WHITEPAPER
October 2024

VOICE AND ACCESS IN AI: GLOBAL AI MAJORITY PARTICIPATION IN ARTIFICIAL INTELLIGENCE DEVELOPMENT AND GOVERNANCE

Authors: Sumaya N. Adan, Robert Trager, Kayla Blomquist, Claire Dennis, Gemma Edom, Lucia Velasco, Cecil Abungu, Ben Garfinkel, Julian Jacobs, Chinasa T. Okolo, Boxi Wu, Jai Vipra

Voice and Access in AI: Global AI Majority Participation in Artificial Intelligence Development and Governance

Sumaya N. Adan*[+6,7], Robert Trager*[+1,3,7], Kayla Blomquist[+7,9,10], Claire Dennis[+3,7], Gemma Edom[+1], Lucia Velasco[+7], Cecil Abungu[8], Ben Garfinkel[3,7], Julian Jacobs[5,7], Chinasa T. Okolo[2], Boxi Wu[9], Jai Vipra[4]

[1] Blavatnik School of Government, University of Oxford [2] Brookings Institution, [3] Centre for the Governance of AI, [4] Cornell University, [5] Department of Politics and International Relations, [6] Institute for AI Policy and Strategy, [7] Oxford Martin School AI Governance Initiative, [8] University of Cambridge, [9] Oxford Internet Institute, [10] Oxford China Policy Lab

*Corresponding authors: sumayan.adan@gmail.com & robert.trager@bsg.ox.ac.uk

[+]Denotes primary authors who contributed most significantly to the direction and content of the paper.

We are grateful to Lennart Heim, Sihao Huang, Benjamin Prudhomme, and Jaime Sevilla for their valuable feedback on this project.

Authorship of this paper indicates contribution but does not imply full agreement with every claim.

Please cite as: Adan, S. N., and Trager, R. F., et al. (2024) "Voice and Access in AI: Global AI Majority Participation in Artificial Intelligence Development and Governance, Oxford Martin AI Governance Initiative," *Oxford Martin School AI Governance Initiative White Paper*.

Abstract

Artificial intelligence (AI) is rapidly emerging as one of the most transformative technologies in human history, with the potential to profoundly impact all aspects of society globally. However, access to AI and participation in its development and governance is concentrated among a few countries with advanced AI capabilities, while the 'Global AI Majority' – defined as the population of countries primarily encompassing Africa, Latin America, South and Southeast Asia, and parts of Eastern Europe – is largely excluded. These regions, while diverse, share common challenges in accessing and influencing advanced AI technologies. This white paper investigates practical remedies to increase voice in and access to AI governance and capabilities for the Global AI Majority, while addressing the security and commercial concerns of frontier AI states. We examine key barriers facing the Global AI Majority, including limited access to digital and compute infrastructure, power concentration in AI development, Anglocentric data sources, and skewed talent distributions. The paper explores the dual-use dilemma of AI technologies and how it motivates frontier AI states to implement restrictive policies. We evaluate a spectrum of AI development initiatives, from domestic model creation at one end to structured access to deployed models at the other, as well as their feasibility for the Global AI Majority. To resolve governance dilemmas, we propose three key approaches: interest alignment, participatory architecture, and safety assurance. The paper recommends near-term steps including expanding internet connectivity, developing national data infrastructures, investing in AI education programs, and strengthening regional bodies to amplify the voice of Global AI Majority countries. It also emphasizes the need for safety assurance mechanisms that balance security concerns with increased access to AI technologies. By thoughtfully expanding participation while addressing potential security risks, we argue that a more equitable global AI ecosystem can be achieved – one that ensures that AI's immense potential benefits humanity as a whole, while the technology's risks are collectively managed through truly global cooperation.

Executive Summary

This white paper investigates practical solutions to increase voice in and access to AI governance and capabilities for the 'Global AI Majority' – populations primarily in Africa, Latin America, South and Southeast Asia, and parts of Eastern Europe. Our goal is to propose actionable recommendations while addressing the security and commercial concerns of frontier AI states.

Current AI Landscape
We begin by examining the current AI landscape, highlighting stark disparities in digital infrastructure, talent distribution, and computational power between frontier AI states and the Global AI Majority. These disparities significantly limit access to and development of AI technologies for Global AI Majority countries.

Need for Inclusive Voice in AI Governance
We argue for the necessity of global inclusion in AI governance, given AI's potential to impact all of humanity. However, current governance efforts are often exclusionary and not representative of global interests.

Challenges to Change
We identify two main obstacles to inclusive AI governance and development:

1. Commercial and security interests of frontier AI states.
2. The dual-use nature of AI technologies and accompanying risks associated with their proliferation.

These factors lead to restrictive policies that consolidate AI power among a few actors.

AI Development Initiatives and Access
To address these challenges, we examine a range of development initiatives aimed at enhancing access to advanced AI among the Global AI Majority. We categorize these initiatives into two main areas: domestic AI development and access to deployed AI models.

Our analysis (figure 1) reveals significant hurdles in domestic AI development, particularly in assembling the necessary 'AI triad' of compute, data, and skilled labour.

Most Global AI Majority countries struggle to independently build the computational infrastructure required for training cutting-edge models.

We identify more promising avenues, including leveraging local knowledge and data to train models on compute hosted elsewhere and fine-tuning existing models for local contexts. These approaches offer potential pathways for Global AI Majority countries to develop AI capabilities while mitigating resource constraints.

Input	Opportunities	Bottlenecks	Level of Challenge
Data	- Develop local datasets for underrepresented languages. - Build community-driven data-collection and management initiatives.	- Limited availability of local data. - Privacy concerns and data misuse. - Difficulties in responsible data governance.	High
Skilled Labour	- Boost AI education and training. - Strengthen international collaborations for skill-building. - Support grassroots AI communities.	- Shortage of specialized expertise. - Brain drain due to better opportunities abroad. - Restrictions on knowledge transfer between geopolitical rivals.	Moderate
Compute (Local training and inference)	- Build local compute clusters. - Foster tech sovereignty and local innovation. - Improve infrastructure for future AI development.	- Extremely high costs and maintenance needs. - High energy consumption. - Likely resistance from frontier AI states due to strategic concerns.	High
Compute (Using Compute Hosted Abroad)	- Leverage external cloud computing resources. - Access advanced compute capabilities through global partnerships. - Cost-effective alternative to local clusters.	- Still costly, though less than local options. - Dependence on foreign infrastructure reduces autonomy. - Possible restrictions from frontier AI states on compute access.	Moderate

Figure 3: Opportunities, Bottlenecks, and Challenges for AI Development in the Global AI Majority

We also investigate access to deployed models as an alternative strategy. While some near state-of-the-art models are open-sourced, the most advanced typically

remain proprietary, accessible only through fee-based services. We highlight an emerging paradigm of 'structured access,' which aims to broaden access to advanced AI models in a controlled manner, often through application programming interfaces (APIs).

Our findings indicate that as AI capabilities and associated risks grow, the connection between effective governance and access to AI technologies is likely to strengthen. We draw parallels with historical governance of dual-use technologies like nuclear power, suggesting that expanding Global AI Majority participation in AI development while addressing safety concerns will require innovative governance frameworks.

Approaches to Resolving Governance Dilemmas

To address these challenges, we propose three key approaches to balance inclusivity with the participation of frontier AI states:

1. **Identify Interest Alignments:** Broad stakeholder governance functions best where actors' perceived interests are relatively aligned. We note the possibility of identifying such areas of alignment, such as in responsible digitization initiatives, and the need to ensure that broadly governed institutions focus only in these areas in order to foster participation of both frontier AI states and Global AI Majority countries.
2. **Participatory Governance Architectures:** We emphasize the importance of designing institutional governance structures that ensure meaningful participation from all states and communities. In some cases, this implies using the structure of governing bodies to protect the interests of actors who would not otherwise support governing institutions. As in other domains, governing institutions must balance minority protections and majority voice.
3. **Safety Assurance:** We underscore the connection between safety assurance and Global AI Majority voice and access. Frontier AI states will be more willing to promote broad voice and access when an international safety regime, including international safety standards and incentives for adoption at the international level, is in place.

Recommendations

To achieve a more inclusive and equitable global AI governance framework, we propose the following key recommendations:

a. *For Global AI Majority countries, we recommend the following actions:*

1. **Expand foundational digital and energy infrastructure:** Prioritize investments in expanding reliable internet connectivity and sustainable energy sources, particularly in underserved areas, to support AI development and broader

digital transformation. This strategic focus on foundational infrastructure will enable more equitable participation in the global AI ecosystem while fostering domestic innovation and economic growth.

2. **Prioritize national data infrastructure and curation:** Invest in developing robust, locally relevant datasets and establish centres for data curation, with a focus on underrepresented languages and cultural contexts. Implement strong data-governance policies to protect privacy while enabling responsible data-collection, curation, and management initiatives, positioning the country as a valuable contributor to the global AI data ecosystem.
3. **Leverage external compute resources strategically:** Rather than heavily investing in local compute infrastructure for AI model development, focus on diversifying compute procurement and training locations through international partnerships. This approach should prioritize adaptability and geopolitical resilience over extensive domestic infrastructure investments, ensuring access to state-of-the-art AI development capabilities while safeguarding national technological autonomy.
4. **Build local AI capacity:** We encourage investing in AI education and talent-development programs, alongside supporting grassroots AI communities and initiatives to nurture local expertise.
5. **Strengthen regional cooperation:** Countries should develop coordinated AI strategies through regional bodies like the African Union and ASEAN, and work towards presenting unified positions in global AI governance discussions in order to amplify their collective voice.
6. **Advocate for inclusive global AI-safety governance:** Countries should advocate for global AI-safety assurance provisions, while ensuring that safety regimes are inclusively governed and promote Global AI Majority development.

b. For all states, we recommend the following actions:

1. **Implement inclusive governance structures**: Design participatory frameworks that ensure meaningful involvement from all states, balancing minority and majority interests.
2. **Focus on responsible digitization:** Recognize that the lack of data in many languages and communities is the most significant bottleneck to broad participation in advanced AI. Cooperate to collect needed data responsibly and in line with local needs.
3. **Facilitate responsible and adaptive AI access:** Develop structured access protocols for advanced AI models through secure APIs, incorporating confidential computing safeguards and fine-tuning capabilities. Implement privacy-preserving oversight mechanisms, such as indicators of potential misuse, to ensure responsible use without compromising data confidentiality.

This balanced approach expands global AI access while enabling Global AI Majority countries to safely leverage and adapt cutting-edge AI technologies to their specific contexts and needs.

4. **Foster international collaboration:** Support knowledge-sharing initiatives to accelerate AI development in Global AI Majority countries, and actively engage in collaborative efforts to address global AI-safety concerns.

Introduction

Artificial intelligence (AI) is rapidly emerging as one of the most transformative technologies in human history, with the potential to profoundly impact all aspects of society and human life globally. However, access to AI, and participation in its global development and governance, is concentrated among just a few countries with advanced AI resources and capabilities, the frontier AI states, while non-frontier AI states – which, alongside underprivileged populations worldwide, we term the Global AI Majority – are largely excluded. This white paper investigates practical remedies to increase voice and access to AI governance and capabilities for the Global AI Majority, while accounting for the concerns of diverse stakeholders who must cooperate to create inclusive global governance of advanced AI.

We examine the challenges and opportunities associated with expanding both access to AI technologies and the voice of the Global AI Majority in AI governance. The global digital divide remains significant, with notable disparities in internet penetration across different regions. While 67% of the global population uses the internet, there is significant disparity: internet usage rates in Europe, the Americas, and the Commonwealth of Independent States are about 90%,[1] while only 37% of people in Africa are online. This stark divide not only reflects uneven access to basic digital technologies but also has profound implications for who can participate in and benefit from advanced AI systems. Beyond basic connectivity, we explore the implications of extreme centralization in advanced AI development capabilities and governance processes. For instance, this centralization is clearly seen in the distribution of advanced AI compute infrastructure. The United States and People's Republic of China (PRC) alone host the majority of the world's 101-GPU-enabled cloud regions, which are critical for training and deploying advanced AI models. Additionally, only 15 countries worldwide currently physically deploy the A100 or H100 GPUs, the most common hardware currently used for cutting-edge AI development.[2] This concentration of resources in just a few countries creates a significant divide, limiting the ability of most nations – the Global AI Majority – to develop AI capabilities and participate effectively in global AI governance.[3]Although cloud-based access to advanced compute provides some pathways for engagement, dependence on external infrastructure often comes with high costs and limited control over data and security. Addressing these barriers requires structured and equitable access solutions

[1] ITU, '2023 AI for Good Global Summit Snapshot Report', 2023, https://s41721.pcdn.co/wp-content/uploads/2021/06/SNAPSHOT-REPORT-2023-FINAL.pdf.
[2] Vili Lehdonvirta, Boxi Wu, and Zoe Hawkins, 'Compute North vs. Compute South: The Uneven Possibilities of Compute-Based AI Governance Around the Globe', in *Proceedings of the 2024 AAAI/ACM Conference on AI, Ethics, and Society (AIES '24)*, 2024, https://osf.io/8yp7z/download.
[3] Lehdonvirta, Wu, and Hawkins, 'Compute North vs. Compute South'.

that empower broader participation without necessitating physical infrastructure investments.

We find that there are two key factors that motivate states with advanced AI capabilities to implement policies restricting global access to advanced AI technologies. First, in some instances, frontier AI states might perceive it to be in their commercial interest not to encourage model development in other countries in order to advance the interests of 'national champion' AI developers. Second, frontier AI states are concerned over the proliferation of a technology with potentially dangerous capabilities. The dual-use potential of AI technologies raises concerns that the more actors with this capability (e.g. Global AI Majority countries), the greater the risk the technology will be misused to cause harm, potentially on a large scale.

In this paper, our main argument is that while significant disparities exist in AI development and governance between frontier AI states and the Global AI Majority, it is both necessary and possible to create a more inclusive global AI governance framework that balances the interests of all stakeholders. We contend that exclusion of the Global AI Majority from AI access and governance is not a sustainable or ethical approach, despite the legitimate security and commercial concerns of frontier AI states. Instead, we argue for a balanced strategy that expands both access to AI technologies and voice in AI governance for the Global AI Majority, while addressing the safety and security concerns of frontier AI states.

To establish this, **Section I** begins by examining the current global AI landscape, highlighting the stark disparities in digital infrastructure, talent distribution, and computational power between frontier AI states and the Global AI Majority. It explores how these disparities limit the ability of Global AI Majority countries to access and develop AI technologies, thus reinforcing their marginalization in global AI governance. In **Section II,** we discuss the need for inclusive voice – the ability of countries and populations to participate meaningfully in decision-making processes related to AI – in the global AI governance conversation. We observe that current critical discussions relating to AI governance are happening at fora such as the G7 and G20 which are dominated by frontier AI states. Then, **Section III** describes the two main challenges to Global AI Majority inclusion: frontier AI state interests, of which we mostly consider commercial and strategic interests and security concerns (malicious attacks, data breaches, and unauthorized access),[4] and safety (unintended

[4] Xiangyu Qi et al., 'AI Risk Management Should Incorporate Both Safety and Security' (arXiv, 29 May 2024), https://doi.org/10.48550/arXiv.2405.19524; Ken Huang, 'AI Safety vs. AI Security: Navigating the Differences', *Cloud Security Alliance* (blog), 19 March 2024, https://cloudsecurityalliance.org/blog/2024/03/19/ai-safety-vs-ai-security-navigating-the-commonality-and-differences.

harms and negative consequences).[5] We argue that, while frontier AI states have legitimate security concerns, exclusion in terms of voice and/or access is not the best approach to addressing them.

Instead, in **Sections IV** and **V** we discuss ways to better address the issues of access and voice respectively that ensure that the benefits of these technologies are distributed while the issues of safety and security are mitigated. **Section IV** focuses on addressing domestic AI development initiatives and access in the Global AI Majority. This mainly includes AI development approaches such as locally developing models from scratch and training models on remote compute facilities. We also look at access to deployed models through both open source and structured access methods. We evaluate opportunities and bottlenecks each option presents to the Global AI Majority.

In **Section V**, we propose that to address the question of voice, it is critical to establish a global AI governance regime with approaches that provide avenues for inclusion – giving voice to all affected states and communities – while simultaneously ensuring the participation of frontier AI states, whose involvement is crucial for the regime to function at all. We highlight three key approaches: (i) Interest Alignment, which focuses on areas where state interests naturally align, thereby building broad governance structures that can be expanded over time; (ii) Participatory Architecture, which entails designing inclusive governance structures that incentivize the participation of frontier AI states while ensuring that all voices are heard; and (iii) Safety Assurance by establishing safety protocols that mitigate risks without overly restricting access to AI technologies, thus creating a balanced governance framework.

