being mostly focusing on data/AI regulation.

SSA could emulate this approach by fostering multi-country partnerships to fund, manage, and share AI infrastructure. Putting this into the perspective of the African declaration on AI, establishing a similar joint undertaking would require about 17% of the planned AI fund.

2) The Asian Model: Robust Foundations and Government Support

Singapore, China, Taiwan

Asia's model for developing robust AI ecosystems—if there exists any—is characterized by strategic foundations and strong governmental support, exemplified by Taiwan, China, and Singapore. Each country has adopted unique approaches to nurture its domestic AI landscape.

Ranked as the best in the world in human capital development,[38] Singapore has established initiatives such as the AI Apprenticeship Programme (AIAP) through AI Singapore,[39] which provides structured, paid training that seamlessly integrates academic learning with practical AI applications. This program exemplifies how to effectively bridge the gap between education and industry demands. Furthermore, entities like SGInnovate[40] play a key role by co-investing in deeptech startups, creating a symbiotic relationship between government, academia, and the private sector. Singapore has continuously made efforts to enhance its AI ecosystem through international collaborations, e.g., the partnership between the National University of Singapore and FPT Corporation for the establishment of a state-of-the-art AI lab.[41]

China's approach to AI, as outlined in its 2017 *New Generation Artificial Intelligence Development Plan*, emphasizes a strategic, state-driven vision to become a global leader in AI by 2030. Its overall strategy is built on several foundational principles including open-source, technology-led growth, massive R&D investment, and sectoral integration. The plan prioritizes investing in core technologies like machine learning (ML), natural language processing (NLP), and robotics, while fostering collaboration between academia, industry, and government [77]. It also focuses on integrating AI into key sectors such as healthcare, transportation, defense, and manufacturing to drive economic growth and social progress. Another key aspect of the country's AI strategy is the establishment of AI research institutes supported by substantial public and private funding [77].
Foundational efforts—especially in talent pipelines—have significantly contributed to the growth of local firms like *Alibaba*, *ByteDance*, *Baidu*, to name a few. The benefits were also noticed in DeepSeek's hiring process, which the firm argued it was primarily targeting fresh graduates from top Chinese universities or PhD candidates nearing completion [78].
In the recent years, China's strategy highlights a strong focus on attracting highly skilled individuals, in a global situation where many other countries, i.e., the US or Canada are tightening policies on immigration.

Taiwan adopts a strategic focus on specialization within its AI ecosystem. Taiwan's AI Taiwan Action Plan (2018-2021) aimed to position the island as a global leader in AI by leveraging its strengths in semiconductors, fostering innovation, and integrating AI into industries. The plan emphasized actions like talent development (training researchers and producing 10,000+ technicians annually), semiconductor leadership by expanding the country's global chip industry, and building an AI innovation hub in order to attract tech giants like Microsoft and Google. By leveraging its strong semiconductor industry, Taiwan has positioned itself as a leader in AI hardware production. In particular, it is worth noting that the world increasingly relies on Taiwan Semiconductor

[36] eurohpc-ju.europa.eu/index_en
[37] www.fz-juelich.de/en/ias/jsc/jupiter
[38] 2020 World Bank Human Capital Index, see https://www.worldbank.org/en/publication/human-capital.
[39] aisingapore.org
[40] www.sginnovate.com
[41] https://news.nus.edu.sg/fpt-nus-join-forces-in-driving-ai-innovation-fostering-talent-development/

Manufacturing Company (TSMC),[42] for advanced chip production serving major tech companies like NVIDIA. However, while this is key advantage for the country, it is also creating significant dependence and posing a threat given the growing global demand for advanced AI-specific hardware and the global geopolitical situation.

Together, these approaches highlight the importance of government involvement in setting clear policies, investing in education, and promoting public-private partnerships.

3) The North American Way: Strong Basic, Applied Research and Innovation

The United States and Canada

In both the United States (US) and Canada, robust ecosystems for AI development have been cultivated through continuous government support, industry collaboration, and academia-led research.

The US has long been a leader in AI innovation, driven by a combination of private sector dynamism and public investment. Key players include tech giants (Meta, Google, Microsoft, NVIDIA, Amazon, tesla, etc.) which invest heavily in AI R&D and support ecosystem growth through their contributions to open research. On the one hand, government agencies such as the National Science Foundation (NSF) and Defense Advanced Research Projects Agency (DARPA) provide substantial grants and support for AI research, fostering cutting-edge advancements. On the other hand, the US has continuously been home to startups that have significant impact on how humanity engages with advanced technologies, hosting firms like *Hugging Face*, *Scale AI*, *OpenAI*, *Anthropic*, *Skild AI*, *Perplexity* and many others.
As reported by OECD's 2022 data on new ICT short-cycle tertiary graduates[43], the US was way ahead producing more graduates at each of the associate, bachelor's, master's, and PhD levels than any other country included in the data set [79]. For instance, Carnegie Mellon University (CMU) secured the top spot among US academic institutions in 2023.[44] Distinguishing itself from others, CMU was one of the few universities providing specialized programs in AI until relatively recently [79]. The university has also been instrumental in supporting the same dynamics in Africa by establishing a regional campus in Kigali, Rwanda.

