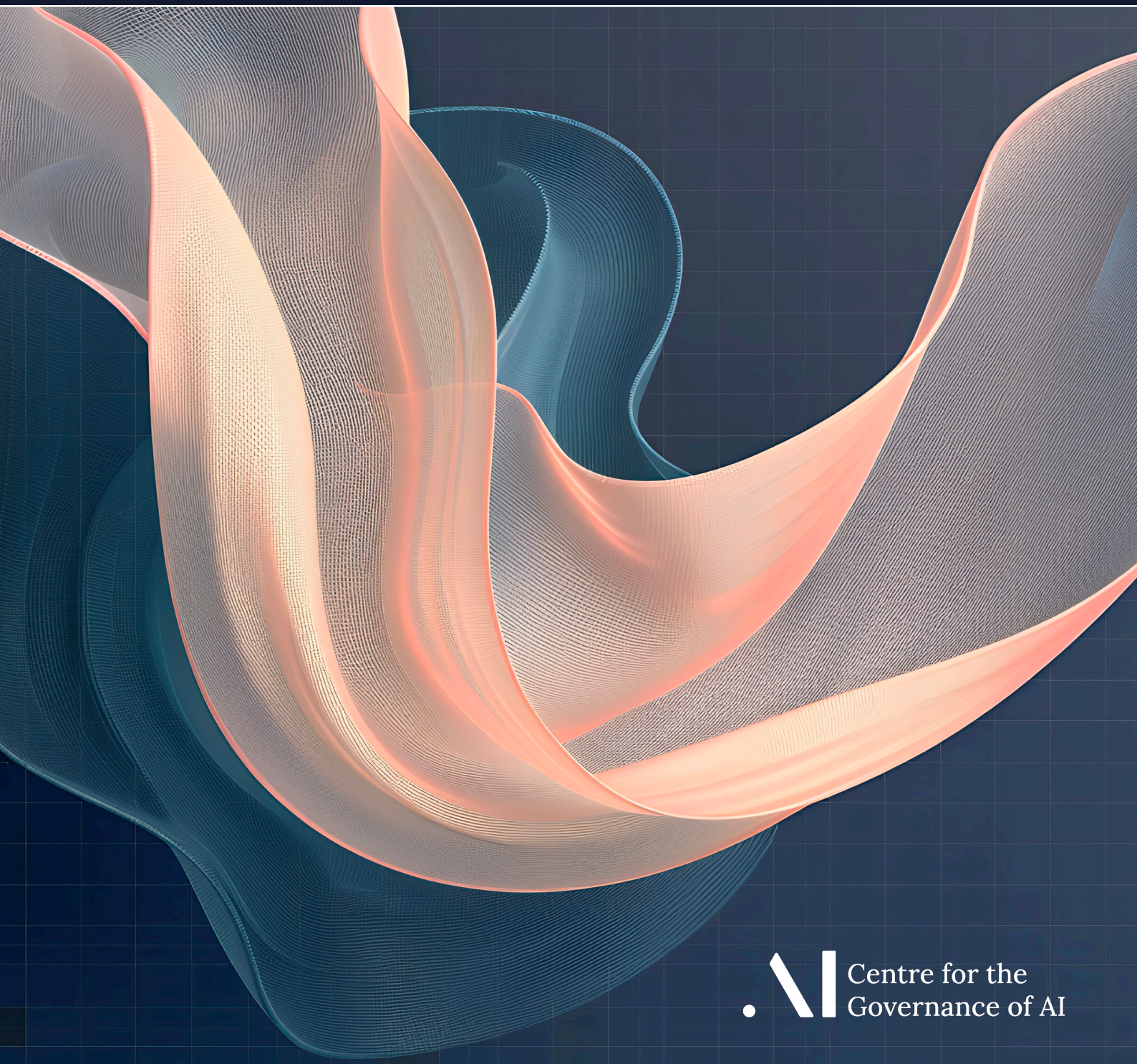
Options and Motivations for International AI Benefit Sharing

Claire Dennis, Sam Manning, Stephen Clare, Boxi Wu, Jake Okechukwu Effoduh, Chinasa T. Okolo, Lennart Heim, Katya Klinova



Centre for the Governance of AI

Authors and affiliations

CLAIRE DENNIS*
Centre for the Governance of AI
Oxford Martin AI Governance Initiative, University of Oxford

SAM MANNING*
Centre for the Governance of AI

STEPHEN CLARE*
Centre for the Governance of AI

BOXI WU
University of Oxford

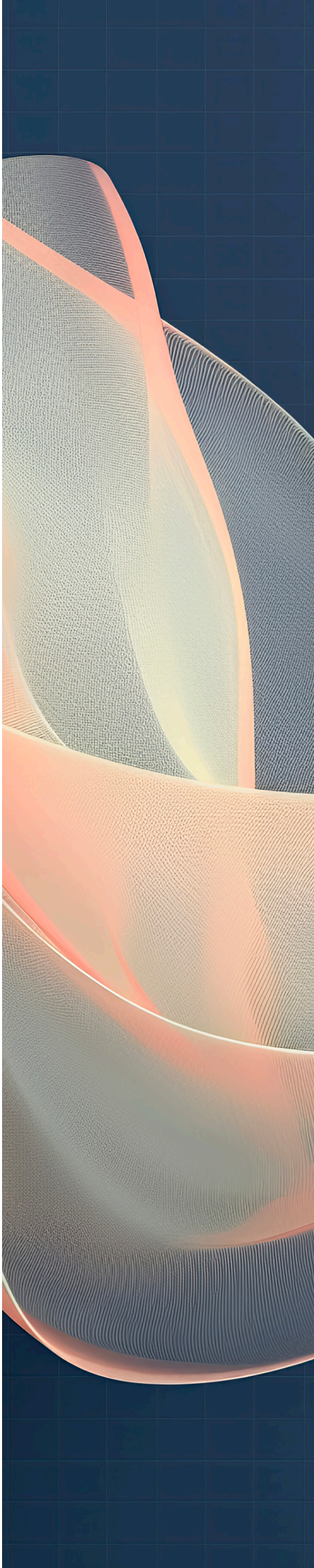
JAKE OKECHUKWU EFFODUH
Lincoln Alexander School of Law

CHINASA T. OKOLO
The Brookings Institution

LENNART HEIM
Centre for the Governance of AI

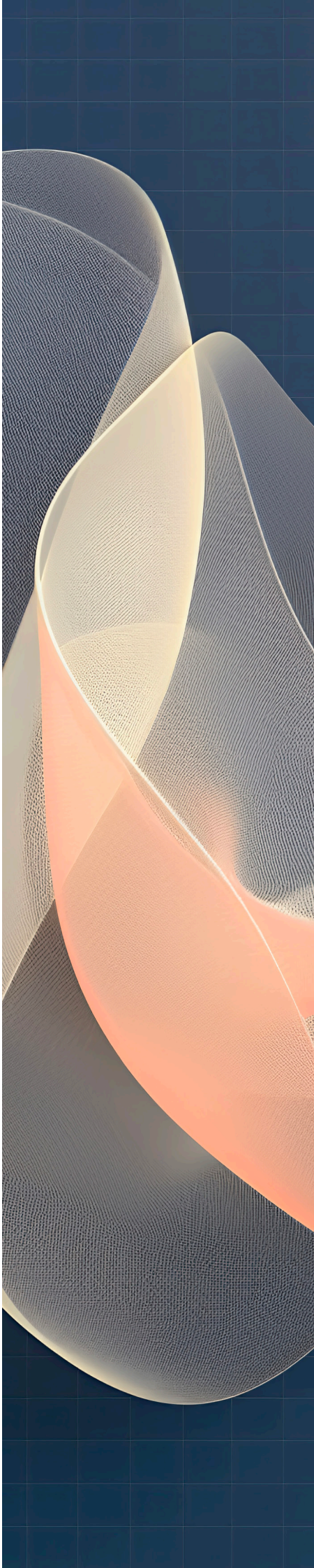
KATYA KLINOVA
United Nations

*Primary authors



Acknowledgements

For helpful conversations, comments and feedback on drafts of this report, we thank: Cullen O'Keefe, Jai Vipra, Sihao Huang, Sam Winter-Levy, Jonas Kgomo, Atoosa Kasirzadeh, Anton Korinek, Ben Garfinkel, Sumaya Nur Adan, Michel Justen, Cecil Abungu, Janet Egan, Casey Mahoney, Neel Alex, Alan Chan, Tomás Aguirre, Joanna Wiaterek, Lucia Velasco, Beth Eakman, Anna Yelizarova, Noemi Dreksler and the rest of the GovAI team. For design help, we thank Taylor Jones and José Luis León Medina.



Contact

Corresponding authors: Claire Dennis (claire.dennis@governance.ai) and Sam Manning (sam.manning@governance.ai)

Please cite as

Dennis, C., Manning, S., Clare, S., et al. (2025). *Options and Motivations for International AI Benefit Sharing.* Centre for the Governance of AI.

Abstract

Advanced AI systems could generate substantial economic and other societal benefits, but these benefits may not be widely shared by default. For a range of reasons, a number of prominent actors and institutions have called for efforts to expand access to AI's benefits. In this report, we define the concept of **international AI benefit sharing** ("benefit sharing") as efforts to support and accelerate international access to AI's economic or broader societal benefits. Calls for benefit sharing typically invoke at least one of three motivations: 1) supporting inclusive economic growth and sustainable development, 2) fostering technological self-determination in low- and middle-income countries, and 3) advancing geopolitical objectives, including strengthening international partnerships on AI governance. Notably, as a subset of the third motive, some powerful actors – like the US government – may support benefit sharing as a tool to further their economic and national security interests. Benefit sharing could be implemented by (1) **sharing AI resources** (e.g., computing power or data), (2) **expanding access to AI systems**, or (3) t**ransferring a portion of the financial proceeds from AI** commercialisation or AI-driven economic growth. Depending on the objective that benefit sharing is intended to achieve, each of these approaches offers distinct opportunities and implementation challenges. These challenges include the potential for some benefit-sharing options to raise security concerns and increase certain global risks. Actors interested in benefit sharing may consider implementing low-risk forms of benefit sharing immediately, while launching cooperative international discussions to develop more comprehensive, mutually-beneficial initiatives.

Contents

I Introduction and Summary

Advanced AI could bring unprecedented economic growth and other societal benefits (Erdil and Besiroglu, 2023; Gosselink et al., 2024). However, these benefits might not be widely accessible by default (Ben-Ishai et al., 2024). This possibility is increasingly prompting calls for **international AI benefit sharing**: *efforts to support and accelerate international access to AI's economic or broader societal benefits*.

These and related calls have been made by a wide range of actors, often with divergent motives and perspectives. For example, the UN's AI Advisory Body (United Nations AI Advisory Body, 2024) has advocated for benefit sharing from a sustainable development and self-determination perspective, while Trump administration officials have emphasised the strategic importance of promoting access to American AI technology abroad in order to counter adversaries' influence in emerging markets (Ruiz, 2024; Oxford Generative AI Summit, 2024). Benefit sharing is also central to the stated missions of several leading AI companies: "broadly distributed benefits" is the first principle listed in OpenAI's charter (OpenAI, 2018) and Google DeepMind's mission is to "ensure AI benefits everyone and helps solve the biggest challenges facing humanity" (Google DeepMind, 2023).

The purpose of this report is to add clarity to these overlapping discussions – which reflect both a diverse set of motives for benefit sharing and a diverse set of concrete ideas about what benefit sharing might entail – and outline paths forward for relevant actors. **Specifically, in this report we: describe the core motivations for international AI benefit sharing, present and categorise concrete options for benefit sharing, analyse the strengths and limitations of these options, and clarify implementation challenges**. Our discussion has four parts.

In part I, we review the various motivations for benefit sharing, dividing them into three categories:

1. **Inclusive Economic Growth:** Benefit sharing could help accelerate economic growth and achieve Sustainable Development Goals in low- and middle-income countries.[1]
2. **Technological Self-Determination:** Benefit sharing could promote national sovereignty and self-determination in AI development.
3. **Geopolitical Motivations:** Benefit sharing could enable international cooperation on AI governance or advance particular states' geopolitical interests.

The first two of these motivations could be described as *altruistic*. They focus on broadening access to AI benefits in order to improve the wellbeing of people in developing countries through economic growth and technological self-determination.

In contrast, the third motivation highlights that frontier AI states[2] – countries that host AI companies developing the most advanced, general-purpose models – may use benefit sharing

[1] Following World Bank classifications, we define *low- and middle-income countries* as those with a GNI per capita of US $14,005 or less in 2023 (this includes economies classified by the World Bank as low-income, lower middle-income, and upper middle-income).

[2] As of January 2025, this primarily includes the US and China, though companies in France, Canada, and the UAE are also pursuing frontier AI capabilities and could eventually join this group.

as a strategic tool to strengthen partnerships and achieve a range of foreign policy goals (both AI-related and not AI-related) (Franke, 2021). This may be especially relevant as AI capabilities advance amid growing competition between leading states – especially the US and China. Such partnerships could help enforce trade policies like export controls, implement security measures to protect data and intellectual property, and incentivise participation in international governance frameworks aimed at reducing cross-border risks. Additionally, some benefit sharing strategies may serve both frontier states and AI companies by expanding their presence in emerging markets, creating long-term commercial opportunities.

In part II, following the discussion of motivations, we examine options for how access to AI's benefits could practically be shared. We distinguish between three approaches:

- **Sharing AI development resources** such as data, computing power, technical talent, and information about training algorithms and procedures
- **Sharing access to advanced AI systems** through user applications, application programming interfaces (APIs), or open-source models
- **Sharing financial proceeds** such as a portion of AI company profits or tax revenue

We consider the advantages, limitations, and some specific implementation options for each approach. These options vary widely, and choosing among them will involve significant trade-offs. Some, for example, offer more flexibility, but also potentially increase risks or involve coordination among multiple diverse stakeholders.

In part III, we consider challenges that apply to all three approaches. We explore five challenges in particular. Benefit sharing may:

1. **Be redundant** if AI benefits are likely to diffuse widely and quickly through market forces alone.
2. **Be intractable** due to competing interests and incentives for key actors.
3. **Increase risks** by broadening access to systems with dangerous capabilities or exacerbating competitive dynamics that make AI development less safe (Scharre, 2021; Armstrong, Bostrom, and Shulman, 2016).
4. **Be constrained by geopolitical and strategic goals of leading states**, which might make states reluctant to share valuable resources or AI access. Particularly for actors who are pursuing benefit sharing primarily as a means to inclusive growth or self-determination, these geopolitical constraints may often present significant barriers.[3]
5. **Be leveraged for coercion**. For example, powerful actors could use benefit sharing as a tool to pressure less powerful actors into certain policy actions that might otherwise run counter to their interests. Particularly for actors who are pursuing benefit sharing as a means to self-determination, this may be a significant concern.[4]

[3]Today, for example, some of the policies of frontier AI states, such as export controls, primarily limit rather than accelerate the spread of AI resources such as advanced chips.

[4]The fact that this potential effect of benefit sharing appears negative from the perspective of those with self-determination motives, but may appear positive from the perspective of those with certain geopolitical motives, highlights the kinds of tensions that can arise between motives for benefit sharing.

I. Motivations for AI benefit sharing

Economic and Societal Benefit Motivations

- Accelerate economic growth and prosperity in low- and middle-income countries
- Distribute the gains from labour automation

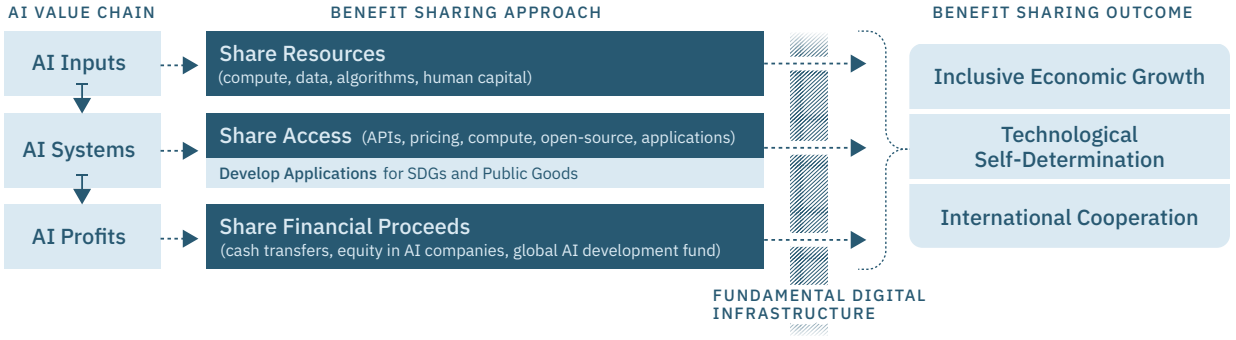
Technological Self-Determination Motivations

- Promote national sovereignty and self-determination in AI development.

Geopolitical Motivations

- Encourage adherence to international safety standards
- Reduce risks from competitive racing to deploy increasingly powerful AI systems
- Enhance collective security
- Strengthen partnerships to limit the influence of strategic competitors

II. Approaches to AI benefit sharing



III. Challenges associated with AI benefit sharing

Challenge 1: Benefit sharing may be redundant.

Challenge 2: Benefit sharing may be intractable.

Challenge 3: Benefit sharing may increase risks.

Challenge 4: Benefit sharing may be constrained by other geopolitical goals.

Challenge 5: Benefit sharing may be leveraged for coercion.

IV. Next steps for benefit sharing

Initiate low-risk, high-potential benefit sharing approaches

Establish dedicated forums to discuss benefit sharing options and implementation strategies

Ensure representation of low- and middle-income countries in AI governance decision-making processes

Figure 1: Executive Summary

Finally, in part IV we outline next steps for advancing benefit-sharing initiatives. We suggest that **remaining uncertainties and open questions do not necessarily need to prevent action today**. For actors and institutions motivated to engage in benefit sharing to promote inclusive economic growth and development, some options, such as funding fundamental digital infrastructure and AI applications aimed at advancing the Sustainable Development Goals (SDGs), are both promising and low-risk. These actors could consider implementing these options immediately. At the same time, though, there remains significant uncertainty about the best approach to more ambitious and mutually-beneficial benefit-sharing initiatives, especially if rapid AI progress continues. Reducing this uncertainty will likely require discussion with a wide range of stakeholders, including representatives from the countries that such initiatives aim to support.

To that end, **establishing a dedicated forum for all stakeholders to discuss benefit sharing options and implementation strategies** could help interested actors identify potential paths forward. The United Nations' (UN) planned "Global Dialogue on AI Governance," recently agreed upon as part of the Global Digital Compact, could potentially host such a process, though it is not the only option (United Nations Office of the Secretary-General's Envoy on Technology, 2024). Our analysis also highlights that benefit-sharing efforts are best understood in connection with broader issues of international AI governance such as risk reduction and ethical considerations. Thus we note the importance of **international representation in broader AI governance discussions**, not only those specifically related to benefit sharing.

