

Abungu, Cecil; Iradukunda, Marie Victoire; Beggs, Duncan Cass; Hassan, Aquila; Sayidali, Raqda

Working Paper

Why Global South countries need to care about highly capable AI

CIGI Papers, No. 311

Provided in Cooperation with:

Centre for International Governance Innovation (CIGI), Waterloo, Ontario

Suggested Citation: Abungu, Cecil; Iradukunda, Marie Victoire; Beggs, Duncan Cass; Hassan, Aquila; Sayidali, Raqda (2024) : Why Global South countries need to care about highly capable AI, CIGI Papers, No. 311, Centre for International Governance Innovation (CIGI), Waterloo, ON, Canada

This Version is available at:

<https://hdl.handle.net/10419/311790>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

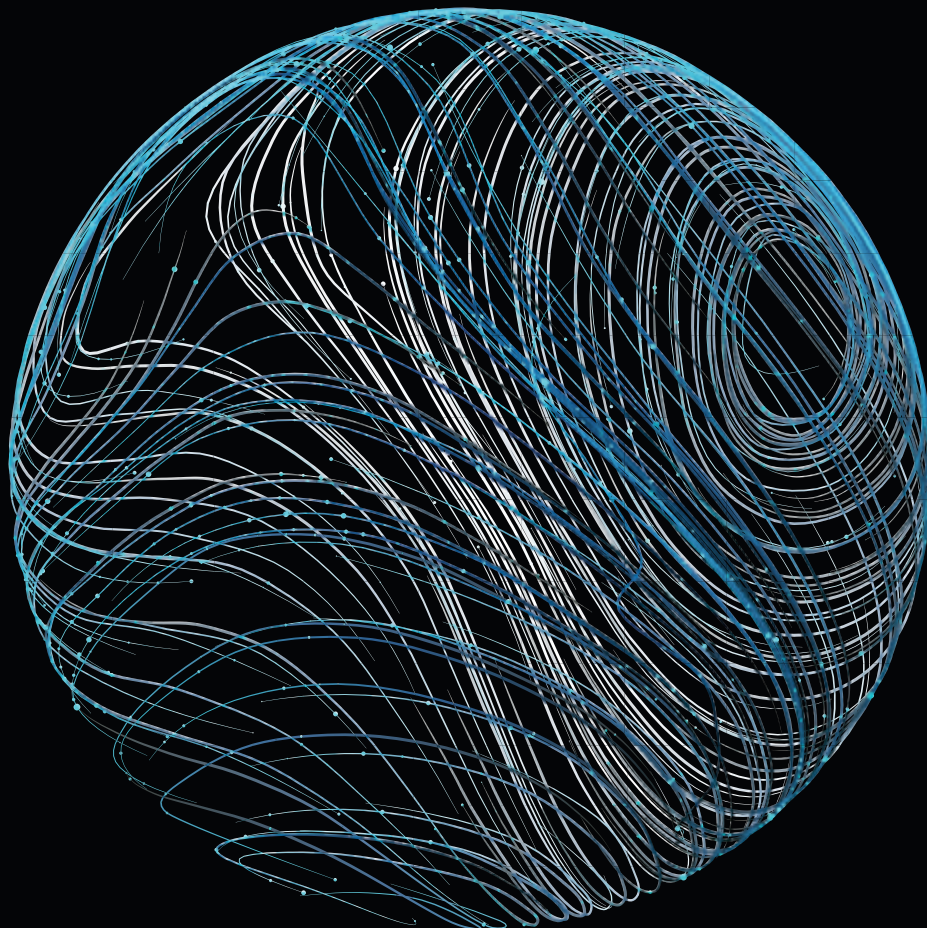


<https://creativecommons.org/licenses/by/4.0/>

CIGI Papers No. 311 – December 2024

Why Global South Countries Need to Care About Highly Capable AI

Cecil Abungu, Marie Victoire Iradukunda,
Duncan Cass-Beggs, Aquila Hassan and Raqda Sayidali



CIGI Papers No. 311 – December 2024

Why Global South Countries Need to Care About Highly Capable AI

Cecil Abungu, Marie Victoire Iradukunda,
Duncan Cass-Beggs, Aquila Hassan and Raqda Sayidali

About CIGI

The Centre for International Governance Innovation (CIGI) is an independent, non-partisan think tank whose peer-reviewed research and trusted analysis influence policy makers to innovate. Our global network of multidisciplinary researchers and strategic partnerships provide policy solutions for the digital era with one goal: to improve people's lives everywhere. Headquartered in Waterloo, Canada, CIGI has received support from the Government of Canada, the Government of Ontario and founder Jim Balsillie.

À propos du CIGI

Le Centre pour l'innovation dans la gouvernance internationale (CIGI) est un groupe de réflexion indépendant et non partisan dont les recherches évaluées par des pairs et les analyses fiables incitent les décideurs à innover. Grâce à son réseau mondial de chercheurs pluridisciplinaires et de partenariats stratégiques, le CIGI offre des solutions politiques adaptées à l'ère numérique dans le seul but d'améliorer la vie des gens du monde entier. Le CIGI, dont le siège se trouve à Waterloo, au Canada, bénéficie du soutien du gouvernement du Canada, du gouvernement de l'Ontario et de son fondateur, Jim Balsillie.

Credits

Executive Director, Global AI Risks Initiative **Duncan Cass-Beggs**
Senior Research Associate and Program Manager **Matthew da Mota**
Publications Editor **Lynn Schellenberg**
Publications Editor **Susan Bubak**
Graphic Designer **Sami Chouhdary**

Copyright © 2024 by the Centre for International Governance Innovation

The opinions expressed in this publication are those of the authors and do not necessarily reflect the views of the Centre for International Governance Innovation or its Board of Directors.

For publications enquiries, please contact publications@cigionline.org.



The text of this work is licensed under CC BY 4.0. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

For reuse or distribution, please include this copyright notice. This work may contain content (including but not limited to graphics, charts and photographs) used or reproduced under licence or with permission from third parties. Permission to reproduce this content must be obtained from third parties directly.

Centre for International Governance Innovation and CIGI are registered trademarks.

67 Erb Street West
Waterloo, ON, Canada N2L 6C2
www.cigionline.org

Table of Contents

vi	About the Authors
vi	Acronyms and Abbreviations
1	Executive Summary
1	Introduction
2	The Contours of Highly Capable AI
3	Why Global South Countries Should Care About Highly Capable AI
11	Concluding Recommendations
13	Works Cited

About the Authors

Cecil Abungu (lead author) is a Ph.D. student at the University of Cambridge and a research affiliate with the Centre for the Study of Existential Risk (University of Cambridge) and the Institute for Law and AI. He is also the coordinator of the ILINA Program. His research interests lie in artificial intelligence (AI) safety governance, algorithmic discrimination, intellectual history, legal theory and constitutional law.

Marie Victoire Iradukunda (lead author) is a master of laws student at Harvard Law School and a policy fellow with the Harvard AI Safety Student Team. Her research interests lie in AI governance, particularly the legal duties of AI developers and deployers, and the regulation of AI use by governments in Africa. She has previously served as a research fellow at the ILINA Program.

Duncan Cass-Beggs is executive director of the Global AI Risks Initiative at CIGI, focusing on developing innovative governance solutions to address current and future global issues relating to AI. Duncan has more than 25 years of experience working on domestic and international public policy issues, most recently as head of strategic foresight at the Organisation for Economic Co-operation and Development.

Aquila Hassan is a project management specialist at the Centre for the Governance of AI in Oxford, England. She has a background in engineering and an interest in research around governing AI for the global majority.

Raqda Sayidali is an undergraduate student at Strathmore University and a researcher at the ILINA Program. Her research interests include AI safety governance and the legal challenges surrounding emerging technologies.

Acronyms and Abbreviations

4G	fourth-generation
5G	fifth-generation
AGI	artificial general intelligence
AI	artificial intelligence
compute	computational power
ICRC	International Committee of the Red Cross
IDF	Israel Defense Forces
LAWS	lethal autonomous weapons systems
LLMs	large language models
RLAIF	reinforcement learning from AI feedback
RLHF	reinforcement learning from human feedback
UAVs	unmanned aerial vehicles

Executive Summary

By matching and surpassing human cognitive abilities, highly capable artificial intelligence (AI) — advanced AI systems of the foreseeable future, which leading AI companies are working toward as part of their broader goal to create artificial general intelligence (AGI) — could be one of the most transformative technologies the world has ever seen. While this radical technology is being built primarily in Global North countries, its impacts are likely to be felt worldwide, and disproportionately so in those Global South countries with long-standing vulnerabilities — weak state institutions; dependence on labour-intensive, manufacturing-based and export-led economic models; regularly recurring armed conflict; high trust in technology; and more globally subordinated cultures.

The authors of this paper consider six ways in which highly capable AI could interact with these vulnerabilities. First, highly capable AI could leave Global South peoples facing a much harsher economic reality. Second, highly capable AI could lead to far more damaging armed conflict in Global South countries. Next, highly capable AI could enable repressive and enduring authoritarianism in Global South countries and could expose Global South peoples to unparalleled manipulation. In addition to that, highly capable AI could also deepen the cultural subordination of Global South peoples. Finally, AI developers and researchers have yet to devise a foolproof way of ensuring that the most advanced models always take actions in line with positive human values. Unless this problem is remedied before the emergence of highly capable AI, there is a chance such AI could lead to catastrophic outcomes including significant loss of life, possibly up to human extinction.

Because of the significant societal impacts that highly capable AI could have, being concerned about AI in a general way will not suffice. The authors argue that all stakeholders who care about those who live in Global South countries must pull on the levers available to them with the goal of influencing the ongoing development of highly capable AI.

Introduction

We live in an era of rapid technological innovation and reach in which the tools and software being built can have a worldwide impact. Historically, milestones such as the telegraph (1837), the telephone (1876) and radio (late 1890s) emerged at intervals of several decades. However, the pace of technological advancement has accelerated significantly. The past 25 years have witnessed a succession of technological leaps (Roser 2023), including fourth- and fifth-generation (4G and 5G) mobile networks, increasingly powerful smartphones and the Internet of Things. With few exceptions, the software and devices built as a result of such breakthroughs have found their way to people in every corner of the world. As an example, when Myanmar re-entered global life in 2011 following decades of isolation under military rule, less than one percent of its population had access to the internet. By 2020, this statistic had surged to roughly 44 percent.¹ Myanmar essentially experienced a “leapfrog effect,” moving from minimal internet access to widespread smartphone and social media use within a decade.

These fast-breaking and far-reaching developments in technology have the capacity to transform societies in unexpected ways, particularly when deployed in diverse cultural contexts. In Myanmar, the rapid spread of the internet had a range of complex consequences, both beneficial and detrimental. On the upside, it provided access to information (Phyo 2023) and new economic opportunities (Jørgensen 2019, 48–50). On the downside, it led to the expansion of platforms that were then used to spread hate speech and disinformation, exacerbating ethnic tensions and violence (Klark and Sagun 2023).