There are, of course, externality risks emanating from increased global inclusion in AI development. Impacts on labour markets, the risk of bias amplification, and the resulting exacerbation of climate change should not be downplayed. The significant climate risks of AI, combined with the disparate climate change impacts many low- and middle-income countries already face, have been well covered by Monserrate (2022)[6] and Valdivia (2022)[7]. However, these are not arguments against global inclusion; they are factors that all states should consider in the development of AI. If anything, they reiterate the case for global inclusion in AI development: the

[5] Huang, 'AI Safety vs. AI Security.'
[6] Steven Gonzalez Monserrate, 'The Cloud Is Material: On the Environmental Impacts of Computation and Data Storage', *MIT Case Studies in Social and Ethical Responsibilities of Computing* Winter 2022 (2022), https://doi.org/10.21428/2c646de5.031d4553.
[7] Ana Valdivia, 'Silicon Valley and the Environmental Costs of AI', *Political Economy Research Centre* (blog), 5 December 2022, https://www.perc.org.uk/project_posts/silicon-valley-and-the-environmental-costs-of-ai/.

externalities from AI will impact all countries; therefore, all countries should have a voice in shaping decisions on AI.

This paper seeks to provide actionable recommendations for policymakers, international organizations, and stakeholders across both frontier and Global AI Majority countries. By expanding participation and addressing existing disparities, we can create a future where AI serves as a tool for collective progress, driving advancements across the Sustainable Development Goals (SDGs) and ensuring that its benefits are shared more equitably across the globe. By examining the current barriers to access, critiquing existing governance models, and proposing innovative solutions, this paper aims to contribute to the development of a global AI governance framework that reflects the diverse needs and values of all nations.

Note that we use 'Global AI Majority' as an umbrella term to aid this analysis, which builds off of works such as Ricaurte (2022)[8] and Amrute et. al. (2022)[9]. This is inspired by the term 'Global Majority' which, as discussed in Amrute et al, originates from 'Shahidul Alam, a photographer, writer, curator, and activist from Bangladesh, to highlight how the majority of the human population of the world lives in geographic regions – variously addressed as the "Developing World", the "Third World", or the "Global South" – that are rendered and remaindered as passive peripheries of ostensibly global problems and developments.' We broadly use this as a term to denote populations, often but not always in the form of countries, which are or are on track to be systematically disadvantaged by acute power and resource concentrations in AI and broader digital infrastructures as well as in institutions that shape the governance of these technologies.

In contrast to this term, we define 'frontier AI states' as countries that contain actors at the forefront of AI development whose participation in self-reinforcing power structures enable acute AI technology concentrations in those states over time. These states roughly include the United States, the PRC, and the United Kingdom, though state inclusion in this category is debatable and may change over time. Frontier AI states include subpopulations which are vulnerable to AI power concentrations in similar ways to populations in non-frontier AI states. Hence, these subgroups may often be considered part of the 'Global AI Majority'. While this binary categorization provides a useful framework for our analysis, we acknowledge that

[8] Paola Ricaurte, 'Ethics for the Majority World: AI and the Question of Violence at Scale', *Media, Culture & Society* 44, no. 4 (1 May 2022): 726–45, https://doi.org/10.1177/01634437221099612.
[9] Sareeta Amrute, Ranjit Singh, and Rigoberto Lara Guzmán, 'A Primer on AI in/from the Majority World: An Empirical Site and a Standpoint' (Data & Society, 2022), https://doi.org/10.2139/ssrn.4199467.

some advanced economies occupy a middle ground with significant AI capabilities and interests that may align with aspects of both groups.

I. The Need for Broad Access

AI has the potential to accelerate development in many low- and middle-income countries. Research suggests that appropriately deployed AI solutions could positively affect over 80% of Sustainable Development Goal targets.[10] This primarily works on two axes:

1. Increased **access to deployed AI models** could strengthen the quality and accessibility of public services and key economic sectors. For example, advanced AI models can be leveraged to improve diagnostic accuracy in healthcare, personalize learning in education, and optimize crop yields in agriculture, even in regions with limited resources.[11]

2. Increased **domestic AI development** enables AI sovereignty. Developing homegrown AI gives countries autonomy over addressing the specific needs of their population, tailoring AI solutions to advance their own interests in their domestic context. Economically, this could attract foreign investment and create new export opportunities, stimulating significant economic growth and job creation.

However, a number of entrenched barriers prevent Global AI Majority countries from accessing the benefits of AI adoption and development. These are detailed below.

Limited Access to Digital and Compute Infrastructure

The global digital divide remains a significant barrier to widespread **AI adoption,** with access to digital infrastructure varying dramatically across countries and regions. This disparity is particularly pronounced between frontier AI states and the Global AI

[10] Jingchen Zhao and Beatriz Gómez Fariñas, 'Artificial Intelligence and Sustainable Decisions', *European Business Organization Law Review* 24, no. 1 (2023): 1–39, https://doi.org/10.1007/s40804-022-00262-2.
[11] Simon Ofori Ametepey et al., 'The Impact of AI in Sustainable Development Goal Implementation: A Delphi Study', *Sustainability* 16, no. 9 (2024): 3858, https://doi.org/10.3390/su16093858.

Majority. In high-income countries, which overlap with frontier AI states, internet usage has reached near universal levels, with 93% of the population online in 2023, up from 90% in 2020.[12] In stark contrast, low-income countries, mostly part of the Global AI Majority, have seen only modest growth in internet penetration, rising from 24% in 2022 to 27% in 2023.[13] This staggering 66 percentage point gap underscores the magnitude of the digital divide. Regional disparities in internet access are stark, with Europe and the Americas enjoying around 90% penetration, while Africa lags at 37% and Asia-Pacific shows a mixed picture at 66%.[14] Even within regions, significant urban-rural divides persist, exemplified by Africa where 42% of people in urban populations use the internet compared to 32% in rural areas.[15] This uneven distribution of digital infrastructure poses a formidable challenge to equitable AI adoption and development, potentially exacerbating global inequalities as AI technologies advance.

Homegrown AI is difficult to **develop** without a stable supply of electricity and internet access. To train and maintain advanced LLMs, powerful computing infrastructure ('compute') is required. While compute can be accessed through cloud-based resources rather than building physical infrastructure, stable internet connectivity and adequate funding are still important to effectively utilise these services. This reliance on remote compute can be particularly challenging in developing countries, where frequent power outages, limited bandwidth, and high costs of cloud access pose barriers to consistent AI development.

Recent data reveals that the compute used to train frontier AI models has been growing at a rate of 4-5 times per year, with notable language models experiencing even faster growth rates of up to 9 times per year between June 2017 and May 2024.[16] This has contributed to significant AI breakthroughs but is highly resource-intensive. Determining the expense of training and maintaining a large language model (LLM) is quite complex, as there is not a simple, direct method for calculating the costs.[17] Essentially, the cost hinges on two primary elements: the computational resources needed and the duration of the training process. The cost of

[12] ITU, 'Measuring Digital Development: Facts and Figures 2023', 2023, 2, https://www.itu.int/dms_pub/itu-d/opb/ind/d-ind-ict_mdd-2023-1-pdf-e.pdf.
[13] ITU, 'Measuring Digital Development: Facts and Figures 2023', 2.
[14] ITU, 'Measuring Digital Development: Facts and Figures 2023', 2.
[15] ITU, 'Measuring Digital Development: Facts and Figures 2023', 3.
[16] Jaime Sevilla and Edu Roldán, 'Training Compute of Frontier AI Models Grows by 4-5x per Year' (Epoch AI, 28 May 2024), https://epochai.org/blog/training-compute-of-frontier-ai-models-grows-by-4-5x-per-year.
[17] Jai Vipra and Sarah Myers West, 'Computational Power and AI' (AI Now Institute, 27 September 2023), https://ainowinstitute.org/publication/policy/compute-and-ai.

developing OpenAI's GPT-4 is estimated at $100 million,[18] not including post-development maintenance costs. These ongoing expenses cover cloud infrastructure for hosting the model, operational costs for scaling and maintaining the system, data storage for the model parameters and logs, and continuous development for updates and improvements.[19]

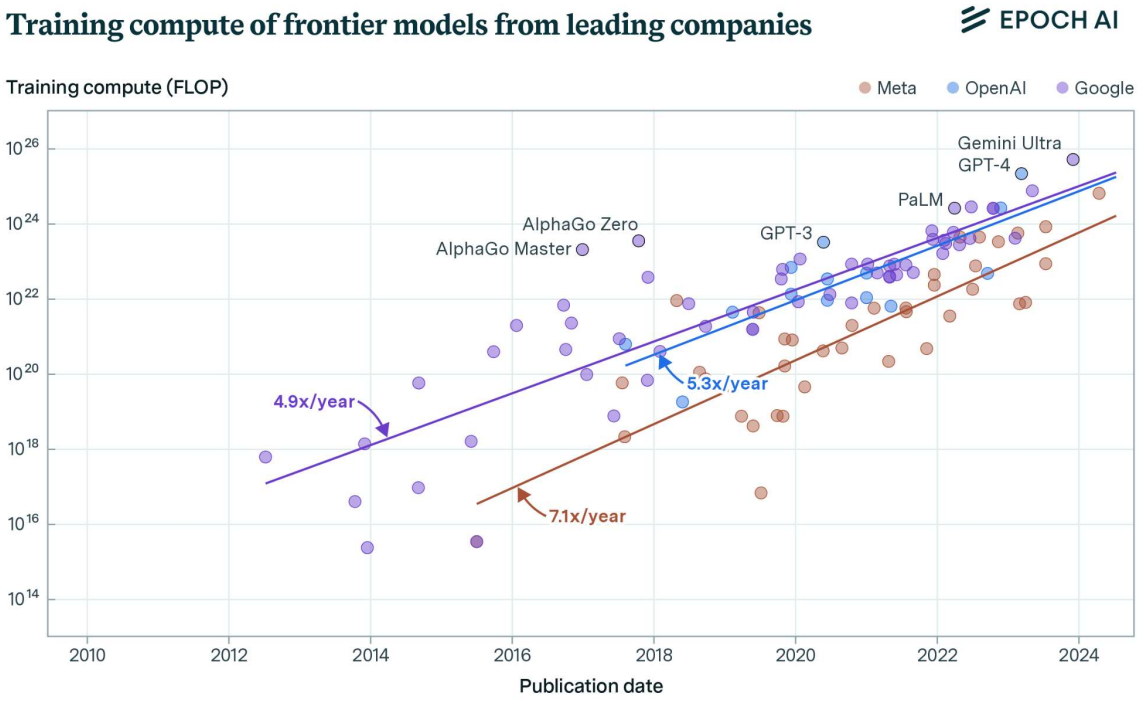

Figure 1

Beyond financial considerations, the environmental impact of AI development is substantial. Data centres and AI chips required for sovereign AI capabilities demand vast quantities of land, water, and electricity, often competing with domestic resources. Monserrate (2022)[20] highlights the extensive environmental impacts of computation and data storage across the supply chain. Additionally, Valdivia (2022)[21] notes the significant resource demands in AI supply chains, particularly for critical

[18] 'Meet the French Startup Hoping to Take on OpenAI', *The Economist*, 26 February 2024, https://www.economist.com/business/2024/02/26/meet-the-french-startup-hoping-to-take-on-openai.
[19] Vipra and West, 'Computational Power and AI'.
[20] Monserrate, 'The Cloud Is Material'.
[21] Valdivia, 'Silicon Valley and the Environmental Costs of AI'.

minerals. These resource requirements put considerable strain on finite water and electricity supplies, especially in contexts where these resources are already scarce, potentially exacerbating existing environmental and social challenges in Global AI Majority countries pursuing AI development.

While cloud computing itself is not a recent innovation, as remote access to computing resources has been available for many years, the more recent 'on-demand' capabilities have been around for over a decade.[22] These capabilities now serve as a critical access point for intensive computational tasks, particularly for Global AI Majority countries that seek to leverage advanced technologies without the need for extensive domestic infrastructure. However, control over cloud computing infrastructures is highly centralized, with major players such as Amazon, Microsoft, and Google holding around 70% of the worldwide cloud market. Alibaba, Huawei, and Tencent own the remaining 30%, which has implications for countries using Chinese data centres.[23] Efforts by governments to establish public equivalents or distributed networks are challenging and still nascent.[24] Nonetheless, there is a growing belief that developing national high-performance computing infrastructures could counter the prevailing influence of dominant cloud providers.

Power Concentration

While training advanced models can technically be done anywhere, the immense computational resources, vast amounts of data, and substantial financial investment required have led to the ownership and control of frontier models being highly concentrated, primarily in regions like the United States where these resources are most accessible.[25] As demonstrated in the graph below (figure 2) the vast majority of AI models originate in the US.

[22] Peter Mell and Tim Grance, 'The NIST Definition of Cloud Computing' (National Institute of Standards and Technology, 28 September 2011), https://doi.org/10.6028/NIST.SP.800-145.
[23] Vili Lehdonvirta, Boxi Wu, and Zoe Hawkins, 'Cloud Empires' Physical Footprint: How Trade and Security Politics Shape the Global Expansion of U.S. and Chinese Data Centre Infrastructures' (SSRN, 20 December 2023), https://doi.org/10.2139/ssrn.4670764.
[24] Lehdonvirta, Wu, and Hawkins, 'Compute North vs. Compute South'.
[25] Roberto Gozalo-Brizuela and Eduardo C. Garrido-Merchan, 'ChatGPT Is Not All You Need. A State of the Art Review of Large Generative AI Models' (arXiv, 11 January 2023), https://doi.org/10.48550/arXiv.2301.04655.

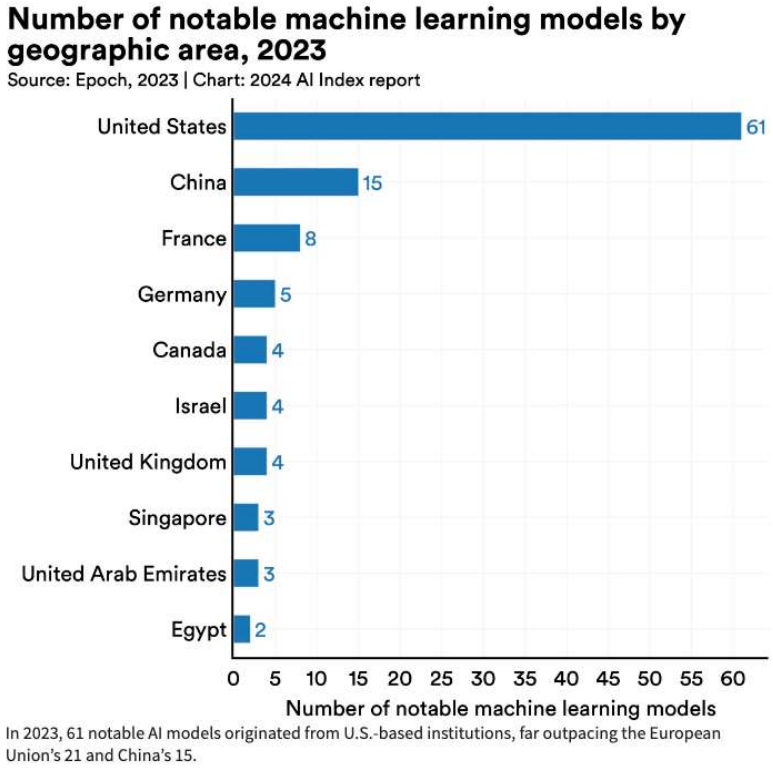

Figure 2

However, the emergence of open source models[26] like Llama 2, Llama 3, and DBRX[27] may reorient the future of AI power concentration. These models, which are freely available to the public, appear to be just behind the state-of-the-art models in terms of performance. The availability of such models could potentially democratise AI development, allowing a wider range of individuals and organizations to participate in creating and adapting advanced AI systems.[28] But as AI models become more powerful and complex, the resources required to train and maintain them may become increasingly prohibitive for open source communities. Additionally, the safety concerns surrounding advanced AI systems may lead to greater reluctance among private companies to share their models and techniques openly, further exacerbating the disparity between open source and proprietary AI development.

---

[26] 'Top Large Language Models Reshaping the Open-Source Arena', *Deci* (blog), 27 March 2024, https://web.archive.org/web/20240609234857/https://deci.ai/blog/list-of-large-language-models-in-open-source/.

[27] 'The Data and AI Company', Databricks, 13 October 2023, https://www.databricks.com/.

[28] Elizabeth Seger et al., 'Open-Sourcing Highly Capable Foundation Models: An Evaluation of Risks, Benefits, and Alternative Methods for Pursuing Open-Source Objectives' (Centre for the Governance of AI, 2023), https://cdn.governance.ai/Open-Sourcing_Highly_Capable_Foundation_Models_2023_GovAI.pdf.

With AI development restricted to well-funded, large-scale institutions in certain locations, these entities have considerable influence over training methods, access to the models, and the resulting profits. This would perpetuate both algorithmic homogenization and skewed talent distributions, discussed below.