Canada has emerged as a notable hub for AI innovation, supported by its forward-thinking policies and investments in education and research. The Canadian government launched the Pan-Canadian Artificial Intelligence Strategy, allocating significant funds to build AI capabilities across universities and research institutions. This strategy emphasizes collaboration between academia, industry, and government, creating an ecosystem that encourages innovation while prioritizing ethical considerations. Canada hosts AI research institutions like the Montreal Institute for Learning Algorithms (MILA), a leading research institute which is renowned for its pioneering work in ML particularly in deep learning. MILA is a hub for cutting-edge AI research and innovation, contributing significantly to global advancements in the field. Alongside Toronto's Vector Institute and Alberta's Machine Intelligence Institute (Amii), it is one of the three national hubs of the Pan-Canadian AI Strategy.[45] *Cohere* is another key player in Canada's AI landscape, acting as a national champion that attracts significant government and private investment. The firm has raised approximately $1.54 billion across 7 funding rounds since its founding in 2019, which makes it one of Canada's most well-funded AI startups.[46]

The US and Canada exemplify the importance of public-private partnerships and investment in AI-focused education as cornerstones for developing sustainable AI ecosystems. Their strategies provide valuable insights into balancing innovation with societal responsibility, making them models worth emulating globally.

E. Enhancing the African Model

Ultimately, Africa's success in the AI era will hinge on its ability to coordinate action across all stakeholders. The AU continental AI strategy [73] and the continent's recent declaration on AI [67] try to formalize collaboration, establish shared resources, and ensure investments in AI align with Africa's inclusive development goals.
In the current landscape, countries are naturally pursuing independent digital strategies, focusing on national priorities. This approach, although necessary for individual advancement, often results in duplication of efforts and underutilization of internal digital resources. Instead, African countries can benefit from effectively shifting toward a coordinated and truly continental strategy for AI adoption, rather than fragmented initiatives. More specifically, a pooled

[42]TSMC is a leading semiconductor manufacturer which pioneered the pure-play foundry business model, manufacturing chips for other firms rather than designing its own.
[43]https://data-explorer.oecd.org
[44]https://nces.ed.gov/ipeds/use-the-data#SearchExistingData
[45]https://ised-isde.canada.ca/site/ai-strategy/en
[46]https://www.canada.ca/en/innovation-science-economic-development/news/2025/03/government-of-canada-finalizes-investment-to-support-canadian-born-ai-leader-cohere.html

multi-country approach can enable African governments and regional blocs (the African Union or regional economic communities) to negotiate more favorable terms with major cloud and AI hardware providers. For instance, joint procurement frameworks, e.g., *EuroHPC*, shared data center investments, and federated cloud platforms can drastically reduce costs and improve general access to modern compute capabilities. Crucially, building trust and interoperability among African countries will be key to enabling long-term, sustainable ecosystem interactions that serve collective interests.

IV. Limitations

This paper broadly analyzes the infrastructure, and policy barriers contributing to the perceived digital divide in African AI development. Nevertheless, several limitations have to be acknowledged.

- We recognize that our reliance on public data may not reflect the full scope, as it does not directly include proprietary infrastructure or unpublished initiatives from the private sector. In fact, many of the deployed AI-ready compute resources either by African governments or local private firms remain under-documented or non verifiable.
- Planned installations or those in construction during the drafting of this paper were not included in the initial version of the Africa compute tracking tool.
- Our analysis is solely based on the NVIDIA ecosystem as it is widely considered the standard for AI research and development globally, primarily due to its dominant GPU hardware and proprietary CUDA[47] software platform [80].
- This work focuses predominantly on compute infrastructure and model scalability as enablers of AI development, and less on ethical, linguistic, and socio-cultural dimensions, which are equally crucial for creating inclusive AI ecosystems in Africa.

V. Conclusion

To close the threatening AI divide, Africa must address infrastructure, accessibility, and affordability in tandem. While strategic and policy frameworks are essential, they alone cannot bridge the gap without tangible, coordinated actions. The *Africa AI Compute Tracker (ACT)* is our contribution to providing data-driven tools as a foundation for targeted investments in compute resources, energy systems, and domestic cloud platforms. The success of these efforts hinges on overcoming systemic barriers, such as fragmented energy grids, unreliable mobile networks, and restrictive data policies, which have historically limited the scalability of nascent AI initiatives. By actually fostering regional collaboration, Africa can align its AI ambitions with global standards while prioritizing local needs. The urgency of this task cannot be overstated: without immediate, action-oriented strategies to accelerate the of digital foundations, Africa risks being left behind in the global AI adoption movement.

