Artificial Intelligence (AI) and Global Justice

Siavosh Sahebi[1] · Paul Formosa[1]

Abstract

This paper provides a philosophically informed and robust account of the global justice implications of Artificial Intelligence (AI). We first discuss some of the key theories of global justice, before justifying our focus on the Capabilities Approach as a useful framework for understanding the context-specific impacts of AI on low- to middle-income countries. We then highlight some of the harms and burdens facing low- to middle-income countries within the context of both AI use and the AI supply chain, by analyzing the extraction of *materials*, which includes mineral extraction and the environmental harms associated with it, and the extraction of *labor*, which includes unethical labor practices, low wages, and the trauma experienced by some AI workers. We then outline some of the potential harms and benefits that AI poses, how these are distributed, and what global justice implications this has for low- to middle-income countries. Finally, we articulate the global justice significance of AI by utilizing the Capabilities Approach. We argue that AI must be considered from a global justice perspective given that, globally, AI puts significant downward pressure on several elements of well-being thereby making it harder for people to achieve threshold levels of the central human capabilities needed for a life of dignity.

Keywords Artificial intelligence (AI) · Global justice · Capabilities · Human dignity · Large language models

✉ Siavosh Sahebi
  siavosh.sahebi@mq.edu.au

  Paul Formosa
  paul.formosa@mq.edu.au

[1] Philosophy, Macquarie University, Sydney, Australia

🍀 Springer

1 Introduction

As Artificial Intelligence (AI) proliferates throughout society, there has been a greater emphasis on the ethical implications of the technology (e.g. Hagendorff, 2020; Borenstein & Howard, 2020; Lauer, 2020; Morley et al., 2021), including as it pertains to low- to middle-income countries (e.g. Kak, 2020; Png, 2022; Ricaurte, 2022; Mannuru et al., 2023). Within the AI pipeline, which includes the supply chain and development of AI, there is a significant need for and usage of resources. This includes the need for natural resources, critical minerals, hardware and infrastructure, energy and power usage, and data and labor. All of these require substantial amounts of extraction, be it minerals, data, or labor, and this extraction has an impact on both the environment and people. Further, much of this resource extraction occurs within low- to middle income countries, and this raises important questions around global justice. This is because despite the significant role that the citizens of these less well-off countries play in the supply chain and development of AI, there is a very real risk that they will not only miss out on the benefits derived from AI, leading to greater global inequality, but also that AI may be used against them in a malevolent way and for malfeasance.

There has been a growing focus in the literature around AI and its ethical implications (Morley et al., 2021), including, but not limited to, the ethical concerns within the AI supply chain (Muldoon et al., 2023), the ethical challenges around datafication, algorithms and automation (Ricaurte, 2022), as well as the ethical dangers of Generative AI (Allen & Weyl, 2024; Nah et al., 2023). However, while *ethical* considerations have been central to the literature, issues of *justice* have been both less explicit and comparatively less prominent. Where justice considerations *are* discussed explicitly in their own right, it is often done in terms of *domestic* justice (e.g. Gabriel, 2022; Sloan & Warner, 2020; Santoni de Sio et al., 2024; Hickok, 2022). Further, where *global* justice and AI is discussed, these discussions are not always deeply embedded in relevant philosophical theories of global justice (e.g. Rafanelli, 2022; Ricaurte, 2022; Tacheva & Ramasubramanian, 2023; Arsenault & Kreps, 2022). Therefore, there is need for work to address this gap by providing a philosophically informed and robust account of the global justice implications of AI. We aim to fill this gap here by focusing on AI as an explicitly global justice issue by utilizing the Capabilities Approach to global justice to flesh out the ways in which individuals' well-being are impacted in the context of AI, both within the AI pipeline and development process, as well as resulting from the global distribution of the harms and benefits of the technology. As we note in more detail in the next section, the Capabilities Approach allows us to assess these impacts across different contexts by investigating how the central capabilities necessary for a life of dignity are impacted, making this approach useful for this analysis. However, we do not rule out that other theories of global justice may also prove fruitful, and we address some of these in the next section.

It is important that we clarify here that there are elements of our analysis which are not unique to AI and apply to other technologies and industries, such as concerns around resource extraction. Throughout our analysis we will emphasize that AI is being developed and deployed within an existing unjust and non-ideal global order.