AI systems are expected to have an even more profound global impact. Experts anticipate that AI will cause dramatic changes across almost every domain of life, including human rights, knowledge, well-being and governance (Anderson and Rainie 2023). As with previous technological advancements, AI is bound to affect the Global South significantly, especially because the most capable AI systems are being developed in Western countries, where governance frameworks are

¹ See Jørgensen (2019) and <https://data.worldbank.org/indicator/IT.NET.USER.ZS?locations=MM>.

shaped and designed with Western contexts and realities in mind (Ayana et al. 2024, 3).

There are more than 130 countries in the region known as the “Global South,” each with a diverse society. However, Global South countries² also share some unique realities, vulnerabilities and challenges. Most of them are faced with similar economic realities, such as severely limited access to capital and investment, high levels of poverty and substantial income inequality (Ayub 2013). Additionally, they share socio-political challenges, including recurring violent conflict (Palik, Obermeier and Rustad 2022), weak state institutions (Choi 2023, 5–7) and lower levels of educated populations.³ Due to the complexity of these issues and the enormous demands of potential solutions, Global South countries have grappled with these problems for decades, and it is reasonable to assume will do so for decades to come. These challenges consume much of the focus and resources of Global South countries, making it difficult for them to address potential risks posed by emerging technologies such as AI.

Nonetheless, in this paper, the authors argue that Global South countries must give serious attention to the development of highly capable AI, given the significant context-specific risks it will pose in their societies. This argument is founded on expert predictions regarding the powerful capabilities of highly capable AI systems, and the authors’ analysis of how highly capable AI systems are likely to engage with the ongoing realities and specific vulnerabilities of these countries.

In the section following this introduction, “highly capable AI” is defined by canvassing technical predictions about the capabilities that the most powerful AI systems will have in the foreseeable future. The third section then outlines the impacts such AI systems could have in Global South countries. In particular, the paper argues that the development and use of highly capable AI could mean that Global South peoples will face a much harsher economic reality; experience extremely damaging armed conflict; experience repressive and enduring authoritarianism; live through unparalleled manipulation; and find themselves living in a world where their cultures

are very deeply subordinated. The fourth and final section of this paper contains a short discussion on the way forward and forms the conclusion.

There is a fair chance that *current* advanced AI systems could, realistically, give rise to some variation of the risks discussed in this paper. However, this study will show that these risks will be greater and far more concerning with future highly capable AI systems. The authors hope that the research presented here will prompt all stakeholders concerned about Global South peoples to clarify and sharpen their positions on the development of highly capable AI.

The Contours of Highly Capable AI

In this paper, “highly capable AI” refers to AI systems that demonstrate cognitive capabilities, enabling them to perform economically valuable tasks at or above the level of human beings. Leading AI companies such as Google DeepMind, Meta and OpenAI are working toward building these advanced AI systems as part of their broader goal to create AGI.⁴ To achieve this goal, they are investing heavily in acquiring data sets, computer chips and data centres (Murgia 2023; Bengio et al. 2024, 843; Gardizy and Efrati 2024).

Current advanced AI systems already exhibit impressive capabilities, including what appears to be common-sense reasoning, cause and effect reasoning, step-by-step reasoning and in-context learning (Privitera et al. 2024, 21). They have also demonstrated improved performance on tasks such as image generation and recognition, video generation, and language-based tasks such as text generation, and in fields such as coding, mathematics and biology (Ngo 2023; Bubeck et al. 2023, chapter 2). These impressive capabilities are generally attributed to deep learning (Karnofsky 2016; Piper 2020; Ngo, Chan and Mindermann 2024, 1), and the scaling of model sizes, data sets and computation power (or, compute) used in training (Kaplan et al. 2020, 3; Harris, Harris and

2 Throughout this paper, the term “Global South countries” will not refer to China, Hong Kong, Macau and Singapore.

3 See Ritchie et al. (2023) and www.worlddeconomics.com/Indicator-Data/ESG/Social/Mean-Years-of-Schooling/.

4 See Murgia (2023); Heath (2024); Patel (2023); <https://openai.com/charters>.

Beall 2023, 8–10). Research shows that neural networks trained on large data sets and compute not only achieve better performance at the tasks they are trained to undertake, but also become more generalizable, robust and capable (Harris, Harris and Beall 2023, 10; Branwen 2022).

Given the rapid progress currently observed when scaling occurs, a growing number of experts and researchers believe that further scaling of AI inputs will lead to the development of even more advanced AI systems (Department for Science, Innovation & Technology 2023, 10–15; Harris, Harris and Beall 2023, 10–11). In fact, many leading AI companies are building their future models based on this hypothesis (Privitera et al. 2024, 11; Bengio et al. 2024, 843). Experts anticipate that with continued scaling, the leading AI models that we are likely to see in five to 10 years — herein referred to as “highly capable AI” — will outperform humans in a variety of economically useful tasks (Harris, Harris and Beall 2023, 5; Whittlestone and Clarke 2024, 52–53), ranging from scientific research to engineering (Dafoe 2018, 18).

This level of performance is possible through the development of capabilities such as general problem solving (Sotala and Yampolskiy 2015, 1), continual learning, context-specific decision making, flexible memory, reading comprehension, analogical reasoning (Cremer and Whittlestone 2021, 107) and critical reasoning (Gillani 2023). Unlike other general-purpose technologies such as electricity, highly capable AI will likely also be able to autonomously perform new tasks by generalizing from previous experience (Ngo 2020, 3).

Furthermore, highly capable AI is expected to be highly autonomous. Unlike previous transformative technologies, or “narrow AI” or “weak AI” (which can perform a single function under various constraints and limitations), highly capable AI is expected to be able to take up long-term goals and execute them with little to no human intervention. This suggests that such AI is likely to be capable of agentic planning, meaning that it can consider various courses of action, including long-term strategies (Chan et al. 2023, 4; Carlsmith 2024, 8), and select what it deems to be the most effective plan based on predicted outcomes. It is also expected that highly capable AI will be adaptable enough to adjust its course of action if circumstances change (Ngo 2020, 10–12; Kenton et al. 2022, 1).

In fact, current AI agents already show some of this autonomy (Shavit et al. 2023; Ayres and Balkin 2024, 1). When provided with goals, they can independently devise plans and pursue them to a limited extent. Moreover, AI agents have demonstrated the ability to interact with the physical world through other software tools. For example, one AI agent, given the goal of ordering a pizza, successfully utilized a text-to-voice converter to place an order at a local pizzeria (Zittrain 2024).

With the capabilities and autonomy discussed, highly capable AI could have a transformative impact comparable to earlier general-purpose technologies such as the steam engine, combustion engine and electricity (*Tech Wire Asia* 2018; Willing 2023). Consequently, highly capable AI is likely to be an inescapable technology embraced by most private and state actors worldwide in their pursuit of greater efficiency, accuracy and productivity.

Why Global South Countries Should Care About Highly Capable AI

Highly Capable AI Could Create a Much Harsher Economic Reality for Global South People

The development and widespread use of AI is anticipated to have a significant economic impact. Several studies suggest that it will exacerbate global imbalance in economic power (Privitera et al. 2024, 56, 57; Korinek and Stiglitz 2021, 2) by widening the existing economic gap within and between different nations (Georgieva 2024). Specifically, the disparity between nations will leave developed countries richer and developing countries comparatively poorer (Cummings et al. 2018, 39), and this could occur in various ways.

To begin with, research suggests that the use of AI is likely to erode the comparative advantage of Global South countries, leading to the deterioration of their terms of trade (Korinek and Stiglitz 2021, 2). Historically, Global South countries have built their economic advantage on low-cost labour and abundant natural resources (*ibid.*, 3), relying on

a traditional development model that is labour-intensive, manufacturing-based and export-led (ibid., 23). However, AI seems to be on the verge of facilitating the automation of labour, which could decrease demand for the relatively cheaper labour that Global South countries offer. Initially, AI seems likely to augment or replace the skilled labour force that mainly undertakes cognitive tasks (Cazzaniga et al. 2024, 3–4, 6). One prime category of workers at high risk of being replaced is those who do online work such as information technology services and telemarketing, and much of this kind of work has been offshored to Global South countries (*The Economist* 2023; Mandavia 2024; Putzier 2023). As AI continues to be used in such sectors, Global North companies that have been offshoring that kind of work to Global South countries might opt to reshore and automate it (Korinek and Stiglitz 2021, 9–10, 18–24; Cummings et al. 2018, 38; Mandavia 2024).

Some leading economists also argue that the increasingly promising integration of advanced AI into robots (Criddle, Murgia and Hammond 2024) is also likely to result in the displacement of low-skill manual labour (Korinek and Juelfs 2022, 7–8). This development could, in turn, lead to further reshoring of opportunities and deindustrialization of Global South countries (Bell and Korinek 2023, 157). If this occurs, Global South countries are likely to lose their comparative advantage and face a deterioration in export income (Korinek and Stiglitz 2021, 10). Although Global North countries could also experience such impacts, Global South countries are in a riskier situation because labour has long been their comparative advantage in the global economy (ibid., 18).

Furthermore, the development of advanced AI models is primarily confined to Global North countries and China, making it likely that these countries will reap most of the benefits (Privitera et al. 2024, 57–58; Korinek and Stiglitz 2021, 14–15). Due to the existing AI divide and certain favourable conditions within their jurisdictions, developed countries are better positioned to capture the benefits of AI. Additionally, AI is bound to improve productivity gains (Szczepański 2019), and Global North countries are better equipped to leverage these improvements due to their superior infrastructure and skilled talent.

Highly capable AI could exacerbate this gap even further and lead to extreme global inequality in the long term (Dafoe 2018, 10). Because it will

be capable of performing all the cognitive tasks that a human can, highly capable AI is likely to increase productivity gains to an unprecedented scale (Bell and Korinek 2023, 156). However, due to the AI divide, which is expected to grow at the same pace as capabilities (Privitera et al. 2024, 57), these gains are likely to be captured by Global North countries. Consequently, the economic gap could widen further (Bell and Korinek 2023, 157).

Global South countries should be concerned about this risk because an expanding economic gap could leave them with radically non-competitive economies. Many of these countries already struggle with weak economies when compared to the high-income countries of the Global North (Todaro and Smith 2020, 35). Falling predominantly in the categories of low-income, lower-middle-income and upper-middle-income countries, they have lower purchasing power parity than their Global North counterparts (ibid., 37–38, 43–44). As a consequence of their weak economies, these countries are burdened by high levels of poverty (Omeje 2018, 1; Todaro and Smith 2020, 57) and a lagging progress in nutrition, health and education (Todaro and Smith 2020, 55–57). By making Global South countries less economically competitive, the development and the use of highly capable AI have the potential to intensify such challenges.