Anglocentric Data Sources

Language

As demonstrated above, the vast majority of leading AI labs are based in the US. It is therefore not surprising that American institutions account for a staggering 54% of the total contributors to new large language and multimodal models, and these contributors are predominantly male.[29] This means that the majority of LLMs are developed by researchers trained in the anglosphere and are tested on English-speaking audiences.[30]

The dominance of English in the development of LLMs has led to a concerning trend of language homogenization and bias. English is the most popular language for web content, representing more than 50% of websites,[31] despite not being the most spoken native language globally.[32] As a result, LLMs acquire linguistic skills primarily from English text collections, resulting in an English-centric proficiency that often reduces effectiveness for non-English or data-scarce languages.[33]

This linguistic bias carries both social and practical safety risks which impede the adoption of AI in non-English-dominant Global AI Majority countries. Such models trained primarily on English data may generate unsafe responses in other languages, potentially exposing users to harm.[34] The need for developing safety alignment for non-English languages is critical to ensure the equitable distribution of language technology benefits and prevent the unintentional perpetuation of harm.[35]

---

[29] Nestor Maslej et al., 'Artificial Intelligence Index Report 2023' (Institute for Human Centered AI, 2023), https://aiindex.stanford.edu/wp-content/uploads/2023/04/HAI_AI-Index-Report_2023.pdf.
[30] Emily M. Bender et al., 'On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?🦜', in *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, 2021, 610–23, https://doi.org/10.1145/3442188.3445922.
[31] 'Most Used Languages Online by Share of Websites 2024', Statista, January 2024, https://www.statista.com/statistics/262946/most-common-languages-on-the-internet/.
[32] Eberhard, David M., Gary F. Simons, and Charles D. Fennig, eds., *Ethnologue: Languages of the World*, 25th ed. (SIL International, 2022), https://www.ethnologue.com/.
[33] Tom B. Brown et al., 'Language Models Are Few-Shot Learners' (arXiv, 22 July 2020), http://arxiv.org/abs/2005.14165.
[34] Wenxuan Wang et al., 'All Languages Matter: On the Multilingual Safety of Large Language Models' (arXiv, 20 June 2024), https://doi.org/10.48550/arXiv.2310.00905.
[35] Wang et al., 'All Languages Matter'.

Translation features have been seen as a potential solution to bridge the language gap, but they face significant challenges and limitations. Machine translation systems rely on vast datasets of matched sentences across languages, which are scarce for low-resource languages.[36] Moreover, current models often extract data from the web, yielding poor-quality translations due to inconsistent and incorrect texts.[37] This limitation in the translation capabilities of Anglocentric LLMs highlights the need for a more inclusive and diverse approach to language model training).[38]

Recent efforts to improve translation features, such as Meta's development of the NLLB-200 model,[39] have made progress in translating a wider range of languages. However, these models still struggle to accurately capture idiomatic expressions, cultural concepts, and common greetings, particularly in African languages like Yorùbá.[40] The incorporation of cultural nuances through prompting strategies shows promise, but the success of this technique heavily relies on the choice of prompts and carries the risk of producing inaccurate, fabricated text outputs known as hallucinations.[41]

The risks of adopting AI models that are not tailored to local languages or cultures demonstrates the need for 'homegrown' AI development in many Global AI Majority countries.

Bias

Beyond the training-data language, bias can originate at nearly every stage of the AI development process.[42]

In particular, the data used for training systems can have varying effects on a model's performance across different groups of people. The use of biased datasets obtained through web scraping can amplify existing social, economic, and political inequalities

[36] Changhan Wang et al., 'VoxPopuli: A Large-Scale Multilingual Speech Corpus for Representation Learning, Semi-Supervised Learning and Interpretation' (arXiv, 27 July 2021), https://doi.org/10.48550/arXiv.2101.00390.
[37] Angela Fan et al., 'Beyond English-Centric Multilingual Machine Translation' (arXiv, 21 October 2020), https://doi.org/10.48550/arXiv.2010.11125.
[38] Fan et al., 'Beyond English-Centric Multilingual Machine Translation.'
[39] NLLB Team et al., 'No Language Left Behind: Scaling Human-Centered Machine Translation' (arXiv, 25 August 2022), https://doi.org/10.48550/arXiv.2207.04672.
[40] Idris Akinade et al., 'Varepsilon Kú Mask: Integrating Yorùbá Cultural Greetings into Machine Translation', in *Proceedings of the First Workshop on Cross-Cultural Considerations in NLP (C3NLP)*, 2023, 1–7, https://doi.org/10.18653/v1/2023.c3nlp-1.1.
[41] Ziwei Ji et al., 'Survey of Hallucination in Natural Language Generation', *ACM Computing Surveys* 55, no. 12 (2023): 248:1-248:38, https://doi.org/10.1145/3571730.
[42] Damián Blasi, Antonios Anastasopoulos, and Graham Neubig, 'Systematic Inequalities in Language Technology Performance across the World's Languages' (arXiv, 13 October 2021), https://doi.org/10.48550/arXiv.2110.06733.

in Global AI Majority countries.[43] Models trained on these datasets may perpetuate discriminatory practices and reinforce power imbalances. LLMs have been found not only to replicate but also to magnify the biases ingrained in their training data. Numerous studies have revealed troubling biases linked to various attributes, including but not limited to gender, race, and religion, deeply embedded within these models.[44] Data obtained without consent or from web scraping poses significant privacy risks, especially for vulnerable populations.[45] The unpredictability of how this data is used in various applications amplifies these concerns.

Limited data availability also impedes the ability of Global AI Majority countries to develop AI models, as these nations lack local datasets large enough to train the most advanced language models.

Homogenization

Increased reliance on foundation models raises the risk of these inherent biases permeating across models.[46] More and more, the AI field is moving away from designing unique models for each specific task, instead leveraging a few highly capable generalist models that serve as the foundation for many applications. This leads to homogenization, as the same models are adapted and reused across numerous tasks and domains.

This homogenization has pros and cons. On the positive side, it allows for the concentration of efforts on improving a small set of models, which can then be deployed across various applications, optimizing resources and potentially enhancing the models' robustness and fairness. This can be likened to developing a form of societal infrastructure within the AI domain, where core models are refined and serve as a resource for other activities, much like public utilities. However, this centralization also presents risks, as these foundation models can quickly become critical points of

---

[43] Maribeth Rauh et al., 'Characteristics of Harmful Text: Towards Rigorous Benchmarking of Language Models' (arXiv, 28 October 2022), https://doi.org/10.48550/arXiv.2206.08325.
[44] Emily Sheng et al., 'The Woman Worked as a Babysitter: On Biases in Language Generation', in *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, 2019, 3407–12, https://doi.org/10.18653/v1/D19-1339; Abubakar Abid, Maheen Farooqi, and James Zou, 'Persistent Anti-Muslim Bias in Large Language Models', in *Proceedings of the 2021 AAAI/ACM Conference on AI, Ethics, and Society*, AIES '21, 2021, 298–306, https://doi.org/10.1145/3461702.3462624.
[45] Eun Seo Jo and Timnit Gebru, 'Lessons from Archives: Strategies for Collecting Sociocultural Data in Machine Learning', in *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, FAT* '20, 2020, 306–16, https://doi.org/10.1145/3351095.3372829; Joy Buolamwini and Timnit Gebru, 'Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification', in *Conference on Fairness, Accountability and Transparency* (PMLR, 2018), 77–91, http://proceedings.mlr.press/v81/buolamwini18a.html.
[46] Rishi Bommasani and Percy Liang, 'Reflections on Foundation Models', *Stanford Human-Centered Artificial Intelligence* (blog), 18 October 2021, https://hai.stanford.edu/news/reflections-foundation-models.

failure. If a foundational model has inherent biases, vulnerabilities, or other issues, these problems can be propagated across all applications that rely on it, potentially causing widespread harm. Any biases in the training or evaluation datasets are massively amplified through homogenization, particularly if these datasets are also used to generate synthetic training data.

This makes it imperative that these models are developed, evaluated, and monitored with utmost care to prevent adverse outcomes across their myriad applications.[47] In addition, it again points to the need for Global AI Majority countries to develop their own foundational AI models in order to prevent bias, and to enable the tailoring of models to domestic contexts.[48]

Skewed Talent Distributions

As highlighted above, the vast majority of frontier AI labs are based in the US (with others in the PRC, EU states, and other Western countries). These labs can attract top AI talent from around the world with annual compensation packages that exceed $1 million dollars.[49] Countries from the Global AI Majority cannot compete: they do not have the academic institutions, job opportunities, or investment to cultivate sufficient numbers of students, researchers, and skilled experts on AI. Those who do become experts are recruited to work in these US-based labs, contributing to a domestic 'brain drain'.[50] While some training programs are in place,[51] there are limited incentives for researchers to stay in their home countries and contribute to domestic AI adoption and development.

Google, Microsoft, and IBM have research labs outside of the US, but these tend to be concentrated in specific geographies (India, Brazil, Ghana, Kenya, and South Africa) and restricted to specific research areas, rarely cutting edge. Although such centres of innovation may appear beneficial for the local workforce, Chan et al

---

[47] Rishi Bommasani et al., 'On the Opportunities and Risks of Foundation Models' (arXiv, 12 July 2022), https://doi.org/10.48550/arXiv.2108.07258.
[48] A. Stevie Bergman et al., 'Representation in AI Evaluations', in *Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency*, FAccT '23, 2023, 519–33, https://doi.org/10.1145/3593013.3594019.
[49] Katherine Bindley, 'The Fight for AI Talent: Pay Million-Dollar Packages and Buy Whole Teams', *The Wall Street Journal*, 27 March 2024, https://www.wsj.com/tech/ai/the-fight-for-ai-talent-pay-million-dollar-packages-and-buy-whole-teams-c370de2b.
[50] Chinasa T Okolo, 'AI in the Global South: Opportunities and Challenges Towards More Inclusive Governance', *International Journal of Innovative Research in Arts, Education and Technology* 2, no. 1 (2021), https://internationalpolicybrief.org/wp-content/uploads/2024/02/ARTICLE-9.pdf.
[51] For example, Nigeria has established partnerships with Microsoft to equip citizens with digital skills, while partnerships with Google have trained nearly eight million people in Latin America in digital skills since 2017.

(2021)[52] found that local representation in the majority of research centres was sorely lacking at both the leadership and general workforce level. As they illustrate, significant numbers of AI researchers and practitioners in these research centres are not from the region, and they displace opportunities for the local workforce.

II. The Need for Inclusive Voice in AI Governance

As AI technologies evolve, it is becoming clear that their effects will not be contained by national borders. Some states are creating AI-safety institutes,[53] which are partly focused on developing ways of evaluating whether an AI system could facilitate the development of a chemical, biological, radiological, nuclear (CBRN) or cyber weapon or tools of mass persuasion, surveillance, and repression.[54] Such technologies could have large impacts outside of the regions where they are developed. AI content generated in one place can affect national discourses and electoral outcomes in another.[55] Integrating AI systems into critical infrastructure such as power grids, water supplies, or communication networks could lead to systemic failures in cities across the world if these systems are compromised, either accidentally due to a lack of robustness or deliberately through cyberattacks.[56] These alone pose enormous national security risks, and could introduce a highly escalatory and catastrophic approach to warfare. Overall, risks stemming from advanced AI could result in significant, even catastrophic, scales of harm that have the potential to affect all countries around the world.

Given that AI has the potential to impact all of humanity, there is a strong moral case for global inclusion in its governance. Potential risks that will arise from the development and deployment of these technologies will have a disproportionate effect on the Global AI Majority. While frontier AI states are at the forefront of these technological developments, they automatically expose the Global AI Majority to risks

---

[52] Alan Chan et al., 'The Limits of Global Inclusion in AI Development' (arXiv, 1 February 2021), https://doi.org/10.48550/arXiv.2102.01265.

[53] To date, the UK, US, Japan among others have all launched national AI-safety institutes. Julian Jacobs, 'National Approaches to AI Safety Diverge in Focus', *OMFIF* (blog), 25 June 2024, https://www.omfif.org/2024/06/national-approaches-to-ai-safety-diverge-in-focus/.

[54] The UK AI Safety Institute prioritizes pre-deployment evaluations for dual-use capabilities, including aiding in cyber-criminality, biological or chemical science, human persuasion, large-scale disinformation campaigns, and weapons acquisition. It will also conduct post-deployment evaluations to investigate psychological impacts, privacy harms, manipulation and persuasion, biased outputs and reasoning, impacts on democracy and trust in institutions, and systemic discrimination. 'Introducing the AI Safety Institute' (AI Safety Institute, 2023), https://www.gov.uk/government/publications/ai-safety-institute-overview/introducing-the-ai-safety-institute.

[55] Alice Dawson and James Ball, 'Generating Democracy AI and the Coming Revolution in Political Communications' (Demos, 2024), https://demos.co.uk/wp-content/uploads/2024/01/Generating-Democracy-Report-1.pdf.

[56] Christian Vasquez, 'DHS Warns of Malicious AI Use Against Critical Infrastructure', *CyberScoop* (blog), 14 September 2023, https://cyberscoop.com/dhs-homeland-threat-assessment/.

that could arise from these technologies. Beyond this, these countries need not be mere bystanders in this era of frontier model development and deployment; many have the potential to play active, agentic roles.[57] However, the current state of AI governance falls short of this ideal. A large portion of the global population is excluded from meaningful participation, as discussions are often confined to a select few countries.

The concentration of advanced AI capabilities in just a few countries gives these nations disproportionate influence over AI governance. Within these countries, advanced technologies are often controlled by a small number of private entities, primarily large tech companies and elite universities. This concentration of ownership among a minority of geographically and physically centralized stakeholders – largely in developed nations – has implications. It shapes the direction of AI research, determines the accessibility of AI advancements, and drives the commercialisation of new technologies. Most critically, it raises significant concerns about the democratisation of AI. For instance, healthcare-related AI tools designed in wealthy nations may overlook diseases prevalent in poorer regions or require infrastructure that is not available in many parts of the world.

Yet, the existing material context is only a part of the reason for the restricted set of voices in the global governance of AI; decisions about *how* to govern AI perpetuate the situation. Thus far, AI policy and governance frameworks are discussed only within frontier AI states or within exclusionary clubs of states. According to one OECD estimate, 58% of the discourse has occurred in Europe and North America, in contrast to only 1.4% in Africa.[58] For instance, G7 countries launched the Hiroshima process in 2023. Such a process, at least in its inception, is highly exclusionary, representing the interests of just 9% of the global population. Leaders from India, Brazil, Indonesia, the African Union, and Vietnam were invited to Hiroshima, but their presence had little impact on the substantive outcomes.[59] In the report prepared for the Hiroshima AI process, the Global South was mentioned only once,[60] while the final statement emphasizes the importance of 'closing digital divides and achieving digital inclusion' without laying out concrete measures for how to do so. Many such international AI governance efforts ultimately result in a 'paradox of participation,' where there is

---

[57] Cecil Abungu, Michelle Malonza, and Sumaya Nur Adan, 'Can Apparent Bystanders Distinctively Shape an Outcome? Global South Countries and Global Catastrophic Risk-Focused Governance of Artificial Intelligence' (arXiv, 7 December 2023), https://doi.org/10.48550/arXiv.2312.04616.
[58] 'OECD's Live Repository of AI Strategies & Policies', OECD, accessed 16 May 2024, https://oecd.ai/en/dashboards.
[59] Chris Olaoluwa Ògúnmọ́dẹdé, 'The G-7's Embrace of the Global South Was All Talk, No Substance', *World Politics Review* (blog), 26 May 2023, https://www.worldpoliticsreview.com/g7-summit-2023-politics-economy-africa-global-south/.
[60] 'G7 Hiroshima Process on Generative Artificial Intelligence (AI)' (OECD, 6 September 2023), https://www.oecd.org/en/publications/2023/09/g7-hiroshima-process-on-generative-artificial-intelligence-ai_8d19e746.html.

surface-level participation of Global Majority stakeholders without any provision of accompanying resources and structural reforms to involve them meaningfully.[61]

Beyond shortcomings in multilateral processes, unilateral actions taken by frontier AI states are poised to profoundly shape the availability of necessary resources for advanced AI development, thus altering the broader trajectory and inclusiveness of AI development and governance. The United States, at times along with its allies, has taken unilateral steps to restrict the spread of advanced computing resources. These measures include tighter controls on the export of semiconductor manufacturing equipment, quantum computing, and artificial intelligence technologies to the PRC.[62] Japan and The Netherlands have also introduced export controls to complement those led by the US.[63] The US has further proposed a 'Know Your Customer' (KYC) regime which would require cloud providers (Infrastructure as a Service, IaaS) to identify and report foreign customers purchasing large amounts of compute.[64] Such steps give these countries exclusive leverage over the development and actions of other states through control of the compute supply chain.

The dominance of Western-led governance efforts has led to increasing backlash from Global AI Majority states to these fora. This follows longstanding patterns of increased dissatisfaction with Western-led governance processes in general, which a large segment of the Global AI Majority believes does not serve their interests. For example, Turkish President Recep Tayyip Erdogan frequently states that 'the world is

---

[61] Marie-Therese Png, 'At the Tensions of South and North: Critical Roles of Global South Stakeholders in AI Governance', in *Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency*, FAccT '22, 2022, 1434–45, https://doi.org/10.1145/3531146.3533200.