This paper seeks to illustrate ways forward for expanding international access to AI's benefits. There is work to be done to clarify how exactly access to AI's benefits should be expanded, and which responsibilities should fall on different global actors. The implementation challenges are also significant, and it will be important to avoid exacerbating the risks raised by widening access to AI systems that may have dangerous capabilities and pose societal risks. However, benefit sharing could also help states make mutually-beneficial bargains that promote access to AI's benefits while reducing shared risks. AI may yet help solve global challenges, including making progress towards the SDGs. Through benefit-sharing efforts, frontier AI states could fulfil that promise while supporting other strategic priorities related to AI development and governance.

II Motivations

Various actors may support international AI benefit sharing for different reasons. We group these motivations into three categories:

1. **Accelerating the spread of economic and societal benefits globally:** Advanced AI systems could generate large economic profits and improve services in industries like healthcare and education, for example. By default, however, these benefits may be concentrated in high-income countries. Benefit-sharing efforts could ensure that they also reach low- and middle-income countries.

2. **Supporting technological self-determination:** AI development is largely concentrated in a few high-income countries and China (Maslej et al., 2024). This may lead to technological

dependency (Lehdonvirta, Wu, and Hawkins, 2023), with most nations reliant on foreign AI systems to drive economic growth and enhance public services while having little influence over their development. Benefit sharing could allow more low- and middle-income economies to obtain the resources needed to shape future AI developments.

3. **Incentivizing international cooperation and advancing geopolitical interests of leading states:** Advanced AI systems may pose global safety and security risks if not developed and used responsibly (Bengio, Mindermann, et al., 2025). Managing these risks likely requires international standards to prevent the development of dangerous systems and to keep threat actors from gaining access to powerful systems they can use to cause harm. Benefit sharing could incentivize participation in international cooperation efforts by tying access to benefits to compliance with safety protocols. Promoting access to AI technologies and resources globally may also align with the strategic interests of frontier AI companies and states, which seek to expand their domestic companies' global market share (Export-Import Bank of the United States, 2023). Furthermore, broadening access to AI benefits internationally could strengthen support among recipient nations for the governance frameworks advocated by leading AI states.

Different actors will weigh these motivations differently. For instance, UN-led initiatives may focus on promoting self-determination to align with precedents set by, for example, the right to self-determination affirmed by the UN Declaration on the Rights of Indigenous Peoples (United Nations, 2007). Similarly, global development actors may adopt the economic and societal benefits motivation as a way to help achieve the SDGs. On the other hand, frontier AI states like the US or China might pursue benefit sharing for their own strategic interests, for example as a way to strengthen international alliances around distinct approaches to AI development and governance.

II.1 Economic and Societal Benefit Motivations

Some actors may support benefit sharing as a way to change how the **economic and societal benefits produced by advanced AI systems are distributed**. Two concerns driving this motivation are that the benefits of AI could primarily accrue to people in higher-income countries rather than those in lower- and middle-income countries, and to owners of capital rather than workers.

Benefit sharing could help accelerate economic growth and achieve development goals in low- and middle-income countries

The first specific motivation for benefit sharing is to **direct resources to developing countries to foster economic growth and improved wellbeing**.

Advanced AI systems could produce significant economic value (Trammell and Korinek, 2020). Although much uncertainty remains regarding their aggregate economic impacts on growth rates and labour markets (Aghion, B. Jones, and C. Jones, 2017; Korinek and Suh, 2024), they will almost certainly raise total factor productivity, and thus economic growth rates, by some factor. Some economists think this factor will be modest (Acemoglu, 2024). Others

argue that scenario planning for far-reaching automation scenarios that drive significant growth within a decade is warranted (Korinek and Suh, 2024). Today, the most advanced AI systems increase labour productivity for software-intensive tasks, but remain less helpful for manual work (Brynjolfsson, Li, and Raymond, 2023; Cui et al., 2024; Eloundou et al., 2024). In the future, though, because AI systems may be able to very cheaply substitute for labour, the possibility of much larger growth effects cannot currently be dismissed (Trammell and Korinek, 2020).[5]

However advanced AI systems affect growth rates, the goods, services, and profits they produce could be used to support a range of societal goals, including in less developed economies. Nearly 700 million people live in extreme poverty around the world (World Bank, 2024b). If the current rate of improvement continues, this number will only fall by 16% to 590 million by 2030 (United Nations, 2024). Just 17% of the targets used to track progress towards the SDGs – of which eliminating global poverty is just one – are on track to be met by 2030 (ibid.). Advanced AI systems, and the revenue they generate, have the potential to help humanity meet more of these goals and others in low- and middle-income countries – for example, through increasing productivity in key sectors or improving the efficiency of resource allocation.

At the same time, there is concern that this will not happen, or will not happen quickly enough, without benefit-sharing efforts. Running AI systems requires computing hardware and reliable internet access, meaning people in countries with lower rates of computer ownership, internet access, cellular coverage, and other digital infrastructure may realise fewer benefits from AI (Okolo, 2023). While there is significant uncertainty about how advanced AI systems will affect productivity across countries, some analyses suggest that AI could benefit high-income countries more substantially, or sooner, potentially driving economic divergence globally (Korinek, Schindler, and Stiglitz, 2021; Nii-Aponsah, Verspagen, and Mohnen, 2023).

Market forces alone might cause the economic and societal benefits of AI systems to reach developing countries over time. Growing populations and rising incomes in many of these countries mean they are increasingly important markets in their own right. Moreover, software, including AI, tends to diffuse more easily than physical technologies. For instance, access to ChatGPT, Gemini, Claude, and Meta's Llama 3.1 are already available in most countries worldwide.[6]

However, benefit-sharing efforts could make it more likely that these economic and societal benefits diffuse globally, and that they diffuse more quickly. Such efforts could take a variety of forms, including financial transfers, investments in digital infrastructure, and increasing access to AI technologies suited for local economies, cultural contexts, and use cases. These options, among others, are discussed in more detail in **Section III**.

[5] For an overview of the arguments for and against the potential of AI to drive rapid economic growth as a result of automation, see Erdil and Besiroglu, 2023.

[6] As of January 5, 2025: For ChatGPT, the excluded UN member states for both website and API access are Belarus, China, North Korea (DPRK), Iran, Russia, Syria, and Venezuela, as well as certain parts of Ukraine under conflict. For Gemini, the excluded countries for both website and API access are Belarus, China, North Korea (DPRK), Iran, Russia, and Syria; additionally, Myanmar is excluded from accessing the Gemini API. For Claude, the excluded countries for both website and API access are Belarus, China, North Korea (DPRK), Iran, Libya, Myanmar, Russia, Syria, and Venezuela, as well as certain parts of Ukraine (Anthropic, 2024b; Google, 2024; Google, 2025; OpenAI, 2024c; OpenAI, 2024a). For Llama 3.1, availability depends on the platform offering access (e.g., AWS, Google Cloud, Hugging Face) but it is also widely available globally.

Benefit sharing could help ensure that the economic benefits from labour-automating AI are broadly distributed

The second motivation for benefit sharing in this category is to **ensure that advanced AI has net benefits for workers whose jobs are automated.**

As AI systems become more generally capable, they may be increasingly able to substitute for human labour. Indeed, OpenAI's mission, in its own words, is to develop systems that can outperform humans at "most economically valuable work" (OpenAI, 2018). In a large survey of leading AI researchers, the aggregate estimate was that there is a 50% chance that AI will outperform humans in every possible task by 2047 (Grace et al., 2024). The accuracy of such long-term forecasts is unclear, and the forecasts from the experts in that survey varied widely. Moreover, significant bottlenecks may limit the degree and pace of labour automation (Svanberg et al., 2024; Erdil and Besiroglu, 2023). However, it is clear that there is at least some chance that advanced AI systems will be highly disruptive to labour markets, with productivity gains coming at the expense of some jobs (Eloundou et al., 2024; Korinek, 2024).

How such disruptions are likely to unfold is highly uncertain (Bengio, Mindermann, et al., 2025). However, historical precedents offer some insight. The effects of previous waves of automation have varied across economic sectors, with positive long-run net economic effects for both workers and consumers. But there have been notable short-term disruptions. For example, one study using data from the U.S. shows that in industries and regions with higher rates of industrial robot adoption, overall productivity rose but the share of income going to workers shrank (Acemoglu and Restrepo, 2017). Other research attributes more than half of the change in wage inequality between 1980 and 2020 in the US to automation-driven reductions in workers' wages (Acemoglu and Restrepo, 2022).

If AI systems come to outperform expert humans on complex tasks across a range of fields, the impacts on labour will be more significant than past automation. Wide-scale automation could accelerate economic growth while reducing demand for labour across a range of jobs (Korinek and Suh, 2024). While controversial, some researchers think that, beyond a certain level of capabilities, automation from advanced AI could largely eliminate the availability of work (Korinek and Suh, 2024; Susskind, 2020). This would shift income away from workers, with the economic returns mostly going to owners of AI capital (Acemoglu and Restrepo, 2018; Acemoglu and Restrepo, 2019; Korinek and Juelfs, 2022). Although higher productivity could reduce the cost of goods and services, which would have many beneficiaries, those workers who struggle to find new work could end up in worse situations (Bengio, Mindermann, et al., 2025). These effects may prove important both within and across countries: labour-automating AI could erode the comparative advantage that many low- and middle-income countries currently have in labour-intensive activities, potentially reducing their national economic competitiveness (Korinek and Stiglitz, 2021; Nii-Aponsah, Verspagen, and Mohnen, 2023).

The social costs of these labour market disruptions could be severe. The wellbeing of displaced workers is likely to suffer, especially if the scale of automation makes it difficult for them to find new employment opportunities (Gedikli et al., 2022; Brand, 2015). There could also be broader risks to social stability, with higher unemployment and inequality leading to political polarisation, a backlash against productivity-enhancing AI progress, and antisocial

behaviours. Benefit sharing could help address these concerns by proactively setting up mechanisms to ensure that the productivity gains from AI are broadly accessible across the globe. This could include sharing of resources to support workers who may otherwise suffer hardship as a result of automation, as well as AI resource and access-sharing initiatives to help low- and middle-income countries maintain their economic competitiveness in an increasingly automated global economy.

II.2 Technological Self-Determination Motivation

Benefit sharing could promote technological self-determination in low- and middle-income countries

The second motivation for benefit sharing is focused on the effects of advanced AI on international power dynamics. The concern stems from two observations: that an understanding of, and competency with, AI systems may be a critical determinant of economic and national power in coming decades; and that AI development is currently concentrated in high-income countries, leaving few opportunities for low- and middle-income countries to develop domestic AI industries.

Some experts contend that this could create a form of technological dependency between high- and low-income nations (Birhane, 2020; Png, 2022). Actors in low-resource settings could become increasingly dependent on AI systems to perform a range of critical functions, and access to those systems could be primarily controlled by high-income nations. Benefit sharing could mitigate this dynamic by cultivating the development of AI talent and systems in low- and middle-income countries, leaving them less reliant on external technology and expertise. As such, **benefit sharing could promote *technological self-determination*** – a nation's capacity to independently develop, adopt, and regulate technologies in alignment with the values, needs, and strategic interests of its people, without undue external influence or control – in a wider range of countries.[7]

The concentration of advanced AI development in a small number of countries (the US, China, the EU, and the UK) (Maslej et al., 2024) threatens to undermine technological self-determination on multiple fronts. First, it could enable dependency relationships with the potential for exploitation. Leading AI states may leverage their control over AI resources to extract geopolitical concessions, just as countries with control over critical resources like oil or financial infrastructure do today. Second, countries without advanced AI capabilities may struggle to shape or regulate AI technologies that are developed externally, but deployed within their own borders.

There are several ways benefit-sharing initiatives may help address these concerns. For example, AI companies, governments of high-income countries, or multilateral institutions

[7] It should be noted that technological self-determination can also refer to the capacity of groups other than nations to adapt technologies to their values and needs. In particular, it can refer to the right of Indigenous communities to shape the influence of technology on their communities, especially data relating to their people, lands, and knowledge systems. For example, this concept underlies the First Nations principles of OCAP (Ownership, Control, Access, and Possession) in Canada for Indigenous data governance, and Māori data sovereignty initiatives in Aotearoa (New Zealand) such as the Te Mana Raraunga network.

could subsidise AI development resources like computing power ("compute"), infrastructure and technical training programs for low- and middle-income countries. This could help these countries build indigenous AI research capacity and develop locally-optimised AI applications (see **Section III**). Increased economic participation could foster greater technical capabilities and increased influence over the development of the technology.

It is worth noting that issues of technological self-determination are particularly sensitive, raising credible concerns about the downsides of benefit-sharing efforts as a whole. They highlight, for example, concerns that some benefit-sharing initiatives could in fact entrench technology dependencies. For example, frontier AI states could renege on benefit-sharing commitments. If recipient countries had been planning on receiving those benefits, frontier AI states suddenly withdrawing them could leave recipient countries more vulnerable and dependent on external support than they were previously. Similarly, if a benefit-sharing arrangement was heavily conditional or came with stringent requirements – for example, restrictions on developing competing AI models or requirements to exclusively use the providing company's cloud computing services – this could potentially undermine a recipient nation's ability to build independent technological capabilities.[8]

Benefit-sharing initiatives aiming to promote technological self-determination will need to be designed with these risks in mind. Commitment and accountability mechanisms will likely be needed.

II.3 Geopolitical Motivations

Finally, a third set of motivations highlights the role that benefit sharing could play in **advancing geopolitical interests of frontier AI states and enabling international cooperation on AI governance.** Unlike the previous two motivations, which focused on fostering broad economic growth and empowering less developed countries, this category views benefit sharing as a diplomatic lever for states to advance their national interests. For instance, access to AI benefits could be tied to participation in international AI safety agreements or frontier AI states could use benefit sharing to establish or strengthen strategic partnerships in line with their vision for AI development and governance. Expanding access to AI products could also be a way for leading companies and states to gain a larger share of emerging markets relative to their global competitors. Here we discuss four specific motivations in this vein.

Benefit sharing could encourage the adoption of international safety standards

First, benefit sharing could incentivise countries to participate in **international efforts to reduce major risks from advanced AI systems.**

[8] The risks of conditionalities in benefit sharing are documented in other sectors. For example, many technology transfer initiatives under the Kyoto Protocol's Clean Development Mechanism included restrictive intellectual property clauses. While these can be standard practice, they have also created significant barriers for recipient countries to adapt or build upon shared green technologies within their local contexts.

Advanced AI systems pose a range of risks, including privacy breaches, cybersecurity threats, unintended biases, and misuse to enable the development of chemical or biological weapons (Slattery et al., 2024; US Department of Homeland Security, 2024). Many of these risks could cause harm that crosses borders, with actions taken in one country having serious effects elsewhere (Dennis et al., 2024). As AI capabilities advance, states may have to cooperate to manage these risks and ensure advanced AI systems are developed and used responsibly (Ho et al., 2023). These efforts may involve establishing common safety standards, implementing risk monitoring systems, and creating channels for information sharing to verify compliance. Some such initiatives are already underway. For instance, the UK AI Safety Institute is developing model evaluation information-sharing channels with various national and international partners (Department for Science, Innovation and Technology, 2024).

Some of these cooperative efforts, however, face a collective action problem. Although all states may benefit in aggregate, individual states may have an incentive to ignore common standards. Many countries, including states with less advanced AI sectors, may seek to circumvent international agreements, or may not join them at all if those states are concerned that participation in such international agreements could undermine efforts to strengthen or develop their domestic AI industries.

Benefit-sharing arrangements could help change these incentives. In particular, access to AI benefits could be made conditional on participation in various international cooperation efforts on AI, including risk mitigation components. Similar "Grand Bargain"–style agreements have been successful in other domains. For example, the US could propose a global AI agreement akin to the "Atoms for Peace" agenda that was focused on promoting the positive uses of nuclear technology while mitigating its risks to global security (O'Keefe, 2024). Such an initiative would form an alliance of countries committed to safe AI development and non-proliferation of high-risk frontier AI models to non-member states.

Benefit sharing could reduce risks from competitive dynamics in AI development

Benefit sharing could also **reduce risks arising from competition between countries and states** racing to deploy increasingly powerful AI systems. Being the first actor to develop and deploy a new AI system can have significant advantages, allowing companies to capture market share and countries to gain geopolitical leverage. But in seeking to gain these advantages, actors may take on more risk in AI development than is socially optimal (Stafford, Trager, and Dafoe, 2022).

Benefit sharing could make actors less likely to race to deploy AI technologies by reducing the perceived costs of "losing" the race. Frontier AI states could commit to sharing AI benefits – such as a portion of financial windfalls and useful technologies – with non-frontier states to prevent them from pursuing risky development strategies of their own (Justen, 2024). If those commitments are credible, the recipient state may stand to gain more from reducing its own AI investments and enjoying the shared benefits than it would from engaging in a risky and uncertain winner-takes-all race.

The strategic dynamics of benefit sharing under adversarial conditions are complex. Some models of AI competition suggest that certain forms of benefit sharing might inadvertently

intensify race dynamics (Stafford, Trager, and Dafoe, 2022). For instance, sharing specific AI technologies could narrow the gap in AI capabilities between leading and lagging states. This, in turn, might increase the likelihood of the lagging state winning an AI race, thereby encouraging them to participate in a race they might have otherwise avoided. This highlights the importance of carefully analysing any given benefit-sharing proposal for its effect on strategic dynamics. For example, it may be important to include monitoring and evaluation procedures in such proposals to verify safe AI development practices. We discuss these considerations for various AI benefit-sharing options more in **sections III** and **IV**.

Benefit sharing could bolster collective security

The third motivation in this category has to do with **leveraging benefit sharing to strengthen collective international security.** For instance, sharing AI with defensive cyber capabilities could help prevent attacks that might spread across borders, ultimately protecting both frontier AI countries and countries receiving benefits. AI systems can also aid in pandemic prevention by processing public health data to detect the spread of potentially dangerous pathogens (Parums, 2023; Brownstein et al., 2023). As the capabilities of general-purpose AI systems improve, they will likely be able to perform other security-enhancing tasks.

Risk-reducing technologies have been shared widely in other fields. For example, under the principle of space safety, the US shares space surveillance data with other governments, including China, to prevent satellite collisions (Silverstein, 2023). In this case, the collective security benefits outweigh the challenges of sharing sensitive data even in a low-trust environment.

Benefit sharing could establish or strengthen strategic partnerships between countries

The final motivation to discuss is the possibility of using benefit sharing to **strengthen strategic partnerships between frontier AI states and international actors.** Frontier AI states can promise to share AI benefits with countries in exchange (either implicitly or explicitly) for diplomatic support on AI and other geopolitical issues. Benefit sharing may thus help align countries behind particular foreign policy goals and visions of international AI governance. For example, the US may view benefit sharing as a way to build a coalition of countries that support its goals for AI governance, such as adopting certain security practices or preventing adversaries from obtaining resources for AI development.