The proposed AI compute tracker is not just a monitoring tool but a call to action—a step toward ensuring that Africa's AI future is both equitable and sustainable.

[47] The Compute Unified Device Architecture (CUDA) is a parallel computing platform developed by NVIDIA for its GPUs.

References

[1] A. B. Rashid and M. A. K. Kausik, "AI revolutionizing industries worldwide: A comprehensive overview of its diverse applications," *Hybrid Advances*, vol. 7, p. 100277, 2024. [Online]. Available: https://doi.org/10.1016/j.hybadv.2024.100277

[2] A. Tsado and C. Lee. (2024, Nov) Only Five Percent of Africa's AI Talent Has the Compute Power It needs. UNDP Digital, Zindi, Alliance4ai. [Online]. Available: https://www.undp.org/digital/blog/only-five-percent-africas-ai-talent-has-compute-power-it-needs

[3] S. Fonn, P. L. Ayiro, P. Cotton, A. Habib, P. M. F. Mbithi, A. M. Mtenje, B. Nawangwe, E. O. Ogunbodede, I. Olayinka, F. Golooba-Mutebi, and A. Ezeh, "Repositioning africa in global knowledge production," *The Lancet*, vol. 392, pp. 1163–1166, 2018.

[4] J. Thondhlana and E. C. Garwe, "Repositioning of africa in knowledge production: shaking off historical stigmas," *Journal of the British Academy*, vol. 9, no. s1, pp. 1–17, 2021.

[5] J. Kaplan, S. McCandlish, T. Henighan, T. B. Brown, B. Chess, R. Child, S. Gray, A. Radford, J. Wu, and D. Amodei, "Scaling Laws for Neural Language Models," *arXiv preprint arXiv:2001.08361*, 2020. [Online]. Available: https://arxiv.org/abs/2001.08361

[6] A. Radford, J. W. Kim, T. Xu, G. Brockman, C. Mcleavey, and I. Sutskever, "Robust Speech Recognition via Large-Scale Weak Supervision," in *Proceedings of the 40th International Conference on Machine Learning*, ser. Proceedings of Machine Learning Research, A. Krause, E. Brunskill, K. Cho, B. Engelhardt, S. Sabato, and J. Scarlett, Eds., vol. 202. PMLR, 23–29 Jul 2023, pp. 28 492–28 518. [Online]. Available: https://proceedings.mlr.press/v202/radford23a.html

[7] N. Benaich, A. Chalmers, and Air Street Capital,

"State of AI Report 2024," 2024. [Online]. Available: https://www.stateof.ai/

[8] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 779–788.

[9] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. u. Kaiser, and I. Polosukhin, "Attention is all you need," in *Advances in Neural Information Processing Systems*, I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, Eds., vol. 30. Curran Associates, Inc., 2017. [Online]. Available: https://proceedings.neurips.cc/paper_files/paper/2017/file/3f5ee243547dee91fbd053c1c4a845aa-Paper.pdf

[10] J. Jumper, R. Evans, A. Pritzel, R. Tibshirani, T. Green, M. Figurnov, N. Rowlands, J. Brown, K. Ledingham, V. Chen, and other authors, "Highly accurate protein structure prediction with Alphafold," *Nature*, vol. 596, pp. 583–589, 2021.

[11] A. Mishra, J. Cha, H. Park, and S. Kim, *Artificial Intelligence and Hardware Accelerators*. Springer Nature Switzerland AG, 2023. [Online]. Available: https://doi.org/10.1007/978-3-031-22170-5

[12] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," in *Advances in Neural Information Processing Systems*, F. Pereira, C. Burges, L. Bottou, and K. Weinberger, Eds., vol. 25. Curran Associates, Inc., 2012. [Online]. Available: https://proceedings.neurips.cc/paper_files/paper/2012/file/c399862d3b9d6b76c8436e924a68c45b-Paper.pdf

[13] D. Ciregan, U. Meier, and J. Schmidhuber, "Multi-column deep neural networks for image classification," in *2012 IEEE Conference on Computer Vision and Pattern Recognition*, 2012, pp. 3642–3649.