Our analysis will also articulate AI specific concerns to ensure our argument is clear in its focus on the AI context and the role AI plays in the global justice concerns we raise. Some AI specific concerns include certain labor practices, the malicious use of AI and AI weaponization, the digital language divide, as well as procedural unfairness in global AI regulation. By addressing these issues, it will become clear that while certain elements of the concerns we raise here are not unique to the AI industry, other elements are specific to AI, and together both elements raise important global justice concerns that demand attention given the rising adoption of AI technology.

The outline of the paper is as follows. First, we lay the groundwork for this analysis by discussing some of the key theories of global justice, before justifying our focus on the Capabilities Approach as a useful framework for understanding the context-specific impacts of AI on low- to middle-income countries. We also note that alternative approaches to global justice should be considered complementary to the account developed here. Next, we analyze the global justice implications of AI in terms of the extraction of *materials*, which includes mineral extraction and the environmental harms associated with it, and the extraction of *labor*, which includes unethical labor practices, low wages, and the trauma experienced by workers. The aim here is to highlight some of the harms and burdens facing low- to middle-income countries within the context of AI and the AI supply chain. This will be followed by a section which outlines some of the potential harms and benefits that AI poses, how they are distributed, and what justice implications this will have for low- to middle-income countries on a global level. The final section will articulate the global justice significance of AI by utilizing the Capabilities Approach. Drawing this together, we argue that AI must also be considered from a global justice perspective, not merely from an ethical or domestic justice perspective, and active steps should be taken to address some of these globally important justice concerns.

2 Theories of Global Justice

Guo et al. (2019) break down theories of global justice into three main theoretical approaches: rights-based, which focus on the basic principles of justice, such as the right to equal treatment; goods-based, which focus on the distribution of primary goods, capabilities and equal opportunities for welfare; and virtue-based, which focus on the internalization of justice as a virtue which is acquired by individuals. We note that many theories of global justice incorporate different elements of these approaches, meaning both that categorizing them into one of the three main approaches outlined here may be difficult and that these three approaches are not mutually exclusive. For example, Rawls' (1971) primary goods approach entails elements of rights-based approaches, such as basic rights and liberties, while Nussbaum's (2000) Capabilities Approach, which will be a focus of this analysis, contains elements of virtue approaches. However, for the purposes of this analysis, these three broad categories provide us with a useful way of differentiating between the core elements of different theories.

All three approaches have relevance in working out the global justice implications of AI. For example, rights-based accounts will be most useful in ideating issues

around AI-driven algorithmic discrimination and unfairness, such as biased algorithms that favor a particular class of individual for a job opportunity (Borenstein & Howard, 2021). Virtue-based accounts also have a role when we consider broader issues around the impacts of AI on humans, such as the ways in which AI models may influence the development of epistemic and moral virtues or vices (Vallor, 2016). On a more global level, all accounts play an important role when considering the implications of AI development being largely driven by for-profit corporations in mostly WEIRD (Western, Educated, Industrialized, Rich and Democratic) countries (McQuillan, 2022) and the implications this has for bias and accessibility (Sloan & Warner, 2020). However, our focus here will be on goods-based approaches to justice, although future work could supplement our work by drawing on these alternative approaches. The reason for this focus is that our primary concern here, given our international focus, is around the global distribution of harms and benefits flowing from AI and the role of the international order in this distribution. We are therefore concerned with analyzing, from a justice perspective, the global distribution of both primary goods and capabilities within the context of the international supply chain, development, and use of AI.