There is already evidence that frontier AI states recognise the strategic value of sharing AI benefits to achieve foreign policy goals. China's Global AI Governance Initiative, for example, names "shared benefits" as one of its three fundamental principles (Belt and Road Forum for International Cooperation, 2023). This commitment includes supporting AI infrastructure development in low- and middle-income countries and providing access to AI products and services. The US's Partnership for Global Inclusivity on AI also indicates a growing interest in benefit sharing, through pledging $100 million in funding for capacity building and expanded access to AI in developing countries (US Department of State, 2024). These commitments could be used by powerful states at the forefront of AI development to strengthen strategic

partnerships with other countries who may then be more inclined to align with them on AI governance issues.

Initiatives to promote domestically developed AI technology abroad can also serve the long-term economic interests of frontier AI states (Export-Import Bank of the United States, 2023). For example, extending access to US technology and digital infrastructure along the AI supply chain may help US companies secure a larger market share in emerging markets relative to China and other geopolitical rivals.

This strategic dimension of expanding access to AI benefits globally could prove particularly important as AI governance goals of frontier states diverge. Both US and Chinese leaders have stated that global leadership in AI is a national priority.[9] The two countries are competing to shape how multiple strategic technologies, including AI, are developed, governed, and deployed globally (US National Security Commission on Artificial Intelligence, 2021; Clare and Ruhl, 2024). One way this competition manifests is in US-led restrictions on exporting certain technologies related to AI chips to China – restrictions that require international cooperation to implement and enforce.[10] Leading actors like the US or China may decide to share access to benefits to secure international support for similar policies that attempt to limit the flow of pivotal AI technologies to key competitors (Winter-Levy, 2024).

As global AI governance initiatives advance, this competition increasingly reflects fundamental differences in values and goals for frontier AI states. The US explicitly frames AI governance as critical to preventing the technology from bolstering authoritarianism and undermining democratic institutions worldwide (The White House, 2024). Many existing multilateral AI governance efforts that include the US, like the G7 Hiroshima AI Process, also explicitly promote "democratic values" and the "rule of law" in AI (G7 Leaders, 2023). This framing may be in tension with the governance goals of the Chinese Communist Party, which prioritise national sovereignty and sociopolitical stability (Hine, 2024).

In scenarios where the US leverages its current AI development lead and supply chain advantages to gain an extremely dominant position in AI, it may also consider engaging in restricted and conditional forms of benefit sharing with competitor states. The purpose of such sharing could be to increase international stability by providing reassurance that the fundamental economic and security interests of competing states will be preserved. However, any benefit-sharing programs targeted at competing states would pose greater political and security challenges.

[9] For the US, see, e.g., the 2024 *Memorandum on Advancing the United States' Leadership in Artificial Intelligence*; for China, see State Council (2017) Xin Yidai Rengong Zhineng Fazhan Guihua *A new generation artificial intelligence development plan*.

[10] Given the international nature of the AI chip supply chain, the US has had to coordinate with states including Taiwan, the Netherlands, Japan, and the Republic of Korea.

III Options for AI Benefit Sharing

We now turn to the question of what different international AI benefit-sharing options could look like in practice. We identify three things that could be shared to accelerate the spread of AI benefits internationally: **1) the resources used to develop AI systems, 2) access to advanced AI systems,** and **3) financial proceeds from AI development.** These correspond roughly to three different stages of the AI value chain (figure 2).

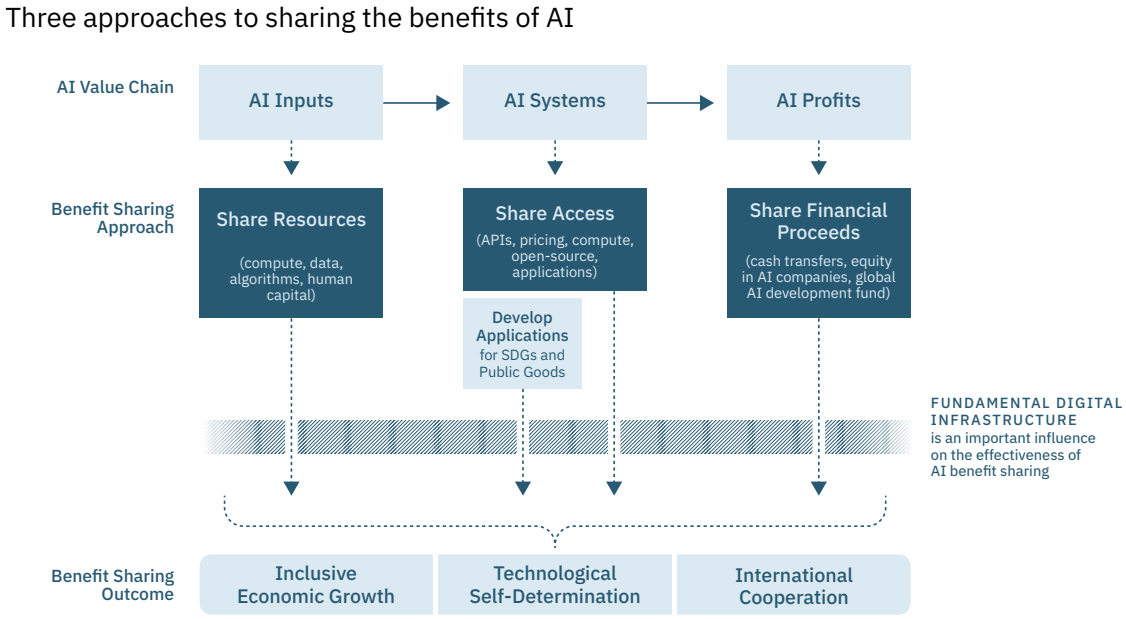
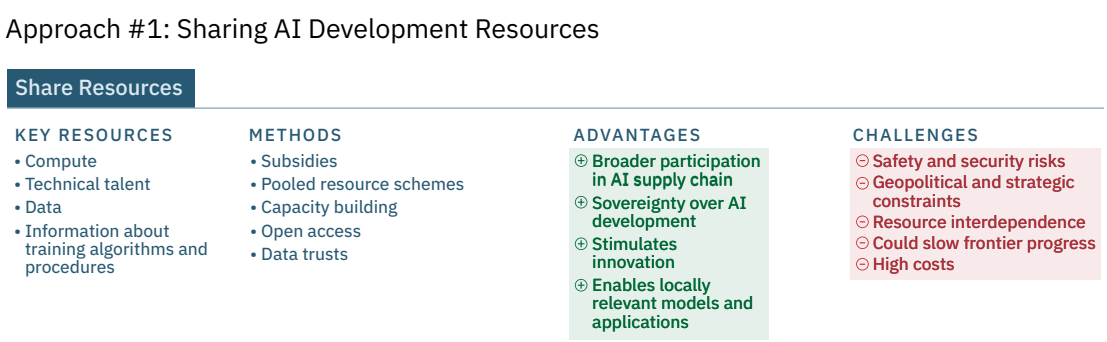


Figure 2: Three approaches to AI benefit sharing. This diagram illustrates three approaches to benefit sharing across the AI value chain: sharing resources, sharing access, and sharing financial proceeds. These approaches aim to promote inclusive economic growth, expand opportunities for technological self-determination, and advance geopolitical interests and strategic cooperation on AI governance.

Choosing among different benefit-sharing options involves making difficult compromises, and some options are better suited to one of the motivations discussed in **section II** than others. They also vary in terms of whom, specifically, they are most likely to benefit. For example, sharing financial benefits may directly benefit those most in need in lower-income nations, while sharing AI development resources may initially directly benefit AI developers in low- and middle-income countries, with flow-through effects for other citizens. A comprehensive, effective system for sharing benefits is likely to combine several of the specific options discussed below.

The range of options and considerations also means that actors seeking to implement benefit sharing are also likely to face challenges. In this section, we briefly discuss challenges specific to each benefit-sharing option at the end of each subsection. Then, in **section IV** we cover some of the major challenges that are relevant to all benefit-sharing options in more detail.

III.1 Sharing AI Development Resources



First, **the resources needed to develop, fine-tune, evaluate, and deploy advanced AI systems could be shared.** These resources include compute, technical talent, data, and information about the algorithms and the methods used to train and fine-tune AI models. In this section we discuss specific ways to share these resources:

1. **Compute**: Creating shared compute pools, subsidizing access to cloud resources, or supporting domestic compute industries.
2. **Technical Talent**: Building global training programs, fostering local AI expertise, and reducing brain drain through local opportunities.
3. **Data**: Expanding open datasets, creating public data trusts, and supporting data-efficient training techniques.
4. **Information about Training Algorithms and Procedures**: Sharing details about algorithms and model training procedures in order to advance collective knowledge about AI development.

Currently, many of these resources are scarce, highly concentrated, or costly, particularly at the frontier of AI development. Efforts to make them more widely-accessible could help accelerate the spread of AI benefits in at least two ways.

The first is by **stimulating economic growth.** Transferring resources and technology could make AI businesses and entrepreneurs in developing countries more competitive. Benefit sharing could create economic opportunities at multiple points along the AI value chain. For example, technology transfer and industrial coordination initiatives could accelerate the development of local industries. These could be related to the various components of the compute supply chain (see Figure 3), from semiconductor manufacturing or packaging to data center construction and operation. Sharing AI development resources may also support industrial development at later stages of the AI supply chain, such as in application development, where access to compute, data, and technical talent is necessary to build and run beneficial AI applications at scale.

The second is by **enabling the development of AI systems that meet local needs.** A broad range of AI applications are currently built on top of a few generally-capable models. This homogenization supports economies of scale and interoperability, but also presents challenges.

For example, most leading foundation models today are optimised for English-speaking users, raising performance and bias issues for use in other languages (Blasi, Anastasopoulos, and Neubig, 2022; Gallegos et al., 2024). Sharing AI resources could enable local developers in low- and middle-income countries to build AI systems that are tailored to a country's specific needs and thus produce more societal benefits (Adan et al., 2024). This could go beyond mere localization of existing AI applications developed in other contexts or languages and include the development of entirely new AI systems and applications. For instance, local developers may be more motivated to prioritise AI development that addresses region-specific health challenges, enhances local agricultural practices, or preserves and applies indigenous knowledge. It could also involve developing AI systems optimised for areas with limited computational resources or intermittent internet connectivity.

In the following subsections, we'll examine each of these key AI resources in more detail. Before examining the specifics of sharing AI development resources, however, it is crucial to acknowledge one supporting condition that influences the potential effectiveness of each approach to AI benefit sharing discussed in this report: **access to fundamental digital infrastructure** (see Box 1).

Box 1: Access to Fundamental Digital Infrastructure

Approximately one-third of the world's population does not have reliable internet access (Ritchie et al., 2024), effectively excluding them from many potential benefits of accessing current AI technologies. This digital divide is not evenly distributed across the globe. While regions such as Europe and the Americas boast internet usage rates of around 90% (International Telecommunication Union (ITU), 2025), Sub-Saharan Africa lags significantly behind with only 37% of its population online as of 2023 (World Bank, 2024a).

The root causes of this disparity are multifaceted. Inadequate access to electricity and insufficient investments in essential internet infrastructure, such as fibre optic cables, subsea cable landings, and cellular towers, play significant roles (Boakye et al., 2023). Moreover, the divide isn't just between countries but also within them. A recent study in Sub-Saharan Africa, for instance, revealed that 47% of urban residents have internet access, versus only 12% in rural areas (Seuyong et al., 2023). Additionally, women in this region as a whole were found to be 37% less likely than men to use mobile internet (Delaporte, 2023).

Addressing this digital divide is fundamental to ensuring that the gains from any benefit-sharing approach can be maximised. While it is beyond the scope of this report to discuss detailed proposals to improve fundamental digital infrastructure in low- and middle-income countries, two general categories of assistance that could complement benefit-sharing initiatives involve:

- **Expanding internet connectivity infrastructure.** Strategic investments are needed to build comprehensive broadband networks across underserved regions. This could involve installing essential physical infrastructure like fibre optic

networks and subsea cables, constructing cellular towers, and deploying satellite internet solutions to reach remote areas. Such an expansion could be funded through collaborative efforts between governments, international development agencies, and technology companies, for example.

- **Ensuring access to AI-capable devices:** For digital infrastructure to enable AI-driven benefits, people need access to computers and smartphones with sufficient processing power to run beneficial AI applications. Though most of the world currently has access to mobile phones (Our World in Data, 2024), not all are able to run such applications efficiently. Making AI-capable devices widely available may require a coordinated approach: global manufacturers could partner with local distributors to ensure supply, while government subsidies and innovative financing models supported by international financial institutions could make devices affordable for low-income populations. A core goal would be to prevent hardware costs from becoming a barrier to AI adoption, even in areas where reliable internet infrastructure exists.

Limited digital infrastructure particularly constrains the effectiveness of benefit-sharing approaches that focus on sharing AI development resources or AI access. But it also limits the effective use of shared financial proceeds by restricting the range of economic and technological solutions that could be implemented in regions without access to fundamental digital infrastructure.

Compute

Advanced AI systems are trained using enormous amounts of compute (Sastry et al., 2024). In general, using more compute in training produces more capable AI models (Kaplan et al., 2020). Compute is also needed to use AI systems after training, and there is evidence that using more compute at the time of "inference" improves AI performance at tasks such as mathematics and general reasoning problems (OpenAI, 2024d). This makes it one of the most critical resources for advanced AI development and deployment.[11]

Compute is also one of the most costly resources needed to develop advanced AI systems. Accessing it requires specialised AI chips such as graphics processing units (GPUs) which are extremely difficult to manufacture. Highly-specialised machinery supporting infrastructure are needed to produce chips, from software stacks and networking equipment to cooling systems and sufficient power sources. As a result, the chip supply chain includes multiple "chokepoints" controlled primarily by single companies (e.g. ASML for lithography machines or NVIDIA for advanced chip design) (Miller, 2022; Sastry et al., 2024). Soaring demand and the concentration of the supply chain mean that actors with fewer resources will likely struggle to access the compute required to develop or use advanced AI systems. A "compute divide" has developed both within and between nations, with the US and China controlling far more computing resources than any other country (Organisation for Economic Co-Operation and Development, 2023; Besiroglu et al., 2024).

[11] For context, an estimated $78 million worth of compute was used to train OpenAI's GPT-4, while Google spent $191 million on compute to train Gemini Ultra.

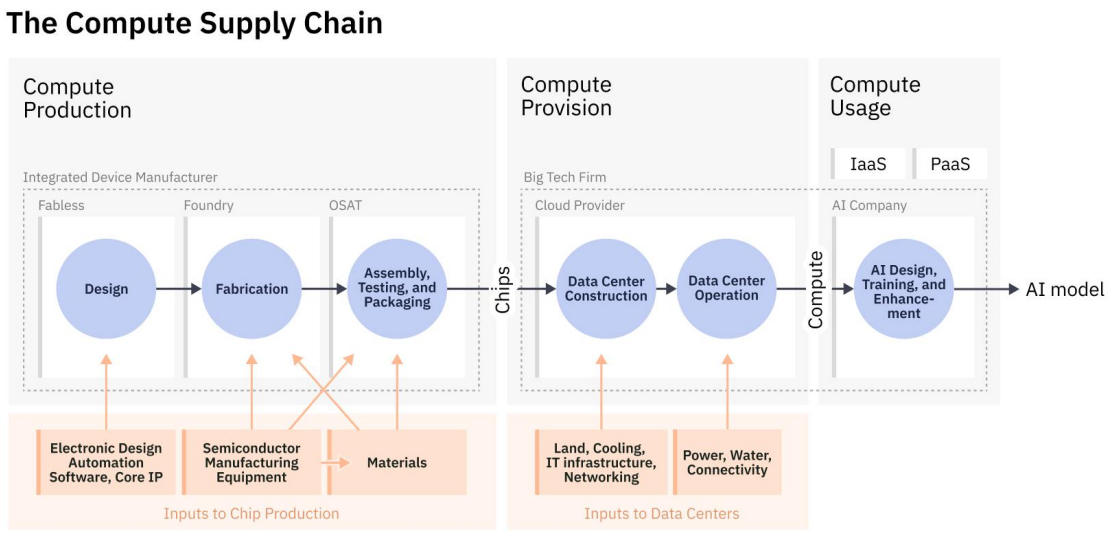


Figure 3: The compute supply chain. Resources could be shared to enable industrial development along various points in the compute supply chain, from the compute production stage (e.g., sharing software used to design AI chips) to the compute provision stage (e.g., sharing resources to build and operate new data centres) to the compute usage stage (e.g., sharing cloud computing credits). Note: Figure adapted from Sastry et al., 2024.

Still, several features of compute make it a potentially attractive resource to share. For one, it is a flexible resource that could be used as an input to train new AI models, fine-tune existing ones, or power specific applications. That means actors in developing countries could use shared compute for a variety of purposes. Another advantage is that compute resources can be readily distributed. Physical chips can be purchased, donated, or reallocated, making it possible to supply hardware directly to specific actors. Additionally, cloud-based compute resources offer flexibility – access to these resources can be granted, adjusted, or withdrawn as needed, enabling tailored support for different users or projects.

How could compute be shared? Given the value frontier states place on accumulating and, in some cases, controlling access to chips, simply purchasing or subsiding purchases of computing resources for other state actors directly will often be in opposition to their incentives. However, compute could still be shared in at least two ways: by creating shared compute pools and sharing access, and by subsidising the provision of compute for certain developers.

First, **states or multilateral organisations could create a shared pool of compute resources**, similar to how multiple states pooled resources to conduct particle physics research at CERN (CERN, 2025). Access to these pooled resources could then be granted to users according to certain criteria. These should be worked out in consultation with the intended users and beneficiaries, but could include things such as their location and whether they intend to use the resources to generate benefits for low- and middle-income countries. Similar initiatives already exist: regional compute "banks" have been proposed, while national governments have funded public supercomputers (Sastry et al., 2024; Boakye et al., 2023).

Second, **various actors could subsidise compute for certain developers**. These could include developers in low-income countries, or developers building AI tools that aim to benefit disadvantaged populations. Such subsidies could be provided by governments, philanthropists, multilateral institutions, or even AI developers or cloud compute providers themselves.

These subsidies could be applied to chip purchases or to cloud compute services. Subsidising compute access via the cloud may give more flexibility to the subsidy providers. Actors in low- and middle-income countries could be given access to compute without the need for large and illiquid hardware investments, and this access could be scaled up or down as their needs change. Cloud compute subsidies could also differentiate between infrastructure-as-a-service (IaaS) providers and platform-as-a-service (PaaS) providers (Heim and Egan, 2023; Heim, 2024a). The former offer direct, virtual access to hardware resources, making it particularly valuable for technically advanced users who want to optimise their compute usage or develop novel AI architectures. In contrast, PaaS providers offer compute access via pre-configured tools and software, simplifying development by letting users focus on building applications rather than managing infrastructure. Sharing cloud compute via IaaS or PaaS providers can extend resources to a broader range of users, democratising access to compute inputs while allowing flexible options based on technical needs and resource management preferences.