[14] R. Raina, A. Madhavan, and A. Y. Ng, "Large-scale deep unsupervised learning using graphics processors," in *Proceedings of the 26th Annual International Conference on Machine Learning*, ser. ICML '09. New York, NY, USA: Association for Computing Machinery, 2009, p. 873–880. [Online]. Available: https://doi.org/10.1145/1553374.1553486

[15] NVIDIA Corporation and P. Glaskowsky, "NVIDIA's Fermi: The First Complete GPU Computing Architecture," 2009. [Online]. Available: https://www.nvidia.com/content/pdf/fermi_white_papers/p.glaskowsky_nvidia's_fermi-the_first_complete_gpu_architecture.pdf

[16] N. Corporation, "NVIDIA Tesla Kepler GPU Computing Accelerators," 2012. [Online]. Available: https://www.nvidia.com/content/tesla/pdf/nv-ds-teslak-family-jul2012-lr.pdf

[17] ——, "GeForce GTX 980: Featuring Maxwell, The Most Advanced GPU Ever Made," 2014.

[18] ——, "NVIDIA Tesla P100: The Most Advanced Datacenter Accelerator Ever Built ," 2016. [Online]. Available: https://images.nvidia.com/content/pdf/tesla/whitepaper/pascal-architecture-whitepaper.pdf

[19] ——, "NVIDIA TESLA V100 GPU ARCHITECTURE: The World's Most Advanced Data Center GPU," 2017. [Online]. Available: https://images.nvidia.com/content/volta-architecture/pdf/volta-architecture-whitepaper.pdf

[20] ——, "NVIDIA Turing GPU Architecture: Graphics Reinvented," 2018. [Online]. Available: https://images.nvidia.com/aem-dam/en-zz/Solutions/design-visualization/technologies/turing-architecture/NVIDIA-Turing-Architecture-Whitepaper.pdf

[21] ——, "NVIDIA A100 Tensor Core GPU Architecture: Unprecedented Acceleration at Every Scale," 2020. [Online]. Available: https://images.nvidia.com/aem-dam/en-zz/Solutions/data-center/nvidia-ampere-architecture-whitepaper.pdf

[22] ——, "NVIDIA H100 Tensor Core GPU Architecture: Exceptional Performance, Scalability, and Security for the Data Center," 2023. [Online]. Available: https://resources.nvidia.com/en-us-hopper-architecture/nvidia-h100-tensor-c

[23] ——, "NVIDIA Blackwell Architecture Technical Brief: Built for the Age of AI Reasoning," 2025. [Online]. Available: https://images.nvidia.com/aem-dam/Solutions/geforce/blackwell/nvidia-rtx-blackwell-gpu-architecture.pdf

[24] J. P. Research. (2025, sep) Q2'25 PC graphics add-in board shipments increased 27.0% from last quarter. [Online]. Available: https://www.jonpeddie.com/news/q225-pc-graphics-add-in-board-shipments-increased-27-0-from-last-quarter/

[25] V. Lehdonvirta, B. Wú, and Z. Hawkins, "Compute North vs. Compute South: The Uneven Possibilities of Compute-based AI Governance Around the Globe," *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society*, vol. 7, no. 1, pp. 828–838, Oct. 2024. [Online]. Available: https://doi.org/10.1609/aies.v7i1.31683

[26] Tony Blair Institute for Global Change, "State of compute access: How to bridge the new digital divide," 2023. [Online]. Available: https://institute.global/insights/tech-and-digitalisation/state-of-compute-access-how-to-bridge-the-new-digital-divide

[27] Africa Green Compute Coalition, "Unlocking Compute in Africa," 2025. [Online]. Available: https://www.aihubfordevelopment.org/green-compute-coalition

[28] GSMA, "The state of the industry report on mobile money 2025," 2025. [Online]. Available:

https://www.gsma.com/solutions-and-impact/connectivity-for-good/mobile-for-development/gsma_resources/the-state-of-the-industry-report-on-mobile-money-2025/

[29] L. Klapper, D. Singer, L. Starita, and A. Norris, *The Global Findex Database 2025: Connectivity and Financial Inclusion in the Digital Economy*. Washington, DC: World Bank, 2025. [Online]. Available: https://doi.org/10.1596/978-1-4648-2204-9

[30] M. Cruz, Ed., *Digital Opportunities in African Businesses*. Washington, DC: World Bank, 2024.

[31] International Labour Organization (ILO), *World Employment and Social Outlook: Trends*. Geneva: International Labour Office, 2023.

[32] H. O. Eromosele, "Currency Devaluation and Performance of Small and Medium Enterprises (SMEs) in Nigeria," *AKSU Journal of Management Sciences (AKSUJOMAS)*, august 2025.

[33] B. Wadinga and A. Ahmed, "Relationship Between Currency Devaluation and Economic Growth in Nigeria: An Empirical Evidence," *International Journal of Economics and Financial Management (IJEFM)*, pp. 117–132, 07 2024.