Before we further explore our focus on goods-based approaches to justice, it is important that we first justify why these approaches are relevant to our current global context. The global order plays a significant role in the distribution of goods, as well as the harms and benefits that flow from that distribution. Consequently, it is this same distribution of goods which underlies historic global inequalities, which subsequently helps to create and maintain current global inequalities. Veneziani and Yoshihara (2024), when examining the structure of the interaction between countries in global markets, found that global inequality in primary goods and well-being derives from differences in the ownership of productive assets which can be used to produce further goods, including commodities and labor. To understand what this means in an empirical sense, Hickel at el. (2021) found that between 1960 and 2018 the appropriation of commodities and unequal exchange led to $62 trillion being drained from low- to middle-income countries, which amounts to $152 trillion when accounting for lost growth. Hickel et al. (2022) further found that between 1990 and 2015, $242 trillion was drained from low- to middle-income countries due to the appropriation of resources and labor. The latter study highlights the significant value that labor plays within the global economy. This drain from low- to middle-income countries is a key feature of the global economy which benefits wealthy, high-income countries. Further, the losses incurred by low- to middle-income nations through this process are 30 times greater than the total aid receipts they received over the same period (Hickel et al., 2022). This clearly has a significant impact on people's ability to meet their basic needs and enjoy decent lives. The AI pipeline and development process is part of this broader global dynamic and thereby threatens to exacerbate these existing inequalities.

Within the context of AI and its pipeline, as we will see in the next section, there is a significant need for resources and labor, and much of this comes from low- to middle-income countries. As Crawford (2021, 31) puts it, "the full-stack supply chain of AI reaches into capital, [labor], and Earth's resources – and from each, it demands an enormous amount". As we have seen above, this trend of resource and labor extrac-

tion from low- to middle-income countries is not new to AI, but it is a continuation of the same global order which has moved into a new domain. The reason these considerations are important for our analysis of global justice within the context of AI, and what helps to justify a goods-based approach here, is that low- to middle-income countries are carrying a significant burden within the AI pipeline as it pertains to the resources and labor that are being appropriated from them to develop AI technologies. Meanwhile, there is a significant risk that these countries and their citizens may miss out on many of the benefits derived from AI's development, an issue we will address in more detail in Sect. 4.

A useful overarching framework to understand much of the global reach of AI is to draw on Tacheva and Ramasubramanian (2023, 2) argument that we are in the "age of AI Empire", which they refer to as a period marked by deep and pervasive influence of AI on many aspects of global societies. According to this framework, the mechanisms under which AI Empire operates are: *extractivism*, the extraction of natural resources and materials, labor and data; the *automation* of physical and cognitive labor; *essentialism*, the reduction of complex and diverse social categories to a set of predefined and measurable traits; *surveillance*, the tracking, management and control of populations; and the *containment* and control of populations both in the physical environment they inhabit and their online environments through access to digital services (Tacheva & Ramasubramanian, 2023; 6–10). While this framework does not play a central role in our analysis, it does helpfully bring together several of the individual elements that we focus on throughout our discussion (and back up with evidence), which helps to highlight how these mechanisms interact and are currently playing out in the context of AI. While our aim is not to argue for the overall validity of this framework, it nonetheless provides a useful theoretical backdrop for the current analysis in that it allows us to better understand how the different mechanisms within AI Empire fit together and are influencing global justice concerns.

There are numerous goods-based approaches to global justice, including Rawls' (1971) primary goods approach which argues that there are several basic goods which every rational individual requires in order to pursue their own ends. This includes the fulfilment of material needs, which requires access to income, wealth, and resources, basic rights and liberties, freedom of movement, and the social bases of self-respect. Another important goods-based approach to global justice is developed by Thomas Pogge. For Pogge, institutions play a significant role when we are considering matters of justice and morality, and humans play an integral role in forming, maintaining, and changing these institutions (Pogge, 1988). He argues that we can fault institutions for the excessive inequalities they perpetuate and that the goal of global justice should be to ameliorate these. For Pogge (1988, 228), justice demands that those who are advantaged by these global institutions, primarily those who live in high-income countries, should "work to reform the institutions that establish or engender such deprivations or excessive inequalities in rights, opportunities, or bargaining positions".

According to Pogge (2001, 14), there are at least three morally significant ways in which those in high-income countries are connected to what he refers to as "the global poor". First, there is a shared single historical process where injustices such as colonialism, genocide, and slavery play a direct role in the economic inequalities

we see globally. Second, there is a shared dependence on the base natural resources available globally which are largely exploited by high-income countries and their corporations. High-income countries, and the corporations they empower, exploit their relationships with those in power within low- to middle- income countries to gain control over these resources in a way that benefits them to the detriment of the majority of the world's citizenry. Third, there is a shared single global economic order which largely maintains and exacerbates economic inequality globally (Pogge, 2001).