Some initiatives in this vein have already been launched. The US established the National AI Research Resource (NAIRR) to provide compute infrastructure for academic researchers, while the UK is developing its AI Research Resource (AIRR). The EU has invested in supercomputing through EuroHPC, and other nations are making similar investments in national compute infrastructure. However, these initiatives remain predominantly concentrated in high-income countries, potentially reinforcing existing global inequalities in compute access rather than addressing them.

While other forms of compute sharing are possible, they may be less feasible than simply subsidising compute access for developers. For example, resources could be shared to support the development of domestic compute industries. This could include investments in collaborative semiconductor research and development, specialised manufacturing facilities, or workforce training for chip design and fabrication. Such initiatives have the potential to be mutually beneficial, providing economic opportunities for recipient countries and alleviating supply chain bottlenecks for frontier states.

At the same time, the extreme concentration of the semiconductor industry suggests that gaining footholds in the industry is far from straightforward. Developing the capacity to manufacture semiconductors, for example, requires substantial capital and a high level of existing expertise, and may only be viable in regions that already have some involvement in advanced manufacturing. Investments are more likely to succeed if they aim to build on established industrial capacities rather than create entirely new industries. While entering into such a highly specialised part of the compute supply chain would be difficult, it could be more feasible for developing economies to break into other areas. For example, they may be able to capitalise on economic opportunities elsewhere in the semiconductor supply chain, such as in semiconductor packaging and assembling, or constructing and managing data centres.

Finally, it is also worth noting that compute-sharing schemes raise a number of other risks, including some related to international security, that need to be considered. We discuss these at length in **section IV**.

Technical talent

A second AI development resource that may be leveraged to support benefit sharing is technical talent: the skilled workforce needed to develop and deploy advanced AI.

This, too, is a scarce and concentrated AI resource (Musser et al., 2023). States and firms currently compete to develop and attract the most talented AI engineers and researchers. The demand for top talent is reflected in the compensation packages offered by leading AI companies, where top researchers can earn over $1 million annually (Bindley, 2024). But even outside of leading firms, technical AI talent is needed to adapt or develop AI applications suited to local needs, help firms in other industries integrate new AI applications into their workflows, and inform government programmes and policies related to AI. AI talent development, though, overwhelmingly takes place in high-income countries like the US, Canada, and the UK (Okolo, 2023). Additionally, experts from developing countries are often incentivised to join well-resourced companies in higher-income nations; for example, in 2022, reportedly only one-fifth of AI researchers trained in India stayed in India (MacroPolo, 2022). This suggests there is a need to not only cultivate AI talent, but also create attractive local employment opportunities for AI researchers and practitioners.

To support global capacity-building and domestic talent development programs in low- and middle-income countries, international organisations, groups of states, or leading AI companies could **fund initiatives to develop and retain AI talent.** These may include university partnerships, technical training centres, research grants, and professional development programs, for example.

This has been done, to an extent, for other critical technologies. For example, the International Atomic Energy Agency supports capacity development in member states through technical cooperation projects related to nuclear energy (International Atomic Energy Agency, 2016). Recently, leading AI companies have taken steps in this direction as well. For example, as part of a $1 billion digital ecosystem initiative for Kenya, Microsoft and G42 are launching an East African Innovation Lab in Nairobi to support local entrepreneurs and organisations developing AI services (Microsoft, 2024).

Data

Data is another key resource for AI development, particularly for training and fine-tuning models. Massive amounts of data are needed to train the most advanced systems. Some of the largest known training datasets include more than 15 trillion tokens of text – equivalent to approximately 128 million books (Epoch AI, 2023). These data come from multiple sources: public web scraping, licensing deals with content providers, private data from third parties, paid data labelling services, and synthetic data generated by AI developers themselves (Anthropic, 2024c; Dubey et al., 2024). Compiling and curating such large, high-quality datasets is costly. This gives well-resourced AI developers an edge in creating or acquiring the most valuable datasets. The growing importance of synthetic data (Villalobos et al., 2022;

Gerstgrasser et al., 2024) – artificially generated content used for training, fine-tuning and evaluating models – may further entrench these inequities, since generating large amounts of high-quality synthetic data requires access to powerful AI systems in the first place. This may create a feedback loop, where those with the best data are able to create better synthetic data, widening the gap between actors.

These factors suggest that sharing data could be an effective way to support equitable access to advanced AI systems by reducing the barriers to training or fine-tuning models. Data could be shared in a number of ways.

Public data trusts have been proposed as a way to enable access to training data while ensuring fair compensation for data providers (Chan, Bradley, and Rajkumar, 2023). In this model, data would be licensed to AI developers with a percentage of revenues from deployment being redistributed back to the trust. This could create sustainable incentives for expanded data sharing with developers while helping to address concerns about data ownership and compensation.

Open datasets, which are freely available for anyone to access and use, could further democratise access to training data. Existing initiatives include LAION, which includes hundreds of millions of text-image pairings, and CommonCrawl, which includes billions of scraped webpages. Other efforts, such as Cohere's Aya project, demonstrate a path forward for creating publicly available datasets for training and fine-tuning models with certain high-value data (Aya collected data from dozens of underrepresented languages around the world). However, while these datasets are large, they may be insufficient on their own for training advanced AI systems.

Training data documentation could increase transparency if companies share detailed information about the data used to train and fine-tune their models. This could help other developers understand what types of data are most valuable for training and potentially guide their own data collection efforts. However, as competition among frontier AI developers intensifies, companies are becoming more restrictive about sharing this information (Bommasani et al., 2023). The most valuable proprietary datasets – whether high-quality labelled data or synthetic data for specific training or fine-tuning purposes – are likely to remain closely guarded.

Research on data-efficient learning techniques focuses on developing improved training methods that require less data overall (Sachdeva et al., 2024). More data-efficient training methods would reduce dependence on vast and specialised datasets, potentially lowering barriers for new entrants and broadening opportunities for innovation. This approach may be particularly promising because it aligns with the interests of leading AI companies while potentially making AI development more accessible to a wider range of organisations and developers.

Information about training algorithms and procedures

Details about training algorithms and procedures used to develop an AI system are yet another key input to AI development. Sharing information about these procedures could make it easier for more people to train these systems, which could accelerate innovation globally and spread benefits more widely.

A range of information about how an AI system was trained could be shared. At the extreme, this could include **sharing the code, model architecture, information about the software and hardware that was used, fine-tuning procedures**, and more**.** In practice, all the information another developer would need to fully replicate the training procedure is rarely shared (Liesenfeld and Dingemanse, 2024). For many companies, this information is a core aspect of their intellectual property and fundamental business model. However, some amount of information about model training is sometimes made available when a new AI system is released (e.g., Meta, 2024a; Anthropic, 2024c; Technology Innovation Institute, 2023). While this transparency can advance collective knowledge, the practical impact of sharing training information is often limited by barriers like limited access to computing resources and high-quality training data. This creates a dynamic where knowledge about AI training algorithms and procedures could be made widely available, but the capability to actually train advanced AI models remains concentrated among organisations with greater access to other resources.

Challenges

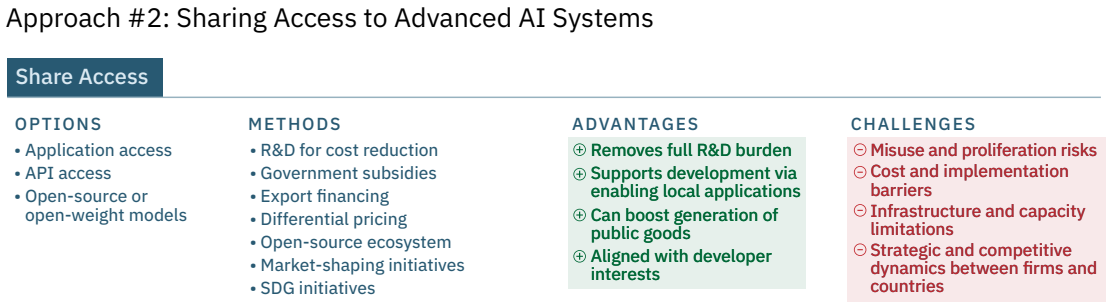
AI resource-sharing initiatives come with several significant challenges and risks.

These include:

- **Safety and security risks** from broader access to the inputs needed to develop powerful AI systems. Distributing AI development resources more widely could enable malicious actors to develop harmful AI applications or accelerate unsafe AI development practices. Risks stemming from this could include more AI-enabled cyberattacks, automated disinformation campaigns, or AI-aided biological weapons development, for example (Bengio, Mindermann, et al., 2025; Hendrycks, Mazeika, and Woodside, 2023). Each of these risks could have international impacts.
- **Geopolitical and strategic constraints** on leading states' willingness to share compute and other AI development resources widely. For example, states may be reluctant to share AI resources if they think it will diminish their competitive advantage in AI development.
- Technical challenges related to **resource interdependence.** Since developing or fine-tuning advanced AI systems requires all of the resources discussed above, the value of sharing large amounts of one resource in isolation may be limited. For example, the benefits of increased access to large amounts of advanced compute may be limited for developers or countries that lack the technical talent to use it effectively.
- Shifting AI development resources away from leading developers may **slow frontier AI progress**, which could, in turn, reduce the rate of scientific advancement and delay the development of beneficial applications. If these frontier advances would have produced substantial benefits that naturally spread widely, then prematurely redistributing resources could ultimately harm even the recipients by delaying major technological and scientific progress.
- The **high costs** of redistributing scarce resources like compute and technical talent may stymie large-scale resource sharing efforts.

These risks and challenges associated with AI resource sharing are discussed in more detail in **section IV**.

III.2 Sharing Access to Advanced AI Systems



A second approach to AI benefit sharing is to **share access to advanced AI systems**. These systems could be helpful in a wide range of ways, and making them more accessible to actors in low- and middle-income countries could generate a variety of benefits. For example, they can be used to develop applications that generate commercial profits, address local issues, or are helpful for a range of diverse users and use cases.

Advanced AI models can be accessed at the application level, via an API, or through direct access to publicly available model weights. Here we discuss three specific ways to share access:

1. Actors could **reduce the cost of deploying and running advanced models**. This could be achieved, for example, through **open-sourcing** models or through **research and development** to make them more compute-efficient.
2. Governments could provide **subsidies** for certain users or applications.
3. AI companies could use **differential pricing** for certain users or applications.

Ensuring that recipients have access to sufficient digital infrastructure and AI development resources may enable the full realisation of benefits from widening AI access.

Reducing the Cost of Running Advanced AI Systems

One obvious way to make advanced AI systems more accessible is to make them less expensive to use. While some advanced AI systems are currently freely available in most countries, users still face a number of costs to access and use the most capable AI systems at scale.[12]

Providing advanced AI models to users involves compute and infrastructure costs. These are typically covered via subscription fees, usage fees, or some combination of these. If an

[12]There are usage limits for the free versions of a number of advanced models. For example, Google's Gemini 1.5 Pro API can be accessed for free but has a rate limit of 50 requests per day and 32,000 tokens per minute (as of January 14, 2025)

advanced model is deployed via an API or a first-party application like ChatGPT or Claude, then users often need to pay an access fee and a fee per token of input or output from the system (though the cost per token is falling rapidly). Some models are publicly-downloadable, with no up-front cost. However, running and developing on top of these models can still involve considerable compute costs. For third-party applications that are built using advanced models at least some portion of the costs to run the underlying AI model are passed onto the end user. Lowering these costs could allow more users, particularly those with less ability to pay, to access these models.

One way to reduce costs would be investing in **research and development (R&D) to improve inference efficiency**. This could include investing in compute efficiency (the number of operations AI chips can perform per dollar) (Hobbhahn, Heim, and Aydos, 2023), model optimization (Liu et al., 2023; Leviathan, Kalman, and Matias, 2022), and datacentre capacity (Evans and Gao, 2016). Such investments could be funded by AI developers, AI chip companies,[13] or even public entities. It is worth noting that private actors already have strong incentives to make these investments and private R&D funding has already led to increased compute efficiency, both for training (Pilz, Heim, and Brown, 2023) and inference (OpenAI, 2024b) purposes, in recent years. This leaves some uncertainty regarding the impact of additional public funding in this space. Additionally, while greater compute efficiency could lead to lower costs to deploy and use an advanced AI system, it could also lead to faster progress on frontier AI development. This could have significant downsides. For example, developing frontier systems faster could accelerate the emergence of new risks from AI faster than effective risk mitigations could be developed (Bengio, Hinton, et al., 2024). R&D that is targeted specifically at making inference more efficient (and therefore more affordable) could help expand access to the benefits of using a system but may not translate into faster progress at the frontier of AI development.

Another way to reduce costs is by **supporting the development of open-source AI models**. Because these models are freely available, they eliminate some access expenses. Open-source availability also enables downstream developers to modify models more flexibly. By lowering development costs, open-source models can thus enable the creation of a range of lower-cost applications. This could be particularly important in lower-income settings. Developing indigenous AI applications using open-source models is a core part of the Indian government's "IndiaAI" initiative, for example (Jeevanandam, 2024).

So far, despite larger training runs driving higher development costs for frontier AI models, market competition and efficiency improvements have allowed per-token inference costs to fall. Additionally, the capabilities of the best open-source models are relatively close to those of the top closed-source models (Cottier, You, et al., 2024). However, continued price decreases and the future viability of open-source models may depend on the advanced AI market remaining competitive. Economies of scale in AI training and certain regulatory proposals could threaten this competition in the future, though the probability of this is uncertain. As development costs grow and the AI market evolves, preserving a competitive marketplace will be an important regulatory goal (Azoulay, Krieger, and Nagaraj, 2024), among others. Low-cost inference and highly-capable open-source models also pose certain

[13] Companies in the semiconductor supply chain already have strong incentives to invest in R&D. According to one estimate, NVIDIA's R&D expenditure in FY 2024 exceeded $8.6 billion, up 365% from annual spending five years earlier.

risks that may prove important to address; these are discussed further at the end of this section and in **section IV**.

Government Subsidies

A second way to make AI systems more accessible is via **public subsidies for certain advanced AI systems, applications, or APIs.** Subsidies could be more targeted than other cost-reducing measures. For example, governments could try to expand subsidies for those users or use cases that are expected to accelerate the spread of AI benefits internationally.

To be effective for benefit sharing, subsidies would likely have to be available internationally. Although this could pose political and logistical challenges, there are precedents for international subsidies in other industries. In global health, for example, donors subsidise the purchase of vaccines in lower-income countries through organisations like the Global Alliance for Vaccines and Immunization (Brugha, Starling, and Walt, 2002).

In the AI context, there may be a strategic element to subsidies in that they incentivize using certain AI systems over others. Governments may want to use subsidies to bring more users to systems developed within their borders. They could complement subsidies with technical assistance and capacity building to accelerate the diffusion of these particular systems globally. In the long-term, building a large and global user base could bring strategic and economic benefits. Such initiatives could mirror aspects of programs like Germany's Energy Solutions Made in Germany Initiative, which supports the export of German renewable energy technologies to developing countries.

One potentially promising avenue for subsidising access to advanced AI would be the **expansion of export financing options provided by leading states to low- and middle-income countries.** For example, the Export Import Bank of the United States (EXIM) provides financing options for the purchase of US AI and computing technologies by low- and middle-income countries. From January 2020 and September 2023, EXIM authorized nearly $8 million dollars worth of financing to enable countries to purchase access to AI technologies developed in the US (Export-Import Bank of the United States, 2023). An increase in export financing for AI access would likely be aligned with the economic competitiveness interests of the US as it seeks to counter Chinese influence and technological expansion in emerging markets.

Another option could involve subsidising access to specific applications that are meant to advance progress towards the SDGs. This option is discussed in more detail in Box 2.

Differential Pricing

Advanced AI companies could implement a private version of public subsidies by directly reducing the price of their products for certain users or use cases.

Currently, individuals pay the same price worldwide for access to leading subscription-based AI systems. For example, as of January 2025, the cost of a ChatGPT Plus account – needed to use OpenAI's o1 model – is $20 USD per month. That represents at least 20% of average annual income per-capita in low-income countries, where gross national income is $1,145 or less (World Bank, 2024c).

Differential pricing based on local economic conditions could help make these systems more accessible globally. This approach is also not without precedent: many digital services, including cloud providers like AWS, streaming platforms like Netflix, and software companies like Microsoft already employ region- or country-based pricing strategies. Notably, these companies do not do this out of charity; it is in their interest as profit-maximising companies to price access differently around the globe. While this suggests that differential pricing could also be in the interest of leading AI companies, making any substantial price reductions in certain regions could present challenges, particularly given recent reporting that some leading AI developers are operating at a loss (Efrati and Holmes, 2024).

One solution could involve significantly increasing the price of access in places with higher per capita incomes to subsidise reduced prices in lesser-resourced parts of the world. Such sliding-scale pricing could be based on the local economic conditions of the user or more directly could be based on the user's willingness to pay for access above a low floor price (as elicited via a digital auction, for example). Of course, the operational feasibility and desirability of this approach is unclear; it is notable that no leading developer has implemented it to date.

It's possible that leading AI companies are currently employing a penetration pricing strategy, prioritising the growth of their user base over immediate profitability (Noble and Gruca, 1999). This would mean they are already, in effect, subsidising access to their AI systems. This arguably makes the prospect of further price cuts in lower-income countries unlikely. Region-based pricing differentials could also be arbitraged by users accessing AI services through virtual private networks (VPNs). On the whole, significantly expanding AI access through differential pricing may be difficult.

Box 2: Access for SDGs + Public Goods

While security concerns may motivate certain restrictions on general-purpose AI proliferation, there is a case for deliberately fostering international access to AI applications that can advance sustainable development in particular (Goh and Vinuesa, 2021).

There is growing evidence that AI could be a powerful tool for major scientific breakthroughs with enormous benefits for humanity (US Department of Energy, 2024; Abramson et al., 2024; Zhou et al., 2023; Lam et al., 2023). AI applications in areas like healthcare, education, and agriculture could help achieve the SDGs (United Nations AI Advisory Body, 2024). AI may also help create new opportunities for economic growth in developing economies linking them to the global digital economy, raising productivity, and improving the provision of public services (Ben-Ishai et al., 2024). This convergence could be accelerated through **targeted efforts to steer AI access and AI development towards addressing development challenges and economic needs in low- and middle-income countries.** This is one path forward for sharing AI access.