[62] Kirti Gupta, Chris Borges, and Andrea Leonard Palazzi, 'Collateral Damage: The Domestic Impact of U.S. Semiconductor Export Controls', Center for Strategic & International Studies, 9 July 2024, https://www.csis.org/analysis/collateral-damage-domestic-impact-us-semiconductor-export-controls; 'Commerce Implements New Export Controls on Advanced Computing and Semiconductor Manufacturing Items to the People's Republic of China (PRC)' (U.S. Department of Commerce, Bureau of Industry and Security, 7 October 2022), https://www.bis.doc.gov/index.php/documents/about-bis/newsroom/press-releases/3158-2022-10-07-bis-press-release-advanced-computing-and-semiconductor-manufacturing-controls-final/file.

[63] Gregory C. Allen, Emily Benson, and Margot Putnam, 'Japan and the Netherlands Announce Plans for New Export Controls on Semiconductor Equipment', Center for Strategic & International Studies, 10 April 2023, https://www.csis.org/analysis/japan-and-netherlands-announce-plans-new-export-controls-semiconductor-equipment.

[64] Robert Stankey et al., 'Commerce Department Proposes Cybersecurity/AI Reporting and "KYC" Requirements for Certain Cloud Providers', *Davis Wright Tremaine* (blog), 14 February 2024, https://www.dwt.com/blogs/artificial-intelligence-law-advisor/2024/02/commerce-department-proposes-kyc-ai-rules-for-iaas.

bigger than five',[65] referring to the permanent members of the United Nations Security Council. Resistance from G7 and OECD agendas has led to renewed enthusiasm for the BRICS grouping – comprising Brazil, Russia, India, the PRC, and South Africa. The PRC in particular is increasing its diplomatic efforts to capitalize on this sense of disenfranchisement and present itself as a leader in providing alternatives to Western AI governance.[66]

Other more inclusive governance initiatives also exist but face challenges to influence, as we discuss below. Thus, the Global AI Majority grapples with the impacts of externally developed AI systems but lacks influence and access to reciprocally shape AI progress. Closing these gaps by elevating excluded voices is essential for ethically aligned innovation. Empowering diverse participation throughout the development and diffusion process facilitates AI solutions that account for marginalized interests and serve local needs.

III. Challenges to Change: State Interests and the Dual-Use Dilemma

At the heart of the difficulties of mitigating inequalities in access and voice are the twin challenges of national commercial and security interests. Frontier AI states have both security and commercial incentives to limit the diffusion of advanced AI capabilities in order to restrict opportunities for AI accidents and misuse by malicious actors as well as to maintain competitive advantages for their firms. To mitigate security risks from widely diffused advanced AI models, these states may pursue protectionist policies that simultaneously nurture 'national champion' firms[67] and maintain national competitiveness. For example, export controls on semiconductors and other dual-use AI technologies prevent proliferation while protecting domestic industries.

The dual-use nature of some technologies has long been a justification for state-imposed barriers to their diffusion, such as export or import control regimes. Advanced AI technologies may be the most dual-use of all given the 'general' nature

---

[65] '"Our Motto 'the World Is Bigger than Five' Is the Biggest-Ever Rise against Global Injustice"', Presidency of the Republic of Türkiye, 1 October 2018, https://www.tccb.gov.tr/en/news/542/89052/our-motto-the-world-is-bigger-than-five-is-the-biggest-ever-rise-against-global-injustice; Berdal Aral, '"The World Is Bigger than Five": A Salutary Manifesto of Turkey's New International Outlook', *Insight Turkey*, 2019, https://doi.org/10.25253/99.2019214.05.

[66] Bill Drexel and Hannah Kelley, 'Behind China's Plans to Build AI for the World', Politico, 30 November 2023, https://www.politico.com/news/magazine/2023/11/30/china-global-ai-plans-00129160.

[67] Will Henshall, 'E.U.'s AI Regulation Could Be Softened After Pushback', *TIME*, 22 November 2023, https://time.com/6338602/eu-ai-regulation-foundation-models/.; Gregory C. Allen, 'Understanding China's AI Strategy' (Center for a New American Security, 6 February 2019), https://www.cnas.org/publications/reports/understanding-chinas-ai-strategy.

of their potential applications and capabilities.[68] As AI systems grow more capable, they could be misused deliberately or accidentally in ways that threaten critical infrastructure, enable novel cyberattacks, facilitate persuasion campaigns that undermine democracy, or empower new biological or nuclear threats. Thus, proliferation raises the possibility of misuse, and such harms cross national boundaries. Indeed, some of the same factors that make global voice in governance necessary – in particular the potential of cross-border impacts – also make it difficult to achieve. Advanced AI also has the potential to directly impact traditional state security interests and the balance of power.[69] As in nuclear industries, and others, such concerns lead frontier AI states to impose export restrictions and classification regimes and take other steps to prevent the global spread of the technology.

States have legitimate security concerns,[70] but these can also be used to advocate for beggar-thy-neighbour commercial interest policies.[71] Ambiguous dual-use classifications reinforce this trend, functioning to advance individual state interests as much as to protect collective global security. Unilateral or club-based restrictions on dual-use AI systems will concentrate capabilities among fewer states. This trend is in direct conflict with moral arguments that all nations should benefit from technological progress and have a voice in processes that directly affect them. Reconciling security imperatives with moral calls for inclusion remains deeply complex, as restrictions meant to mitigate genuine risks of misuse also frequently serve to consolidate the power and advantage of a limited group of actors.

Other features of the technology also have significant governance implications. These include its 'black box' nature, continued disagreements over the risks it poses, and limited reliability of safety assurance techniques including evaluations,

---

[68]As defined in the Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence, the term 'dual-use foundation model' means 'an AI model that is trained on broad data; generally uses self-supervision; contains at least tens of billions of parameters; is applicable across a wide range of contexts; and that exhibits, or could be easily modified to exhibit, high levels of performance at tasks that pose a serious risk to security, national economic security, national public health or safety, or any combination of those matters'. See: 'Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence' (The White House, 30 October 2023), https://www.whitehouse.gov/briefing-room/presidential-actions/2023/10/30/executive-order-on-the-safe-secure-and-trustworthy-development-and-use-of-artificial-intelligence/.

[69] Ilaria Carrozza, Nicholas Marsh, and Gergory M. Reichberg, 'Dual-Use AI Technology in China, the US and the EU: Strategic Implications for the Balance of Power' (Peace Research Institute Oslo, 2022), https://www.prio.org/publications/13150;; Stefka Schmid, Thea Riebe, and Christian Reuter, 'Dual-Use and Trustworthy? A Mixed Methods Analysis of AI Diffusion Between Civilian and Defense R&D', *Science and Engineering Ethics* 28, no. 2 (2022): 12, https://doi.org/10.1007/s11948-022-00364-7.

[70] Jane Vaynman and Tristan A. Volpe, 'Dual Use Deception: How Technology Shapes Cooperation in International Relations', *International Organization* 77, no. 3 (2023): 599–632, https://doi.org/10.1017/S0020818323000140.

[71]Adam Smith, *The Wealth of Nations*, 1776.

interpretability, and red-teaming.[72] The still limited understanding of these systems makes establishing universal risk assessments, safety protocols, and effective oversight mechanisms difficult. Further, any established governance and safety frameworks may be quickly outpaced by the rapid evolution of the technology itself. These factors necessitate a nimble, adaptable governance structure that facilitates swift, informed action while constructing a delicate balance between mitigating security risks and including the interests of a wide range of actors.

IV. AI Development Initiatives and Access

A range of development initiatives have been suggested to enhance access to advanced AI among the Global AI Majority. It is important to note that these are not merely efforts to facilitate Global AI Majority access to models developed elsewhere. As we note above, they are also attempts to gain access to and shape development processes and to create advanced AI systems that truly represent the diversity of cultural contexts. Thus, we divide these initiatives into two categories: **domestic AI development** and **access to deployed AI models**. The steps required and the prospects for success are substantially different in each category.

We discuss these initiatives and some of their challenges. Drawing firm conclusions about the viability of different approaches is beyond the scope of this paper. The success of development initiatives is dependent upon the path that these technologies take in the future as well as upon the actions of broad sets of stakeholders. We do not presume to have all of these answers. Some companies in which savvy venture capitalists invested hundreds of millions have foundered, while others that initially received little notice have been successful.[73] Thus, our goal in this section is to canvas opportunities and challenges as they currently appear.

---

[72] Tom Wheeler, 'The Three Challenges of AI Regulation', Brookings, 15 June 2023, https://www.brookings.edu/articles/the-three-challenges-of-ai-regulation/

[73] Recently, for instance, the CEO of Stability AI, previously valued at $1 billion, left the company along with other departures of top management, seemingly signaling that the company's business model might not prove successful, at least in the short term. Similarly, Microsoft recently hired most of the staff of Inflection AI, which was previously valued at $4 billion. See Kenrick Cai and Iain Martin, 'Stability AI Founder Emad Mostaque Tanked His Billion-Dollar Startup', 30 March 2024, https://www.forbes.com/sites/kenrickcai/2024/03/29/how-stability-ais-founder-tanked-his-billion-dollar-startup/ and Shirin Ghaffary and Rachel Metz, 'Microsoft to Pay Inflection AI $650 Million After Scooping Up Most of Staff', Bloomberg, 21 March 2024, https://www.bloomberg.com/news/articles/2024-03-21/microsoft-to-pay-inflection-ai-650-million-after-scooping-up-most-of-staff.

Domestic AI Development

Access to basic digital services is fundamental to access to advanced AI. Any focus on the latter must not come at the expense of the former. Initiatives such as the Alliance for Affordable Internet (A4AI) are working to drive down the cost of internet access in low- and middle-income countries through policy and regulatory reform. Plans are underway to improve connectivity in the Global South, starting with the African continent. 2Africa, for instance, the longest subsea internet cable ever designed, currently has 46 connections to land-based networks around 33 countries within Africa, Asia, and Europe.[74]

Yet, there are trade-offs in the development of basic digital infrastructure that relate to countries' approaches to increasing access to advanced AI. Many existing initiatives related to the SDGs and global access to the benefits of advancements in technology seek to broaden access to stable internet in rural and hard-to-reach areas. These efforts are critical for broader economic development and inclusion throughout more rural areas. On the other hand, should countries aim to become significant players in the realm of advanced AI development, they would likely need to invest large sums into locally based computing clusters and, more broadly, would need to rapidly develop high-quality digital infrastructure concentrated in urban areas.[75] This choice would not be without significant trade-offs for the former priority of expanding basic connectivity to the most underserved populations within countries.

Developing Systems Locally from Scratch

Developing advanced AI systems from scratch, particularly general-purpose systems, requires immense amounts of computing power and data as well as access to high-skilled labour with knowledge of the latest techniques. This is often called the 'AI triad' of inputs:[76] compute, data, and algorithms (embodied in high-skilled labour). No country encompasses the entirety of the supply chains that produce these systems. Very few states have the computing capacity, in public or private hands, to single-handedly produce the most advanced general-purpose models. Indeed, as we have seen, only a few companies have the computing resources to train frontier models – and even those computing resources are distributed across national borders.[77]

---

[74] '2Africa Deployment Is Underway', 2Africa Cable, accessed 2 May 2024, https://www.2africacable.net.
[75] Ben Buchanan, 'The AI Triad and What It Means for National Security Strategy' (Center for Security and Emerging Technology, 2020), https://cset.georgetown.edu/publication/the-ai-triad-and-what-it-means-for-national-security-strategy/.
[76] Buchanan, 'The AI Triad and What It Means for National Security Strategy'.
[77] Lehdonvirta, Wu, and Hawkins, 'Compute North vs. Compute South.'

Further, the cost of training frontier AI models is staggering. A single training run for a state-of-the-art AI model can cost more than $1 billion.[78] This expense is primarily driven by the enormous amount of **computing power** required, which consumes significant amounts of energy. For example, training OpenAI's GPT-3 model is estimated to have cost around $12 million in computing resources alone.[79] Building the capacity to train advanced AI models from scratch on local compute is impractical for many countries, including European nations that have recently invested in sovereign compute initiatives.[80]

Compute

Among the AI inputs triad, the proliferation of massive amounts of training compute poses the greatest threat to frontier AI state interests, both commercial and security. Compute is increasingly seen as a key governance tool due to its excludability, quantifiability, and supply chain concentration.[81] Frontier AI states are likely to resist the broad spread of advanced AI computing power, viewing it as a critical input akin to radioactive material for nuclear technologies. Frontier AI states may also prioritize investments in domestic computing infrastructure and talent development to reduce reliance on foreign suppliers and maintain technological leadership.[82]

However, creating a more structured, rules-based compute governance regime could potentially mitigate these threats while guaranteeing equitable access. Computing infrastructure providers could serve as intermediaries between AI firms and governments, acting as verifiers and enforcers of governance regimes.[83] Hardware mechanisms for governing chips,[84] for instance enabling remote shutoff, while controversial, could also enable safer proliferation of chips internationally. The potential of governance regimes to facilitate the diffusion of technology is an important factor for furthering development, and a theme to which we shall return.

---

[78] Girish Sastry et al., 'Computing Power and the Governance of Artificial Intelligence' (arXiv, 13 February 2024), https://doi.org/10.48550/arXiv.2402.08797.
[79] Chuan Li, 'OpenAI's GPT-3 Language Model: A Technical Overview', *Lambda* (blog), 3 June 2020, https://lambdalabs.com/blog/demystifying-gpt-3.
[80] Keith Strier, Jack Clark, and Sana Khareghani, 'Measuring Compute Capacity: A Critical Step to Capturing AI's Full Economic Potential', *OECD.AI Policy Observatory* (blog), 8 February 2022, https://oecd.ai/en/wonk/ai-compute-capacity.
[81] Sastry et al., 'Computing Power and the Governance of Artificial Intelligence'.
[82] Paul Scharre, 'Future-Proofing Frontier AI Regulation: Projecting Future Compute for Frontier AI Models' (Center for a New American Security, March 2024), https://www.cnas.org/publications/reports/future-proofing-frontier-ai-regulation; Samuel Hammond, 'The Scramble for AI Computing Power', *American Affairs* 8, no. 2 (2024), https://americanaffairsjournal.org/2024/05/the-scramble-for-ai-computing-power/.
[83] Gabriel Kulp et al., 'Hardware-Enabled Governance Mechanisms: Developing Technical Solutions to Exempt Items Otherwise Classified Under Export Control Classification Numbers 3A090 and 4A090' (RAND Corporation, 18 January 2024), https://www.rand.org/pubs/working_papers/WRA3056-1.html.
[84] Kulp et al., 'Hardware-Enabled Governance Mechanisms'.

Skilled Labour

Skilled labour for AI development is in high demand, but specialized expertise is limited. States can promote workforce development by nurturing top innovators to create domestic AI systems and by enhancing digital literacy across the board. This approach should help to build a robust domestic AI ecosystem that can effectively navigate the risks and opportunities of AI.

Locally based initiatives using locally based labour exist in Global AI Majority countries. These include grassroots efforts such as Deep Learning Indaba, KHIPU, AI Saturdays Lagos, and Data Science Africa.[85] These initiatives are building communities of local AI researchers and developers. In Indonesia, for example, investments in AI education have led to the creation of AI research centres at leading universities, as well as industry collaborations.[86] In India, collaboration between educational institutions and industries has been strengthened through initiatives like Digital India and Skill India, which integrate AI education and technology into curricula and provide training tailored to industry needs. These programs encompass practical training and internships, ensuring that graduates possess the skills demanded by the AI job market. Beyond the public sector, tech giants are supporting the digital skills transformation. Google, for instance, has committed to train 20,000 Nigerian women and young people in AI and digital skills.[87]

While the development of skilled AI labour is progressing globally, some restrictions on knowledge and talent transfer exist, particularly between geopolitical rivals. The United States, United Kingdom, and other frontier states in AI development have adopted a targeted approach to restricting AI technology transfers and collaborations, primarily focused on countries like the PRC and Russia.[88] These restrictions, which have previously included visa restrictions and increased scrutiny of

---

[85] Shakir Mohamed, Marie-Therese Png, and William Isaac, 'Decolonial AI: Decolonial Theory as Sociotechnical Foresight in Artificial Intelligence', *Philosophy & Technology* 33, no. 4 (1 December 2020): 659–84, https://doi.org/10.1007/s13347-020-00405-8.

[86] Examples include the AI Research and Big Data Analytics Center (Pusat Riset AI dan Big Data) at Padjadjaran University, as well as projects at Bandung Institute of Technology (ITB) and the University of Brawijaya, and collaborations with industries players including NVIDIA, Tokopedia, and Bukalapak. See 'Pusat Riset Kecerdasan Artifisial dan Big Data', The University of Padjadjaran, n.d., https://informatika.unpad.ac.id/pusat-riset-kecerdasan-artifisial-dan-big-data/; 'Tokopedia-UI AI Center', n.d., https://tokopedia-ai.cs.ui.ac.id/; Webmaster Team ITB Direktorat Sistem dan Teknologi Informasi, 'Pusat Artificial Intelligence', Institut Teknologi Bandung, n.d., https://www.itb.ac.id/pusat-artificial-intelligence.