One area where Pogge's institutional account of global justice may prove fruitful in our context is around AI regulation. Bradford (2023) highlights, regulation within the digital economy is largely dominated by three regulatory approaches from three dominant governing institutions: the American market-driven model, the Chinese state-driven model, and the European rights-driven model. Bradford (2023) argues, these approaches dominate institutions globally and they shape the digital economy and digital society with global consequences. These developments will therefore influence the regulatory landscape globally, including in low- to middle-income countries which do not have institutions which are as resource rich or robust as those of the dominating global powers. This creates procedural justice concerns, as the EU, US, and to a lesser extent China, write the *de facto* global AI regulatory rules that all nations will be largely bound by in practice, even though these other nations had little or no input into the creation or content of these regulatory regimes.

However, it is important to consider the reality that low- to middle-income countries and high-income countries are not monolithic. They each have their own historical, cultural, social, and political contexts which situate their state of development. Throughout this paper, we bring together cases from countries across the global spectrum to articulate our argument and, while these cases highlight a global trend, to treat each case as identical is not only inaccurate but would also be a disservice to the citizens of different countries. While our analysis necessarily involves a degree of abstraction to highlight global trends in the context of AI, we do not wish to abstract away everything to the point where we treat all cases as being the same. Therefore, we require an account which gives us a framework to assess the global justice implications of AI in a way that addresses this reality.

Among the various goods-based approaches to global justice, one of the most prominent is the Capabilities Theory of justice. Sen (1979) and Nussbaum (2000) argue for a Capabilities Approach to justice where we can assess whether the thresholds for the central human capabilities needed for justice and human well-being are achieved. Sen (1993) argues that the capabilities approach provides a relevant framework to assess well-being, in the form of achievement and freedom, as well as living standards, whereas Nussbaum focuses on a list of capabilities needed to achieve "a life that is worthy" of the "dignity of the human being" (Nussbaum, 2003, p. 40). The need for a Capabilities Approach to justice is due to interpersonal and intersocial variations which are often not captured by metrics of primary goods or simple metrics such as GDP, justifying a focus on what individuals can *be* and *do* in their contexts, i.e. their capabilities (Sen, 1993; Guo et al., 2019). Differences in circumstances mean that individuals have differing abilities to convert resources into levels of well-being (Santoni de Sio et al., 2024). The capabilities approach allows us

to gain an objective understanding of people's lives by evaluating their well-being, including the important functions of human agency and choice (Santoni de Sio et al., 2024). This approach allows us to assess how individuals are being impacted within the AI pipeline and development process, as well as explore how the global distributions of harms and benefits caused by this process impacts on their well-being.

To assess this over different contexts, Nussbaum (2000) provides a list of ten human capabilities which are integral to human well-being and dignity on the individual level, as well as ten principles which can be applied to global structures to assess their promotion of capabilities in the face of global inequality (Nussbaum, 2004). On the individual level these central capabilities are:

1. Life;
2. Bodily health;
3. Bodily integrity;
4. Senses, imagination and thought;
5. Emotions;
6. Practical reason;
7. Affiliation;
8. Other species;
9. Play;
10. Control over one's environment (Nussbaum, 2000).

The ten principles for the global structure are:

1. Over-determination of responsibility to promote capabilities;
2. Respect for national sovereignty within the constraints of promoting human capabilities;
3. Responsibility of prosperous nations to give substantial portions of their GDP to poorer nations;
4. Responsibility of multinational corporations to promote human capabilities in the regions they operate;
5. A global economic system designed to be fair to poor and developing countries;
6. Cultivation of a thin, decentralized, yet forceful global public sphere;
7. Institutional and individual focus on the problems of the disadvantaged in each nation and region;
8. Care for the ill, the elderly and the disabled as a prominent focus of the world community;
9. The family sphere as precious but not private;
10. A responsibility by all institutions and individuals to support education as a key to empowering currently disadvantaged people (Nussbaum, 2004).

In articulating her principles for the global structure, Nussbaum (2004) argues that the Capabilities Approach, by suggesting a set of basic human entitlements which form a minimum requirement for justice, provides a valuable way of considering goals of development given the interconnectedness of our world.