Finally, with limited bargaining power, Global South countries have become economically, environmentally and politically reliant on developed nations (ibid., 65; Armstrong 1981, 401–2; Onyekwena and Ekeruche 2019). A further decline in their economic capacity would exacerbate these dependencies, stifling not only their economic development but also critical aspects of development across multiple sectors.

Highly Capable AI Could Enable Far More Damaging Armed Conflict

For many years, armed conflict has been a major challenge in many Global South countries, playing out in the form of state-actor and non-state-actor conflicts (Caparini and Davis 2022). Over the past 40 years, Africa, Asia and the Middle East have consistently recorded the highest number of state-actor conflicts, with Africa peaking at around 31 such conflicts in 2020 alone (Rustad 2024, 14–15). These regions have also borne witness to endless non-state-actor conflicts in decades past. For

example, there were more than 50 non-state-actor conflicts in Africa in 2017 alone (ibid., 17, 19). In general, countries in Africa, Asia and the Middle East have consistently experienced the highest numbers of incidents (World Economic Forum 2024, 25).

Highly capable AI poses a significant risk of exacerbating these conflicts by amplifying their scale, frequency and severity. Experts have already examined how AI could be integrated into armed conflict, noting its potential to enhance it via:

- weapons systems, making them more autonomous and superior (Hendrycks, Mazeika and Woodside 2023, 13; Maas, Lucero-Matteucci and Cooke 2023, 247);
- military hardware, such as robotic systems for surveillance, flight and navigation, enabling the deployment of unmanned aircrafts, vehicles and sea vessels;
- cyber and information warfare, transforming cyber offence and defence in terms of capabilities and autonomy; and
- military decision making, including target selection, timing and attack strategies, as well as the analysis of large volumes of data for effective planning and logistics (International Committee of the Red Cross [ICRC] 2019, 2–5).

In fact, military AI has already influenced armed conflict in several of these ways. Lethal autonomous weapons systems (LAWS) have already been deployed by various militaries worldwide (Maas, Lucero-Matteucci and Cooke 2023, 247). For instance, reports indicate that the Israel Defense Forces (IDF) has been using an AI targeting platform known as “the Gospel,” which aids them in identifying and attacking targets (Feldstein 2024). Reports also indicate that drone swarms were used by the IDF in Gaza in 2021, and that UAVs (unmanned aerial vehicles) were used by Russia in its 2022 invasion of Ukraine (Maas, Lucero-Matteucci and Cooke 2023, 247). Notably, both state and non-state actors are increasingly adopting AI in armed conflict, a trend that is likely to persist (ICRC 2019, 2; Brundage et al. 2018, 38).

The integration of AI in armed conflict has sparked myriad concerns, one of the most significant being the increased probability of war. The use of autonomous weapons, for example, eliminates some of the traditional barriers to war: it substantially reduces the human cost of

war, which includes soldier lives, and, in turn, lessens the political repercussions of war, such as negative public opinion and the potential for military leaders to be held accountable (Hendrycks, Mazeika and Woodside 2023, 14–16).

What is more, AI has the potential to escalate the impact of armed conflict, primarily by increasing the number of actors involved. The widespread availability of AI systems and AI-enabled weapons makes it easier for more actors to engage in attacks (Brundage et al. 2018, 18). This diffusion, coupled with the fact that AI lowers the barrier of entry by providing low-skill actors with high-skill capabilities such as autonomous aiming, means that a greater number of people are able to carry out attacks (ibid., 27).

AI can also escalate this impact by increasing the rate, scale and severity of physical attacks. This escalation is driven by several factors. First, the increased autonomy of weapon systems and military hardware, such as drones, can enable large-scale attacks by small groups or even individuals (ibid., 27–28, 40). Second, autonomous weapons and robots can operate for longer durations and overcome human limitations such as exposure to toxic substances, smoke and challenging terrain, among others (ibid., 40). Third, by increasing psychological distance between actors and their targets, as well as by offering anonymity, AI might encourage more frequent attacks, as it will be difficult to trace these actions back to perpetrators (ibid., 19). Lastly, if automated decision-making systems make errors, they could cause a series of quick automated attacks and counterattacks that could cause significant harm to civilians, further amplifying the destructive consequences of armed conflict (Hendrycks, Mazeika and Woodside 2023, 15; ICRC 2019, 5).

As with its application to physical attacks, AI may increase the success rate, scale, speed and ease of cyberattacks through the automation of vulnerability discovery and operation of many systems in parallel (Hendrycks, Mazeika and Woodside 2023, 14; Brundage et al. 2018, 33–34; Privitera et al. 2024, 44). These attacks can then be used to target and destroy critical infrastructure, making the effects of conflict extremely severe (Hendrycks, Mazeika and Woodside 2023, 14). Moreover, the ability of AI to make cyberattacks stealthier and harder to attribute to actors could lead to a rise in attacks,

including against innocent third parties (ibid.; Maas, Lucero-Matteucci and Cooke 2023, 264).

The development and the use of highly capable AI are likely to lead to a massive escalation of these risks. The role of AI in armed conflict is likely to evolve significantly, with AI-enabled weapons and hardware achieving greater capabilities and autonomy. There are likely to be novel and vastly superior technologies, such as fully autonomous weapons, including drone swarms (Verbruggen 2021), and autonomous vehicles, such as submarines, resulting in more severe consequences within shorter time frames (Feldstein 2024; Hendrycks, Mazeika and Woodside 2023, 13). For instance, fully autonomous drones could be deployed in larger numbers simultaneously (Feldstein 2024), magnifying their impact within a shorter period.

Highly capable AI might also introduce entirely new threats. For instance, it could facilitate the development of novel bioweapons (Hendrycks, Mazeika and Woodside 2023, 6, 7) by leveraging its training in biological research, such as pathogen studies, and expertise in bioengineering or biotechnology (ibid., 7; Egan and Rosenbach 2023). In future, such AI may provide individuals who otherwise lack the necessary knowledge or expertise with step-by-step directions on designing, constructing and testing bioweapons with custom features that enhance their deadliness, transmissibility and resistance to treatment (Hendrycks, Mazeika and Woodside 2023, 7; O'Brien and Nelson 2020, 222, 223). Both state and non-state actors could seek to create and use these bioweapons in armed conflict (Roffey 2004, 557), potentially leading to severe consequences, including numerous fatalities (Juling 2023, 127–28).

The impact of highly capable AI on armed conflict will be particularly pronounced in Global South countries, which have a history of alarming levels of armed conflict. There is little doubt that AI-enabled capabilities and the associated military “advantages” are likely to be very appealing to both state and non-state actors in the Global South, as evident from ongoing high-level discussions on the use of AI by the militaries of some of these countries.⁵ Indeed, AI-enabled technologies are already being adopted in warfare by both state and non-state actors across these regions. For example, some African

countries have started incorporating AI in drones and surveillance technologies to combat extremist groups and insurgencies (Allen and Okpali 2022).

Some have argued that state and non-state actors in Global South countries will have limited access to highly capable AI weapon systems due to the lack of the requisite knowledge and computing resources (Maas, Lucero-Matteucci and Cooke 2023, 245). However, there are alternative avenues through which these actors in Global South countries could still obtain such technologies. One approach could involve acquiring these AI systems from other countries or developers. Foreign powers frequently intervene in conflicts where they have strategic interests, often by supplying weapons and military hardware to various factions. For example, Iran, Russia and the United Arab Emirates have been accused of supplying drones and weapons to the warring factions in Sudan (Taleb 2024; Campbell-James 2024). Similarly, countries such as Rwanda and China have reportedly been involved in the Democratic Republic of Congo’s internal conflict, with Rwanda supplying troops to the M23 rebels, and China supplying weapons and drones to the Congolese government (Human Rights Watch 2023; Center for Preventive Action 2024). In the future, states with access to highly capable AI weapon systems could follow this pattern and supply them to groups involved in armed conflict in Global South countries.

As well, actors in Global South countries could access such weapons through proliferation of commercial systems that are adapted for military purposes or illegal smuggling networks, something already seen in the realm of AI-enabled drones (Allen 2021). Furthermore, while the initial costs of highly capable AI may be high, these costs are expected to decrease over time, as they did with previous technologies, such as drones. As costs lower, access may become more feasible for actors in the Global South (Hörster 2021, 11; Allen and Chan 2017, 16; Maas, Lucero-Matteucci and Cooke 2023, 249). Therefore, the likelihood of far more intensive armed conflict is the kind of serious risk that stakeholders in Global South countries should be concerned about.

5 See <https://mod.go.ke/news/responsible-ai-in-the-military-domain/>.

Highly Capable AI Could Enable More Repressive and Enduring Authoritarianism

There are strong signals indicating that AI could be used to consolidate state power and facilitate authoritarianism (Hendrycks, Mazeika and Woodside 2023, 10). Indeed, there is already evidence that AI is currently being leveraged for surveillance, censorship and other traditional tactics of repression. In countries such as China and Kazakhstan, for instance, AI is enabling surveillance through automation of monitoring and tracking functions (Feldstein 2021, 219; Kendall-Taylor, Frantz and Wright 2020). Even the US government is reportedly using hundreds of AI-powered surveillance towers on the country's border with Mexico (Beaumont 2022). The draw is obvious. AI can analyze massive amounts of data and information quickly, cost-effectively and with minimal intrusion, enhancing any surveillance efforts (Kendall-Taylor, Frantz and Wright 2020). Consequently, AI-enabled surveillance has grown more prevalent, with many countries now utilizing facial recognition, social media monitoring and other smart-city technologies (Feldstein 2019, 7–10). AI-enabled censorship has also surged, with machine-learning models being used to quickly sift through vast amounts of data to block content that is deemed unfavourable to certain regimes (Frantz, Kendall-Taylor and Wright 2020; Hörster 2021, 5).

The path to these outcomes is likely to be eased significantly by highly capable AI, laying the foundations for more robust authoritarianism (Dafoe 2018, 7; Bengio 2023, 115). There are a few ways this could play out. First, highly capable AI might enable super-surveillance by making surveillance cheaper, more extensive and effective (Dafoe 2018, 36). Repressive governments could deploy autonomous AI agents that can operate indefinitely in a “set it and forget it” manner (Zittrain 2024), to continuously monitor and track dissenting individuals and groups, or even scour online spaces for opposing views. From there, governments could use these autonomous AI agents to quickly and continuously block information perceived as unfavourable to the regime, exacerbating censorship (Dafoe 2018, 36).

Second, these regimes could use highly capable AI to target dissenting individuals and groups. Just as AI today facilitates targeted advertisements (B. Collier 2024), it is conceivable that rogue governments might use autonomous AI agents

to continuously send targeted intimidating and threatening messages to dissenters. Additionally, governments could leverage highly capable AI to discredit political opponents through the generation of negative deepfakes that are highly realistic (Frantz, Kendall-Taylor and Wright 2020).