[34] E. Humeau, "Bridging the New Digital Divide: How Africa Can Boost Its Compute Capacity for AI," 2025. [Online]. Available: https://www.gsma.com/solutions-and-impact/connectivity-for-good/mobile-for-development/blog/bridging-the-new-digital-divide-how-africa-can-boost-its-compute-capacity-for-ai/

[35] O. V. Babasanmi and J. Chavula, "Measuring Cloud Latency in Africa," in *Proceedings of the IEEE 11th International Conference on Cloud Networking (CloudNet)*, Paris, France, 2022, pp. 61–66. [Online]. Available: https://pubs.cs.uct.ac.za/id/eprint/1704/1/Measuring_Cloud_Latency_in_Africa.pdf

[36] J. Chavula, A. Phokeer, A. Formoso, and N. Feamster, "Insight into Africa's Country-level Latencies," in *IEEE Africon 2017 Proceedings*. IEEE, 2017, pp. 938–944. [Online]. Available: https://afrinic.net/ast/pdf/research/insight-latency-africa.pdf

[37] M. Dayarathna, Y. Wen, and R. Fan, "Data Center Energy Consumption Modeling: A Survey," *IEEE Communications Surveys & Tutorials*, pp. 732–794, 2016.

[38] "Energy and AI," 2025. [Online]. Available: https://www.iea.org/reports/energy-and-ai

[39] H. E. Lee, W. Y. Kim, H. Kang, and K. Han, "AD514: Still lacking reliable electricity from the grid, many Africans turn to other sources," *Afrobarometer Dispatch No. 514*, 2022. [Online]. Available: https://www.afrobarometer.org/publication/ad514-still-lacking-reliable-electricity-from-the-grid-many-africans-turn-to-other-sources/

[40] D. Msafiri and R. Adjadeh, "AD793: Energy gaps: Slight, uneven progress still leaves many Africans without electricity," *Afrobarometer Dispatch No. 793*, 2024. [Online]. Available: https://www.afrobarometer.org/publication/ad793-energy-gaps-slight-uneven-progress-still-leaves-many-africans-without-electricity/

[41] "Regulatory Indicators for Sustainable Energy (RISE)," 2025.

[42] B. Johnston, L. Timm, D. Macleod, and J. Poole, "Ten Years of the HPC Ecosystems Project - Transforming HPC in Africa for the past decade," in *Practice and Experience in Advanced Research Computing 2024: Human Powered Computing*, ser. PEARC '24. New York, NY, USA: Association for Computing Machinery, 2024. [Online]. Available: https://doi.org/10.1145/3626203.3670537

[43] Galsenai, "xTTS-v2-wolof," *HuggingFace Model Hub*, 2024. [Online]. Available: https://huggingface.co/galsenai/xTTS-v2-wolof

[44] T. Dao, D. Y. Fu, S. Ermon, A. Rudra, and C. Ré, "Flashattention: fast and memory-efficient exact attention with io-awareness," in *Proceedings of the 36th International Conference on Neural Information Processing Systems*, ser. NIPS '22. Red Hook, NY, USA: Curran Associates Inc., 2022.

[45] T. Dao, "FlashAttention-2: Faster attention with better parallelism and work partitioning," in *International Conference on Learning Representations (ICLR)*, 2024.

[46] J. Shah, G. Bikshandi, Y. Zhang, V. Thakkar, P. Ramani, and T. Dao, "FlashAttention-3: fast and accurate attention with asynchrony and low-precision," in *Proceedings of the 38th International Conference on Neural Information Processing Systems*, ser. NIPS '24. Red Hook, NY, USA: Curran Associates Inc., 2025.

[47] DeepSeek-AI *et al.*, "DeepSeek-V2: A Strong, Economical, and Efficient Mixture-of-Experts Language Model," 2024. [Online]. Available: https://arxiv.org/abs/2405.04434

[48] Microsoft. (2025, Jun) Comparing GPU types in Azure Container Apps. [Online]. Available: https://learn.microsoft.com/en-us/azure/container-apps/gpu-types

[49] Cassava Technologies, "NVIDIA Partners with Cassava Technologies to Build Africa's Largest AI Infrastructure," April 2024. [Online]. Available: https://www.cassavatechnologies.com/cassava-to-upgrade-its-data-centres-with-nvidia-supercomputers-to-drive-africas-ai-future/

[50] Microsoft, "Microsoft and g42 partner to accelerate ai innovation in uae and beyond," https://blogs.microsoft.com/blog/2024/04/15/microsoft-and-g42-partner-to-accelerate-ai-innovation-in-uae-and-beyond/, April 2024, accessed May 2025. [Online].