As will become clear below, the AI pipeline and development process is an example of the interconnected nature of our current global order. There is a strong demand for resources, including minerals and labor, from less-developed parts of the world, which are largely appropriated by wealthy corporations and governments for their own AI development. As Santoni de Sio et al. (2024) point out, the Capabilities Approach provides a powerful conceptual framework in the context of AI because technology can directly impact both people's well-being and their opportunities to realize their own ends. Given the differences in circumstances that people find themselves in, evaluating individuals' well-being and achievement of human dignity in terms of capabilities gives us a more robust understanding of whether we are meeting the demands of global justice in the age of AI.

It is important to note that there are also criticisms of the Capabilities Approach. For example, Sen argues against Nussbaum's use of a pre-determined list of central capabilities because such a task needs to be context specific and demands public participation and reasoning (Sen, 2005). Others have criticized Nussbaum's list of capabilities as lacking individual and cultural differences, and therefore not having universal application (Qizilbash, 2002). Okin (2003) further argues that the list of capabilities relates more to western, liberal, and educated conceptions of human life and dignity, and lacks meaningful consideration of conceptions found elsewhere in the world. While we acknowledge these criticisms (and others; e.g., Formosa & Mackenzie, 2014), Nussbaum's list of central capabilities nonetheless provides a useful starting point for an analysis of the global justice implications of AI as its list-based approach clearly articulates a range of important impact areas to investigate. Further, as noted earlier, we believe that this current analysis can be - and should be - complimented by other theories of global justice, as these may bring forth other justice implications. Nonetheless, despite its potential limitations, the Capabilities Approach remains a powerful framework to begin assessing the emerging global impact of AI on human dignity. We will begin that process in the next section by taking a deeper look at the AI pipeline.

The basic structure of the argument below is to outline the various levels of harms and lack of benefits globally imposed on different groups by the development and use of AI, before showing how this puts downward pressure on the ability of many people, especially those already in vulnerable positions, to meet the threshold levels of the capacities needed for a dignified life. As well as outlining the negative impacts of the AI industry on the ten central human capabilities, we also draw out further global justice implications, including global procedural unfairness in AI regulation regimes. Several clarifications are required to properly understand the scope of this argument. First, AI is not unique in this regard. Many other technologies also put downward pressure on the capabilities of vulnerable people around the globe, and this can often depend on similar processes of extracting resources and labor as we outline below. But while the AI industry is not unique in this regard, it is still significant and becoming more significant over time as the adoption of the technology grows. The global justice implications of this growth need to be part of a broader assessment of the technology, and while some of the harms the AI industry leads to are common to other forms of technology (such as those that require mineral extraction and resource use), there are also harms that are produced that are more specific to AI technology

(such as malicious AI use and the digital language divide). Whether the AI industry is better or worse in this regard than a range of other comparative industries is a further issue beyond our scope or focus. Second, given our global justice lens, we are focused here on the overall global distribution of the benefits and harms of AI. There will likely also be many benefits of AI technology, and some of these will help to lift some people up to higher levels of functioning in certain central human capabilities. But, as we argue below, these benefits are unlikely to be distributed evenly, and this uneven distribution means that, overall, downward pressure will still likely be exerted on members of some globally vulnerable groups, thereby raising the global justice concerns we focus on here. This does not imply that no one, or no members of globally vulnerable groups, will receive any benefits whatsoever from AI technology. What matters from a global justice point of view is the overall pattern of distribution and how this impacts the extent people can live dignified lives.

Finally, we are engaged in a project here of non-ideal theory that takes seriously the facts of global non-compliance with the principles of justice (Valentini, 2012). The AI industry, as with all other industries, operates within this broader non-ideal context. This means that, while the AI industry is hardly responsible for creating the globally unjust environment in which it finds itself, we need to assess it in terms of the global environment in which it actually operates, especially given its potential to exacerbate already existing global inequalities and injustices.

3 Resource Extraction within the AI Supply Chain and Development Process

This section will explore the *resource extraction* present within the context of AI, including labor practices, highlighting not only the crucial role that low- to middle-income countries and their citizens play in the AI supply chain and development process, but also the harms that can be attributed to these extractive processes. Before we begin our analysis of the AI supply chain, it is important to note that resource extraction and the outsourcing of labor has occurred for decades within many different industries. The concerns outlined here are not unique to the AI supply chain. However, given the transformative nature of AI and the range of possible harms and benefits it could pose, including the possible exacerbation of current inequalities, it is still important to provide an analysis of the impacts of the AI supply chain if we are to understand fully the global justice issues raised by this technology.