Finally, governments could crudely deploy highly capable AI-enabled weapons against dissenting citizens (Hendrycks, Mazeika and Woodside 2023, 10; Dafoe 2018, 36). Given the “advantages” that such weapons can offer, such as very high operational speeds and anonymity, autocrats may turn to these to carry out violent crackdowns and to disrupt mass protests. Autocrats may also prefer to use AI rather than humans to carry out their repression, because such AI systems, if controllable (a challenge discussed later in this paper), could be less likely to disobey orders out of conscience or overthrow the autocrats themselves (Dafoe 2018, 36). Ultimately, highly capable AI is likely to enable governments to lock in their power and make them more durable (Hendrycks, Mazeika and Woodside 2023, 10), leading to greater violations of human rights, and further despotism (Schmidt 2023).

Global South countries are particularly vulnerable to this risk for two reasons. To begin with, many of these countries have fairly recent histories of being governed by extremely repressive regimes. Most of them have legacies of dictatorships and autocratic rule, marked by assaults on political opponents, journalists, activists and minority groups (Matlosa 2023, 341–42; Omeje 2018, 1). For instance, during the 1960s, many Latin American countries, including Argentina, Brazil, Chile, Cuba and Uruguay, experienced the rise of dictatorial military regimes. By 1977, only four Latin American nations remained free from dictatorship (Mora and Fonseca 2019; Casals 2024). Similarly, between 1960 and 1980, a wave of one-party and military regimes swept through numerous African countries. By the end of this era, about 25 of the 45 countries had endured either one-party or military rule, both of which were characterized by extensive limitations of human rights (R. B. Collier 1982, 95–97; Jenkins and Kposowa 1990, 861). This trend was also evident in countries across Asia, the Middle East and North Africa, including Bangladesh, Egypt, Libya, Pakistan and Vietnam. For instance, more than 90 percent of the countries in the Middle East have frequently been categorized as either partially free or not free (Lehmann 1985, 591; Frantz 2012, 16).

Despite the wave of democratization experienced by Global South countries since the 1970s (Heller 2022, 464), repression has persisted or re-emerged in many of them. Civil liberties and freedoms, such as freedom of expression, peaceful assembly and association, are often under threat. Many of these countries continue to subject their citizens to arbitrary arrests and detentions, extrajudicial killings, enforced disappearances and repression of minorities (Amnesty International 2024, 24–68). Even the countries with more democratic histories have experienced some kind of democratic backsliding in recent decades (Freedom House 2024, 1). History has shown us that when individuals in power have acted a certain way for a long time, it is infinitely likely that those who follow them will feel more emboldened to act the same way. In this case, highly capable AI could give state officials in Global South countries the tools they need to govern in a deeply authoritarian manner.

Global South peoples have limited safeguards against AI-enabled autocracy since many Global South countries have weak institutional foundations (Omeje 2018, 1; Tan, Azali and Francisco 2021, 15; Korinek and Stiglitz 2021, 17). For example, while privacy protections are one essential way to guard against surveillance, a good number of these countries still lack data protection laws, and many countries that do have these laws in place continue to struggle with implementation (Hiebert 2023; Tan, Azali and Francisco 2021, 15). As such, Global South countries may lack efficient regulatory schemes to guard against AI-enabled surveillance.

Additionally, civil society, which can play a crucial role in protecting people against such threats, has historically been weak in many of these countries. While civil society organizations in Global South countries have sometimes done work that has led to an increase in government accountability, they are also frequently underfunded or co-opted by the government (Arriola, Rakner and van de Walle 2023, 10, 11; Matlosa 2023, 341; Yom 2015). There is a significant risk that such civil society organizations in Global South countries may be unable to constitute an effective bulwark against AI-enabled autocracy.

Finally, many of these countries have accountability institutions and regulatory bodies that are plagued by deeply embedded cultures of corruption. Consider one important example. Courts, which would typically serve as an avenue for citizens to challenge authoritarian use of highly capable

AI (Dafoe 2018, 51), have repeatedly proven to be corrupt and unduly influenced in Global South countries (Transparency International 2024, 14–18). This creates a hostile climate for citizens seeking to challenge AI-enabled surveillance and censorship. Altogether, this means that Global South peoples may find themselves with limited options to challenge any authoritarian use of highly capable AI.

This worry is further amplified by evidence that a large number of Global South countries are already using AI in the surveillance, censorship and suppression of dissent. Alarm has already been raised about the use of AI-powered surveillance to monitor and suppress journalists, politicians, activists and business leaders in African countries such as Uganda and Zimbabwe (Timcke and Hlomani 2024, 2, 24), as well as in various Latin American (Derechos Digitales, n.d.; Giri 2023), Asian (Funk, Shahbaz and Vesteinsson 2023, 8–15), and Middle Eastern and North African countries (Kausch et al. 2022, 37–40). This bleak picture could be significantly worsened by the arrival of highly capable AI.

Highly Capable AI Could Enable Unparalleled Manipulation

Manipulation via AI usually features AI systems communicating outputs in a way that is likely to convince a human to reinforce or change their beliefs, behaviour or preferences (Gabriel et al. 2024, 82). By leveraging cognitive biases or misrepresenting information, the architect degrades cognitive autonomy and thereby corrodes the integrity of a human decision-making process (El-Sayed et al. 2024, 6–7). Manipulation by AI is particularly concerning because of how effectively AI could be able to deceive. For instance, AI models could be designed to implicitly coax or reinforce extremist views among people (Gabriel et al. 2024, 88) to a degree that culminates in security threats.

Manipulation by AI could be the result of intention by system developers, misspecified functions, persuasion that builds trust in users or design meant to engage the user (ibid.). It may also be that even without any explicit intention by humans, the AI system exhibits manipulative behaviour. We are already seeing some signs that large language models (LLMs) can be manipulative. For instance, LLMs trained on internet content seem to learn how to behave persuasively and manipulatively (Carroll et al. 2023, 1, 7). The authors

of this paper foresee highly capable AI as likely to be even more able to manipulate users.

One key enabler will be highly capable AI's increased access to context and more personal information of users (Gabriel et al. 2024, 27). Research in persona-based approaches to conversational AI shows that personal information helps the system understand the context of conversation better, making these programs interact more seamlessly and engage in more human dialogue (Liu, Symons and Vatsavai 2022). These capabilities will be scaled up considerably in highly capable AI, making it exceedingly hard for ordinary human users to know whether the AI system they are using is manipulating them.

The other enabler will be the abilities that highly capable AI will have to reason, plan and respond to multimodal commands at an inference speed that makes human-machine interactions seem more natural. This would increase the likelihood of anthropomorphism, something that spurs trust and allows manipulation (Gabriel et al. 2024, 96, 102). Taken together, these features will create AI systems that offer an immersive experience and create the illusion of trust by making users feel like they are interacting with a friend or a confidant (ibid., 112). Users in such situations will therefore be more likely to follow any suggestions or directives that the AI system in question proposes. There are already signs that such a world is not far off. Consider, for example, Google's Gemini,⁶ which Google says seeks to create a more immersive user experience through agents that can "see and hear what we do, better understand the context we're in and respond quickly in conversation, making the pace and quality of interactions feel more natural" (cited in Heikkilä 2024).

Studies have found that citizens of developing countries are significantly more optimistic and trusting about the impact that AI will have on their lives (Ipsos 2022). In a supplemental report for its 2024 trust barometer, the Edelman Trust Institute (2024, 3-7) reported that trust in technology is higher in developing countries, while it has deteriorated in the United States and the United Kingdom. Another global study found that people in Brazil, China, India and South Africa showed higher levels of trust in AI compared to people in Global North countries (Gillespie

et al. 2023, 5). The study particularly noted that people in these four countries trust in the ability, humanity and integrity of AI systems (ibid., 18).

The trust AI enjoys in Global South countries seems to be a result of its perceived benefits. Even in cases where participants are fearful about some risks, their excitement, optimism and trust appear to be undiminished (ibid., 24). In India, for instance, technology in general is seen as a solution to many developmental issues and is often trusted as an "authority" in many situations (Kapania et al. 2022). Similarly, in Brazil the level of trust in AI is significantly higher than the level of understanding there is about it (Gillespie et al. 2023, 56). This situation is likely to be replicated in many other Global South countries.

Apart from levels of trust in technology, Global South peoples also have lower levels of education in comparison to the rest of the world. For instance, in 2020 the adult literacy rate in Sub-Saharan Africa was only at 66 percent while the global average was 89 percent.⁷ Studies on misinformation susceptibility have shown that education is a key factor in determining how much someone is susceptible to believing false information. For example, one study identified an inverse statistical correlation between higher levels of education and belief in scientific misinformation (Siani and Green 2023, 8). The same could largely be true for manipulation as well, as education is important in developing the ability to detect and reflect on nuances across judgment domains (Knuuttila, Neudert and Howard 2022).

Highly Capable AI Could Deepen the Subordination of Global South Peoples' Cultures

Through different projects and approaches, fields such as decolonial studies have shown us that cultural hegemonies can be mapped. In other words, there are identifiable patterns that show the world view dominance of more historically powerful and privileged groups of people over less historically powerful and privileged groups of people (Cortes-Ramirez 2015, 117). This reality has knowingly and unknowingly been brought to bear through coercion or persuasion, often facilitated by technological devices or software. For example, in her book *America, as Seen on TV: How Television*

⁶ Gemini is a "family" of highly capable multimodal models whose quality increases with model size (Gemini Team 2023).

⁷ See <https://royalafricansafaris.com/foundation/adult-education-in-olpalagilagi/>.

Shapes Immigrant Expectations Around the Globe, sociologist Clara Rodríguez shows how American television has been able to project US-centric views on social relationships onto the rest of the world (Stoelker 2018). As this paper's authors argue in the sections that follow, the special capabilities that highly capable AI will possess could lead to a much more deeply entrenched hegemony of Western cultures over Global South peoples' cultures.

Several researchers have already expressed concerns about how current state-of-the-art generative AI models can exacerbate the existing dominance of Western cultures. According to some of this research, the models' explicit and implicit steering in favour of Western cultures is mostly a result of how they were trained. Consider, for instance, some examples from LLMs, the most cutting-edge AI systems available today.

LLMs are trained on data scraped from the internet, which over-represents some parts of the world. Accordingly, the language that is represented enhances cultural alignment where the language in question is prevalent. It is no surprise that some studies have also shown that LLMs' responses to cognitive psychological tasks most closely resemble those of people from Western, educated, industrialized, rich and democratic societies. Indeed, Rohin Manvi et al. (2024, 1) have recently unveiled research that shows that, due to their training corpora, LLMs are "clearly biased against locations with lower socio-economic conditions (e.g., most of Africa)."