[87] Felix Onuah, 'Google to Train 20,000 Nigerians in Digital Skills', *Reuters*, 15 August 2023, https://www.reuters.com/world/africa/google-train-20000-nigerians-digital-skills-2023-08-15/.

[88] 'The Global AI Talent Tracker 2.0', MacroPolo, accessed 11 May 2024, https://macropolo.org/digital-projects/the-global-ai-talent-tracker/.

research partnerships, aim to prevent unauthorized transfer of sensitive AI knowledge.[89] While these constraints are largely targeted at specific nations, other dimensions, such as significantly higher salaries, access to cutting-edge research facilities, and proximity to leading tech hubs in frontier AI states, lead to strong concentration of talent. Despite these challenges, there is potential to leverage existing talent and invest in local skilled labour development in many Global AI Majority countries, as the initiatives above demonstrate.

Data

The third pillar of the triad of AI inputs is **data** – and for the Global AI Majority, local data is often in particularly short supply. This appears to be an important area of development efforts, and a variety of organizations have focused on it. This includes Lelapa AI, Lesan AI,[90] Masakhane,[91] and Ghana NLP,[92] which develop datasets and machine translation tools to expand access to low-resource African languages.[93] In an attempt to revitalize and protect the *te reo* Māori language, local entrepreneurs, through the help of their community members, collected over 300 hours of annotated audio for building language tech tools.[94] This was partly motivated by the frustration of having big tech companies working with Māori speakers to develop language-learning tools that would be marketed back to the same communities for a price.[95] With the data, the locals were able to build a speech-to-text engine with an initial error rate of 14%, which eventually was reduced to 10%.

Developing datasets of less-digitized cultures responsibly appears to be an essential area for research and development efforts. According to GlobalData, the

---

[89] Remco Zwetsloot, Roxanne Heston, and Zachary Arnold, 'Strengthening the U.S. AI Workforce' (Center for Security and Emerging Technology, 2019), https://doi.org/10.51593/20190003.
[90] Abdullahi Tsanni, 'This Company Is Building AI for African Languages', MIT Technology Review, 17 November 2023, https://www.technologyreview.com/2023/11/17/1083637/lelapa-ai-african-languages-vulavula/.
[91] 'Masakhane', n.d., https://www.masakhane.io/.
[92] 'Ghana Natural Language Processing (NLP)', n.d., https://ghananlp.org/.
[93] The AI for Development Programme, announced at the UK's AI Safety Summit in November 2023, is an $80 million collaboration to support homegrown AI expertise and computing power in Africa. It seeks to support the continent's long-term development. The initiative will help to fund training and fellowships in AI in African universities, foster responsible AI governance, and invest in 'homegrown' datasets that accurately represent the African continent. See Foreign, Commonwealth & Development Office, 'UK Unites with Global Partners to Accelerate Development Using AI', GOV.UK, 1 November 2023, https://www.gov.uk/government/news/uk-unites-with-global-partners-to-accelerate-development-using-ai..
[94] Donavyn Coffey, 'Māori Are Trying to Save Their Language from Big Tech', *Wired*, 20 April 2021, https://www.wired.com/story/maori-language-tech/.
[95] Karen Hao, 'A New Vision of Artificial Intelligence for the People', MIT Technology Review, 22 April 2022, https://www.technologyreview.com/2022/04/22/1050394/artificial-intelligence-for-the-people/.

data-collection and -labelling industry is expected to grow to over $14 billion by 2030, with much of the labelling work conducted in east Africa, India, and the Philippines. Such efforts are closely related to attempts to preserve national heritages, yet they also present novel challenges. Chief among those are privacy concerns and ensuring that any data collected is not misused for political and other ends.[96] In the case of the Māori language mentioned above, a data licence has been established. The licence outlines the guidelines for future partnerships and is rooted in the Māori concept of *kaitiakitanga*, or guardianship. Access to data will only be permitted to organizations that commit to honouring Māori principles, adhere to consent boundaries, and ensure that any benefits obtained are returned to the Māori community.

States attempt to restrict foreign use of their own citizens' data, but among the elements of the AI triad, states generally restrict data flows the least for commercial and strategic reasons. Nonetheless, the UK and US implement specific regulations such as the UK's GDPR alignment with EU standards for data protection and the US's sector-specific protections for sensitive data like health and financial information.[97] As discussed above, large datasets are available, but they do not cover many of the world's languages well. When it comes to Global AI Majority states' own data, there is little opportunity or incentive for frontier AI states to attempt to restrict its use.

These considerations suggest a mixed conclusion on the viability of development efforts to assist Global AI Majority states in training advanced AI systems from scratch. Most Global AI Majority states, on their own, will not be able to assemble the compute resources for training models. Assembling this compute is extremely costly, and frontier AI states may well place barriers to the diffusion of the most advanced AI chips. Development efforts to leverage local knowledge and talent, as well as local data, to train models from scratch on compute hosted elsewhere appear more promising, though still present challenges as skilled talent and comprehensive datasets remain difficult for non-frontier AI states to acquire.

---

[96] There is also a danger of exploitation in the data-labelling industry itself. Frontier AI firms often rely on low-wage workers from sub-Saharan Africa and Southeast Asia; Madhumita Murgia, 'AI's New Workforce: The Data-Labelling Industry Spreads Globally', *Financial Times*, 24 July 2019, https://www.ft.com/content/56dde36c-aa40-11e9-984c-fac8325aaa04. The risks of exploitation are significant; Samasource content moderators in Nairobi recently called for an investigation into the exploitative conditions they faced while reviewing content to train OpenAI's ChatGPT; Niamh Rowe, '"It's Destroyed Me Completely": Kenyan Moderators Decry Toll of Training of AI Models', *The Guardian*, 2 August 2023, https://www.theguardian.com/technology/2023/aug/02/ai-chatbot-training-human-toll-content-moderator-meta-openai. Moderators were underpaid (between $1.46 and $3.74 an hour) and forced to review violent, often explicit texts and images with no psychological support.

[97] 'International Data Transfers', Information Commissioner's Office (ICO, 19 October 2023), https://ico.org.uk/for-organisations/data-protection-and-the-eu/data-protection-and-the-eu-in-detail/the-uk-gdpr/international-data-transfers/.; 'US Data Transfers - Update', *Dentons* (blog), 28 October 2022, https://www.dentons.com/en/insights/articles/2022/october/28/us-data-transfers.

Training Models on Compute Hosted Abroad and Modifying Models Developed Abroad

Another option for Global AI Majority countries and industries is to train models on compute hosted elsewhere. This is still expensive, as discussed previously, even if only a small number of models are trained. Yet, it is much less expensive than assembling local compute clusters and is also likely to encounter less resistance from frontier AI states. One notable example is the Masakhane project, which aims to build NLP models for African languages by utilizing computing resources from cloud services primarily in the Global North, such as Google's, through compute credits.[98]

Fine-tuning a model developed elsewhere is another approach that might sometimes serve Global AI Majority actors well. Fine-tuning is a machine learning technique where a pre-trained model is further trained on a specific dataset or task, allowing it to adapt its general knowledge to more specialized applications or local contexts.[99] This process can significantly enhance a model's performance on targeted tasks while requiring less computational resources and data than training from scratch would. It requires the same elements as training models from scratch does – but considerably less of them. This is a natural approach for some important tasks, such as translation to and from highly digitized languages.[100] Fine-tuning relies on access to the model to be fine-tuned, which is possible when models are open-sourced or when their proprietors enable fine-tuning access. Meta's 'No Language Left Behind (NLLB)' family of models is an example of a useful set of open source models. They are able to translate between any of 202 languages (compared to more than 7,000 languages in the world). These models can be fine-tuned to add new languages for translation,[101] tailoring them to local contexts.

---

[98] 'Masakhane – Pioneering Participatory Approaches to Building African Language Technologies, for Africans, by Africans', *Ars Electronica Festival 2023 - Who Owns the Truth?* (blog), 2 September 2023, https://ars.electronica.art/who-owns-the-truth/en/masakhane/.
[99] 'What Is Fine-Tuning?', IBM, 15 March 2024, https://www.ibm.com/topics/fine-tuning.
[100] Jiawei Zheng et al., 'Fine-Tuning Large Language Models for Domain-Specific Machine Translation' (arXiv, 22 February 2024), https://doi.org/10.48550/arXiv.2402.15061.
[101] David Dale, 'How to Fine-Tune a NLLB-200 Model for Translating a New Language', *Medium* (blog), 17 October 2023, https://cointegrated.medium.com/how-to-fine-tune-a-nllb-200-model-for-translating-a-new-language-a37fc706b865.

Figure 3: Opportunities, Bottlenecks, and Challenges for AI Development in the Global AI Majority

Input	Opportunities	Bottlenecks	Level of Challenge
Data	- Develop local datasets for underrepresented languages. - Implement community-driven data-collection and management initiatives.	- Limited availability of local data. - Privacy concerns and data misuse. - Difficulties in responsible data governance.	High
Skilled Labour	- Boost AI education and training. - Strengthen international collaborations for skill-building. - Support grassroots AI communities.	- Shortage of specialized expertise. - Brain drain due to better opportunities abroad. - Restrictions on knowledge transfer between geopolitical rivals.	Moderate
Compute (Local training and inference)	- Build local compute clusters. - Foster tech sovereignty and local innovation. - Improve infrastructure for future AI development.	- Extremely high costs and maintenance needs. - High energy consumption. - Likely resistance from frontier AI states due to strategic concerns.	High
Compute (Using Compute Hosted Abroad)	- Leverage external cloud computing resources. - Access advanced compute capabilities through global partnerships.	- Still costly, though less than local options. - Dependence on foreign infrastructure reduces autonomy. - Possible restrictions from frontier AI states on compute access.	Moderate

The table above presents a detailed analysis of the opportunities, bottlenecks, and levels of challenge associated with key inputs necessary for AI development within the Global AI Majority. Each category outlines the potential opportunities for enhancing AI capabilities, identifies significant bottlenecks that may impede progress, and assesses the overall level of challenge.

Access to Deployed Models

Access to Open Source Frontier Models

Today, some near state-of-the-art models are open-sourced, meaning they can be used and modified by those with the skills and resources to do so. Significant open source models include Meta's Llama 3[102] and Databricks' DBRX[103] models. As we have discussed, such models, trained and developed in particular contexts, do not account well for many Global AI Majority concerns and cultural spaces. Even if they came to do so, however, the safety concerns of frontier AI states are pushing against open sourcing future generations of models. Commercial incentives of companies also appear to incentivize against open sourcing the most advanced models, though this is an actively evolving issue.[104]

Advanced language models, such as Anthropic's Claude 3 Opus and OpenAI's GPT-4, are not open-sourced, but are available to individual users for a fee. Such services also attempt to regulate the uses to which these systems can be put. In addition, the companies monitor the use of the models post deployment and regulate the *forms* of access that users have. A company may allow forms of fine-tuning access to some users, for instance, or it may not.

Structured Access

These approaches are a form of *structured access* – an emerging paradigm in AI governance that aims to broaden access to advanced AI models in a safe and controlled manner. The goal is to prevent the proliferation of dangerous AI

---

[102] 'Llama 3.1', Meta Llama, n.d., https://www.llama.com/.
[103] The Mosaic Research Team, 'Introducing DBRX: A New State-of-the-Art Open LLM', *Databricks* (blog), 27 March 2024, https://www.databricks.com/blog/introducing-dbrx-new-state-art-open-llm.
[104] Mistral AI's first model open-sourced its model weights, but after a $16 million investment from Microsoft, access to its latest model, Mistral Large, will remain restricted. This is motivated chiefly by commercial interests. Mistral cofounder Arthur Mensch said in an interview with *Le Monde*, 'We started with open source models, which anyone can deploy for free because that's a way of distributing them widely and creating demand. But from the outset, we have provided a business model with [the most powerful] optimised models'. He added that 'commercial activity' will allow the company 'to finance the costly research required for model development'. https://www.euronews.com/next/2024/02/26/microsoft-partners-with-french-ai-start-up-mistral-ai-as-it-unveils-new-ai-tool

capabilities while facilitating access to AI capabilities that can be used safely.[105] In practice, structured access is typically accomplished through the interfaces, such as an application programming interface (API), that provide users with certain forms of access and not others. When the private companies that control such access allow it, these forms of access may enable researchers in Global AI Majority countries to adapt these technologies to local contexts and challenges, fostering the development of AI solutions that are more suited to their specific needs and environments – for example, by fine-tuning LLMs to improve their functioning in local languages.

Many current structured access interfaces do not allow for fine-tuning of the model, however. This raises an important weakness of this approach as a dependable mode of access for the Global AI Majority – model owners control the extent and type of access a user has. This means that when designing an API for structured access, the developer is able to decide what functionality is included and what is not, in effect setting the bounds within which users will operate. This could introduce a power imbalance between the API provider and external user that is not present if a wide array of actors are able to directly download the model weights. The user would also be subject to any terms and conditions set by the developer if they want any access to the model at all. While important for protecting intellectual property, this could result in unnecessarily restricting what users can do with the model. On the other hand, structured access might allow Global AI Majority countries to participate in using models developed elsewhere, even without access to substantial AI-triad resources, when basic digital infrastructure allows.

Structured access to AI models through APIs offers a controlled way for Global AI Majority countries to engage with advanced technologies, albeit with limitations that may not fully address localized needs. This access model, while restrictive, provides a significant step toward bridging technological gaps without requiring extensive local resources. Effective governance will be crucial in ensuring that this access is both safe and fair, paving the way for more inclusive global participation in AI development.

Here again, as in the discussion of access to computing infrastructure above, we see the connection between effective governance and access. As model capabilities and safety risks multiply, this connection is likely to grow even tighter. This has occurred in other fields, with nuclear technology being a primary example. Advanced nuclear

[105] Toby Shevlane, 'Structured Access: An Emerging Paradigm for Safe Ai Deployment' (arXiv, 11 April 2022), https://doi.org/10.48550/arXiv.2201.05159; Benjamin S Bucknall and Robert F. Trager, 'Structured Access for Third-Party Research on Frontier AI Models: Investigating Researchers' Model Access Requirements' (AI Governance Initiative, 2023), https://www.oxfordmartin.ox.ac.uk/publications/structured-access-for-third-party-research-on-frontier-ai-models-investigating-researchers-model-access-requirements.

states facilitated the spread of peaceful nuclear technology through the International Atomic Energy Agency (IAEA) alongside safeguards to ensure that nuclear technologies would be applied peacefully.[106] Similarly, in AI, the expansion of deeper model access such as fine-tuning, integration into other products, or access to model internals, will likely depend on the establishment of governance frameworks that ensure safety.

V. Approaches to Resolving Governance Dilemmas

Throughout this paper, we have identified two primary challenges to achieving an inclusive global AI governance framework: first, the divergence of interests between frontier AI states and the Global AI Majority, and second, the significant concerns regarding the risks associated with unrestricted development and deployment of advanced AI technologies. These challenges highlight a central dilemma in global AI governance: how to create a framework that is truly inclusive – one that provides a meaningful voice to all states and communities affected by AI – while also ensuring the active participation of frontier AI states, whose involvement is essential for effective governance.

To address this complex governance challenge, this paper proposes three key approaches:

1. **Interest Alignment:** Focusing on areas where both frontier AI state and Global AI Majority interests align, thereby building broad governance structures that can be expanded over time.
2. **Participatory Governance Architecture:** Designing inclusive governance structures that incentivize the participation of frontier AI states, while ensuring that all voices are heard.
3. **Safety Assurance:** Establishing safety protocols that mitigate risks without overly restricting access to AI technologies, thus creating a balanced governance framework.

These approaches offer potential pathways to navigate the governance dilemmas posed by divergent interests and safety concerns in the rapidly evolving landscape of AI technology.

---

[106] Christoph Bluth et al., 'Civilian Nuclear Cooperation and the Proliferation of Nuclear Weapons', *International Security* 35, no. 1 (2010): 184–200, https://www.jstor.org/stable/40784651; This literature also highlights the danger that attempts to spread peaceful technology could further weapons programs; Matthew Fuhrmann, *Atomic Assistance: How "Atoms for Peace" Programs Cause Nuclear Insecurity* (Cornell University Press, 2012).