Available: https://blogs.microsoft.com/blog/2024/04/15/microsoft-and-g42-partner-to-accelerate-ai-innovation-in-uae-and-beyond/
[51] United Nations Secretary-General's High-level Advisory Body on Artificial Intelligence, "Governing AI for Humanity: Final Report," United Nations, Tech. Rep., September 2024. [Online]. Available: https://www.un.org/sites/un2.un.org/files/governing_ai_for_humanity_final_report_en.pdf
[52] R. Adams. (2024, May) What is AI Governance? An African Response. [Online]. Available: https://www.globalcenter.ai/research/what-is-ai-governance-an-african-response
[53] Africa Data Protection Association (ADP), "Governance of Artificial Intelligence in Africa," 2025. [Online]. Available: https://www.africadataprotection.org/ia/
[54] S. Abdella and A. Alayande. (2025, April) African Countries Are Racing to Create AI Strategies - But Are They Putting the Cart Before the Horse? [Online]. Available: https://www.globalcenter.ai/research/african-countries-are-racing-to-create-ai-strategies-but-are-they-putting-the-cart-before-the-horse
[55] Centre for Intellectual Property and Information Technology Law (CIPIT), "The State of AI in Africa Report 2023," 2023. [Online]. Available: https://cipit.org/wp-content/uploads/2023/12/Final-Report-The-State-of-AI-in-Africa-Report-2023-3.pdf
[56] Eke, Damian Okaibedi and Wakunuma, Kutoma and Akintoye, Simisola, Ed., *Responsible AI in Africa: Challenges and Opportunities*. Cham: Palgrave Macmillan, 2023.
[57] A. Alayande, S. Segun, and L. Junck, "Emerging technology policies and democracy in Africa," 2025. [Online]. Available: https://www.atlanticcouncil.org/in-depth-research-reports/report/emerging-technology-policies-and-democracy-in-africa-south-africa-kenya-nigeria-ghana-and-zambia-in-focus/
[58] African Union Commission, Infrastructure and Energy Department, "AU Data Policy Framework," Online PDF, Jul. 2022, endorsed by the AU Executive Council in February 2022, published July 28, 2022. [Online]. Available: https://au.int/en/documents/20220728/au-data-policy-framework
[59] F. Adeleke. (2024, May) What Does it Look Like to Develop AI Policies in Africa? [Online]. Available: https://www.globalcenter.ai/research/what-does-it-look-like-to-develop-ai-policies-in-africa
[60] M. Musoni, P. Karkare, and C. Teevan, "Cross-Border Data Flows in Africa: Continental Ambitions and Political Realities," ECDPM, Tech. Rep. Discussion Paper No. 379, 2024. [Online]. Available: https://ecdpm.org/work/cross-border-data-flows-africa-continental-ambitions-and-political-realities
[61] Qhala and QUBIT HUB, "AI Talent Readiness Index for Africa," 2025. [Online]. Available: https://talentindex.ai/overall-performance
[62] Deep Learning Indaba. (2024, April) Yebetumi: Annual Report of the Deep Learning Indaba 2023. [Online]. Available: https://deeplearningindaba.com/wp-content/uploads/2024/05/Indaba-2023-Report-web.pdf
[63] ——. (2025, March) Xam Xamlé: Annual Report of the Deep Learning Indaba 2024. [Online]. Available: https://deeplearningindaba.com/wp-content/uploads/2025/04/Impact-Report-Deep-Learning-Indaba-2024-03.pdf
[64] D. Vernon, "Robotics in Africa is trending upward and has a bright future," *Science Robotics*, vol. 10, no. 101, p. eadx2410, 2025. [Online]. Available: https://www.science.org/doi/abs/10.1126/scirobotics.adx2410
[65] A. Grattafiori *et al.*, "The Llama 3 Herd of Models," 2024. [Online]. Available: https://arxiv.org/abs/2407.21783
[66] DeepSeek-AI *et al.*, "DeepSeek-V3 Technical Report," 2025. [Online]. Available: https://arxiv.org/abs/2412.19437
[67] "Africa declaration on artificial intelligence," Centre for the Fourth Industrial Revolution Rwanda and Ministry of ICT & Innovation, 2025. [Online]. Available: https://c4ir.rw/docs/Africa-Declaration-on-Artificial-Intelligence.pdf
[68] W. Chen, J. Tian, Y. Peng, B. Yan, C.-H. H. Yang, and S. Watanabe, "OWLS: Scaling laws for multilingual speech recognition and translation models," in *Proceedings of the 42nd International Conference on Machine Learning*, ser. Proceedings of Machine Learning Research, A. Singh, M. Fazel, D. Hsu, S. Lacoste-Julien, F. Berkenkamp, T. Maharaj, K. Wagstaff, and J. Zhu, Eds., vol. 267. PMLR, 13–19 Jul 2025, pp. 9121–9145. [Online]. Available: https://proceedings.mlr.press/v267/chen25bj.html
[69] CIPIT and partners, "The State of AI in Africa Report," 2023. [Online]. Available: https://aiconference.cipit.org/documents/the-state-of-ai-in-africa-report.pdf
[70] GSMA Mobile for Development, "AI for Africa: Enabling a Responsible and Inclusive Future," 2024. [Online]. Available: https://www.gsma.com/solutions-and-impact/connectivity-for-good/mobile-for-development/wp-content/uploads/2024/07/AI_for_Africa.pdf
[71] Z. Ratsoga and M. Primus, "Factors influencing post-hackathon project continuation in an African corporate setting," *The African Journal of Information and Communication (AJIC)*, vol. 31, 06 2023.
[72] B. A. Soglo, K. Hach, S. Beckmann, B. Bridges, and AfriLabs, "Evaluating the African Deep-Tech