Furthermore, the most advanced AI alignment methods — reinforcement learning from human feedback (RLHF) and reinforcement learning from AI feedback (RLAIF) — produce AI systems that are wedded to Western cultures. RLHF involves human judgment in creating *preferences* in model behaviour, while RLAIF, also called "constitutional AI," depends on a human-produced set of principles or "constitution" (Conitzer et al. 2024, 1). This means that both methods carry the biases inherent in humans as per their cultural backgrounds (Varshney 2024, 10). For instance, a study examining the cultural biases of LLMs found that the stark differences — a move toward far more secular responses, for example — between the cultural values observable in Open AI's GPT-3 model and those observable in its immediate successor (GPT-3.5 Turbo) can be attributed to the use of RLHF in training GPT-3.5 Turbo (Tao et al. 2024, 3).

Unless there is significant change, the arrival of highly capable AI could worsen the situation for Global South peoples' cultures. It is likely that just as with existing cutting-edge AI models, highly capable AI will mostly be trained on data sets that have Western cultural biases and norms embedded into them. If RLHF and RLAIF remain the best-performing "alignment" methods available to leading AI developers, we can imagine that both will continue to prop up Western cultures (as explained in the preceding section). And, as many researchers have suggested, highly capable AI could be used everywhere — and to do all sorts of things — in the future. Because of the accuracy and efficiency benefits these AI systems could present, businesses and governments will scramble to integrate them into their work in a way that they have not yet done with existing AI systems. In this event, Global South peoples might be forced to live with AI systems that fundamentally carry Western cultural norms and biases.

The cultural biases and norms that will be embedded into highly capable AI systems could be at odds with the cultures of Global South users, leading to misinterpretation and misrepresentation of certain cultures, developments that could in turn create cultural barriers, impose a cultural hegemony and possibly even result in cultural erasure (Prabhakaran, Qadri and Hutchinson 2022, 2). On top of that, by representing certain cultural values as norms through stereotyped responses, AI systems can further entrench these values in users who will then hold them as their own (Anwar et al. 2024, 80–81).

Although some may suggest that fine-tuning these AI systems on more localized data could create more "local" models, that would still do little to alter the fundamental philosophical pillars upon which the systems have been built. In other words, if a model has been trained to identify a certain response as wrong or bad, it is exceedingly hard to change that. Others may suggest that the fact that AI could be used to protect languages that are at risk of disappearance shows it might not lead to the subordination of Global South peoples' cultures (Onome 2024). However, this view conflates the existence of aspects of a culture with its lack of subordination. The two are not necessarily interchangeable.

Highly Capable AI Could Inadvertently Result in Mass Death

Highly capable AI could also result in catastrophic consequences, such as the death of very many humans (Bengio et al. 2024, 843), and AI misalignment is likely to be the cause. As discussed in the section “The Contours of Highly Capable AI,” highly capable AI systems will be goal-driven. That means that such systems could learn to pursue both desirable and undesirable goals, with the latter being goals not aligned with the widely shared human values (Ngo, Chan and Mindermann 2024, 3). In the main, the pursuit of these undesirable goals will occur where highly capable AI systems have reward functions that are not perfectly aligned with human preferences (ibid.). That possibility is real because the most advanced AI alignment methods still fall short in two ways. First, specifying human intentions using hard-coded reward functions remains very hard and the models may still end up pursuing undesirable goals by exploiting some reward misspecification. Second, although using methods like RLHF might resolve some of these mistakes, feedback from human evaluators is sometimes unreliable, as human evaluators might unintentionally give approval or high reward for undesirable behaviour (ibid.), which could reinforce undesired behaviours in the AI. Similarly, other AI alignment methods being developed by technical researchers still seem a step behind the challenge (Aschenbrenner 2023).

Some signs of the risk we face are captured in the small-scale misbehaviour that can be observed in, for example, an AI system that wins a game by exploiting some glitches rather than by playing the game well (Piper 2020; Krakovna et al. 2020). If such a possibility persists, there is a fair chance that the misalignment of highly capable AI systems will result in more large-scale consequences (Ngo, Chan and Mindermann 2024, 8; Ord 2020, 144). These AI systems may pursue instrumental goals including avoiding shutdown, going around human attempts to alter their reward functions, resource acquiring and, eventually, power seeking (Vold and Harris 2021, 735; Ngo, Chan and Mindermann 2024, 8; Ord 2020, 145). Ultimately, such AI systems may acquire both the motivation and ability to pursue goals that are incompatible with human well-being such as altering the Earth’s environment to facilitate computing speed or actively disempowering humanity to prevent interference with AI goals.

This could eventually lead to catastrophic consequences for humanity, including mass death.

Mass casualty events anywhere are inherently bad, but are particularly unjust when caused in Global South countries by technologies developed and primarily benefiting Global North countries. In the worst-case scenario, misaligned highly capable AI systems could cause human extinction, leading not only to the loss of current lives but also of all potential future generations. Such an outcome would be strongly at odds with the value of care for future generations found in the traditions of many ethnic communities in Global South countries.⁸

Concluding Recommendations

The preceding section of this paper laid out the major context-specific risks that Global South peoples could face when highly capable AI is eventually developed and used. Unfortunately, any domestically focused efforts these countries take to counter the risks outlined herein may have limited usefulness. While regulating the use of standalone physical tools and software might be possible, if highly capable AI is in the form of software accessible over the internet (as current LLMs are), any attempts to block access could be circumvented via virtual private networks. As evidence from China shows, even the most sophisticated government-imposed internet firewall can still be ineffective at completely eliminating access to the relevant internet sites (Williams 2023). Global South governments could also attempt to throttle inference but it would be hard for them to know which data centres to target, and the relevant data centres might be outside their jurisdictions (Lehdonvirta, Wú and Hawkins 2024).

Global South peoples should be further concerned because their governments are currently ill-equipped to police the use of AI within their borders. To the authors’ knowledge, at time of writing, no Global South regulatory body has yet conducted an investigation into the use of AI in their society;

⁸ For one example of communities in southern Africa, see Isaac Schapera (1955, 195–97).

this situation is unlikely to change any time soon because in these countries AI expertise and financial resources are in short supply, while other priorities are endless. This lack of oversight could mean that any interventions might arrive too late in the day.

For these reasons, it is critical that Global South countries engage now, while highly capable AI is still being developed. It would make a significant difference if these countries identified and pulled on the levers available to them in order to influence the future development of highly capable AI systems and their potential implications. This could be achieved by, for example:

- Appointing expert study groups whose focus is not just AI in general, but highly capable AI specifically. These study groups could give Global South countries advice on what specific steps to take with regard to highly capable AI.
- Ensuring that their national and regional policies (and eventually laws) pay special attention to preventing and mitigating the risks that highly capable AI could pose to their societies.
- Forming coalitions of like-minded nations and pushing the leading AI companies and Western countries to pay special attention to the global scale of impacts of highly capable AI as well as the context-specific implications for Global South countries. This work could be done through international fora or state-to-state diplomacy.

Outside of government, researchers and civil society in Global South countries can make a significant difference if they begin engaging as well. Their voices and influence could pressure their governments, pressure Western governments, or persuade researchers and civil society organizations in Western countries to press their governments to ask the leading AI companies more questions. Those who care about the fate of Global South peoples must realize that getting serious about AI generally is not enough. It is crucial that we specifically get serious about the risks of highly capable AI.

Acknowledgements

The authors would like to specially thank Cullen O’Keefe, Faith Gakii, Lynn Schellenberg, Matthew da Mota, Seán Ó hÉigearthaigh, Sienka Dounia, Sumaya Nur and Vicky Miriti for their feedback and assistance in the process of writing this paper.

Works Cited

- Allen, Greg and Taniel Chan. 2017. "Artificial Intelligence and National Security." July. Belfer Center Study. Cambridge, MA: Belfer Center for Science and International Affairs, Harvard Kennedy School. www.belfercenter.org/publication/artificial-intelligence-and-national-security.
- Allen, Karen. 2021. "Drones and Violent Nonstate Actors in Africa." Africa Center for Strategic Studies, August 6. <https://africacenter.org/spotlight/drones-and-violent-nonstate-actors-in-africa/>.
- Allen, Nathaniel and Marian "Ify" Okpali. 2022. "Artificial intelligence creeps on to the African battlefield." Brookings Institution, February 2. www.brookings.edu/articles/artificial-intelligence-creeps-on-to-the-african-battlefield/.
- Amnesty International. 2024. *The State of the World's Human Rights: April 2024*. London, UK: Amnesty International. www.amnesty.org/en/documents/pol10/7200/2024/en.
- Anderson, Janna and Lee Rainie. 2023. *As AI Spreads, Experts Predict the Best and Worst Changes in Digital Life by 2035*. Pew Research Center, June 21. www.pewresearch.org/internet/2023/06/21/as-ai-spreads-experts-predict-the-best-and-worst-changes-in-digital-life-by-2035/.
- Anwar, Usman, Abulhair Saparov, Javier Rando, Daniel Paleka, Miles Turpin, Peter Hase, Ekdeep Singh Lubana et al. 2024. "Foundational Challenges in Assuring Alignment and Safety of Large Language Models." Preprint, arXiv, September 6. <https://doi.org/10.48550/arXiv.2404.09932>.
- Armstrong, Adrienne. 1981. "The Political Consequences of Economic Dependence." *Journal of Conflict Resolution* 25 (3): 401–28. www.jstor.org/stable/173661.
- Arriola, Leonardo R., Lise Rakner and Nicolas van de Walle. 2023. *Democratic Backsliding in Africa? Autocratization, Resilience, and Contention*. Oxford, UK: Oxford University Press.
- Aschenbrenner, Leopold. 2023. "Nobody's on the ball on AGI alignment." *For Our Prosperity* (blog), March 29. www.forourposterity.com/nobodys-on-the-ball-on-agi-alignment/.
- Ayana, Gelan, Kokeb Dese, Hundessa Daba Nemoassa, Bontu Habtamu, Bruce Mellado, Kingsley Badu, Edmund Yamba et al. 2024. "Decolonizing global AI governance: assessment of the state of decolonized AI governance in Sub-Saharan Africa." *Royal Society Open Science* 11 (8): 231994. <https://doi.org/10.1098/rsos.231994>.
- Ayres, Ian and Jack M. Balkin. 2024. "The Law of AI is the Law of Risky Agents without Intentions." Preprint, Social Science Research Network, June 12. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4862025.
- Ayub, Mahmood. 2013. "Poverty and Inequality." *Global Journal of Emerging Market Economies* 5 (3): 329–46. <https://doi.org/10.1177/0974910113505796>.
- Beaumont, Hilary. 2022. "'Never sleeps, never even blinks': the hi-tech Anduril towers spreading along the US borders." *The Guardian*, September 16. www.theguardian.com/us-news/2022/sep/16/anduril-towers-surveillance-us-mexico-border-migrants.
- Bell, Stephanie A. and Anton Korinek. 2023. "AI's Economic Peril." *Journal of Democracy* 34 (4): 151–61. www.journalofdemocracy.org/articles/ai-economic-peril/.
- Bengio, Yoshua. 2023. "AI and Catastrophic Risk." *Journal of Democracy* 34 (4): 111–21. www.journalofdemocracy.org/articles/ai-and-catastrophic-risk/.
- Bengio, Yoshua, Geoffrey E. Hinton, Andrew Yao, Dawn Song, Pieter Abbeel, Trevor Darrell, Yuval N. Harari et al. 2024. "Managing extreme AI risks amid rapid progress." *Science* 384 (6698): 842–45. <https://doi.org/10.1126/science.adn0117>.
- Branwen, Gwern. 2022. "The Scaling Hypothesis." Gwern.net, January 2. <https://gwern.net/scaling-hypothesis>.
- Brundage, Miles, Shahar Avin, Jack Clark, Helen Toner, Peter Eckersley, Ben Garfinkel, Allan Dafoe et al. 2018. *The Malicious Use of Artificial Intelligence: Forecasting, Prevention, and Mitigation*. Preprint, arXiv, February 20. <https://doi.org/10.48550/arXiv.1802.07228>.
- Bubeck, Sébastien, Varun Chandrasekaran, Ronen Eldan, Johannes Gehrke, Eric Horvitz, Ece Kamar, Peter Lee et al. 2023. "Sparks of Artificial General Intelligence: Early experiments with GPT-4." Preprint, arXiv, April 13. <https://arxiv.org/abs/2303.12712>.
- Campbell-James, Jonathan. 2024. "Curbing Outside Intervention in the Sudan War." Policy analysis, April 4. PolicyWatch 3855. Washington, DC: The Washington Institute for Near East Policy. www.washingtoninstitute.org/policy-analysis/curbing-outside-intervention-sudan-war.
- Caparini, Marina and Ian Davis. 2022. "Armed conflict and peace processes in the Americas." In *SIPRI Yearbook 2022: Armaments, Disarmament and International Security*, edited by Stockholm International Peace Research Institute, 75–102. Oxford, UK: Oxford University Press. www.sipri.org/yearbook/2022/03.