International institutions like the UN and the World Health Organization (WHO) are already developing initiatives to promote AI applications with significant potential

to advance sustainable development outcomes globally (International Telecommunication Union, 2024; UN Global Pulse, 2023; Department of Digital Health and Innovation, 2024). Some leading AI developers already have programs that provide access and support to users who are leveraging their AI tools for social good (OpenAI, 2023; The International Rescue Committee, 2024). Additionally, some AI companies have shared technology with scientific institutions to advance scientific progress on specific global challenges (see, e.g., (The Francis Crick Institute, 2022)). Such initiatives could be scaled up to accelerate the global distribution of AI's benefits.

Other approaches are also possible. Institutions could use **market-shaping mechanisms like prizes and advance market commitments** to incentivize the creation of new AI applications for sustainable development (Mazzucato, 2023), or they could establish a **global fund with commitments from leading AI developers, frontier AI states, and/or philanthropists.** Such a fund could support R&D for AI applications specifically tailored to address development challenges.

Inspiration could be taken from multilateral and multisectoral efforts to fund the production of public goods in other sectors. For example, the Coalition for Epidemic Preparedness Innovations, the Global Fund for AIDS, Tuberculosis, and Malaria, and Gavi, The Vaccine Alliance. One step in this direction is the recent commitment from the UK, Canada, the US, the Bill and Melinda Gates Foundation and partners in Africa to fund £80 million ($100 million) worth of programming to accelerate AI-driven economic development on the continent (Foreign, Commonwealth & Development Office and The Rt Hon James Cleverly MP, 2023).

Challenges

Although AI systems may help achieve benefit-sharing goals, approaches to AI benefit sharing that focus on widening access to such systems also have important limitations and pose risks. These include:

- **Misuse and proliferation risks**. While increasing access could make AI more accessible globally, it also increases the threat from malicious use of powerful systems by adversarial actors (Brundage et al., 2018). There is a tension between sharing capable models to drive progress on societal challenges and ensuring responsible deployment of dual-use technologies. While some risks could be mitigated through governance measures, careful consideration is needed about which models to share and how to structure access (Shevlane, 2022).
- **Cost and implementation barriers**. Despite falling compute costs, scaling up access globally could require complex subsidy schemes or differential pricing mechanisms. These could be expensive to implement and face technical challenges like arbitrage between different pricing tiers. Even modest usage costs can represent a substantial portion of average income in many countries.
- **Infrastructure and capacity limitations** that constrain the benefits of expanded access. Many countries lack sufficient broadband access (Shanahan and Bahia, 2024), affordable

devices and data, digital literacy (Bandura and Méndez Leal, 2024), and consistent electricity supply (World Bank, 2023a). This may limit the benefits users in these countries can realise simply by gaining access to AI systems. Even if AI tools become widely accessible, the primary benefits may still accrue disproportionately to those who possess the skills and resources to leverage AI most effectively.

- **Strategic and competitive dynamics** could limit leading AI companies' willingness to share access to their most capable systems, particularly if maintaining exclusive access provides competitive advantages.

These and other challenges associated with sharing access to AI systems are discussed further in **section IV**.

III.3 Sharing Financial Proceeds

A third way to share the benefits of AI is to **distribute the financial proceeds generated by AI development**. It is plausible that, by default, the bulk of these financial proceeds will flow to people in the few countries that currently dominate the advanced AI industry (O'Keefe et al., 2019). Benefit-sharing efforts in this category would aim to redistribute some portion of these profits internationally.

An advantage of this approach is that cash transfers are more fungible than both AI development resources or access to AI systems. Financial proceeds could be used to purchase AI resources or to access AI systems, but they could also be used to make other purchases or investments that are unrelated to AI. This flexibility means that recipients may find financial transfers more valuable than equivalent in-kind transfers of specific goods or services.

Financial proceeds could be distributed in multiple ways. In this section, we discuss two specific options:

1. **AI companies could directly distribute proceeds**. For example, they could fund cash transfer programs funded by a portion of their profits.
2. A **public body could collect and redistribute these assets.**

International redistribution of financial resources through aid and impact investing programs is not uncommon. However, focusing large-scale distribution of financial resources on a single technology or sector would be unusual. There are numerous unresolved legal and logistical questions about how it could work. There are at least two broad possibilities: distribution by companies in the AI supply chain or by public institutions.

Financial Proceeds Distributed by Companies in the AI Supply Chain

First, financial proceeds could be shared by companies in the AI supply chain. This would involve **companies themselves reserving a portion of their financial assets or profits and transferring them to a set of beneficiaries directly or through an intermediary**

organisation. These companies could include semiconductor suppliers (e.g., NVIDIA) or private companies developing leading AI systems (e.g., OpenAI, GoogleDeepmind, Anthropic, etc.). This option could understandably face significant resistance from companies and shareholders, as it could appear to conflict with their financial interests and companies lack strong incentives for large-scale, voluntary implementation of profit-sharing.

One proposal, the **Windfall Clause** (O'Keefe et al., 2019), suggests that AI companies should commit to donating a meaningful portion of their profits once they exceed a predefined threshold. This threshold could be high, such as a tangible fraction of all global economic output. The distribution of profits could be managed by a trust, a global oversight body, or independent organisations that specialise in delivering financial assets to individuals worldwide.[14]

A key feature of the Windfall Clause is that AI developers would commit to it before generating substantial profits. This pre-commitment could increase the likelihood that companies make clear, public promises to distribute windfall profits. While such pledges may be aligned with a company's fiduciary duties by improving its public image, the significant financial obligations of surpassing a windfall threshold could still outweigh these reputational benefits, posing a legal challenge for profit-maximising AI companies. Companies may be more likely to fulfil their promises if these commitments are legally binding from the start.

OpenAI initially launched with a capped-profit business model, similar in spirit to a Windfall Clause. Under this structure, any shareholder returns beyond a certain (high) rate of return were to be directed to a non-profit organisation, which could then distribute these funds as it saw fit. However, recent reporting suggests that OpenAI is now planning to transition away from this model to make the company more attractive to investors (Holmes, Mascarenhas, and Hornstein, 2024). No other company in the AI supply chain has a corporate structure that mandates profit distribution above a specified threshold. Anthropic is registered as a Delaware Public Benefit Corporation, which allows it to balance its stockholders' financial interests with the public benefit goal of "the responsible development and maintenance of advanced AI for the long-term benefit of humanity". However, this structure does not require Anthropic to redistribute profits above any threshold.

Many corporations in the AI supply chain do already engage in **Corporate Social Responsibility (CSR) initiatives** as well, where they direct a portion of their profits to support social causes.[15] Financial proceeds distribution could in theory resemble a scaled up version of this. Industry efforts along these lines could also be coordinated by existing industry coalitions, such as the Partnership on AI, that already work to develop standards and facilitate coordination across the AI industry. In principle, this sort of organisation could be leveraged to also establish coordinated action around expanding global access to AI's benefits. However, even with coordinated action, it is unclear whether scaling up the CSR model would be an effective or sufficiently ambitious approach to benefit sharing.

[14] An example of one such organisation is GiveDirectly.
[15] For example, Microsoft's CSR initiatives are detailed here.

Financial Proceeds Distributed by Public Institutions

Public institutions could also distribute AI's financial proceeds. Governments have existing public infrastructure, such as institutions that have the authority to collect taxes and distribute tax revenues, that positions them as natural facilitators of financial benefit sharing. However, they may also face challenges and incentive problems in trying to distribute AI proceeds internationally, such as political pressure from domestic constituencies or competing foreign policy goals.

One simple way national governments, especially in frontier AI states, could distribute benefits is by **taxing AI-driven profits and increasing their foreign aid budgets.** If AI companies come to generate a substantial fraction of a country's gross domestic product, or if advanced AI leads to an overall acceleration of economic growth, then it may be sufficient for governments to just leverage existing infrastructure and expand foreign aid. This approach, while straightforward, exemplifies the challenges of convincing domestic constituents on the value of generous benefit sharing: in 2023, only five of 31 OECD countries exceeded the UN foreign aid spending target of 0.7% of gross national income (OECD, 2024).

An alternative to increasing foreign aid could be the **creation of a multilateral global fund.** Such a fund could distribute AI-driven financial gains internationally in the form of grants, potentially as assistance towards causes like the UN SDGs or towards AI-complementing infrastructure investments in low-resource settings.

The fund could be established with a blend of financial commitments from leading AI developers and frontier AI states. A coordinated international effort could alleviate the burden on individual companies or countries to design and implement their own aid programs. It might also carry greater legitimacy and more strongly incentivise industry-wide participation compared to disparate efforts.

Alternatively, it could be created as **a multilateral version of an AI Growth Bond** (Casey, Roy, and Rockall, 2024). A public entity, or perhaps an intergovernmental group, could acquire shares in companies along the AI supply chain and distribute dividends internationally. One benefit of this approach is that the value of the fund would be correlated with the value of AI companies and thus, presumably, the economic disruption caused by powerful AI systems. Through its investments, an AI Growth Bond approach could both support AI progress via investment and provide insurance against increased inequality that might result from that progress.

Challenges

While sharing financial proceeds offers flexibility in how recipients can use resources, this approach faces multiple challenges. These include:

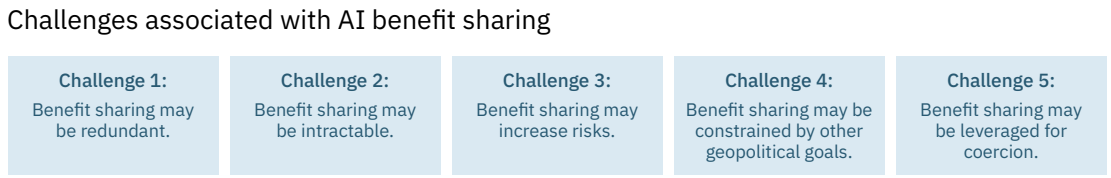
- **Legal and fiduciary constraints** that may limit companies' ability to share financial proceeds. For companies bound by fiduciary duties, voluntarily distributing profits to non-shareholders could conflict with their obligation to maximise shareholder value. Partly in anticipation of this and similar challenges, some AI development companies like OpenAI and Anthropic were originally set up with unconventional corporate governance structures (a capped-profit and public benefit corporation model, respectively) (Tallarita, 2023).

However, these governance arrangements have been stress-tested as the value of advanced AI has increased, and it is unclear whether such structures will allow AI development companies to implement broad sharing of financial proceeds in practice.

- **Implementation and logistical challenges.** There are complex questions about both the mechanics and timing of financial proceeds distribution. It remains unclear whether proceeds could or should be shared directly from companies to individuals globally, or channelled through governments or other intermediaries. If an intermediary is used, it's unclear what kind of legal status such an entity would need, or how it would coordinate with beneficiaries in different countries worldwide. Then, there are questions about who should receive the financial proceeds within a recipient country. Proceeds could be distributed directly to individuals via methods like direct cash transfers (see, e.g., Bastagli et al., 2016). However, legitimate concerns may be raised about external actors deciding how to allocate resources within a country, especially if the scale of the benefits being shared is large. At the same time, distributing those resources via local governments may raise concerns about transparency and political interference. There is also uncertainty about when financial redistribution should be initiated and how much redistribution would be optimal. Compared to sharing access to AI systems through software or APIs, distributing cash or company shares globally could face significantly more logistical and legal hurdles.

These and other challenges associated with sharing financial proceeds from AI are discussed in more detail in **section IV**.

IV Challenges Associated with Benefit Sharing



Our analysis considers the ways in which AI benefit sharing presents both opportunities for progress and significant challenges to implementation. Within each section above, we identified limitations or obstacles specific to each approach. In this section, we discuss five fundamental challenges that are present in all three approaches:

1. Benefit sharing may be **redundant.**
2. It may be **intractable.**
3. It may increase **certain risks to safety and security.**
4. Its potential may be limited by **geopolitical and strategic considerations** that make leading actors reluctant to share benefits.

5. It may be **leveraged for coercion,** pressuring recipients to act in the strategic interests of more powerful actors.

None of these challenges necessarily makes benefit sharing impossible or undesirable. However, they do show that moving from proposal to implementation of benefit-sharing mechanisms will be complicated. Not all benefit-sharing plans will justify their downsides and costs. It is important to be aware of these challenges as international calls for benefit sharing increase.

IV.1 Benefit sharing may be redundant

First, it is **possible that benefit-sharing mechanisms will be redundant** if AI benefits are large and diffuse widely or quickly without proactive intervention. If most people can access and benefit from AI systems by default, or benefit in other ways from broadly-accelerated economic growth, market forces may prove sufficient for distributing AI benefits.

While several of the key inputs needed to train and develop AI systems are scarce and expensive, evidence suggests that AI capabilities are diffusing naturally. For instance, publicly downloadable language models like Llama 2 have been downloaded hundreds of millions of times across over 150 countries (Meta, 2024b). API costs for leading models have fallen over time (DeepLearning.AI, 2024). Compute efficiency has also been increasing over time, reducing the cost of developing or running an AI model of a given capability (Hobbhahn, Heim, and Aydos, 2023).

Access costs can add up when models are used at scale, however. Moderate monthly usage of advanced AI APIs (e.g., 1 million input tokens and 1 million output tokens using Claude 3.5 Sonnet) currently costs approximately $18 (Anthropic, 2024a). This is a small fraction of average monthly income in the US, but is approximately 9% of monthly GDP per capita in India and 17% in Ethiopia, for example (World Bank, 2023b). While this suggests a substantial accessibility gap, per token costs to use advanced AI systems are declining rapidly (DeepLearning.AI, 2024; OpenAI, 2024b).

Additionally, most people almost certainly will not need to develop their own systems from scratch in order to benefit from AI. Although high costs may prevent many lower-income users from fully taking advantage of some AI systems, it remains to be seen how large a barrier this will be. Today, most of the world can use freely downloadable models near the frontier of AI capabilities.

IV.2 Benefit sharing may be intractable

Second, international AI **benefit sharing may be intractable**. One reason this could be the case is if benefit-sharing approaches require multiple actors to take costly actions. For example, it could require governments to transfer resources internationally or cooperate with other states, or it could require companies to pay higher taxes, donate part of their profits, or offer their products at reduced prices or at no cost. Implementing benefit-sharing

mechanisms could require significant financial and political investments that will often run counter to the incentives of individual companies or states.

The world already has a substantial foreign aid apparatus, and its effectiveness is hampered by the frequent failure of countries to fulfil their commitments. For instance, the longstanding UN target for developed countries to allocate 0.7% of their Gross National Income (GNI) to official development assistance has been met by only a handful of nations (United Nations Economic Commission for Europe, 2025). Partly as a result, potential recipient countries or entities would likely be sceptical of promises made by frontier AI states or companies. Unlike in domestic contexts where laws can be enforced through established judicial systems and legal frameworks, international commitments often lack robust enforcement mechanisms. This creates a significant challenge in ensuring that parties follow through on their benefit-sharing promises.

An additional challenge to the tractability of benefit sharing is the potentially high cost of these initiatives. Leading AI developers may spend as much as $1 billion dollars on the compute resources needed for a single training run by 2027 (Cottier, Rahman, et al., 2024). Developers, or even governments, may be reluctant to divert resources to benefit-sharing initiatives if they are worried about funding future scaling.

The interdependence of AI development resources poses another challenge to tractability. The benefits that can be derived from any single resource often depend on access to others. For instance, access to training data and algorithms provides limited value without sufficient compute and talent resources to actually train models with that data. Similarly, while sharing access to advanced AI systems could potentially drive benefits in many countries, approximately one-third of the world's population still lacks reliable internet access, and many more lack the digital infrastructure needed to effectively deploy AI applications at scale. This interdependence means that meaningful benefit sharing may require coordinating the distribution of multiple resources simultaneously, while also ensuring that recipients have the fundamental digital infrastructure needed to make use of them. This complicates implementation and may make certain benefit-sharing approaches ineffective if basic infrastructure is not built first.

There are also many important but unanswered implementation questions associated with any benefit-sharing approach. For example, should all states get equal shares of the redistributed profits or equity? Should AI resources be allocated per capita? Should the least developed states get more or quicker aid? Which institutions should be entrusted with equitably distributing AI benefits within a country, especially in countries with weaker governance institutions? Additional work is needed to make progress on key logistical, legal, and normative questions around the implementation and fairness of any international benefit-sharing scheme. We discuss some of these questions further in **section V**.

IV.3 Benefit sharing may increase risks

Third, since AI systems are dual-use technologies (Ueno, 2023), the benefit-sharing approaches that work by expanding access to develop or use highly-capable systems also **have the potential to exacerbate AI risks** that affect all countries (Bengio, Hinton, et al., 2024).

The first way benefit sharing could increase risk is by making it easier for people to misuse AI to cause harm. Advanced AI systems may enable state actors to implement intrusive surveillance regimes, for example. They may also help both state and non-state actors develop new weapons, including biological agents that could cause severe global harm (US Department of Homeland Security, 2024). The likelihood that AI systems will, or even could, be used in such ways is a contentious, but worrying, possibility (Brundage et al., 2018; Bengio, Mindermann, et al., 2025; Hendrycks, Mazeika, and Woodside, 2023). If such risks materialised, they would cross borders, potentially causing greater harm in low- and middle-income countries that have fewer resources for mitigation and response.

Benefit sharing could meaningfully raise these risks by both expanding the set of actors with access to advanced AI systems and by complicating governance and oversight efforts. Maintaining uniform security standards becomes more challenging in a more diverse landscape, and the accelerated development of advanced AI systems by a broader range of actors raises concerns about adherence to safety and ethical standards. This challenge is particularly acute because once AI capabilities are widely distributed, under some modes of access, it becomes practically impossible to restrict their use or recall them if safety issues emerge (Seger et al., 2023).

A second way benefit sharing can increase risk is by providing more resources to actors seeking to compete in the frontier AI market. Of course, this is also one of the desired outcomes for some benefit-sharing measures: allowing more developers, particularly those in low- and middle-income countries, to develop advanced AI systems that can solve local challenges and boost growth. But some researchers have raised concerns that in a more competitive market, AI developers will feel more pressure to forego extensive safety testing in order to release new products faster, or risk losing market share to more risk-tolerant competitors (Stafford, Trager, and Dafoe, 2022). By expanding the number of developers at the frontier, certain benefit-sharing initiatives may strengthen these dynamics and raise risks overall. These concerns may apply between firms *and* states.

For both of these concerns – misuse and risk-inducing competition – the potential harms could even be large enough to outweigh the international benefits entirely. This means that safety and security measures must be considered as part of any benefit-sharing plan.

For resource-sharing initiatives, for example, these safety measures could include regular auditing, safety protocols,[16] or reporting requirements, among others (Raji et al., 2022; Kolt et al., 2024). Such efforts would seek to make it harder for malicious actors to use the shared resources to develop risky systems undetected.