Interest Alignment

The first approach to resolving governance dilemmas is to focus on identifying and leveraging areas where the interests of frontier AI states and the Global AI Majority naturally align. While the diversity of national interests and capabilities in AI development presents challenges, there are numerous domains where global interests converge. By prioritizing these areas of common ground, such as responsible digitization, ethical AI development, and the use of AI to achieve Sustainable Development Goals (SDGs), we can foster broad-based cooperation and build the foundation for inclusive global governance.[107]

Leveraging Shared Interests

Focusing on shared goals such as responsible AI use and sustainable development can facilitate cooperative governance. Frontier AI states and the Global AI Majority share mutual interests in areas like mitigating climate change, improving global health outcomes, and addressing economic inequalities – domains where AI has transformative potential. Collaborative initiatives that leverage AI for these purposes can help bridge gaps between states with advanced AI capabilities and those seeking to develop them, encouraging a more equitable distribution of AI's benefits.[108]

For example, programs like the United Nations' 'AI for Good' initiative demonstrate how AI can be applied to global challenges, fostering a cooperative governance environment that involves all stakeholders, including those from the Global AI Majority.[109] Similarly, the United Nations Development Programme's (UNDP) work on digitizing local knowledge and integrating it into broader development strategies is an example of how interest alignment can be operationalized to build trust and cooperation across diverse groups (United Nations, 2023a).[110]

Institutionalizing Interest Alignment

---

[107] Jonas Tallberg et al., 'The Global Governance of Artificial Intelligence: Next Steps for Empirical and Normative Research', *International Studies Review* 25, no. 3 (2023): viad040, https://doi.org/10.1093/isr/viad040; Nathan Benaich and Ian Hogarth, 'State of AI Report 2022', 11 October 2022, https://www.stateof.ai/2022.

[108] Francesca Mazzi, Mariarosaria Taddeo, and Luciano Floridi, 'AI in Support of the SDGs: Six Recurring Challenges and Related Opportunities Identified Through Use Cases', in *The Ethics of Artificial Intelligence for the Sustainable Development Goals*, ed. Francesca Mazzi and Luciano Floridi (Springer Verlag, 2023), 9–33.

[109] UNESCO, 'Recommendation on the Ethics of Artificial Intelligence', 2021, https://www.unesco.org/en/articles/recommendation-ethics-artificial-intelligence.

[110] 'United Nations Development Programme (UNDP)', AI for Good, n.d., https://aiforgood.itu.int/about-ai-for-good/un-ai-actions/undp/.

To institutionalize interest alignment, governance frameworks must prioritize inclusivity by focusing on areas where both frontier AI states and the Global AI Majority benefit. Recent efforts, such as the UN's AI Advisory Body, have aimed to include diverse geographical and professional backgrounds, ensuring a wide range of perspectives in shaping global AI policies.[111] These initiatives help establish a governance foundation that encourages the participation of all relevant stakeholders and builds the trust necessary for tackling more contentious governance issues over time.

Further, targeted initiatives can promote collaboration on specific projects that align with both global development goals and frontier AI states' interests. For example, creating global partnerships focused on using AI to tackle shared challenges like healthcare, education, and sustainable agriculture can foster cooperation without immediately confronting more sensitive areas of governance, such as military applications or proprietary commercial technologies.[112]

Building on Existing Frameworks

The strength of the interest alignment approach lies in its ability to build on existing governance frameworks, expanding their scope to include more states and stakeholders over time. This gradual approach prevents the pitfalls associated with attempting to cover all aspects of AI governance under a single mechanism, which could lead to failures that might be misinterpreted as evidence against the feasibility of inclusive governance.[113]

Recent UN-led efforts, such as the Global Digital Compact and the UNESCO Recommendation on the Ethics of Artificial Intelligence, illustrate the potential for creating inclusive frameworks that address the concerns of both frontier AI states and the Global AI Majority.[114] These initiatives offer platforms for cooperation on aligned interests and provide avenues for gradually expanding the scope of governance to more complex and contentious areas.

---

[111] United Nations, 'Governing AI for Humanity' (United Nations AI Advisory Body, September 2024), https://www.un.org/en/ai-advisory-body.
[112] Francesca Mazzi and Luciano Floridi, eds., *The Ethics of Artificial Intelligence for the Sustainable Development Goals* (Springer Verlag, 2023); Josh Cowls et al., 'The AI Gambit: Leveraging Artificial Intelligence to Combat Climate Change—Opportunities, Challenges, and Recommendations', *AI & Society* 38, no. 1 (1 February 2023): 283–307, https://doi.org/10.1007/s00146-021-01294-x.
[113] Shevlane, 'Structured Access'; René von Schomberg, 'A Vision of Responsible Research and Innovation', in *Responsible Innovation*, ed. Richard Owen, John Bessant, and Maggy Heintz (John Wiley & Sons, Ltd, 2013), 51–74, https://doi.org/10.1002/9781118551424.ch3.
[114] UNESCO, 'Recommendation on the Ethics of Artificial Intelligence'.

By leveraging these shared interests, a governance framework can progressively expand to address more complex challenges. This incremental strategy builds a strong foundation of mutual understanding and shared responsibility, essential for navigating the evolving landscape of AI governance and ensuring that all voices are heard and represented.

Participatory Governance Architecture

Participatory architecture focuses on creating governance structures that incentivize frontier AI state participation while still giving voice to all affected states and communities. Current governance frameworks, often dominated by frontier AI states, fail to adequately include the voices of the Global AI Majority, thus perpetuating exclusion. To remedy this, governance structures must be designed to balance the participation of both majority and minority stakeholders, recognizing that while frontier AI states play a critical role, their cooperation cannot come at the expense of broader inclusivity.

A key aspect of this approach is the careful design of institutional structures that encourage participation from all relevant actors. Organizations like the United Nations General Assembly (UNGA) offer the most universal governance platform, but their one-country, one-vote principle may deter frontier AI states, which are few in number, from engaging in governance forms where their interests could easily be outvoted. For example, areas such as auditing the procedures of the most advanced AI firms represent perceived divergences of interest. Frontier AI states may resist governance mechanisms that allow other states to access their firms' proprietary processes, even through international monitoring organizations. These differences are likely to become more pronounced when discussions move from abstract principles to the specifics of AI governance. Frontier AI states are likely to prefer forms of oversight that protect their proprietary techniques, while other states have an interest in gaining access to those techniques. One solution could involve designing governance mechanisms that safeguard frontier AI state and non-frontier AI state interests by devolving certain functions to local authorities, as suggested by Trager et al. (2023).[115]

Several strategies can enhance participatory governance to achieve these goals:

1. **Multi-Stakeholder Models:** Adopting governance frameworks that include diverse actors – governments, private entities, civil society, and international

[115] Robert Trager et al., 'International Governance of Civilian AI: A Jurisdictional Certification Approach' (arXiv, 11 September 2023), https://doi.org/10.48550/arXiv.2308.15514.

organizations – ensures that all affected groups have a meaningful role. Such models would remove governance in certain areas from majority votes of members, thus protecting perceived core interests of actors whose participation is critical to the regime's success. This balance is particularly valuable in domains like data governance and ethical AI standards, where both the frontier AI states and the Global AI Majority have vested interests.

2. **Tiered Membership Structures and Weighted Voting Systems:** Establishing tiered or weighted voting mechanisms could provide greater influence to states with advanced AI capabilities, while still ensuring that Global AI Majority countries meaningfully participate. For instance, regional bodies like the African Union or Association of Southeast Asian Nations (ASEAN) could have formal advisory roles within larger governance frameworks, ensuring that their perspectives are incorporated into global AI policymaking. Additionally, this structure could balance power by giving additional influence to more advanced states, while still retaining an inclusive approach where all voices are heard.
3. **Leveraging Existing Minilateral Fora:** Existing minilateral platforms, such as the GPAI, G7, BRICS, G20, and OECD, can serve as stepping stones toward more inclusive governance. These groups facilitate initial negotiations among like-minded nations to form foundational governance frameworks for AI, which can later be expanded to include more states. These foras may prove useful in creating regulatory consistency or interoperability across different markets, easing the compliance burden for AI companies. However, they are club-based cohorts that, in seeking to develop global governance models, often lack equal representation for non-member states. A governance framework developed by the G7 with support from the OECD, for instance, may become a global regime, as the Financial Action Task Force (FATF) did. If this occurs in the context of AI, mechanisms must be developed to allow voices from the Global AI Majority to influence outcomes, perhaps through regional consultations or rotating advisory roles for non-member states. Another approach is to adopt multi-stakeholder models that remove the governance of certain areas from majority votes of members, thereby protecting perceived core interests of actors whose participation is considered essential to the regime's success. Another approach could involve creating tiered membership structures or weighted voting systems that give additional influence to states with more advanced AI capabilities, while still ensuring that all members have a voice.

It is also important to recognize that frontier AI states have an interest in global governance, particularly as technologies diffuse, and yet they also have alternatives:

club models, bilateral fora, and even unilateral actions. These are not always mutually exclusive – often, bilateral and minilateral arrangements, along with influential unilateral actions, coexist with and complement global governance efforts. However, these alternative frameworks must not undermine or replace truly inclusive global governance processes.

Safety Assurance

A crucial aspect of creating an inclusive global AI governance framework is ensuring that safety concerns are adequately addressed without severely restricting access to AI technologies for the Global AI Majority. Safety assurance focuses on developing governance mechanisms that mitigate the risks associated with advanced AI technologies while promoting their responsible use and equitable access. Frontier AI states are more likely to support broader participation in AI governance when a robust international safety regime is in place, including clearly defined safety standards, protocols, and incentives for compliance at the international level.[116]

To balance these safety concerns with the need for inclusive access, several strategies can be employed:

1. **Structured Access to AI Technologies:** A central component of safety assurance is the implementation of structured access mechanisms. This approach involves creating controlled pathways for access to advanced AI technologies, such as through application programming interfaces (APIs) or other regulated access points. Structured access can provide graded or tiered levels of engagement with AI models, based on a country or actor's compliance with international safety standards. For instance, frontier AI states might be more willing to support broader access if they retain control over critical aspects of AI deployment while allowing Global AI Majority countries to leverage AI tools for local development without the need for full access to sensitive technologies.[117] This approach is analogous to export control regimes used in dual-use technologies, like nuclear technology, where access is carefully calibrated to balance security with cooperation.[118]
2. **International Oversight and Certification:** Establishing international bodies dedicated to overseeing the deployment of AI technologies can provide an essential layer of safety assurance. These bodies would be responsible for

[116] Tallberg et al., 'The Global Governance of Artificial Intelligence'; Bucknall and Trager, 'Structured Access for Third-Party Research on Frontier AI Models'.
[117] Bucknall and Trager, 'Structured Access for Third-Party Research on Frontier AI Models'.
[118] Bluth et al., 'Civilian Nuclear Cooperation and the Proliferation of Nuclear Weapons'; Fuhrmann, *Atomic Assistance*.

conducting rigorous safety assessments and certifying compliance with global standards. Drawing from successful models in other high-risk technologies, an AI oversight body could function similarly to harmonize domestic regulatory oversight and ensure that AI systems are developed and deployed in ways that minimize risks and enhance trust among all stakeholders.[119] Recent international convenings, such as the 'AI Safety Summit' and the 'Global Partnership on AI' (GPAI), have highlighted the need for international cooperation on AI-safety standards.[120]

3. **Global Safety Protocols for AI Development:** Developing and enforcing international safety protocols is critical to maintaining a balanced governance framework. Such protocols could include requirements for pre-deployment testing, ongoing monitoring, and robust oversight of AI systems to ensure they meet established safety standards. They should also include measures for evaluating the safety of AI models in non-English languages and local contexts to prevent biased outcomes.[121] Protocols for monitoring the use of large-scale computing resources, implementing checks for compliance with international norms, and requiring cloud providers to report suspicious activities can help prevent the misuse of AI technologies while allowing for responsible development and deployment.[122]
4. **Tiered Safety Frameworks:** To facilitate broader access without compromising safety, a tiered framework for access to AI technologies could be established. Such a framework would allow for graduated levels of access based on demonstrated adherence to safety protocols and responsible use commitments. For example, different tiers could offer varying levels of model access, from basic API calls to more advanced integration capabilities. This would allow countries or organizations that have demonstrated their commitment to safe AI practices to access more powerful AI tools while maintaining safeguards against misuse.[123]
5. **Collaborative Safety Research and Development:** Encouraging collaborative safety research between frontier AI states and Global AI Majority countries can help build a common understanding of risks and develop shared safety

---

[119] Trager et al., 'International Governance of Civilian AI.'
[120] Department for Science, Innovation and Technology, 'AI Safety Summit: Introduction', GOV.UK, 31 October 2023, https://www.gov.uk/government/publications/ai-safety-summit-introduction.; Markus Anderljung et al., 'Towards Publicly Accountable Frontier LLMs: Building an External Scrutiny Ecosystem under the ASPIRE Framework' (arXiv, 15 November 2023), https://doi.org/10.48550/arXiv.2311.14711.
[121] Wang et al., 'All Languages Matter'; Jo and Gebru, 'Lessons from Archives.'
[122] Kulp et al., 'Hardware-Enabled Governance Mechanisms'.
[123] Shevlane, 'Structured Access'; Eoghan Stafford and Robert F. Trager, 'The IAEA Solution: Knowledge Sharing to Prevent Dangerous Technology Races' (Centre for the Governance of AI, 2022), https://www.governance.ai/research-paper/knowledge-sharing-to-prevent-dangerous-technology-races.

standards. This collaboration could focus on areas such as safety frameworks for AI applications in underrepresented languages, culturally specific safety concerns, or context-specific safety alignment challenges, ensuring that safety standards are globally representative.[124]

Safety assurance must be viewed not as a means of exclusion but as a pathway to inclusive governance. By building a governance framework that emphasizes safety, frontier AI states can be reassured that the expansion of AI access will not come at the cost of increased risk. Meanwhile, Global AI Majority countries can gain broader access to AI technologies, provided they adhere to established safety protocols, ensuring that the benefits of AI are more equitably distributed.

VI. Conclusions & Recommendations

The rapid advancement of artificial intelligence presents unprecedented opportunities and challenges for global development and governance. This white paper has examined the critical disparities in AI access and voice between frontier AI states and the Global AI Majority, highlighting the urgent need for more inclusive approaches to AI development and governance.

Our analysis reveals several key findings:

1. Significant barriers exist for Global AI Majority countries in accessing and developing advanced AI capabilities, including limited digital infrastructure, concentrated compute resources, and skewed talent distributions.
2. Current AI governance frameworks largely exclude Global AI Majority voices, despite the technology's potential for widespread global impact.
3. Balancing the interests of frontier AI states with the needs of the Global AI Majority is complex but essential for effective and equitable global AI governance.

To address these challenges, we propose a nuanced approach that balances the needs of the Global AI Majority with the concerns of frontier AI states. Our strategy centres on three key pillars: Interest Alignment, Participatory Architecture, and Safety Assurance.

By focusing first on areas where interests naturally align across states, such as using AI to combat climate change or improve healthcare, we can build a foundation for broader cooperation. This Interest Alignment approach allows us to develop

[124] Bommasani et al., 'On the Opportunities and Risks of Foundation Models'; Okolo, 'AI in the Global South: Opportunities and Challenges Towards More Inclusive Governance'.

governance structures incrementally, starting with less contentious areas and gradually expanding as trust and collaboration grow.

Our Participatory Architecture concept aims to create governance frameworks that incentivize frontier AI state participation while ensuring all voices are heard. This could involve innovative multi-stakeholder models or tiered membership structures that balance influence with inclusivity, ensuring that Global AI Majority countries have a meaningful say in AI governance.

The Safety Assurance pillar addresses the critical need to expand access to AI technologies without compromising security. We envision frameworks that include structured access protocols and rigorous pre-deployment testing, allowing for broader participation in AI development while mitigating potential risks.

We emphasize that exclusion is not a viable long-term strategy. The global nature of AI's potential impacts necessitates truly inclusive governance. Our proposed approaches aim to directly address and help resolve the tension between broadening access and addressing legitimate security concerns.

Near-Term Steps

To address the challenges of AI access and governance for the Global AI Majority, we propose a series of near-term steps that focus on both development initiatives and governance structures. These steps are designed to bridge the access gap and enhance voice in governance, building on existing initiatives and research findings.

Expanding internet connectivity and stable electricity supply particularly in rural and underserved areas, is a critical first step. The sources of the energy should be sustainable to avoid negative externalities and ensure long-term supply. Expanding internet access is also an area with near universal support, offering an ideal starting point for cooperation. This should include investments in fundamental infrastructure such as fibre optic cables, cellular towers, and base stations. As highlighted by the Alliance for Affordable Internet,[125] innovative solutions like satellite internet and community networks through public-private partnerships can extend connectivity to hard-to-reach areas.

Developing robust national data infrastructures in Global AI Majority countries is a critical early step. This includes investing in data localization efforts, cloud computing centres, and training programs for data management personnel. Countries should be encouraged to participate in initiatives like the G20's Data Gaps Initiative 3 (DGI-3)

[125] 'The A4AI Affordability Report', Alliance for Affordable Internet, 10 August 2022, https://a4ai.org/research/affordability-report/.

and the Digital Public Infrastructure (DPI) strategy to foster innovation and promote accessibility.[126] Simultaneously, it is crucial to develop strong data privacy and security regulations to protect individual rights and maintain trust in the AI ecosystem, as emphasized by Jo and Gebru (2020)[127] in their work on responsible data-collection strategies.