Startup Ecosystem," Intel Corporation, Santa Clara, CA, Tech. Rep., 2022, includes analysis of 200+ startups and 400+ hubs across Africa, focusing on AI, ML, and other deep-tech sectors.

[73] African Union, "AU continental artificial intelligence strategy," 2024, official policy document. [Online]. Available: https://au.int/en/documents/20240809/continental-artificial-intelligence-strategy

[74] Smart Africa, Deutsche Gesellschaft für Internationale Zusammenarbeit (GIZ) and GFA Consulting, "Artificial Intelligence for Africa," 2021, in collaboration with the Republic of South Africa. [Online]. Available: https://smartafrica.org/knowledge/artificial-intelligence-for-africa/

[75] The United Nations, "Global Digital Compact," 2024. [Online]. Available: https://www.un.org/global-digital-compact/en

[76] International Telecommunication Union, "State of digital development and trends in the Africa region: Challenges and opportunities," International Telecommunication Union, Tech. Rep., 2025. [Online]. Available: https://www.itu.int/hub/publication/d-ind-sddt_afr-2025/

[77] G. Webster, R. Creemers, E. Kania, and P. Triolo, "Full Translation: China's New Generation Artificial Intelligence Development Plan," 2017. [Online]. Available: https://digichina.stanford.edu/work/full-translation-chinas-new-generation-artificial-intelligence-development-plan-2017/

[78] J. Schneider, A. Shen, I. Zhang *et al.* (2024, Nov) Deepseek: The Quiet Giant Leading China's AI Race. [Online]. Available: https://www.chinatalk.media/p/deepseek-ceo-interview-with-chinas

[79] N. Maslej, L. Fattorini, R. Perrault, Y. Gil, V. Parli, N. Kariuki, E. Capstick, A. Reuel, E. Brynjolfsson, J. Etchemendy, K. Ligett, T. Lyons, J. Manyika, J. C. Niebles, Y. Shoham, R. Wald, T. Walsh, A. Hamrah, L. Santarlasci, J. B. Lotufo, A. Rome, A. Shi, and S. Oak, "The AI Index 2025 Annual Report," AI Index Steering Committee, Institute for Human-Centered AI, Stanford University, Stanford, CA, Tech. Rep., April 2025. [Online]. Available: https://hai.stanford.edu/ai-index/2025-ai-index-report

[80] Bicaj, Bleona. (2025, August) Why NVIDIA dominates despite low developer program scores. [Online]. Available: https://www.slashdata.co/post/why-nvidia-dominates-despite-low-developer-program-scores

VI. Appendix

A. Definitions adopted

Artificial intelligence (AI) mainly refers to machine learning (ML) or deep learning-based techniques used to enhance data-driven processes and automate high-value tasks.

A ***graphics processing unit (GPU)*** is a specialized computer processor designed for parallel computation, and commonly used for rendering graphics and accelerating ML computations.

AI Compute refers specifically to computing capabilities or the computing power required to run advanced AI algorithms, usually in a distributed manner.

Node is the term used for a single computer or server in a network that can perform tasks independently but can also communicate with other units for coordinated functioning in a distributed system.

Cluster denotes a group of connected and interoperable nodes working together to achieve common goals such as enhancing computational power, and facilitating large-scale processing.

AI Talent is a group of skilled individuals with demonstrated—not potential—expertise in AI-related topics and the ability to design, develop, and implement AI solutions; highly sought after for their innovative problem-solving abilities in diverse domains.

AI model training refers to the process of teaching machine learning models from (large) datasets to automatically learn to solve tasks of interest.