Technical governance measures, particularly relating to controlling access or verifying usage, may be able to mitigate some safety and security concerns related to sharing compute. For example, cloud compute providers could classify users' computing workloads to verify that shared computing resources are not used to train potentially high-risk models (Heim, Fist, et al., 2024). If more stringent oversight or security guarantees are required, hardware-level solutions could be considered. Flexible hardware-enabled governance mechanisms (HEMs) could potentially be incorporated into future AI chip designs, allowing for adaptable controls as the understanding of AI risks evolves (Kulp et al., 2024). However, implementing such

[16] Several leading AI companies have released Responsible Scaling Policies (see: Anthropic) or Frontier Safety Frameworks (see: Google DeepMind), which publicly (but voluntarily) commit them to specific safety measures that must be met before deploying advanced AI systems beyond certain capability thresholds.

hardware-level controls requires substantial additional research to understand their technical feasibility and potential drawbacks (Heim, 2024b).

Sharing AI Access could similarly be approached in ways that can help widely distribute access while limiting potential risks. For example, instead of openly disseminating frontier AI systems, developers can, at least initially, facilitate controlled interactions with their systems through "structured access" approaches (Shevlane, 2022). These approaches can mediate access to models through APIs that put limits on how much a downstream developer can modify the model's behaviours. Structured access can also help with the enforcement of usage policies. For example, models shared with limited access could also have stricter safeguards to further reduce the likelihood that they could be used to generate harmful or illegal content (Mu et al., 2024).

IV.4 Benefit sharing may be constrained by other geopolitical goals

Fourth, the strategic importance of advanced AI means that **geopolitical considerations** could significantly constrain benefit-sharing possibilities. Frontier AI states, motivated by national security concerns and economic interests, have already adopted policies that selectively restrict access to advanced AI development resources. For example, the US has implemented export controls on advanced AI chips and semiconductor manufacturing equipment, with allies including Japan and the Netherlands collaborating to enforce export controls (Bureau of Industry and Security, 2024; Allen, Benson, and Putnam, 2023). New regulation has also been proposed in the US that would put conditions on the export of advanced AI model weights (in addition to new security conditions and limits on the export of advanced chips to certain countries) (Office of the Federal Register, National Archives and Records Administration, 2025).

These restrictions reflect a broader shift towards AI resource nationalism amid rising geopolitical tensions. Countries where leading AI systems are developed may prove increasingly reluctant to share what they view as strategically valuable AI resources or capabilities. They may not want to divert financial resources away from leading AI firms or share access to the most advantageous AI systems. For this reason, benefit sharing may prove most contentious when AI capabilities are most strategically important.

These strategic challenges are further complicated by the growing technology competition between major powers, particularly the US and China. Countries may view benefit-sharing initiatives through the lens of this competition, potentially treating them as tools for expanding technological influence internationally rather than genuine efforts at equitable distribution of AI's benefits. These goals – strengthening geopolitical influence and improving the wellbeing of people in low- and middle-income countries – may yet prove compatible or, indeed, complementary. But this is not a given, and if they are perceived as competing then the potential influence of benefit-sharing efforts could be severely curtailed.

There are also concerns that benefit sharing could slow progress at the frontier of AI development. The limited supply of key inputs to frontier AI development means that reallocating resources away from leading developers might slow the pace of advancement, which would

again counter the strategic interest of leading states and companies. This slowdown could also potentially delay AI-driven benefits, such as economic growth or innovations in areas like healthcare and education, which could be particularly impactful for developing countries. However, some argue that a slower pace of AI development could be beneficial in the long-term by allowing more time for the development of safety measures and governance frameworks (Bengio, 2023; Future of Life Institute, 2023).

It might be possible to share access to AI systems or resources while excluding capabilities most relevant to national security. However, this approach could be challenging to implement in practice, and the benefits of such restricted sharing could be limited. Any meaningful access-sharing framework will need to carefully balance strategic interests with genuine sharing of valuable capabilities.

IV.5 Benefit sharing may be leveraged for coercion

Fifth, AI governance negotiations are shaped by historic and contemporary power dynamics between frontier AI states and other countries (Png, 2022). If these dynamics are not considered, benefit sharing may become a tool that frontier AI states can use **coercively to force low- and middle-income countries to agree to deals that are sub-optimal** or, in the long-term, even harmful to their interests.

How could this occur in practice? For one, frontier AI states could use the *prospect* of benefit sharing to achieve other goals. For example, they could introduce new controls on trade of technologies that otherwise would have diffused via market mechanisms. They could then offer to ease these controls in exchange for agreements that include some benefit-sharing provisions, but primarily serve to maintain their technological advantage.

Alternatively, benefit sharing could be initially favourable, but become a coercive tool over time. For example, if benefit sharing involves sharing access to AI models, it could create the technological dependency that some advocates worry about. Recipient states could come to rely on these AI systems to drive economic growth and enhance public services, for example. Once established, this dependency could then be exploited. Frontier AI states could then threaten to revoke AI access if the dependent countries refuse to adopt certain policies.

Even well-intentioned safety measures could be coercive to some extent. Restrictions meant to mitigate known risks of misuse can also, intentionally or not, serve to consolidate power among a limited group of actors. The Nuclear Non-Proliferation Treaty illustrates this challenge: while creating important safety restrictions that non-nuclear states accepted to improve global security, these same restrictions have created enduring inequalities between nuclear and non-nuclear powers that have become increasingly contentious over time (Lee, Nacht, and Nacht, 2020).

Finally, in some cases, AI resource-sharing schemes may involve conditions on how the shared resources may be used. Technical requirements – such as the incorporation of visibility mechanisms into compute resources to enable external monitoring – could be components of these conditions. Such requirements threaten to limit the ability of actors in low- and middle-income countries to use shared resources autonomously and in line with their own goals (Vipra et al., 2024). This dynamic creates a fundamental tension: access to AI benefits

would be contingent on accepting external controls, undermining the very self-determination that benefit sharing aims to promote. To overcome this challenge, it may prove important to involve marginalized populations, particularly from low-and-middle-income countries, in benefit-sharing design and implementation. This consideration is further discussed in **section V**.

V Next Steps for Benefit Sharing

This report has identified several potential approaches to sharing the benefits of AI internationally. Implementing any of these options will likely be complex, requiring various public and private actors to coordinate and overcome some considerable challenges. To date, progress towards implementation has arguably been slow. None of the example initiatives mentioned in **Section III**, for example, is sufficiently large or inclusive enough to be considered a comprehensive global benefit-sharing scheme.

In this section, we suggest three practical steps that could be taken to advance international benefit sharing. These steps, which would likely involve AI companies, governments, and multilateral organisations working in partnership, are:

1. Initiate **low-risk, high-potential** benefit-sharing options
2. Establish **dedicated forums** to discuss benefit-sharing options and implementation strategies
3. Ensure **representation of low- and middle-income countries** in AI governance decision-making processes

V.1 Initiate low-risk, high-potential benefit-sharing options immediately

The complexity of implementation does not imply that nothing can be done to share AI benefits in the near-term. Some initiatives with clear benefits and minimal risks can be pursued now. These efforts can improve the wellbeing of beneficiaries, build trust between participating states, serve as valuable learning opportunities, and enable more comprehensive and ambitious benefit-sharing efforts in the future.

For example, frontier AI states could immediately increase funding to develop **digital infrastructure in low- and middle-income countries**. Such infrastructure will be a prerequisite for participation in many AI benefit-sharing efforts, can enable opportunities for growth and development outside of the AI industry, and may pose fewer risks and downsides. Expanding access to digital infrastructure in emerging markets may align with the strategic interests of frontier AI states, which aim to secure a larger share of the global market relative to geopolitical rivals. Promoting domestically produced technology abroad could not only advance their national interests but also lay the groundwork for more widely shared benefits as AI capabilities progress.

Frontier AI states, as well as actors like AI companies and multilateral institutions, could also **support the development and deployment of AI applications targeted at achieving SDGs.** Sharing access to general-purpose AI systems can raise complex security and governance considerations. However, narrow AI applications can improve public service delivery, healthcare services, agricultural productivity, and other key outcomes in developing countries while presenting fewer dual-use concerns. Several ongoing efforts already demonstrate how this can work in practice, from AI companies providing specialised access for social impact applications (OpenAI, 2023), to "AI for Good" open datasets and code repositories (*Expand Opportunity - AI for Good* 2024), to early-stage funds supporting AI solutions for development challenges (Foreign, Commonwealth & Development Office and The Rt Hon James Cleverly MP, 2023). Efforts like these could be amplified through public-private partnerships that use prizes, advanced-market commitments, and other mechanisms to accelerate beneficial AI innovations. They could be complemented by programs to provide technical assistance to governments, organisations, and individuals in developing countries aimed at accelerating the uptake of AI services that generate public benefits. Here, the US in particular has a strong platform to build on in the Partnership for Global Inclusivity on AI (US Department of State, 2024). This partnership may facilitate sharing AI benefits now while strengthening partnerships to support even more ambitious efforts in future.

V.2 Establish dedicated forums to discuss benefit-sharing options and implementation strategies

Some forms of benefit sharing could be implemented today. However, a number of fundamental questions must be answered before more ambitious, comprehensive benefit-sharing initiatives could be established. To answer these questions, internationally-inclusive and representative discussions will need to take place, and legitimate forums will be needed to host these discussions. An additional next step, which could happen in parallel with the implementation of the kinds of benefit sharing discussed immediately above, is to **establish a productive and collaborative decision-making forum on AI benefit sharing**.

Discussions in these forums would need to cover several issues, including:

- **How to maximise mutual benefits:** Global perspectives on AI benefits will vary both across and within countries. Some groups may prefer getting access to AI resources, while others may prefer subsidised model access, depending on their economic context. Others may prefer financial benefits that can be used to fund other programs and projects not related to AI. Preferences may also vary within a country; benefit-sharing schemes that focus on providing access to APIs seem likely to provide more benefits for people who already have technical skills, for example. Some benefit-sharing options will be more aligned with the interests of frontier AI states and leading developers than others. Certain approaches could actually favor their strategic self-interest while delivering desired benefits to lesser resources actors. Despite these varied needs and goals, no dedicated forums currently allow for representatives from different countries and groups to discuss practical paths forward and identify areas of alignment between stakeholders.
- **What conditions allow for meaningful participation and benefits while minimising shared risks:** Certain forms of benefit sharing could plausibly lead to the proliferation

of AI systems with dangerous capabilities, or exacerbate AI race dynamics by enabling more intense international competition between AI developers. More work is needed to understand which potential sharing approaches could provide significant benefits without raising these risks. International dialogue will also be needed to understand what conditions or safeguards might be introduced to benefit-sharing initiatives to assuage concerns about these risks without undermining the sovereignty of participating countries.

- **What assurances frontier AI states and AI companies can offer:** Given their control over key AI capabilities and resources, frontier AI governments and companies hold significant bargaining power to set the terms of benefit-sharing arrangements. However, relying on these actors risks reinforcing existing power imbalances, neglecting the specific needs of underrepresented communities, and developing suboptimal solutions. To overcome these concerns, the actors currently shaping AI development trajectories and governance frameworks will need to deeply engage intended beneficiaries in the design process for benefit-sharing initiatives and ensure that credible assurance mechanisms accompany any benefit-sharing commitments.

Failing to discuss these issues, among others, in inclusive, international forums could lead to benefit-sharing proposals that (1) lack buy-in from many states, limiting their appeal; and (2) fail to account for the range of benefits, costs, and incentives participants face.

To advance these discussions, the international community may be able to build on existing efforts. The UN's planned "Global Dialogue on AI Governance," agreed to through the Global Digital Compact, could be a promising venue for this work (United Nations Office of the Secretary-General's Envoy on Technology, 2024). Expected to launch within the next year, this dialogue could bring together governments, AI developers, civil society organisations, and experts from underrepresented regions to discuss their specific needs and aims. The success of an initiative like this will depend on multi-stakeholder participation, including leading AI companies, and meaningful representation, especially from low- and middle-income countries. Additionally, traditional UN processes often move too slowly to keep pace with rapid AI developments, which presents a challenge.

The global AI Summit Series could serve as another promising platform. The series has uniquely succeeded in bringing major AI powers, including China, to the table on safety issues and securing concrete safety commitments from major AI companies (Department for Science, Innovation & Technology, 2024; Prime Minister's Office, 10 Downing Street et al., 2023). While currently more focused on safety, the Series could eventually become a productive forum for benefit-sharing discussions given its proven ability to drive meaningful coordination among key stakeholders. However, this would require significantly expanding the participation of low- and middle-income countries.[17]

[17]The third global AI summit is scheduled for February of 2025 in Paris, following previous summits in Seoul and at Bletchley Park in the UK.

V.3 Ensure representation of low- and middle-income countries in AI governance decision-making processes

To complement these international dialogues focused on benefit sharing, inclusive AI governance processes could be established to **support international decision-making about the development of advanced AI systems more broadly.** Many low- and middle-income countries currently lack the influence, access, and resources to shape AI progress or participate in global governance discussions about advanced AI (Adan et al., 2024). For example, a recent report from the UN found that of a sample of major recent international governance instruments[18] outside the UN only the members of the G7 are parties to all of them, while 118 countries are in none (United Nations AI Advisory Body, 2024).[19]

When decision-making processes only include a narrow set of actors, the resulting governance frameworks often fail to account for diverse perspectives and needs. Because AI systems have global impacts, these blindspots could pose significant risks for all countries, including frontier AI states. Limited representation in decision-making processes also raises ethical questions due to the potentially transformative economic and social impacts of advanced AI systems, as discussed in **section II** on motivations for benefit sharing.

"Sharing Governance" could therefore be an additional form of benefit sharing, ensuring that the interests of citizens of developing countries are represented in international agreements that will shape the broader AI ecosystem, including shared safety protocols and economic development trajectories.

VI Conclusion

As advanced AI systems continue to reshape economies, ensuring that their substantial benefits are equitably shared becomes increasingly critical. However, despite a growing consensus on the importance of benefit sharing, practical strategies for implementation remain underdeveloped and questions persist about the risks and trade-offs associated with specific approaches. Effective benefit-sharing mechanisms will be crucial for harnessing AI's potential to address global challenges and encouraging meaningful participation in international governance.

This report seeks to contribute to and advance discussions about benefit sharing. It has outlined three potential options for AI benefit sharing: **1) sharing AI resources, 2) sharing access to advanced AI systems, and 3) sharing financial proceeds.** Each approach presents unique advantages and challenges in promoting inclusive economic growth, enabling more widespread technological self-determination, and mitigating shared risks through stronger international cooperation.

[18]This sample included seven initiatives: the OECD AI Principles (2019); G20 AI Principles (2019); Council of Europe AI Convention drafting group (2022-24); GPAI Ministerial Declaration (2022); G7 Ministers' Statement (2023); Bletchley Declaration (2023); Seoul Ministerial Declaration (2024).

[19]The G7 represents just 9% of the global population.

If implemented thoughtfully, international AI benefit sharing could help ensure that advanced AI technologies foster widely shared prosperity. Additionally, it could support international cooperation in managing AI-related risks, an essential factor for effective governance.

Still, this could prove challenging. A wide range of interests must be considered while the trajectory of AI development remains highly uncertain. States will need to cooperate to better understand perceived risks and opportunities and find mutually-agreeable options. Some of these options may require certain actors to give up some leverage over decisions or key resources in order to build trust. Security and safety concerns related to benefit sharing will also demand careful consideration. Advancing these proposals will require not only further research but also coordinated international action to navigate competing interests, assess feasibility and risks, and build the trust needed for successful implementation.

References

Abramson, J. et al. (2024). "Accurate structure prediction of biomolecular interactions with AlphaFold 3". In: *Nature* 630 (8016), pp. 493–500. DOI: 10.1038/s41586-024-07487-w.

Acemoglu, D. (2024). *The simple macroeconomics of AI*. Research rep. w32487. Cambridge, MA: National Bureau of Economic Research. DOI: 10.3386/w32487.

Acemoglu, D. and P. Restrepo (2017). *Robots and jobs: Evidence from US labor markets*. Research rep. Cambridge, MA: National Bureau of Economic Research. DOI: 10.3386/w23285.

– (2018). *Artificial Intelligence, Automation and Work*. Research rep. Cambridge, MA: National Bureau of Economic Research. DOI: 10.3386/w24196.

– (2019). *The Wrong Kind of AI? Artificial Intelligence and the Future of Labor Demand*. Research rep. 25682. Cambridge, MA: National Bureau of Economic Research, w25682. DOI: 10.3386/w25682.

– (2022). "Tasks, Automation, and the Rise in U.S. Wage Inequality". In: *Econometrica: journal of the Econometric Society* 90 (5), pp. 1973–2016. DOI: 10.3982/ECTA19815.

Adan, S. N. et al. (2024). *Voice and Access in AI: Global AI Majority Participation in Artificial Intelligence Development and Governance*. Research rep. Oxford Martin School, University of Oxford. URL: https://www.oxfordmartin.ox.ac.uk/publications/voice-and-access-in-ai-global-ai-majority-participation-in-artificial-intelligence-development-and-governance (visited on 12/02/2024).

Aghion, P., B. Jones, and C. Jones (2017). *Artificial intelligence and economic growth*. Research rep. w23928. Cambridge, MA: National Bureau of Economic Research. DOI: 10.3386/w23928.

Allen, G. C., E. Benson, and M. Putnam (2023). *Japan and the Netherlands Announce Plans for New Export Controls on Semiconductor Equipment*. URL: https://www.csis.org/analysis/japan-and-netherlands-announce-plans-new-export-controls-semiconductor-equipment (visited on 01/18/2025).

Anthropic (2024a). *Pricing*. URL: https://www.anthropic.com/pricing#anthropic-api (visited on 12/02/2024).

– (2024b). *Supported countries and regions*. URL: https://web.archive.org/web/20250104052408/https://www.anthropic.com/supported-countries (visited on 01/12/2025).

– (2024c). *The Claude 3 Model Family: Opus, Sonnet, Haiku*. Research rep. Anthropic. URL: https://www-cdn.anthropic.com/de8ba9b01c9ab7cbabf5c33b80b7bbc618857627/Model_Card_Claude_3.pdf.

Armstrong, S., N. Bostrom, and C. Shulman (2016). "Racing to the precipice: a model of artificial intelligence development". In: *AI & society* 31 (2), pp. 201–206. DOI: 10.1007/s00146-015-0590-y.

Azoulay, P., J. Krieger, and A. Nagaraj (2024). *Old moats for new models: Openness, control, and competition in generative AI*. Research rep. Cambridge, MA. DOI: 10.3386/w32474.

Bandura, R. and E. I. Méndez Leal (2024). *The Digital Literacy Imperative*. Research rep. Center for Strategic and International Studies. URL: https://www.csis.org/analysis/digital-literacy-imperative (visited on 01/17/2025).

Bastagli, F. et al. (2016). *Cash transfers: what does the evidence say? A rigorous review of programme impact and of the role of design and implementation features*. Research rep. Overseas Development Institute. URL: https://odi.org/documents/5301/11316.pdf (visited on 01/17/2025).

Belt and Road Forum for International Cooperation (2023). *Global AI Governance Initiative*. URL: http://www.beltandroadforum.org/english/n101/2023/1019/c127-1231.html (visited on 01/17/2025).