Capitalizing on the widespread availability of mobile devices in developing countries is another crucial step. This involves ensuring the affordability of both devices and data plans, as well as encouraging the development of mobile-friendly AI applications that address local challenges. The GSMA's State of Mobile Internet Connectivity Report (2023)[128] underscores the importance of mobile access in bridging the digital divide.

Developing 'homegrown' AI models, through either private investment or public funding or a blend of two, is a next step Global AI Majority countries can take once a sufficient level of infrastructure and talent has been developed. The Nigerian government, for example, has secured private partnerships to boost its national computing capacity and develop AI projects of national interest, including a Nigerian LLM trained on five low-resource languages and accented English to ensure stronger language representation in existing datasets.[129]

Strengthening regional bodies like the African Union and Association of Southeast Asian Nations (ASEAN) to develop coordinated AI strategies and present unified positions in global governance discussions can amplify the voice of Global AI Majority countries. The African Union's upcoming AI strategy and national AI strategies from countries like Rwanda, Nigeria, and South Africa underscore the growing role of regional entities in shaping AI governance.[130]

Increasing access to advanced AI models and compute resources is important to enable domestic AI development. However, it presents significant political feasibility

---

[126] 'G20 Digital Ministers Recognize Digital Public Infrastructure as an Accelerator of the SDGs', UNDP, 19 August 2023, https://www.undp.org/india/press-releases/g20-digital-ministers-recognize-digital-public-infrastructure-accelerator-sdgs; 'The Third Phase of the G20 Data Gaps Initiative (DGI-3) Starts to Deliver Insights for Action', IMF, 8 November 2023, https://www.imf.org/en/News/Articles/2023/11/08/pr23385-the-third-phase-of-the-g20-data-gaps-initiative-dgi-3-starts-deliver-insights-action.

[127] Jo and Gebru, 'Lessons from Archives'.

[128] Matthew Shanahan and Kalvin Bahia, 'The State of Mobile Internet Connectivity Report 2023' (GSMA, October 2023), https://www.gsma.com/r/somic/.

[129] 'Bosun Tijani, 'Nigeria Announces Multillingual LLM, Computing Infrastructure and AI Collective', LinkedIn, 19 April 2024, https://www.linkedin.com/pulse/nigeria-announces-multillingual-llm-computing-ai-dr-bosun-tijani-fy8ve/.

[130] 'Continental Artificial Intelligence Strategy' (African Union, July 2024), https://au.int/en/documents/20240809/continental-artificial-intelligence-strategy.

challenges. Global AI Majority countries' desire for increased access often conflicts with the security and commercial concerns of leading AI nations and companies. To bridge this gap, we propose developing tiered safety frameworks that allow for graduated access based on demonstrated capabilities and adherence to safety protocols. These could include structured access protocols for AI models, mandatory pre-deployment safety evaluations, and monitoring mechanisms for cloud computing resources. By implementing such frameworks, we can mitigate safety concerns without imposing blanket restrictions, potentially making frontier AI states more amenable to broader participation.

References

2Africa Cable. '2Africa Deployment Is Underway'. Accessed 2 May 2024. https://www.2africacable.net.

Abid, Abubakar, Maheen Farooqi, and James Zou. 'Persistent Anti-Muslim Bias in Large Language Models'. In *Proceedings of the 2021 AAAI/ACM Conference on AI, Ethics, and Society*, 298–306. AIES '21, 2021. https://doi.org/10.1145/3461702.3462624.

Abungu, Cecil, Michelle Malonza, and Sumaya Nur Adan. 'Can Apparent Bystanders Distinctively Shape an Outcome? Global South Countries and Global Catastrophic Risk-Focused Governance of Artificial Intelligence'. arXiv, 7 December 2023. https://doi.org/10.48550/arXiv.2312.04616.

AI for Good. 'United Nations Development Programme (UNDP)', n.d. https://aiforgood.itu.int/about-ai-for-good/un-ai-actions/undp/.

Akinade, Idris, Jesujoba Alabi, David Adelani, Clement Odoje, and Dietrich Klakow. 'Varepsilon Kú Mask: Integrating Yorùbá Cultural Greetings into Machine Translation'. In *Proceedings of the First Workshop on Cross-Cultural Considerations in NLP (C3NLP)*, 1–7, 2023. https://doi.org/10.18653/v1/2023.c3nlp-1.1.

Allen, Gregory C. 'Understanding China's AI Strategy'. Center for a New American Security, 6 February 2019. https://www.cnas.org/publications/reports/understanding-chinas-ai-strategy.

Allen, Gregory C., Emily Benson, and Margot Putnam. 'Japan and the Netherlands Announce Plans for New Export Controls on Semiconductor Equipment'. Center for Strategic & International Studies, 10 April 2023. https://www.csis.org/analysis/japan-and-netherlands-announce-plans-new-export-controls-semiconductor-equipment.

Alliance for Affordable Internet. 'The A4AI Affordability Report', 10 August 2022. https://a4ai.org/research/affordability-report/.

Ametepey, Simon Ofori, Clinton Aigbavboa, Wellington Didibhuku Thwala, and Hutton Addy. 'The Impact of AI in Sustainable Development Goal Implementation: A Delphi Study'. *Sustainability* 16, no. 9 (2024): 3858. https://doi.org/10.3390/su16093858.

Amrute, Sareeta, Ranjit Singh, and Rigoberto Lara Guzmán. 'A Primer on AI in/from the Majority World: An Empirical Site and a Standpoint'. Data & Society, 2022. https://doi.org/10.2139/ssrn.4199467.

Anderljung, Markus, Everett Thornton Smith, Joe O'Brien, Lisa Soder, Benjamin Bucknall, Emma Bluemke, Jonas Schuett, Robert Trager, Lacey Strahm, and Rumman Chowdhury. 'Towards Publicly Accountable Frontier LLMs: Building an External Scrutiny Ecosystem under the ASPIRE Framework'. arXiv, 15 November 2023. https://doi.org/10.48550/arXiv.2311.14711.

Aral, Berdal. '"The World Is Bigger than Five": A Salutary Manifesto of Turkey's New International Outlook'. *Insight Turkey*, 2019. https://doi.org/10.25253/99.2019214.05.

Ars Electronica Festival 2023 - Who Owns the Truth? 'Masakhane – Pioneering Participatory Approaches to Building African Language Technologies, for Africans, by Africans', 2 September 2023. https://ars.electronica.art/who-owns-the-truth/en/masakhane/.

Benaich, Nathan, and Ian Hogarth. 'State of AI Report 2022', 11 October 2022. https://www.stateof.ai/2022.

Bender, Emily M., Timnit Gebru, Angelina McMillan-Major, and Shmargaret Shmitchell. 'On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?🦜'. In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, 610–23, 2021. https://doi.org/10.1145/3442188.3445922.

Bergman, A. Stevie, Lisa Anne Hendricks, Maribeth Rauh, Boxi Wu, William Agnew, Markus Kunesch, Isabella Duan, Iason Gabriel, and William Isaac. 'Representation in AI Evaluations'. In *Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency*, 519–33. FAccT '23, 2023. https://doi.org/10.1145/3593013.3594019.

Bindley, Katherine. 'The Fight for AI Talent: Pay Million-Dollar Packages and Buy Whole Teams'. *The Wall Street Journal*, 27 March 2024. https://www.wsj.com/tech/ai/the-fight-for-ai-talent-pay-million-dollar-packages-and-buy-whole-teams-c370de2b.

Blasi, Damián, Antonios Anastasopoulos, and Graham Neubig. 'Systematic Inequalities in Language Technology Performance across the World's Languages'. arXiv, 13 October 2021. https://doi.org/10.48550/arXiv.2110.06733.

Bluth, Christoph, Matthew Kroenig, Rensselaer Lee, William C. Sailor, and Matthew Fuhrmann. 'Civilian Nuclear Cooperation and the Proliferation of Nuclear Weapons'. *International Security* 35, no. 1 (2010): 184–200. https://www.jstor.org/stable/40784651.

Bommasani, Rishi, Drew A. Hudson, Ehsan Adeli, Russ Altman, Simran Arora, Sydney von Arx, Michael S. Bernstein, et al. 'On the Opportunities and Risks of Foundation Models'. arXiv, 12 July 2022. https://doi.org/10.48550/arXiv.2108.07258.

Bommasani, Rishi, and Percy Liang. 'Reflections on Foundation Models'. *Stanford Human-Centered Artificial Intelligence* (blog), 18 October 2021. https://hai.stanford.edu/news/reflections-foundation-models.

Brown, Tom B., Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared Kaplan, Prafulla Dhariwal, Arvind Neelakantan, et al. 'Language Models Are Few-Shot Learners'. arXiv, 22 July 2020. http://arxiv.org/abs/2005.14165.

Buchanan, Ben. 'The AI Triad and What It Means for National Security Strategy'. Center for Security and Emerging Technology, 2020. https://cset.georgetown.edu/publication/the-ai-triad-and-what-it-means-for-national-security-strategy/.

Bucknall, Benjamin S, and Robert F. Trager. 'Structured Access for Third-Party Research on Frontier AI Models: Investigating Researchers' Model Access Requirements'. AI Governance Initiative, 2023. https://www.oxfordmartin.ox.ac.uk/publications/structured-access-for-third-party-research-on-frontier-ai-models-investigating-researchers-model-access-requirements.

Buolamwini, Joy, and Timnit Gebru. 'Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification'. In *Conference on Fairness, Accountability and Transparency*, 77–91. PMLR, 2018. http://proceedings.mlr.press/v81/buolamwini18a.html.

Cai, Kenrick, and Iain Martin. 'Stability AI Founder Emad Mostaque Tanked His Billion-Dollar Startup', 30 March 2024. https://www.forbes.com/sites/kenrickcai/2024/03/29/how-stability-ais-founder-tanked-his-billion-dollar-startup/.

Carrozza, Ilaria, Nicholas Marsh, and Gergory M. Reichberg. 'Dual-Use AI Technology in China, the US and the EU: Strategic Implications for the Balance of Power'. Peace Research Institute Oslo, 2022. https://www.prio.org/publications/13150.

Chan, Alan, Chinasa T. Okolo, Zachary Terner, and Angelina Wang. 'The Limits of Global Inclusion in AI Development'. arXiv, 1 February 2021. https://doi.org/10.48550/arXiv.2102.01265.

Coffey, Donavyn. 'Māori Are Trying to Save Their Language from Big Tech'. *Wired*, 20 April 2021. https://www.wired.com/story/maori-language-tech/.

'Commerce Implements New Export Controls on Advanced Computing and Semiconductor Manufacturing Items to the People's Republic of China (PRC)'. U.S. Department of Commerce, Bureau of Industry and Security, 7 October 2022. https://www.bis.doc.gov/index.php/documents/about-bis/newsroom/press-releases/3158-2022-10-07-bis-press-release-advanced-computing-and-semiconductor-manufacturing-controls-final/file.

'Continental Artificial Intelligence Strategy'. African Union, July 2024. https://au.int/en/documents/20240809/continental-artificial-intelligence-strategy.

Cowls, Josh, Andreas Tsamados, Mariarosaria Taddeo, and Luciano Floridi. 'The AI Gambit: Leveraging Artificial Intelligence to Combat Climate Change—Opportunities, Challenges, and Recommendations'. *AI & Society* 38, no. 1 (1 February 2023): 283–307. https://doi.org/10.1007/s00146-021-01294-x.

Dale, David. ‘How to Fine-Tune a NLLB-200 Model for Translating a New Language’. *Medium* (blog), 17 October 2023. https://cointegrated.medium.com/how-to-fine-tune-a-nllb-200-model-for-translating-a-new-language-a37fc706b865.

Databricks. ‘The Data and AI Company’, 13 October 2023. https://www.databricks.com/.

Dawson, Alice, and James Ball. ‘Generating Democracy AI and the Coming Revolution in Political Communications’. Demos, 2024. https://demos.co.uk/wp-content/uploads/2024/01/Generating-Democracy-Report-1.pdf.

Deci. ‘Top Large Language Models Reshaping the Open-Source Arena’, 27 March 2024. https://web.archive.org/web/20240609234857/https://deci.ai/blog/list-of-large-language-models-in-open-source/.

Dentons. ‘US Data Transfers - Update’, 28 October 2022. https://www.dentons.com/en/insights/articles/2022/october/28/us-data-transfers.

Department for Science, Innovation and Technology. ‘AI Safety Summit: Introduction’. GOV.UK, 31 October 2023. https://www.gov.uk/government/publications/ai-safety-summit-introduction.

Drexel, Bill, and Hannah Kelley. ‘Behind China’s Plans to Build AI for the World’. Politico, 30 November 2023. https://www.politico.com/news/magazine/2023/11/30/china-global-ai-plans-00129160.

Eberhard, David M., Gary F. Simons, and Charles D. Fennig, eds. *Ethnologue: Languages of the World*. 25th ed. SIL International, 2022. https://www.ethnologue.com/.

‘Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence’. The White House, 30 October 2023. https://www.whitehouse.gov/briefing-room/presidential-actions/2023/10/30/executive-order-on-the-safe-secure-and-trustworthy-development-and-use-of-artificial-intelligence/.

Fan, Angela, Shruti Bhosale, Holger Schwenk, Zhiyi Ma, Ahmed El-Kishky, Siddharth Goyal, Mandeep Baines, et al. ‘Beyond English-Centric Multilingual Machine Translation’. arXiv, 21 October 2020. https://doi.org/10.48550/arXiv.2010.11125.

Foreign, Commonwealth & Development Office. ‘UK Unites with Global Partners to Accelerate Development Using AI’. GOV.UK, 1 November 2023. https://www.gov.uk/government/news/uk-unites-with-global-partners-to-accelerate-development-using-ai.

Fuhrmann, Matthew. *Atomic Assistance: How “Atoms for Peace” Programs Cause Nuclear Insecurity*. Cornell University Press, 2012.

‘G7 Hiroshima Process on Generative Artificial Intelligence (AI)’. OECD, 6 September 2023. https://www.oecd.org/en/publications/2023/09/g7-hiroshima-process-on-generative-artificial-intelligence-ai_8d19e746.html.

Ghaffary, Shirin, and Rachel Metz. ‘Microsoft to Pay Inflection AI $650 Million After Scooping Up Most of Staff’. Bloomberg, 21 March 2024. https://www.bloomberg.com/news/articles/2024-03-21/microsoft-to-pay-inflection-ai-650-million-after-scooping-up-most-of-staff.

‘Ghana Natural Language Processing (NLP)’, n.d. https://ghananlp.org/.

Gozalo-Brizuela, Roberto, and Eduardo C. Garrido-Merchan. ‘ChatGPT Is Not All You Need. A State of the Art Review of Large Generative AI Models’. arXiv, 11 January 2023. https://doi.org/10.48550/arXiv.2301.04655.

Gupta, Kirti, Chris Borges, and Andrea Leonard Palazzi. ‘Collateral Damage: The Domestic Impact of U.S. Semiconductor Export Controls’. Center for Strategic & International Studies, 9 July 2024. https://www.csis.org/analysis/collateral-damage-domestic-impact-us-semiconductor-export-controls.

Hammond, Samuel. ‘The Scramble for AI Computing Power’. *American Affairs* 8, no. 2 (2024). https://americanaffairsjournal.org/2024/05/the-scramble-for-ai-computing-power/.

Hao, Karen. ‘A New Vision of Artificial Intelligence for the People’. MIT Technology Review, 22 April 2022. https://www.technologyreview.com/2022/04/22/1050394/artificial-intelligence-for-the-people/.

Henshall, Will. ‘E.U.’s AI Regulation Could Be Softened After Pushback’. *TIME*, 22 November 2023. https://time.com/6338602/eu-ai-regulation-foundation-models/.

Huang, Ken. ‘AI Safety vs. AI Security: Navigating the Differences’. *Cloud Security Alliance* (blog), 19 March 2024. https://cloudsecurityalliance.org/blog/2024/03/19/ai-safety-vs-ai-security-navigating-the-commonality-and-differences.

IBM. ‘What Is Fine-Tuning?’, 15 March 2024. https://www.ibm.com/topics/fine-tuning.

IMF. ‘The Third Phase of the G20 Data Gaps Initiative (DGI-3) Starts to Deliver Insights for Action’, 8 November 2023. https://www.imf.org/en/News/Articles/2023/11/08/pr23385-the-third-phase-of-the-g20-data-gaps-initiative-dgi-3-starts-deliver-insights-action.

Information Commissioner’s Office. ‘International Data Transfers’. ICO, 19 October 2023. https://ico.org.uk/for-organisations/data-protection-and-the-eu/data-protection-and-the-eu-in-detail/the-uk-gdpr/international-data-transfers/.

‘Introducing the AI Safety Institute’. AI Safety Institute, 2023. https://www.gov.uk/government/publications/ai-safety-institute-overview/introducing-the-ai-safety-institute.