B. A Snapshot of Supercomputers in the TOP500 List

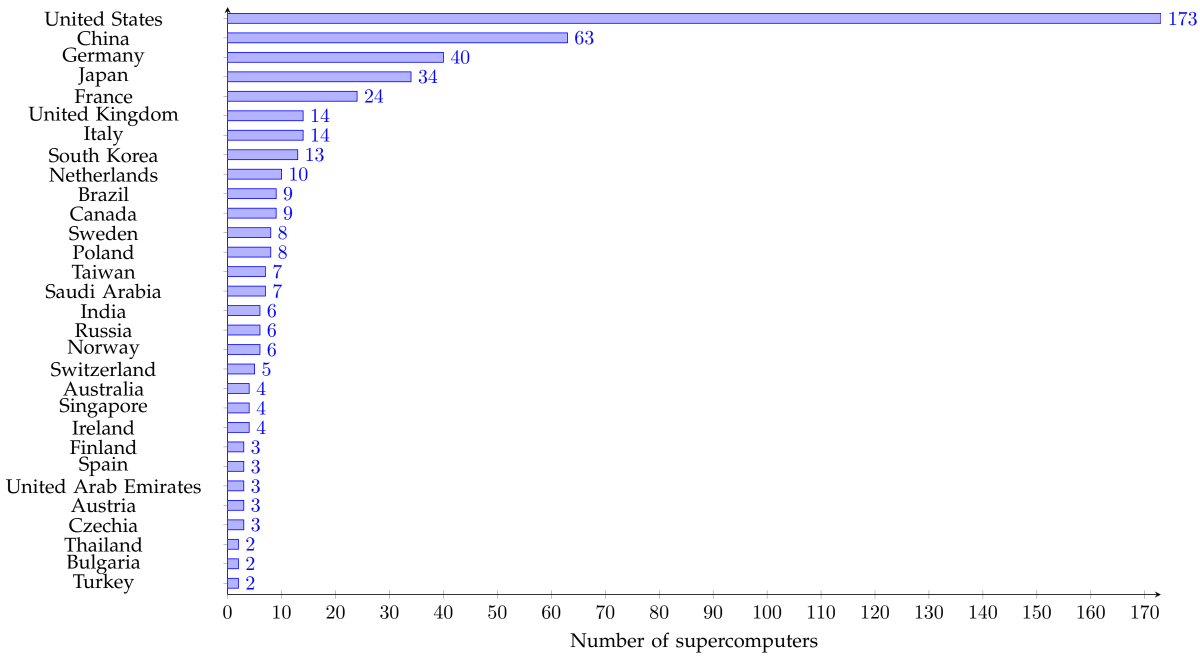
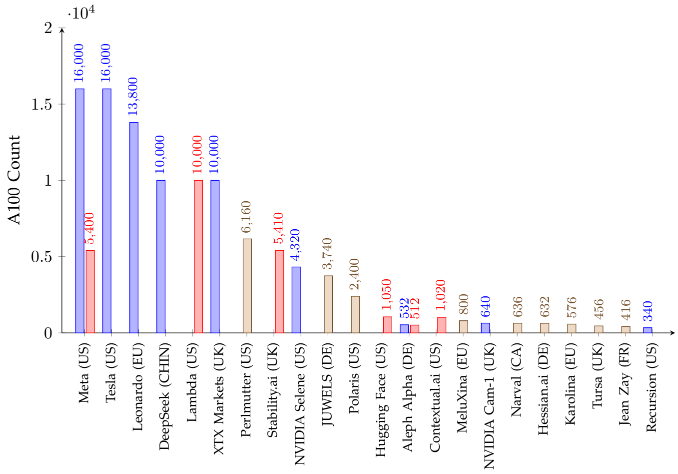


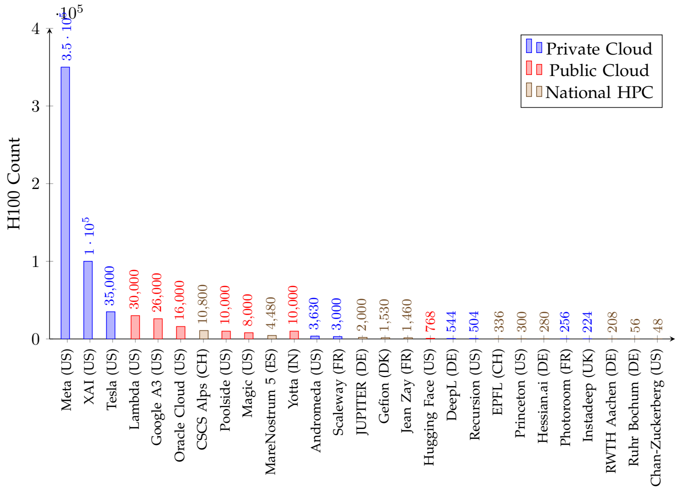
Fig. 5: Distribution of supercomputers in the TOP500 list by country (as per November 2024 data)[48].

[48]https://top500.org/lists/top500/list/2024/11/

C. *Details about AI Compute Distribution*



(a) Global distribution of A100 GPUs



(b) Global distribution of H100 GPUs

Fig. 6: NVIDIA A100 and H100 GPUs across public, private, and national HPC infrastructure (redrawn from the state of AI Report - compute index 2024).