Ben-Ishai, G. et al. (2024). *AI and the opportunity for shared prosperity: Lessons from the history of technology and the economy*. arXiv: 2401.09718 [econ.GN].

Bengio, Y. (2023). *Slowing down development of AI systems passing the Turing test*. URL: https://yoshuabengio.org/2023/04/05/slowing-down-development-of-ai-systems-passing-the-turing-test/ (visited on 01/18/2025).

Bengio, Y., G. Hinton, et al. (2024). "Managing extreme AI risks amid rapid progress". In: *Science*, eadn0117. DOI: 10.1126/science.adn0117.

Bengio, Y., S. Mindermann, et al. (2025). *International AI Safety Report*. DSIT 2025/001. URL: https://www.gov.uk/government/publications/international-ai-safety-report-2025.

Besiroglu, T. et al. (2024). *The Compute Divide in Machine Learning: A Threat to Academic Contribution and Scrutiny?* arXiv: 2401.02452 [cs.CY].

Bindley, K. (2024). *The Fight for AI Talent: Pay Million-Dollar Packages and Buy Whole Teams*. URL: https://www.wsj.com/tech/ai/the-fight-for-ai-talent-pay-million-dollar-packages-and-buy-whole-teams-c370de2b (visited on 01/17/2025).

Birhane, A. (2020). "Algorithmic colonization of Africa". In: *SCRIPT-ed* 17 (2), pp. 389–409. DOI: 10.2966/scrip.170220.389.

Blasi, D., A. Anastasopoulos, and G. Neubig (2022). "Systematic Inequalities in Language Technology Performance across the World's Languages". In: *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. Ed. by S. Muresan, P. Nakov, and A. Villavicencio. Dublin, Ireland: Association for Computational Linguistics, pp. 5486–5505. DOI: 10.18653/v1/2022.acl-long.376.

Boakye, B. et al. (2023). *State of Compute Access: How to Bridge the New Digital Divide*. Research rep. Tony Blair Institute. URL: https://institute.global/insights/tech-and-digitalisation/state-of-compute-access-how-to-bridge-the-new-digital-divide (visited on 12/01/2024).

Bommasani, R. et al. (2023). *The Foundation Model Transparency Index*. Research rep. Center for Research on Foundation Models (CRFM) and Institute on Human-Centered Artificial Intelligence (HAI). URL: http://arxiv.org/abs/2310.12941.

Brand, J. E. (2015). "The far-reaching impact of job loss and unemployment". In: *Annual review of sociology* 41 (1), pp. 359–375. DOI: 10.1146/annurev-soc-071913-043237.

Brownstein, J. S. et al. (2023). "Advances in artificial intelligence for infectious-disease surveillance". In: *The New England journal of medicine* 388 (17), pp. 1597–1607. DOI: 10.1056/NEJMra2119215.

Brugha, R., M. Starling, and G. Walt (2002). "GAVI, the first steps: lessons for the Global Fund". In: *Lancet* 359 (9304), pp. 435–438. DOI: 10.1016/S0140-6736(02)07607-9.

Brundage, M. et al. (2018). *The Malicious Use of Artificial Intelligence: Forecasting, Prevention, and Mitigation*. arXiv: 1802.07228 [cs.AI].

Brynjolfsson, E., D. Li, and L. Raymond (2023). *Generative AI at Work*. Research rep. w31161. Cambridge, MA: National Bureau of Economic Research, w31161. DOI: 10.3386/w31161.

Bureau of Industry and Security (2024). *Commerce Strengthens Export Controls to Restrict China's Capability to Produce Advanced Semiconductors for Military Applications*. URL: https://www.bis.gov/press-release/commerce-strengthens-export-controls-restrict-chinas-capability-produce-advanced (visited on 01/18/2025).

Casey, E., H. Roy, and E. Rockall (2024). *Designing an AI Bond for Growth and Shared Prosperity in the UK - UK Day One*. Research rep. UK Day One. URL: https://ukdayone.org/briefings/ai-bond-for-growth-and-shared-prosperity (visited on 12/02/2024).

CERN (2025). *About CERN*. URL: https://home.cern/about (visited on 01/17/2025).

Chan, A., H. Bradley, and N. Rajkumar (2023). "Reclaiming the digital commons: A public data trust for training data". In: *Proceedings of the 2023 AAAI/ACM Conference on AI, Ethics, and Society*. AIES '23: AAAI/ACM Conference on AI, Ethics, and Society (Montréal QC Canada). Vol. 1442. New York, NY, USA: ACM, pp. 855–868. DOI: 10.1145/3600211.3604658.

Clare, S. and C. Ruhl (2024). *Great Power Competition and Transformative Technologies*. Research rep. Founders Pledge. URL: https://dkqj4hmn5mktp.cloudfront.net/High_Risk_Technology_Competition_01d3b6538a.pdf.

Cottier, B., R. Rahman, et al. (2024). *The rising costs of training frontier AI models*. arXiv: 2405.21015 [cs.CY].

Cottier, B., J. You, et al. (2024). *How Far Behind Are Open Models?* URL: https://epoch.ai/blog/open-models-report (visited on 12/01/2024).

Cui, K. Z. et al. (2024). "The Productivity Effects of Generative AI: Evidence from a Field Experiment with GitHub Copilot". In: *An MIT Exploration of Generative AI*. URL: https://mit-genai.pubpub.org/pub/v5iixksv/release/2.

DeepLearning.AI (2024). *Weekly Issues: Issue 264*. URL: https://www.deeplearning.ai/the-batch/issue-264/ (visited on 12/02/2024).
Delaporte, A. (2023). *New insights on mobile internet connectivity in Sub-Saharan Africa*. URL: https://www.gsma.com/solutions-and-impact/connectivity-for-good/mobile-for-developmentregion/sub-saharan-africa-region/new-insights-on-mobile-internet-connectivity-in-sub-saharan-africa/ (visited on 01/17/2025).
Dennis, C. et al. (2024). *What Should Be Internationalised in AI Governance?* Research rep. Published in partnership with the Centre for the Governance of AI, ICFG, Institute for Law and AI, Stanford Intelligent Systems Laboratory, and SaferAI. Oxford Martin; AI Governance Initiative. URL: https://oms-www.files.svdcdn.com/production/downloads/What%20should%20be%20internationalised%20in%20AI%20Governance-final.pdf?dm=1731486256.
Department for Science, Innovation & Technology (2024). *Frontier AI Safety Commitments, AI Seoul Summit 2024*. Research rep. GOV.UK. URL: https://www.gov.uk/government/publications/frontier-ai-safety-commitments-ai-seoul-summit-2024/frontier-ai-safety-commitments-ai-seoul-summit-2024 (visited on 06/10/2024).
Department for Science, Innovation and Technology (2024). *AI Safety Institute approach to evaluations*. Research rep. GOV.UK. URL: https://www.gov.uk/government/publications/ai-safety-institute-approach-to-evaluations/ai-safety-institute-approach-to-evaluations (visited on 06/10/2024).
Department of Digital Health and Innovation (2024). *Artificial Intelligence for Health*. Research rep. World Health Organization. URL: https://cdn.who.int/media/docs/default-source/digital-health-documents/who_brochure_ai_web.pdf?sfvrsn=aa4f4e3b_3&download=true.
Dubey, A. et al. (2024). *The Llama 3 herd of models*. Research rep. Meta. URL: https://ai.meta.com/research/publications/the-llama-3-herd-of-models/ (visited on 11/04/2024).
Efrati, A. and A. Holmes (2024). *Why OpenAI Could Lose $5 Billion This Year*. URL: https://www.theinformation.com/articles/why-openai-could-lose-5-billion-this-year (visited on 12/01/2024).
Eloundou, T. et al. (2024). "GPTs are GPTs: Labor market impact potential of LLMs". In: *Science* 384 (6702), pp. 1306–1308. DOI: 10.1126/science.adj0998.
Epoch AI (2023). *Key trends and figures in machine learning*. URL: https://epochai.org/trends.
Erdil, E. and T. Besiroglu (2023). *Explosive growth from AI automation: A review of the arguments*. arXiv: 2309.11690 [econ.GN].
Evans, R. and J. Gao (2016). *DeepMind AI reduces Google data centre cooling bill by 40%*. URL: https://deepmind.google/discover/blog/deepmind-ai-reduces-google-data-centre-cooling-bill-by-40/ (visited on 12/01/2024).
Expand Opportunity - AI for Good (2024). URL: https://www.microsoft.com/en-us/research/project/expand-opportunity-ai-for-good/ (visited on 01/18/2025).
Export-Import Bank of the United States (2023). *Report to Congress on the Transformational Export Areas (TEAs)*. URL: https://img.exim.gov/s3fs-public/documents/2023.12.20-exim-report-on-transformational-export-areas.pdf.
Foreign, Commonwealth & Development Office and The Rt Hon James Cleverly MP (2023). *UK unites with global partners to accelerate development using AI*. URL: https://www.gov.uk/government/news/uk-unites-with-global-partners-to-accelerate-development-using-ai (visited on 01/17/2025).
Franke, U. (2021). *Artificial Intelligence diplomacy: Artificial Intelligence governance as a new European Union external policy tool*. Requested by the AIDA committee. URL: https://www.europarl.europa.eu/RegData/etudes/STUD/2021/662926/IPOL_STU(2021)662926_EN.pdf.
Future of Life Institute (2023). *Policymaking In The Pause*. URL: https://futureoflife.org/document/policymaking-in-the-pause/ (visited on 01/18/2025).
G7 Leaders (2023). *G7 Leaders' Statement on the Hiroshima AI Process*. URL: https://digital-strategy.ec.europa.eu/en/library/g7-leaders-statement-hiroshima-ai-process (visited on 05/22/2024).

Gallegos, I. O. et al. (2024). "Bias and fairness in large language models: A survey". In: *Computational linguistics (Association for Computational Linguistics)* 50 (3), pp. 1–83. DOI: 10.1162/coli_a_00524.

Gedikli, C. et al. (2022). "The relationship between unemployment and wellbeing: an updated meta-analysis of longitudinal evidence". In: *European journal of work and organizational psychology*, pp. 1–17. DOI: 10.1080/1359432x.2022.2106855.

Gerstgrasser, M. et al. (2024). *Is Model Collapse Inevitable? Breaking the Curse of Recursion by Accumulating Real and Synthetic Data*. arXiv: 2404.01413 [cs.LG].

Goh, H.-H. and R. Vinuesa (2021). "Regulating artificial-intelligence applications to achieve the sustainable development goals". In: *Discover sustainability* 2 (1), p. 52. DOI: 10.1007/s43621-021-00064-5.

Google (2024). *Available regions for Google AI Studio and Gemini API*. URL: https://web.archive.org/web/20250103103657/https://ai.google.dev/gemini-api/docs/available-regions (visited on 01/12/2025).

– (2025). *Where you can use the Gemini web app*. URL: https://web.archive.org/web/20250106064229/https://support.google.com/gemini/answer/13575153?hl=en (visited on 01/12/2025).

Google DeepMind (2023). *Build AI responsibly to benefit humanity*. URL: https://deepmind.google/about/ (visited on 11/27/2024).

Gosselink, B. H. et al. (2024). *AI in action: Accelerating progress towards the Sustainable Development Goals*. Research rep. Google. URL: https://static.googleusercontent.com/media/publicpolicy.google/en//resources/research-brief-ai-and-SDG.pdf (visited on 11/28/2024).

Grace, K. et al. (2024). *Thousands of AI Authors on the Future of AI*. arXiv: 2401.02843 [cs.CY].

Heim, L. (2024a). *AI Benefit Sharing Options*. URL: https://blog.heim.xyz/ai-benefit-sharing-options/ (visited on 12/01/2024).

– (2024b). *Considerations and Limitations for AI Hardware-Enabled Mechanisms*. URL: https://blog.heim.xyz/considerations-and-limitations-for-ai-hardware-enabled-mechanisms/ (visited on 12/02/2024).

Heim, L. and J. Egan (2023). "Accessing controlled AI chips via infrastructure-as-a-service (IaaS): Implications for export controls". URL: https://cdn.governance.ai/Accessing_Controlled_AI_Chips_via_Infrastructure-as-a-Service.pdf (visited on 12/01/2024).

Heim, L., T. Fist, et al. (2024). *Governing Through the Cloud: The Intermediary Role of Compute Providers in AI Regulation*. Research rep. Oxford Martin AI Governance Initiative. URL: https://cdn.governance.ai/Governing-Through-the-Cloud_The-Intermediary-Role-of-Compute-Providers-in-AI-Regulation.pdf.

Hendrycks, D., M. Mazeika, and T. Woodside (2023). *An Overview of Catastrophic AI Risks*. arXiv: 2306.12001 [cs.CY].

Hine, E. (2024). "Governing silicon valley and Shenzhen: Assessing a New Era of artificial intelligence governance in the United States and China". In: *Digital society: ethics, socio-legal and governance of digital technology* 3 (3). DOI: 10.1007/s44206-024-00138-7.

Ho, L. et al. (2023). *International Institutions for Advanced AI*. arXiv: 2307.04699 [cs.CY].

Hobbhahn, M., L. Heim, and G. Aydos (2023). *Trends in machine learning hardware*. Research rep. Epoch AI. URL: https://epochai.org/blog/trends-in-machine-learning-hardware.

Holmes, A., N. Mascarenhas, and J. Hornstein (2024). *OpenAI CEO Says Company Could Become Benefit Corporation Akin to Rivals Anthropic, xAI*. URL: https://www.theinformation.com/articles/openai-ceo-says-company-could-become-benefit-corporation-akin-to-rivals-anthropic-xai (visited on 12/02/2024).

International Atomic Energy Agency (2016). *How IAEA Technical Cooperation projects work*. URL: https://www.iaea.org/services/technical-cooperation-programme/how-it-works (visited on 01/17/2025).

International Telecommunication Union (2024). *United Nations Activities on Artificial Intelligence (AI)*. Research rep. ITU. URL: https://www.itu.int/dms_pub/itu-s/opb/gen/S-GEN-UNACT-2023-PDF-E.pdf.

International Telecommunication Union (ITU) (2025). *Internet use: Global offline population steadily declines to 2.6 billion people in 2023*. URL: https://www.itu.int/itu-d/reports/statistics/2023/10/10/ff23-internet-use (visited on 01/18/2025).

Jeevanandam, N. (2024). *IndiaAI Meta launch Srijan YuvAI propel open source AI innovation, research, skill development India*. URL: https://indiaai.gov.in/article/indiaai-and-meta-launch-srijan-and-yuvai-to-propel-open-source-ai-innovation-research-and-skill-development-in-india.

Justen, M. (2024). *Sharing the AI Windfall: A Strategic Approach to International Benefit-Sharing*. URL: https://wrtaigovernance.substack.com/p/sharing-the-ai-windfall-a-strategic (visited on 11/30/2024).

Kaplan, J. et al. (2020). *Scaling Laws for Neural Language Models*. arXiv: 2001.08361 [cs.LG].

Kolt, N. et al. (2024). *Responsible Reporting for Frontier AI Development*. arXiv: 2404.02675 [cs.CY].

Korinek, A. (2024). *Economic policy challenges for the age of AI*. Research rep. w32980. Cambridge, MA: National Bureau of Economic Research. DOI: 10.3386/w32980.

Korinek, A. and M. Juelfs (2022). *Preparing for the (non-existent?) future of work*. Research rep. w30172. Cambridge, MA: National Bureau of Economic Research. DOI: 10.3386/w30172.

Korinek, A., M. Schindler, and J. E. Stiglitz (2021). *Technological Progress, Artificial Intelligence, and Inclusive Growth*. Research rep. Research in progress to elicit comments and encourage debate. International Monetary Fund. URL: https://www.imf.org/en/Publications/WP/Issues/2021/07/30/Technological-Progress-Artificial-Intelligence-and-Inclusive-Growth-462086.

Korinek, A. and J. E. Stiglitz (2021). *Artificial Intelligence, Globalization, and Strategies for Economic Development*. Research rep. National Bureau of Economic Research. DOI: 10.3386/w28453.

Korinek, A. and D. Suh (2024). *Scenarios for the Transition to AGI*. Research rep. 32255. National Bureau of Economic Research. DOI: 10.3386/w32255.

Kulp, G. et al. (2024). *Hardware-Enabled Governance Mechanisms: Developing Technical Solutions to Exempt Items Otherwise Classified Under Export Control Classification Numbers 3A090 and 4A090*. Santa Monica, CA: RAND Corporation. DOI: 10.7249/WRA3056-1.

Lam, R. et al. (2023). "Learning skillful medium-range global weather forecasting". In: *Science (New York, N.Y.)* 382 (6677), pp. 1416–1421. DOI: 10.1126/science.adi2336.

Lee, M., M. Nacht, and M. Nacht (2020). "Challenges to the Nuclear Non- Proliferation Treaty". In: *Strategic Studies Quarterly* 14 (3), pp. 95–120. URL: https://www.jstor.org/stable/26937413?seq=18 (visited on 12/02/2024).

Lehdonvirta, V., B. Wu, and Z. Hawkins (2023). *Weaponized interdependence in a bipolar world: How economic forces and security interests shape the global reach of U.S. and Chinese cloud data centres*. DOI: 10.2139/ssrn.4670764.

Leviathan, Y., M. Kalman, and Y. Matias (2022). *Fast inference from Transformers via speculative decoding*. arXiv: 2211.17192 [cs.LG].

Liesenfeld, A. and M. Dingemanse (2024). "Rethinking open source generative AI: open washing and the EU AI Act". In: *The 2024 ACM Conference on Fairness, Accountability, and Transparency*. FAccT '24: The 2024 ACM Conference on Fairness, Accountability, and Transparency (Rio de Janeiro Brazil). New York, NY, USA: ACM. DOI: 10.1145/3630106.3659005.

Liu, S.-Y. et al. (2023). "LLM-FP4: 4-Bit Floating-Point Quantized Transformers". In: *The 2023 Conference on Empirical Methods in Natural Language Processing*. URL: https://openreview.net/forum?id=wiI8ycNfgJ.

MacroPolo (2022). *The Global AI Talent Tracker 2.0*. URL: https://macropolo.org/interactive/digital-projects/the-global-ai-talent-tracker/ (visited on 12/01/2024).

Maslej, N. et al. (2024). *The AI Index 2024 Annual Report*. Research rep. Institute for Human-Centered AI, Stanford University. URL: https://aiindex.stanford.edu/report/ (visited on 07/17/2024).

Mazzucato, M. (2023). *A collective response to our global challenges: a common good and 'market-shaping' approach*. Research rep. 2023/01. UCL Institute for Innovation and Public Purpose. URL: https://www.ucl.ac.uk/bartlett/public-purpose/publications/2023/jan/collective-response-our-global-challenges-common-good-and-market-shaping (visited on 12/02/2024).