ITB, Webmaster Team, Direktorat Sistem dan Teknologi Informasi. ‘Pusat Artificial Intelligence’. Institut Teknologi Bandung, n.d. https://www.itb.ac.id/pusat-artificial-intelligence.

ITU. ‘2023 AI for Good Global Summit Snapshot Report’, 2023. https://s41721.pcdn.co/wp-content/uploads/2021/06/SNAPSHOT-REPORT-2023-FINAL.pdf.

———. ‘Measuring Digital Development: Facts and Figures 2023’, 2023. https://www.itu.int/dms_pub/itu-d/opb/ind/d-ind-ict_mdd-2023-1-pdf-e.pdf.

Jacobs, Julian. ‘National Approaches to AI Safety Diverge in Focus’. *OMFIF* (blog), 25 June 2024. https://www.omfif.org/2024/06/national-approaches-to-ai-safety-diverge-in-focus/.

Ji, Ziwei, Nayeon Lee, Rita Frieske, Tiezheng Yu, Dan Su, Yan Xu, Etsuko Ishii, Ye Jin Bang, Andrea Madotto, and Pascale Fung. ‘Survey of Hallucination in Natural Language Generation’. *ACM Computing Surveys* 55, no. 12 (2023): 248:1-248:38. https://doi.org/10.1145/3571730.

Jo, Eun Seo, and Timnit Gebru. ‘Lessons from Archives: Strategies for Collecting Sociocultural Data in Machine Learning’. In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, 306–16. FAT* ’20, 2020. https://doi.org/10.1145/3351095.3372829.

Kulp, Gabriel, Daniel Gonzales, Everett Smith, Lennart Heim, Prateek Puri, Michael J. D. Vermeer, and Zev Winkelman. ‘Hardware-Enabled Governance Mechanisms: Developing Technical Solutions to Exempt Items Otherwise Classified Under Export Control Classification Numbers 3A090 and 4A090’. RAND Corporation, 18 January 2024. https://www.rand.org/pubs/working_papers/WRA3056-1.html.

Lehdonvirta, Vili, Boxi Wu, and Zoe Hawkins. ‘Cloud Empires’ Physical Footprint: How Trade and Security Politics Shape the Global Expansion of U.S. and Chinese Data Centre Infrastructures’. SSRN, 20 December 2023. https://doi.org/10.2139/ssrn.4670764.

———. ‘Compute North vs. Compute South: The Uneven Possibilities of Compute-Based AI Governance Around the Globe’. In *Proceedings of the 2024 AAAI/ACM Conference on AI, Ethics, and Society (AIES ’24) (Forthcoming)*, 2024. https://osf.io/8yp7z/download.

Li, Chuan. ‘OpenAI’s GPT-3 Language Model: A Technical Overview’. *Lambda* (blog), 3 June 2020. https://lambdalabs.com/blog/demystifying-gpt-3.

MacroPolo. ‘The Global AI Talent Tracker 2.0’. Accessed 11 May 2024. https://macropolo.org/digital-projects/the-global-ai-talent-tracker/.

‘Masakhane’, n.d. https://www.masakhane.io/.

Maslej, Nestor, Loredana Fattorini, Erik Brynjolfsson, John Etchemendy, Katrina Ligett, Terah Lyons, James Manyika, et al. ‘Artificial Intelligence Index Report 2023’. Institute for Human Centered AI, 2023. https://aiindex.stanford.edu/wp-content/uploads/2023/04/HAI_AI-Index-Report_2023.pdf.

Mazzi, Francesca, and Luciano Floridi, eds. *The Ethics of Artificial Intelligence for the Sustainable Development Goals*. Springer Verlag, 2023.

Mazzi, Francesca, Mariarosaria Taddeo, and Luciano Floridi. ‘AI in Support of the SDGs: Six Recurring Challenges and Related Opportunities Identified Through Use Cases’. In *The Ethics of Artificial Intelligence for the Sustainable Development Goals*, edited by Francesca Mazzi and Luciano Floridi, 9–33. Springer Verlag, 2023.

Mell, Peter, and Tim Grance. ‘The NIST Definition of Cloud Computing’. National Institute of Standards and Technology, 28 September 2011. https://doi.org/10.6028/NIST.SP.800-145.

Meta Llama. ‘Llama 3.1’, n.d. https://www.llama.com/.

Mohamed, Shakir, Marie-Therese Png, and William Isaac. ‘Decolonial AI: Decolonial Theory as Sociotechnical Foresight in Artificial Intelligence’. *Philosophy & Technology* 33, no. 4 (1 December 2020): 659–84. https://doi.org/10.1007/s13347-020-00405-8.

Monserrate, Steven Gonzalez. ‘The Cloud Is Material: On the Environmental Impacts of Computation and Data Storage’. *MIT Case Studies in Social and Ethical Responsibilities of Computing* Winter 2022 (2022). https://doi.org/10.21428/2c646de5.031d4553.

Murgia, Madhumita. ‘AI’s New Workforce: The Data-Labelling Industry Spreads Globally’. *Financial Times*, 24 July 2019. https://www.ft.com/content/56dde36c-aa40-11e9-984c-fac8325aaa04.

Nations, United. ‘Governing AI for Humanity’. United Nations AI Advisory Body, September 2024. https://www.un.org/en/ai-advisory-body.

NLLB Team, Marta R. Costa-jussà, James Cross, Onur Çelebi, Maha Elbayad, Kenneth Heafield, Kevin Heffernan, et al. ‘No Language Left Behind: Scaling Human-Centered Machine Translation’. arXiv, 25 August 2022. https://doi.org/10.48550/arXiv.2207.04672.

OECD. ‘OECD’s Live Repository of AI Strategies & Policies’. Accessed 16 May 2024. https://oecd.ai/en/dashboards.

Ògúnmọ́dẹdé, Chris Olaoluwa. ‘The G-7’s Embrace of the Global South Was All Talk, No Substance’. *World Politics Review* (blog), 26 May 2023. https://www.worldpoliticsreview.com/g7-summit-2023-politics-economy-africa-global-south/.

Okolo, Chinasa T. ‘AI in the Global South: Opportunities and Challenges Towards More Inclusive Governance’. *International Journal of Innovative Research in Arts, Education and Technology* 2, no. 1 (2021). https://internationalpolicybrief.org/wp-content/uploads/2024/02/ARTICLE-9.pdf.

Onuah, Felix. ‘Google to Train 20,000 Nigerians in Digital Skills’. *Reuters*, 15 August 2023. https://www.reuters.com/world/africa/google-train-20000-nigerians-digital-skills-2023-08-15/.

Png, Marie-Therese. ‘At the Tensions of South and North: Critical Roles of Global South Stakeholders in AI Governance’. In *Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency*, 1434–45. FAccT ’22, 2022. https://doi.org/10.1145/3531146.3533200.

Presidency of the Republic of Türkiye. ‘“Our Motto ‘the World Is Bigger than Five’ Is the Biggest-Ever Rise against Global Injustice”’, 1 October 2018. https://www.tccb.gov.tr/en/news/542/89052/our-motto-the-world-is-bigger-than-five-is-the-biggest-ever-rise-against-global-injustice.

Qi, Xiangyu, Yangsibo Huang, Yi Zeng, Edoardo Debenedetti, Jonas Geiping, Luxi He, Kaixuan Huang, et al. ‘AI Risk Management Should Incorporate Both Safety and Security’. arXiv, 29 May 2024. https://doi.org/10.48550/arXiv.2405.19524.

Rauh, Maribeth, John Mellor, Jonathan Uesato, Po-Sen Huang, Johannes Welbl, Laura Weidinger, Sumanth Dathathri, et al. ‘Characteristics of Harmful Text: Towards Rigorous Benchmarking of Language Models’. arXiv, 28 October 2022. https://doi.org/10.48550/arXiv.2206.08325.

Ricaurte, Paola. ‘Ethics for the Majority World: AI and the Question of Violence at Scale’. *Media, Culture & Society* 44, no. 4 (1 May 2022): 726–45. https://doi.org/10.1177/01634437221099612.

Rowe, Niamh. ‘“It’s Destroyed Me Completely”: Kenyan Moderators Decry Toll of Training of AI Models’. *The Guardian*, 2 August 2023. https://www.theguardian.com/technology/2023/aug/02/ai-chatbot-training-human-toll-content-moderator-meta-openai.

Sastry, Girish, Lennart Heim, Haydn Belfield, Markus Anderljung, Miles Brundage, Julian Hazell, Cullen O’Keefe, et al. ‘Computing Power and the Governance of Artificial Intelligence’. arXiv, 13 February 2024. https://doi.org/10.48550/arXiv.2402.08797.

Scharre, Paul. ‘Future-Proofing Frontier AI Regulation: Projecting Future Compute for Frontier AI Models’. Center for a New American Security, March 2024. https://www.cnas.org/publications/reports/future-proofing-frontier-ai-regulation.

Schmid, Stefka, Thea Riebe, and Christian Reuter. ‘Dual-Use and Trustworthy? A Mixed Methods Analysis of AI Diffusion Between Civilian and Defense R&D’. *Science and Engineering Ethics* 28, no. 2 (2022): 12. https://doi.org/10.1007/s11948-022-00364-7.

Schomberg, René von. ‘A Vision of Responsible Research and Innovation’. In *Responsible Innovation*, edited by Richard Owen, John Bessant, and Maggy Heintz, 51–74. John Wiley & Sons, Ltd, 2013. https://doi.org/10.1002/9781118551424.ch3.

Seger, Elizabeth, Noemi Dreksler, Richard Moulange, Emily Dardaman, Jonas Schuett, K. Wei, Christoph Winter, et al. ‘Open-Sourcing Highly Capable Foundation Models: An Evaluation of Risks, Benefits, and Alternative Methods for Pursuing Open-Source Objectives’. Centre for the Governance of AI, 2023. https://cdn.governance.ai/Open-Sourcing_Highly_Capable_Foundation_Models_2023_GovAI.pdf.

Sevilla, Jaime, and Edu Roldán. ‘Training Compute of Frontier AI Models Grows by 4-5x per Year’. Epoch AI, 28 May 2024. https://epochai.org/blog/training-compute-of-frontier-ai-models-grows-by-4-5x-per-year.

Shanahan, Matthew, and Kalvin Bahia. ‘The State of Mobile Internet Connectivity Report 2023’. GSMA, October 2023. https://www.gsma.com/r/somic/.

Sheng, Emily, Kai-Wei Chang, Premkumar Natarajan, and Nanyun Peng. ‘The Woman Worked as a Babysitter: On Biases in Language Generation’. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, 3407–12, 2019. https://doi.org/10.18653/v1/D19-1339.

Shevlane, Toby. ‘Structured Access: An Emerging Paradigm for Safe Ai Deployment’. arXiv, 11 April 2022. https://doi.org/10.48550/arXiv.2201.05159.

Smith, Adam. *The Wealth of Nations*, 1776.

Stafford, Eoghan, and Robert F. Trager. ‘The IAEA Solution: Knowledge Sharing to Prevent Dangerous Technology Races’. Centre for the Governance of AI, 2022. https://www.governance.ai/research-paper/knowledge-sharing-to-prevent-dangerous-technology-races.

Stankey, Robert, K.C. Halm, Michael T. Borgia, Andrew M. Lewis, and Assaf Ariely. ‘Commerce Department Proposes Cybersecurity/AI Reporting and “KYC” Requirements for Certain Cloud Providers’. *Davis Wright Tremaine* (blog), 14 February 2024.

https://www.dwt.com/blogs/artificial-intelligence-law-advisor/2024/02/commerce-department-proposes-kyc-ai-rules-for-iaas.

Statista. 'Most Used Languages Online by Share of Websites 2024', January 2024. https://www.statista.com/statistics/262946/most-common-languages-on-the-internet/.

Strier, Keith, Jack Clark, and Sana Khareghani. 'Measuring Compute Capacity: A Critical Step to Capturing AI's Full Economic Potential'. *OECD.AI Policy Observatory* (blog), 8 February 2022. https://oecd.ai/en/wonk/ai-compute-capacity.

Tallberg, Jonas, Eva Erman, Markus Furendal, Johannes Geith, Mark Klamberg, and Magnus Lundgren. 'The Global Governance of Artificial Intelligence: Next Steps for Empirical and Normative Research'. *International Studies Review* 25, no. 3 (2023): viad040. https://doi.org/10.1093/isr/viad040.

The Economist. 'Meet the French Startup Hoping to Take on OpenAI'. 26 February 2024. https://www.economist.com/business/2024/02/26/meet-the-french-startup-hoping-to-take-on-openai.

The Mosaic Research Team. 'Introducing DBRX: A New State-of-the-Art Open LLM'. *Databricks* (blog), 27 March 2024. https://www.databricks.com/blog/introducing-dbrx-new-state-art-open-llm.

The University of Padjadjaran. 'Pusat Riset Kecerdasan Artifisial dan Big Data', n.d. https://informatika.unpad.ac.id/pusat-riset-kecerdasan-artifisial-dan-big-data/.

Tijani, 'Bosun. 'Nigeria Announces Multillingual LLM, Computing Infrastructure and AI Collective'. LinkedIn, 19 April 2024. https://www.linkedin.com/pulse/nigeria-announces-multillingual-llm-computing-ai-dr-bosun-tijani-fy8ve/.

'Tokopedia-UI AI Center', n.d. https://tokopedia-ai.cs.ui.ac.id/.

Trager, Robert, Ben Harack, Anka Reuel, Allison Carnegie, Lennart Heim, Lewis Ho, Sarah Kreps, et al. 'International Governance of Civilian AI: A Jurisdictional Certification Approach'. arXiv, 11 September 2023. https://doi.org/10.48550/arXiv.2308.15514.

Tsanni, Abdullahi. 'This Company Is Building AI for African Languages'. MIT Technology Review, 17 November 2023. https://www.technologyreview.com/2023/11/17/1083637/lelapa-ai-african-languages-vulavula/.

UNDP. 'G20 Digital Ministers Recognize Digital Public Infrastructure as an Accelerator of the SDGs', 19 August 2023. https://www.undp.org/india/press-releases/g20-digital-ministers-recognize-digital-public-infrastructure-accelerator-sdgs.

UNESCO. 'Recommendation on the Ethics of Artificial Intelligence', 2021. https://www.unesco.org/en/articles/recommendation-ethics-artificial-intelligence.

United Nations. ‘Our Common Agenda: Report of the Secretary-General’, 2021. https://www.un.org/en/content/common-agenda-report/.

Valdivia, Ana. ‘Silicon Valley and the Environmental Costs of AI’. *Political Economy Research Centre* (blog), 5 December 2022. https://www.perc.org.uk/project_posts/silicon-valley-and-the-environmental-costs-of-ai/.

Vasquez, Christian. ‘DHS Warns of Malicious AI Use Against Critical Infrastructure’. *CyberScoop* (blog), 14 September 2023. https://cyberscoop.com/dhs-homeland-threat-assessment/.

Vaynman, Jane, and Tristan A. Volpe. ‘Dual Use Deception: How Technology Shapes Cooperation in International Relations’. *International Organization* 77, no. 3 (2023): 599–632. https://doi.org/10.1017/S0020818323000140.

Vipra, Jai, and Sarah Myers West. ‘Computational Power and AI’. AI Now Institute, 27 September 2023. https://ainowinstitute.org/publication/policy/compute-and-ai.

Wang, Changhan, Morgane Rivière, Ann Lee, Anne Wu, Chaitanya Talnikar, Daniel Haziza, Mary Williamson, Juan Pino, and Emmanuel Dupoux. ‘VoxPopuli: A Large-Scale Multilingual Speech Corpus for Representation Learning, Semi-Supervised Learning and Interpretation’. arXiv, 27 July 2021. https://doi.org/10.48550/arXiv.2101.00390.

Wang, Wenxuan, Zhaopeng Tu, Chang Chen, Youliang Yuan, Jen-tse Huang, Wenxiang Jiao, and Michael R. Lyu. ‘All Languages Matter: On the Multilingual Safety of Large Language Models’. arXiv, 20 June 2024. https://doi.org/10.48550/arXiv.2310.00905.

Wheeler, Tom. ‘The Three Challenges of AI Regulation’. Brookings, 15 June 2023. https://www.brookings.edu/articles/the-three-challenges-of-ai-regulation/.

Zhao, Jingchen, and Beatriz Gómez Fariñas. ‘Artificial Intelligence and Sustainable Decisions’. *European Business Organization Law Review* 24, no. 1 (2023): 1–39. https://doi.org/10.1007/s40804-022-00262-2.

Zheng, Jiawei, Hanghai Hong, Xiaoli Wang, Jingsong Su, Yonggui Liang, and Shikai Wu. ‘Fine-Tuning Large Language Models for Domain-Specific Machine Translation’. arXiv, 22 February 2024. https://doi.org/10.48550/arXiv.2402.15061.

Zwetsloot, Remco, Roxanne Heston, and Zachary Arnold. ‘Strengthening the U.S. AI Workforce’. Center for Security and Emerging Technology, 2019. https://doi.org/10.51593/20190003.