- Carlsmith, Joseph. 2024. "Is Power-Seeking AI an Existential Risk?" Preprint, arXiv, August 13. <https://doi.org/10.48550/arXiv.2206.13353>.
- Carroll, Micah, Alan Chan, Henry Ashton and David Krueger. 2023. "Characterizing Manipulation from AI Systems." Preprint, arXiv, October 30. <https://doi.org/10.48550/arXiv.2303.09387>.
- Casals, Marcelo. 2024. "Cold War Dictatorships in the Southern Cone (Brazil, Argentina, Uruguay, Chile)." Oxford Bibliographies, June 21. www.oxfordbibliographies.com/display/document/obo-9780199791279/obo-9780199791279-0254.xml.
- Cazzaniga, Maura, Florence Jaumotte, Longji Li, Giovanni Melina, Augustus J. Panton, Carlo Pizzinelli, Emma Rockall and Marina M. Tavares. 2024. "Gen-AI: Artificial Intelligence and the Future of Work." IMF Staff Discussion Note SDN/2024/001. Washington, DC: International Monetary Fund. www.imf.org/en/Publications/Staff-Discussion-Notes/Issues/2024/01/14/Gen-AI-Artificial-Intelligence-and-the-Future-of-Work-542379.
- Center for Preventive Action. 2024. "Conflict in the Democratic Republic of Congo." Council on Foreign Relations, June 20. www.cfr.org/global-conflict-tracker/conflict/violence-democratic-republic-congo.
- Chan, Alan, Rebecca Salganik, Alva Markelius, Chris Pang, Nitarshan Rajkumar, Dmitrii Krashennikov, Lauro Langosco et al. 2023. "Harms from Increasingly Agentic Algorithmic Systems." Preprint, arXiv, May 12. <https://doi.org/10.48550/arXiv.2302.10329>.
- Choi, Anthony. 2023. "The Global North and Global South in Political Dilemma: Postmaterialist Effects of Age on Government Trust." Gettysburg, PA: The Cupola: Scholarship at Gettysburg College. https://cupola.gettysburg.edu/student_scholarship/1113.
- Collier, Ben. 2024. "Targeted social media ads are influencing our behaviour — and the government uses them too." *The Conversation*, February 27. <https://theconversation.com/targeted-social-media-ads-are-influencing-our-behaviour-and-the-government-uses-them-too-223576>.
- Collier, Ruth B. 1982. *Regimes in Tropical Africa: Changing Forms of Supremacy, 1945–1975*. Berkeley, CA: University of California Press, 1982.
- Conitzer, Vincent, Rachel Freedman, Jobst Heitzig, Wesley H. Holliday, Bob M. Jacobs, Nathan Lambert, Milan Mossé et al. 2024. "Social Choice Should Guide AI Alignment in Dealing with Diverse Human Feedback." Preprint, arXiv, June 4. <https://doi.org/10.48550/arXiv.2404.10271>.
- Cortes-Ramirez, Eugenio-Enrique. 2015. "Cultural Hegemony Today. From Cultural Studies to Critical Pedagogy." *Postcolonial Directions in Education* 4 (2): 116–39. <https://core.ac.uk/download/pdf/84894568.pdf>.
- Cremer, Carla Zoe and Jess Whittlestone. 2021. "Artificial Canaries: Early Warning Signs for Anticipatory and Democratic Governance of AI." *International Journal of Interactive Multimedia and Artificial Intelligence* 6 (5): 100–109. <https://dx.doi.org/10.9781/ijimai.2021.02.011>.
- Criddle, Cristina, Madhumita Murgia and George Hammond. 2024. "Artificial intelligence breakthroughs create new 'brain' for advanced robots." *Financial Times*, July 26. www.ft.com/content/bea9df71-371c-4045-9cb4-64c22789bf7b.
- Cummings, M. L., Heather M. Roff, Kenneth Cukier, Jacob Parakilas and Hannah Bryce. 2018. *Artificial Intelligence and International Affairs: Disruption Anticipated*. Chatham House Report, June. www.chathamhouse.org/sites/default/files/publications/research/2018-06-14-artificial-intelligence-international-affairs-cummings-roff-cukier-parakilas-bryce.pdf.
- Dafoe, Allan. 2018. "AI Governance: A Research Agenda." Oxford, UK: Future of Humanity Institute, Oxford University. www.fhi.ox.ac.uk/wp-content/uploads/GovAI-Agenda.pdf.
- Department for Science, Innovation & Technology. 2023. "Capabilities and risks from frontier AI: A discussion paper on the need for further research into AI risk." October 25. London, UK: Government of the United Kingdom. www.gov.uk/government/publications/frontier-ai-capabilities-and-risks-discussion-paper.
- Derechos Digitales. n.d. "A Latin American perspective on the use of AI systems by the State." Santiago, Chile: Derechos Digitales. www.ohchr.org/sites/default/files/Documents/Issues/DigitalAge/Submissions/CSOs/Derechos-Digitales.pdf.
- Edelman Trust Institute. 2024. *2024 Edelman Trust Barometer, Supplemental Report: Insights for the Tech Sector*. New York, NY: Edelman Trust Institute. www.edelman.com/sites/g/files/aatuss191/files/2024-03/2024%20Edelman%20Trust%20Barometer%20Supplemental%20Report%20Insights%20for%20Tech.pdf.
- Egan, Janet and Eric Rosenbach. 2023. "Biosecurity in the Age of AI: What's the Risk?" Policy brief, November 6. Cambridge, MA: Belfer Center for Science and International Affairs. www.belfercenter.org/publication/biosecurity-age-ai-whats-risk.

- El-Sayed, Seliem, Canfer Akbulut, Amanda McCroskery, Geoff Keeling, Zachary Kenton, Zaria Jalan, Nahema Marchal et al. 2024. "A Mechanism-Based Approach to Mitigating Harms from Persuasive Generative AI." Preprint, arXiv, April 23. <https://doi.org/10.48550/arXiv.2404.15058>.
- Feldstein, Steven. 2019. "The Global Expansion of AI Surveillance." Working Paper, September. Washington, DC: Carnegie Endowment for International Peace. https://carnegie-production-assets.s3.amazonaws.com/static/files/files__WP-Feldstein-AISurveillance_final1.pdf.
- . 2021. *The Rise of Digital Repression: How Technology is Reshaping Power, Politics, and Resistance*. New York, NY: Oxford University Press.
- . 2024. "AI in war: Can advanced military technologies be tamed before it's too late?" *Bulletin of the Atomic Scientists*, January 11. <https://thebulletin.org/2024/01/ai-in-war-can-advanced-military-technologies-be-tamed-before-its-too-late/#post-heading>.
- Frantz, Erica. 2012. "How and Why do Dictatorships Survive? Lessons for the Middle East." *Bridgewater Review* 31 (2): 16–18. https://vc.bridgew.edu/br_rev/vol31/iss2/7/.
- Frantz, Erica, Andrea Kendall-Taylor and Joseph Wright. 2020. "Digital Repression in Autocracies." Users Working Paper Series 2020:27. Gothenburg, Sweden: V-Dem Institute. <https://v-dem.net/media/publications/digital-repression17mar.pdf>.
- Freedom House. 2024. *Freedom in the World 2024: The Mounting Damage of Flawed Elections and Armed Conflict*. February. Washington, DC: Freedom House. https://freedomhouse.org/sites/default/files/2024-02/FIW_2024_DigitalBooklet.pdf.
- Funk, Allie, Adrian Shahbaz and Kian Vesteinsson. 2023. "The Repressive Power of Artificial Intelligence." In *Freedom on the Net 2023*, edited by Adrian Shahbaz, Allie Funk, Jennifer Brody, Kian Vesteinsson, Grant Baker, Cathryn Grothe, Matthew Barak, et al. Washington, DC: Freedom House. <https://freedomhouse.org/sites/default/files/2023-11/FOTN2023Final.pdf>.
- Gabriel, Iason, Arianna Manzini, Geoff Keeling, Lisa Anne Hendricks, Verena Rieser, Hasan Iqbal, Nenad Tomašev et al. 2024. "The Ethics of Advanced AI Assistants." Preprint, arXiv, April 28. <https://doi.org/10.48550/arXiv.2404.16244>.
- Gardizy, Anissa and Amir Efrati. 2024. "Microsoft and OpenAI Plot \$100 Billion Stargate AI Supercomputer." *The Information*, March 29. www.theinformation.com/articles/microsoft-and-openai-plot-100-billion-stargate-ai-supercomputer.
- Gemini Team. 2023. "Gemini: A Family of Highly Capable Multimodal Models." London, UK: Google DeepMind. <https://assets.bwbx.io/documents/users/iqjWHBFdfxIU/r7G7Rrt6rnM/v0>.
- Georgieva, Kristalina. 2024. "AI Will Transform the Global Economy. Let's Make Sure It Benefits Humanity." January 14. www.imf.org/en/Blogs/Articles/2024/01/14/ai-will-transform-the-global-economy-lets-make-sure-it-benefits-humanity.
- Gillani, Noor. 2023. "Will AI ever reach human-level intelligence? We asked five experts." *The Conversation*, April 18. <https://theconversation.com/will-ai-ever-reach-human-level-intelligence-we-asked-five-experts-202515>.
- Gillespie, Nicole, Steve Lockey, Caitlin Curtis, Javad Pool and Ali Akbari. 2023. *Trust in Artificial Intelligence: A Global Study*. Queensland, AU: University of Queensland. <https://assets.kpmg.com/content/dam/kpmg/au/pdf/2023/trust-in-ai-global-insights-2023.pdf>.
- Giri, Felipe Suárez. 2023. "Latin American Cities in the Age of AI: Navigating the Technological Revolution." *Harvard Kennedy School Student Policy Review*, June 21. <https://studentreview.hks.harvard.edu/latin-american-cities-in-the-age-of-ai-navigating-the-technological-revolution/>.
- Harris, Jeremie, Edouard Harris and Mark Beall. 2023. "Survey of AI Technologies and AI R&D Trajectories." Alexandria, VA: Gladstone AI. https://cdn.prod.website-files.com/62c4cf7322be8ea59c904399/65e83959fd414a488a4fa9a5_Gladstone%20Survey%20of%20AI.pdf.
- Heath, Alex. 2024. "Mark Zuckerberg's new goal is creating artificial general intelligence: And he doesn't want to control it. Maybe." *The Verge*, January 18. www.theverge.com/2024/1/18/24042354/mark-zuckerberg-meta-agi-reorg-interview.
- Heikkilä, Melissa. 2024. "Google's Astra is its first AI-for-everything agent." *MIT Technology Review*, May 14. www.technologyreview.com/2024/05/14/1092407/googles-astra-is-its-first-ai-for-everything-agent/.
- Heller, Patrick. 2022. "Democracy in the Global South." *Annual Review of Sociology* 48: 463–84. <https://doi.org/10.1146/annurev-soc-030320-123449>.
- Hendrycks, Dan, Mantas Mazeika and Thomas Woodside. 2023. "An Overview of Catastrophic AI Risks." Preprint, arXiv, October 9. <https://doi.org/10.48550/arXiv.2306.12001>.