Meta (2024a). *Introducing Llama 3.1: Our most capable models to date*. URL: https://ai.meta.com/blog/meta-llama-3-1/ (visited on 10/31/2024).
– (2024b). *With 10x growth since 2023, Llama is the leading engine of AI innovation*. URL: https://ai.meta.com/blog/llama-usage-doubled-may-through-july-2024/ (visited on 12/02/2024).
Microsoft (2024). *Microsoft and G42 announce $1 billion comprehensive digital ecosystem initiative for Kenya*. URL: https://news.microsoft.com/2024/05/22/microsoft-and-g42-announce-1-billion-comprehensive-digital-ecosystem-initiative-for-kenya/ (visited on 11/28/2024).
Miller, C. (2022). *Chip War: The Fight for the World's Most Critical Technology*. New York, NY: Scribner. URL: https://www.simonandschuster.com/books/Chip-War/Chris-Miller/9781982172008 (visited on 04/23/2024).
Mu, T. et al. (2024). *Rule Based Rewards for language model safety*. Research rep. Open AI. (Visited on 12/02/2024).
Musser, M. et al. (2023). *"The Main Resource is the Human": A Survey of AI Researchers on the Importance of Compute*. Research rep. Center for Security and Emerging Technology. DOI: 10.51593/20210071.
Nii-Aponsah, H., B. Verspagen, and P. Mohnen (2023). *Automation-induced reshoring and potential implications for developing economies*. Research rep. UNU-MERIT. URL: https://ideas.repec.org/p/unm/unumer/2023018.html.
Noble, P. M. and T. S. Gruca (1999). "Industrial pricing: Theory and managerial practice". In: *Marketing science* 18 (3), pp. 435–454. DOI: 10.1287/mksc.18.3.435.
O'Keefe, C. (2024). *Chips for peace: How the U.s. and its allies can lead on safe and beneficial AI*. URL: https://www.lawfaremedia.org/article/chips-for-peace--how-the-u.s.-and-its-allies-can-lead-on-safe-and-beneficial-ai (visited on 11/30/2024).
O'Keefe, C. et al. (2019). *The Windfall Clause: Distributing the benefits of AI for the common good*. Research rep. Future of Humanity Institute, University of Oxford. URL: https://cdn.governance.ai/Windfall-Clause-Report.pdf (visited on 12/02/2024).
OECD (2024). *International aid rises in 2023 with increased support to Ukraine and humanitarian needs*. URL: https://www.oecd.org/en/about/news/press-releases/2024/04/international-aid-rises-in-2023-with-increased-support-to-ukraine-and-humanitarian-needs.html (visited on 12/02/2024).
Office of the Federal Register, National Archives and Records Administration (2025). *90 FR 4544 - Framework for Artificial Intelligence Diffusion*. URL: https://www.govinfo.gov/content/pkg/FR-2025-01-15/pdf/2025-00636.pdf.
Okolo, C. T. (2023). *AI in the Global South: Opportunities and challenges towards more inclusive governance*. URL: https://www.brookings.edu/articles/ai-in-the-global-south-opportunities-and-challenges-towards-more-inclusive-governance/ (visited on 04/23/2024).
OpenAI (2018). *OpenAI Charter*. URL: https://openai.com/charter (visited on 04/22/2024).
– (2023). *For-good use cases*. URL: https://openai.com/form/for-good-use-cases/ (visited on 01/17/2025).
– (2024a). *ChatGPT Supported Countries*. URL: https://web.archive.org/web/20250103040402/https://help.openai.com/en/articles/7947663-chatgpt-supported-countries (visited on 05/07/2024).
– (2024b). *GPT-4o mini: advancing cost-efficient intelligence*. URL: https://openai.com/index/gpt-4o-mini-advancing-cost-efficient-intelligence/ (visited on 11/03/2024).
– (2024c). *OpenAI API - Supported Countries and Territories*. URL: https://web.archive.org/web/20240911005416/https://help.openai.com/en/articles/5347006-openai-api-supported-countries-and-territories (visited on 01/12/2025).
– (2024d). *OpenAI o3 and o3-mini—12 Days of OpenAI: Day 12*. URL: https://www.youtube.com/watch?v=SKBG1sqdyIU.
Organisation for Economic Co-Operation and Development (2023). *A blueprint for building national compute capacity for artificial intelligence*. Research rep. OECD. DOI: 10.1787/876367e3-en.

Our World in Data (2024). *Mobile phone subscriptions per 100 people*. URL: https://ourworldindata.org/grapher/mobile-cellular-subscriptions-per-100-people (visited on 01/17/2025).

Oxford Generative AI Summit (2024). *Michael Kratsios, Managing Director, Scale AI | Oxford Generative AI Summit 2024*. YouTube video. URL: https://www.youtube.com/watch?v=ODw372xkZh0.

Parums, D. V. (2023). "Editorial: Infectious disease surveillance using artificial intelligence (AI) and its role in epidemic and pandemic preparedness". In: *Medical science monitor: international medical journal of experimental and clinical research* 29, e941209. DOI: 10.12659/MSM.941209.

Pilz, K., L. Heim, and N. Brown (2023). *Increased Compute Efficiency and the Diffusion of AI Capabilities*. arXiv: 2311.15377 [cs.CY].

Png, M.-T. (2022). "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)*. New York, NY, USA: Association for Computing Machinery, pp. 1434–1445. DOI: 10.1145/3531146.3533200.

Prime Minister's Office, 10 Downing Street et al. (2023). *Countries agree to safe and responsible development of frontier AI in landmark Bletchley Declaration*. URL: https://www.gov.uk/government/news/countries-agree-to-safe-and-responsible-development-of-frontier-ai-in-landmark-bletchley-declaration (visited on 01/18/2025).

Raji, I. D. et al. (2022). "Outsider Oversight: Designing a Third Party Audit Ecosystem for AI Governance". In: *Proceedings of the 2022 AAAI/ACM Conference on AI, Ethics, and Society (AIES '22)*. New York, NY, USA: Association for Computing Machinery, pp. 557–571. DOI: 10.1145/3514094.3534181.

Ritchie, H. et al. (2024). "Internet". In: *Our World in Data*. URL: https://ourworldindata.org/internet (visited on 12/01/2024).

Ruiz, S. (2024). *How to Rebuild the Arsenal of Democracy*. Podcast Interview Transcript. URL: https://www.statecraft.pub/p/how-to-rebuild-the-arsenal-of-democracy.

Sachdeva, N. et al. (2024). *How to Train Data-Efficient LLMs*. arXiv: 2402.09668 [cs.LG].

Sastry, G. et al. (2024). *Computing Power and the Governance of Artificial Intelligence*. arXiv: 2402.08797 [cs.CY].

Scharre, P. (2021). "Debunking the AI arms race theory". In: *Texas National Security Review* 4 (3), pp. 121–132. DOI: 10.26153/TSW/13985.

Seger, E. et al. (2023). *Open-Sourcing Highly Capable Foundation Models: An evaluation of risks, benefits, and alternative methods for pursuing open-source objectives*. arXiv: 2311.09227 [cs.CY].

Seuyong, F. T. et al. (2023). *The Size and Distribution of Digital Connectivity Gaps in Sub-Saharan Africa*. Research rep. World Bank. URL: https://documents.worldbank.org/en/publication/documents-reports/documentdetail/099241003142325200/IDU0cb2e42f3050260484d0b8370b84eee303ecf (visited on 01/17/2025).

Shanahan, M. and K. Bahia (2024). *The State of Mobile Internet Connectivity Report 2024*. Research rep. GSMA. URL: https://www.gsma.com/r/somic/ (visited on 01/17/2025).

Shevlane, T. (2022). *Structured access: an emerging paradigm for safe AI deployment*. arXiv: 2201.05159 [cs.AI].

Silverstein, B. (2023). "Promoting international cooperation to avoid collisions between satellites". In: *Carnegie Endowment for International Peace*. URL: https://carnegieendowment.org/research/2023/09/promoting-international-cooperation-to-avoid-collisions-between-satellites?lang=en (visited on 12/01/2024).

Slattery, P. et al. (2024). *The AI Risk Repository: A comprehensive meta-review, database, and taxonomy of risks from Artificial Intelligence*. arXiv: 2408.12622 [cs.AI].

Stafford, E., R. F. Trager, and A. Dafoe (2022). *Safety not guaranteed: International races for risky technologies*. Research rep. Center for the Governance of AI. URL: https://cdn.governance.ai/International_Races_for_Risky_Technologies_DRAFT_NOV_2022.pdf (visited on 11/30/2024).

Susskind, D. (2020). *A World Without Work: Technology, Automation, and How We Should Respond*. Ed. by M. Books. Macmillan. URL: https://us.macmillan.com/books/9781250173515/aworldwithoutwork/ (visited on 11/28/2024).

Svanberg, M. et al. (2024). *Beyond AI exposure: Which tasks are cost-effective to automate with computer vision?* DOI: 10.2139/ssrn.4700751.

Tallarita, R. (2023). *AI Is Testing the Limits of Corporate Governance*. URL: https://hbr.org/2023/12/ai-is-testing-the-limits-of-corporate-governance (visited on 12/02/2024).

Technology Innovation Institute (2023). *Falcon-180B Model Card*. URL: https://huggingface.co/tiiuae/falcon-180B (visited on 12/01/2024).

The Francis Crick Institute (2022). *The Francis Crick Institute and DeepMind join forces to apply machine learning to biology*. URL: https://www.crick.ac.uk/news/2022-07-06_the-francis-crick-institute-and-deepmind-join-forces-to-apply-machine-learning-to-biology (visited on 01/17/2025).

The International Rescue Committee (2024). *OpenAI x International Rescue Committee: Leveraging AI to Scale Ed-Tech in Crisis Affected Settings*. URL: https://www.rescue.org/press-release/openai-x-international-rescue-committee-leveraging-ai-scale-ed-tech-crisis-affected (visited on 01/17/2025).

The White House (2024). *Memorandum on Advancing the United States' Leadership in Artificial Intelligence; Harnessing Artificial Intelligence to Fulfill National Security Objectives; and Fostering the Safety, Security, and Trustworthiness of Artificial Intelligence*. URL: https://www.whitehouse.gov/briefing-room/presidential-actions/2024/10/24/memorandum-on-advancing-the-united-states-leadership-in-artificial-intelligence-harnessing-artificial-intelligence-to-fulfill-national-security-objectives-and-fostering-the-safety-security/ (visited on 01/12/2025).

Trammell, P. and A. Korinek (2020). *Economic growth under transformative AI*. Research rep. Global Priorities Institute. URL: https://globalprioritiesinstitute.org/philip-trammell-and-anton-korinek-economic-growth-under-transformative-ai/ (visited on 11/28/2024).

Ueno, H. (2023). "Artificial intelligence as dual-use technology". In: *Intelligent Systems Reference Library*. Intelligent systems reference library. Cham: Springer International Publishing, pp. 7–32. DOI: 10.1007/978-3-031-22371-6_2.

UN Global Pulse (2023). *Disha*. URL: https://disha.unglobalpulse.org/ (visited on 12/02/2024).

United Nations (2007). *United Nations Declaration on the Rights of Indigenous Peoples (UNDRIP)*. Adopted by the United Nations General Assembly. URL: https://www.un.org/development/desa/indigenouspeoples/wp-content/uploads/sites/19/2018/11/UNDRIP_E_web.pdf.

– (2024). *The Sustainable Development Goals Report 2024*. Research rep. Includes data and insights on achieving the Sustainable Development Goals (SDGs). New York: United Nations. URL: https://unstats.un.org/sdgs/report/2024/The-Sustainable-Development-Goals-Report-2024.pdf.

United Nations AI Advisory Body (2024). *Governing AI for Humanity*. URL: https://www.un.org/sites/un2.un.org/files/governing_ai_for_humanity_final_report_en.pdf.

United Nations Economic Commission for Europe (2025). *Net official development assistance (ODA) as a percentage of OECD-DAC donors GNI (grant equivalent methodology)*. URL: https://w3.unece.org/SDG/en/Indicator?id=72 (visited on 01/30/2025).

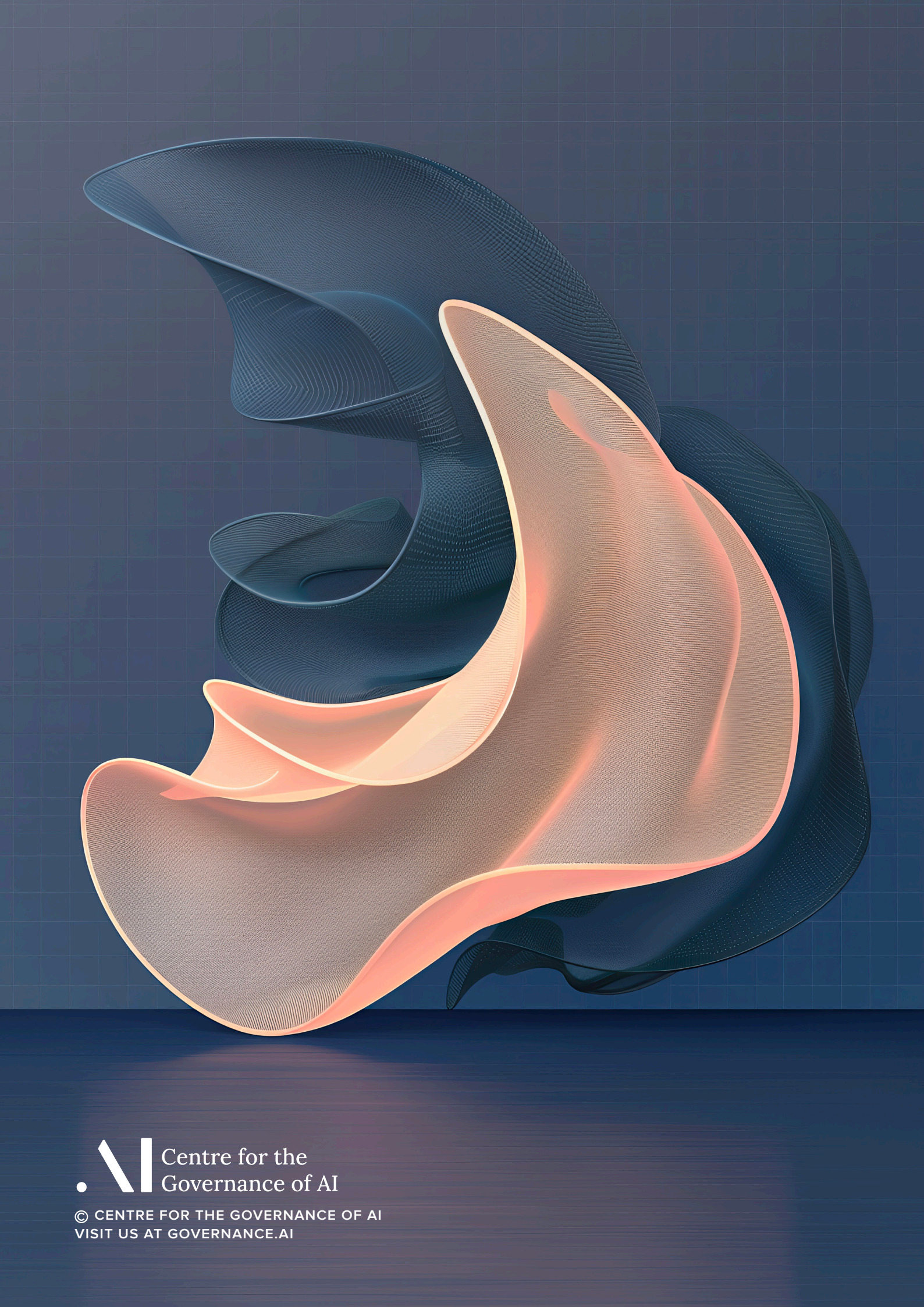
United Nations Office of the Secretary-General's Envoy on Technology (2024). *Global Digital Compact*. URL: https://www.un.org/techenvoy/global-digital-compact (visited on 05/22/2024).

US Department of Energy (2024). *DOE Announces Roadmap for New Initiative for Artificial Intelligence in Science, Security and Technology*. URL: https://www.energy.gov/articles/doe-announces-roadmap-new-initiative-artificial-intelligence-science-security-and (visited on 12/01/2024).

US Department of Homeland Security (2024). *Department of Homeland Security Report on Reducing the Risks at the Intersection of Artificial Intelligence and Chemical, Biological, Radiological, and Nuclear Threats*. Research rep. URL: https://www.dhs.gov/sites/default/files/2024-06/24_0620_cwmd-dhs-cbrn-ai-eo-report-04262024-public-release.pdf.

US Department of State (2024). *Secretary Antony J. Blinken at the Advancing Sustainable Development Through Safe, Secure, and Trustworthy AI Event*. URL: https://www.state.gov/secretary-

antony-j-blinken-at-the-advancing-sustainable-development-through-safe-secure-and-trustworthy-ai-event/ (visited on 12/02/2024).
US National Security Commission on Artificial Intelligence (2021). *The Final Report*. Research rep. NSCAI. URL: https://reports.nscai.gov/final-report/ (visited on 12/01/2024).
Villalobos, P. et al. (2022). *Will we run out of data? Limits of LLM scaling based on human-generated data*. arXiv: 2211.04325 [cs.LG].
Vipra, J. et al. (2024). *Policy Brief: Governing Computational Infrastructure for Strong and Just AI Economies*. Research rep. G20. URL: https://itforchange.net/governing-computational-infrastructure-for-strong-and-just-ai-economies (visited on 12/02/2024).
Winter-Levy, S. (2024). *The Emerging Age of AI Diplomacy*. URL: https://www.foreignaffairs.com/united-states/emerging-age-ai-diplomacy (visited on 01/12/2025).
World Bank (2023a). *Access to electricity (% of population)*. URL: https://data.worldbank.org/indicator/EG.ELC.ACCS.ZS (visited on 01/17/2025).
– (2023b). *GDP per capita (current US$)*. URL: https://data.worldbank.org/indicator/NY.GDP.PCAP.CD (visited on 12/02/2024).
– (2024a). *Individuals using the Internet (% of population) - Sub-Saharan Africa*. URL: https://data.worldbank.org/indicator/IT.NET.USER.ZS?locations=ZG (visited on 01/18/2025).
– (2024b). *Poverty, Prosperity, and Planet Report: Pathways Out of the Polycrisis*. Research rep. World Bank. DOI: 10.1596/978-1-4648-2123-3.
– (2024c). *World Bank Country and Lending Groups*. URL: https://datahelpdesk.worldbank.org/knowledgebase/articles/906519-world-bank-country-and-lending-groups (visited on 12/01/2024).
Zhou, Y. et al. (2023). "A foundation model for generalizable disease detection from retinal images". In: *Nature* 622 (7981), pp. 156–163. DOI: 10.1038/s41586-023-06555-x.



Centre for the
Governance of AI

VISIT US AT GOVERNANCE.AI