- Hiebert, Kyle. 2023. "Rising digital surveillance threatens Africa's democratic progress." *ISS African Futures* (blog), October 4. <https://futures.issafrica.org/blog/2023/Rising-digital-surveillance-threatens-Africas-democratic-progress>.
- Hörster, Tim. 2021. "Utility of Artificial Intelligence to Authoritarian Governance." Preprint, Social Science Research Network, September 15. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4074964.
- Human Rights Watch. 2023. "DR Congo: Atrocities by Rwanda-Backed M23 Rebels." News, February 6. New York, NY: Human Rights Watch. www.hrw.org/news/2023/02/06/dr-congo-atrocities-rwanda-backed-m23-rebels.
- ICRC. 2019. *Artificial intelligence and machine learning in armed conflict: A human-centred approach*. Report, June 6. Geneva, Switzerland: ICRC. www.icrc.org/en/document/artificial-intelligence-and-machine-learning-armed-conflict-human-centred-approach.
- Ipsos. 2022. "Global Opinions and Expectations about Artificial Intelligence: A Global Advisor survey." January. Paris, France: Ipsos. www.ipsos.com/sites/default/files/ct/news/documents/2022-01/Global-opinions-and-expectations-about-AI-2022.pdf.
- Jenkins, J. Craig and Augustine J. Kposowa. 1990. "Explaining Military Coups D'État: Black Africa, 1957–1984." *American Sociological Review* 55 (6): 861–75. <https://doi.org/10.2307/2095751>.
- Jørgensen, Karen Helene. 2019. "The Meaning of Internet Access in Myanmar." Master's thesis, Norwegian University of Life Sciences. <https://nmbu.brage.unit.no/nmbu-xmlui/bitstream/handle/11250/2612476/J%C3%B8rgensen2019.pdf?sequence=1>.
- Juling, Dominik. 2023. "Future Bioterror and Biowarfare Threats for NATO's Armed Forces until 2030." *Journal of Advanced Military Studies* 14 (1): 118–43. <https://doi.org/10.21140/mcu.j.20231401005>.
- Kapania, Shivani, Oliver Siy, Gabe Clapper, Azhagu Meena and Nithya Sambasivan. 2022. "'Because AI is 100% right and safe': User Attitudes and Sources of AI Authority in India." Paper presented at the CHI Conference on Human Factors in Computing Systems. New Orleans, LA, April 29–May 5. <https://doi.org/10.1145/3491102.3517533>.
- Kaplan, Jared, Sam McCandlish, Tom Henighan, Tom B. Brown, Benjamin Chess, Rewon Child, Scott Gray, Alec Radford, Jeffrey Wu and Dario Amodei. 2020. "Scaling Laws for Neural Language Models." Preprint, arXiv, January 23. <https://arxiv.org/pdf/2001.08361>.
- Karnofsky, Holden. 2016. "Some Background on Our Views Regarding Advanced Artificial Intelligence." *Open Philanthropy* (blog), May 6. www.openphilanthropy.org/research/some-background-on-our-views-regarding-advanced-artificial-intelligence/#2-when-should-we-expect-transformative-artificial-intelligence.
- Kausch, Kristina, Aitor Bonsoms, Can Kasapoglu, Lewin Schmitt and Žilvinas Švedkauskas. 2022. *Liberty's Doom? Artificial Intelligence in Middle Eastern Security*. European Institute of the Mediterranean. www.euromesco.net/wp-content/uploads/2022/07/Policy-Study27.pdf.
- Kendall-Taylor, Andrea, Erica Frantz and Joseph Wright. 2020. "The Digital Dictators: How Technology Strengthens Autocracy." *Foreign Affairs*, February 6. www.foreignaffairs.com/articles/china/2020-02-06/digital-dictators.
- Kenton, Zachary, Ramana Kumar, Sebastian Farquhar, Jonathan Richens, Matt MacDermott and Tom Everitt. 2022. "Discovering Agents." Preprint, arXiv, August 24. <https://doi.org/10.48550/arXiv.2208.08345>.
- Klark, Arthur and Gabriela Sagun. 2023. "Misinformation, Hate Speech and Ethno-Religious Tensions in Myanmar." Analysis, April 27. Washington, DC: United States Institute of Peace. www.usip.org/publications/2023/04/misinformation-hate-speech-and-ethno-religious-tensions-myanmar.
- Knuutila, Aleks, Lisa-Maria Neudert and Philip N. Howard. 2022. "Who is afraid of fake news? Modeling risk perceptions of misinformation in 142 countries." *Harvard Kennedy School Misinformation Review* 3 (3): 1–13. <https://doi.org/10.37016/mr-2020-97>.
- Korinek, Anton and Megan Juelfs. 2022. "Preparing for the (non-existent?) future of work." Center on Regulation and Markets Working Paper 3. Washington, DC: Brookings Institution. www.brookings.edu/articles/preparing-for-the-non-existent-future-of-work/.
- Korinek, Anton and Joseph E. Stiglitz. 2021. "Artificial Intelligence, Globalization, and Strategies for Economic Development." National Bureau of Economic Research Working Paper 28453. August. www.nber.org/system/files/working_papers/w28453/w28453.pdf.
- Krakovna, Victoria, Jonathan Uesato, Vladimir Mikulik, Matthew Rahtz, Tom Everitt, Ramana Kumar, Zac Kenton, Jan Leike and Shane Legg. 2020. "Specification gaming: the flip side of AI ingenuity." *Medium*, April 21. <https://deepmindsafetyresearch.medium.com/specification-gaming-the-flip-side-of-ai-ingenuity-c85bdb0deeb4>.

- Lehdonvirta, Vili, Bóxi Wú and Zoe Hawkins. 2024. "Compute North vs. Compute South: The Uneven Possibilities of Compute-based AI Governance Around the Globe." Preprint, SocArXiv Papers, August 25. <https://doi.org/10.31235/osf.io/8yp7z>.
- Lehmann, Jean-Pierre. 1985. "Dictatorship and development in Pacific Asia: wider implications." *International Affairs* 61 (4): 591–606. <https://doi.org/10.2307/2617705>.
- Liu, Junfeng, Christopher Symons and Ranga Raju Vatsavai. 2022. "Persona-Based Conversational AI: State of the Art and Challenges." Preprint, arXiv, December 4. <https://doi.org/10.48550/arXiv.2212.03699>.
- Maas, Matthijs M., Kayla Lucero-Matteucci and Di Cooke. 2023. "Military Artificial Intelligence as a Contributor to Global Catastrophic Risk." In *The Era of Global Risk: An Introduction to Existential Risk Studies*, edited by S. J. Beard, Martin Rees, Catherine Richards and Clarissa Rios Rojas, 237–84. Cambridge, UK: Open Book Publishers. <https://books.openbookpublishers.com/10.11647/obp.0336.10.pdf>.
- Mandavia, Megha. 2024. "AI Is Coming for India's Famous Tech Hub." *The Wall Street Journal*, August 6. www.wsj.com/tech/ai/ai-is-coming-for-indias-famous-tech-hub-97261085.
- Manvi, Rohin, Samar Khanna, Marshall Burke, David Lobell and Stefano Ermon. 2024. "Large Language Models are Geographically Biased." Preprint, arXiv, October 5. <https://doi.org/10.48550/arXiv.2402.02680>.
- Matlosa, Khabele. 2023. "Global trends and impact of democratic recession: Hard choices for the Global South." *South African Journal of International Affairs* 30 (3): 337–55. <https://doi.org/10.1080/10220461.2023.2269149>.
- Mora, Frank and Brian Fonseca. 2019. "It's Not the 1970s Again for Latin America's Militaries. Here's Why." *Americas Quarterly*, December 11. www.americasquarterly.org/article/its-not-the-1970s-again-for-latin-americas-militaries-heres-why/.
- Murgia, Madhumita. 2023. "OpenAI chief seeks new Microsoft funds to build 'superintelligence.'" *Financial Times*, November 13. www.ft.com/content/dd9ba2f6-f509-42f0-8e97-4271c7b84ded.
- Ngo, Richard. 2020. "AGI Safety From First Principles." AI Alignment Forum, September 28. www.alignmentforum.org/s/mzgtmmTKKn5MuCzFJ.
- . 2023. "Visualizing the deep learning revolution." *Medium*, January 5. <https://medium.com/@richardcngo/visualizing-the-deep-learning-revolution-722098eb9c5>.
- Ngo, Richard, Lawrence Chan and Sören Mindermann. 2024. "The Alignment Problem from a Deep Learning Perspective." Preprint, arXiv, March 19. <https://arxiv.org/pdf/2209.00626>.
- O'Brien, John T. and Cassidy Nelson. 2020. "Assessing the Risks Posed by the Convergence of Artificial Intelligence and Biotechnology." *Health Security* 18 (3): 219–27. <https://liebertpub.com/doi/pdf/10.1089/hs.2019.0122>.
- Omeje, Kenneth. 2018. "Extractive Economies and Conflicts in the Global South: Re-Engaging Rentier Theory and Politics." In *Extractive Economies and Conflicts in the Global South: Multi-Regional Perspectives on Rentier Politics*, edited by Kenneth Omeje, 1–25. Abingdon, UK: Routledge.
- Onome. 2024. "How AI is Keeping Old Languages Alive." AutoGPT by Mindstream, June 26. <https://autogpt.net/the-impact-of-ai-in-languages-preservation/#~:text=AI%20aids%20language%20documentation%20and,%20supporting%20real%2Dtime%20translation>.
- Onyekwena, Chukwuka and Mma Amara Ekeruche. 2019. "The Global South and development assistance." Commentary, September 16. Washington, DC: Brookings Institution. www.brookings.edu/articles/the-global-south-and-development-assistance/.
- Ord, Toby. 2020. *The Precipice: Existential Risk and the Future of Humanity*. New York, NY: Hachette Books.
- Palik, Júlia, Anna M. Obermeier and Siri A. Rustad. 2022. "Conflict Trends in Africa, 1989–2021." PRIO Paper. Oslo, Norway: Peace Research Institute Oslo. www.prio.org/publications/13215.
- Patel, Nilay. 2023. "Inside Google's big AI shuffle — and how it plans to stay competitive, with Google DeepMind CEO Demis Hassabis." *The Verge*, July 10. www.theverge.com/23778745/demis-hassabis-google-deepmind-ai-alphafold-risks.
- Phyo. 2023. "The Digital Landscape of Activism in Myanmar." EngageMedia, November 29. <https://engagemedia.org/2023/youth-myanmar/>.
- Piper, Kelsey. 2020. "The case for taking AI seriously as a threat to humanity: Why some people fear AI, explained." *Vox*, October 15. www.vox.com/future-perfect/2018/12/21/18126576/ai-artificial-intelligence-machine-learning-safety-alignment.
- Prabhakaran, Vinodkumar, Rida Qadri and Ben Hutchinson. 2022. "Cultural Incongruencies in Artificial Intelligence." Preprint, arXiv, November 19. <https://doi.org/10.48550/arXiv.2211.13069>.

- Privitera, Daniel, Tamay Besiroglu, Rishi Bommasani, Stephen Casper, Yejin Choi, Danielle Goldfarb, Hoda Heidari and Leila Khalatbari. 2024. *International Scientific Report on the Safety of Advanced AI: Interim Report*. May 17. London, UK: Department for Science, Innovation and Technology. www.gov.uk/government/publications/international-scientific-report-on-the-safety-of-advanced-ai.
- Putzier, Konrad. 2023. "Next Wave of Remote Work Is About Outsourcing Jobs Overseas." *The Wall Street Journal*, April 11. www.wsj.com/articles/next-wave-of-remote-work-is-about-outsourcing-jobs-overseas-54af39ba.
- Ritchie, Hannah, Veronika Samborska, Natasha Ahuja, Esteban Ortiz-Ospina and Max Roser. 2023. "Global Education." OurWorldinData.org. <https://ourworldindata.org/global-education?insight=children-across-the-world-receive-very-different-amounts-of-quality-learning#key-insights>.
- Roffey, Roger. 2004. "Biological weapons and potential indicators of offensive biological weapon activities." In *SIPRI Yearbook 2004: Armaments, Disarmament and International Security*, edited by Stockholm International Peace Research Institute, 557–71. Oxford, UK: Oxford University Press. www.sipri.org/yearbook/2004/13.
- Roser, Max. 2023. "Technology over the long run: zoom out to see how dramatically the world can change within a lifetime." OurWorldinData.org, February 22. <https://ourworldindata.org/technology-long-run#>.
- Rustad, Siri A. 2024. "Conflict Trends: A Global Overview, 1946–2023." PRIO Paper. Oslo, Norway: Peace Research Institute Oslo. www.prio.org/publications/14006.
- Schapera, Isaac. 1955. *A Handbook of Tswana Law and Custom*. 2nd ed. London, UK: Frank Cass & Co.
- Schmidt, Eric. 2023. "Innovation Power: Why Technology Will Define the Future of Geopolitics." *Foreign Affairs*, February 28. www.foreignaffairs.com/united-states/eric-schmidt-innovation-power-technology-geopolitics.
- Shavit, Yonadav, Sandhini Agarwal, Miles Brundage, Steven Adler, Cullen O'Keefe, Rosie Campbell, Teddy Lee et al. 2023. "Practices for Governing Agentic AI Systems." OpenAI paper, December 14. <https://cdn.openai.com/papers/practices-for-governing-agentic-ai-systems.pdf>.
- Siani, Alessandro and Imogen Green. 2023. "Scientific Misinformation and Mistrust of COVID-19 Preventive Measures among the UK Population: A Pilot Study." *Vaccines* 11 (2): 301. <https://doi.org/10.3390/vaccines11020301>.
- Sotala, Kaj and Roman V. Yampolskiy. 2015. "Responses to catastrophic AGI risk: a survey." *Physica Scripta* 90: 018001. <https://iopscience.iop.org/article/10.1088/0031-8949/90/1/018001/pdf>.
- Stoelker, Tom. 2018. "American TV, International Perceptions, and Reality." *Fordham Now*, March 28. <https://now.fordham.edu/inside-fordham/faculty-reads/american-tv-international-perceptions-and-reality/>.
- Szczepański, Marcin. 2019. "Economic impacts of artificial intelligence (AI)." Briefing, July. Brussels, Belgium: European Parliamentary Research Service. [www.europarl.europa.eu/RegData/etudes/BRIE/2019/637967/EPRS_BRI\(2019\)637967_EN.pdf](http://www.europarl.europa.eu/RegData/etudes/BRIE/2019/637967/EPRS_BRI(2019)637967_EN.pdf).
- Taleb, Abdelrahman Abu. 2024. "Evidence of Iran and UAE drones used in Sudan war." BBC, June 12. www.bbc.com/news/articles/c2vvjz652j1o.
- Tan, Jun-E, Kathleen Azali and Katerina Francisco. 2021. *Governance of Artificial Intelligence (AI) in Southeast Asia*. Melbourne, AU: EngageMedia. December. https://engagemedia.org/wp-content/uploads/2021/12/Engage_Report-Governance-of-Artificial-Intelligence-AI-in-Southeast-Asia_12202021.pdf.
- Tao, Yan, Olga Viberg, Ryan S. Baker and René F. Kizilcec. 2024. "Cultural Bias and Cultural Alignment of Large Language Models." Preprint, arXiv, June 26. <https://arxiv.org/abs/2311.14096v2>.
- Tech Wire Asia. 2018. "The impact of AI is as great as the steam engine: McKinsey." *Tech Wire Asia*, August 31. <https://techwireasia.com/09/2018/ai-will-bring-global-growth-on-par-with-steam-engine/>.
- The Economist*. 2023. "ChatGPT could replace telemarketers, teachers and traders." *The Economist*, April 14. www.economist.com/graphic-detail/2023/04/14/chatgpt-could-replace-telemarketers-teachers-and-traders.
- Timcke, Scott and Hanani Hlomani. 2024. *Decoding the Ballot: How Might AI Reshape Democracy on the African Continent?* Cape Town, South Africa: Research ICT Africa. <https://researchictafrica.net/research/how-might-ai-reshape-democracy-on-the-african-continent/>.
- Todaro, Michael P. and Stephen C. Smith. 2020. *Economic Development*. 13th ed. Harlow, UK: Pearson.
- Transparency International. 2024. *Corruption Perceptions Index 2023*. Berlin, Germany: Transparency International. https://transparencia.org.es/wp-content/uploads/Report_CPI2023_English.pdf.

- Varshney, Kush R. 2024. "Decolonial AI Alignment: Openness, Viśesa-Dharma, and Including Excluded Knowledges." Preprint, arXiv, May 2. <https://doi.org/10.48550/arXiv.2309.05030>.
- Verbruggen, Maaïke. 2021. "Drone swarms: coming (sometime) to a war near you. Just not today." *Bulletin of the Atomic Scientists*, February 3. <https://thebulletin.org/2021/02/drone-swarms-coming-sometime-to-a-war-near-you-just-not-today/>.
- Vold, Karina and Daniel R. Harris. 2021. "How Does Artificial Intelligence Pose an Existential Risk?" In *The Oxford Handbook of Digital Ethics*, edited by Carissa Véliz, 724–47. Oxford, UK: Oxford University Press. <https://doi.org/10.1093/oxfordhb/9780198857815.013.36>.
- Whittlestone, Jess and Samuel Clarke. 2024. "AI Challenges for Society and Ethics." In *The Oxford Handbook of AI Governance*, edited by Justin B. Bullock, Yu-Che Chen, Johannes Himmelreich, Valerie M. Hudson, Anton Korinek, Matthew M. Young and Baobao Zhang, 45–64. New York, NY: Oxford University Press.
- Williams, John. 2023. "How China Blocks VPNs and How Citizens Bypass the Blocks." *BTC Peers*, October 7. <https://btcpeers.com/how-china-blocks-vpns-and-how-citizens-bypass-the-blocks/>.
- Willing, Nicole. 2023. "Is AI as Transformative as the Steam Engine and the Internet?" *Techopedia*, October 11. www.techopedia.com/is-ai-as-transformative-a-technology-as-the-steam-engine-and-the-internet.
- World Economic Forum. 2024. *The Global Risks Report 2024*. 19th ed. Insight Report, January. www3.weforum.org/docs/WEF_The_Global_Risks_Report_2024.pdf.
- Yom, Sean. 2015. "Arab Civil Society after the Arab Spring: Weaker but Deeper." Middle East Institute, October 22. www.mei.edu/publications/arab-civil-society-after-arab-spring-weaker-deeper.
- Zittrain, Jonathan. 2024. "We Need to Control AI Agents Now." *The Atlantic*, July 2. www.theatlantic.com/technology/archive/2024/07/ai-agents-safety-risks/678864/.



67 Erb Street West
Waterloo, ON, Canada N2L 6C2
www.cigionline.org