3.1 Mineral Extraction

To build and maintain the infrastructure required for AI to operate, AI corporations need access to vast amounts of critical mineral resources. In Sect. 2 we outlined the interdependence of the global economic order and the inequalities that arise from it. This leads to a global power asymmetry which can be exploited by corporations which are empowered by largely high-income countries. Sultana (2022, 4) calls this the "colonial logics of extractivism" which is part of the AI Empire (Tacheva & Ramasubramanian, 2023). This means that multinational corporations can extract

resources from low- to middle-income countries and direct them to high-income countries, where their power is legitimized, at the direct expense of low- to middle-income countries and their citizens (Png, 2022).

To train and develop AI models, and to power AI inferences, a large amount of computing power, especially Graphics Processing Units (GPUs), are needed. GPUs and other core components for computing require large amounts of minerals and resources (Crawford, 2021), as well as power to run them (McQuillan, 2022; Bender et al., 2021). These computing resources in turn require certain basic building blocks, most prominently minerals such as cobalt, copper, lithium, nickel, tin, and tantalum (Crawford, 2021). The European Parliamentary Research Service's Panel for the Future of Science and Technology warned that the extraction of critical minerals such as nickel, cobalt and graphite, which are vital for lithium battery production, will only be exacerbated by the demands posed by AI development, leading to further depletion of critical minerals and causing further environmental degradation (Bird et al., 2020). Furthermore, countries such as Bolivia, Mongolia, Indonesia and central Congo which have lithium rich sites are also impacted by the extraction taking place within the AI supply chain (Crawford, 2021). Tapia and Peña (2020, 160) also highlight the destructive impact of lithium extraction in the "lithium triangle" comprising of Argentina, Bolivia and Chile, particularly the ecological damage to the environment and the impacts this has on Indigenous populations. The extraction of tin - used in semiconductors – impacts locals by damaging their farms and forests, and by destroying their coral reefs and local fish supplies (Crawford, 2021). This is the case in the Indonesian Islands of Bangka and Belitung, and NVIDIA Corporation, which is a leader in chipmaking and was estimated to control 95% of the market share for AI servers in 2023 (de Vries, 2023), reports smelters and refiners from these areas as part of their supply chain (NVIDIA Corporation, 2024).

The extraction of these minerals also comes with direct harms impacting citizens in low- to middle- income countries, including geopolitical conflict and violence, as well as environmental harms (Crawford, 2021; Dauvergne, 2022; Png, 2022). For example, there has been significant reporting on the conflict and violence attributed to the mining sector in the Congo where thousands have died and millions have been displaced (Eichstaedt, 2011; Diemel & Hilhorst, 2019; Crawford, 2021). This is relevant to the AI supply chain given that NVIDIA's 2024 Conflict Minerals Report discloses that 37 smelters and refiners out of the 237 worldwide processing facilities within their supply chain are sourcing from the Democratic Republic of the Congo or an adjoining country (NVIDIA Corporation, 2024). As it pertains to environmental harms, there are environmental damages associated with mineral extraction (Sultana, 2022; Tapia & Peña, 2020), with low- to middle- income countries being used as dumping grounds for electronic and chemical waste and mining refuse, as well as polluting of local water supplies and diminishing natural resources (Crawford, 2021; Dauvergne, 2022; Png, 2022).

A further concern pertains to the energy consumption and usage by AI models and data centers (IEA, 2024; EPRI, 2024). For example, a traditional Google search uses approximately 0.3 W-hours of electricity, while a request to ChatGPT uses 2.9 W-hours, a 10-fold increase in electricity usage (IEA, 2024; EPRI, 2024). Big Tech companies, such as Google and Microsoft, have seen significant increases in

their recent greenhouse gas emissions largely due to their build out of data centers to train and run AI models, with Google's emissions increasing by 48% in the last five years and Microsoft's by almost a third since 2020 (Hodgson and Morris, 2024). Given that the global impacts of climate change tend to be worse for people in low- to middle-income countries (Sultana, 2022; Odeku, 2022; Ngcamu, 2023), this is yet another example of unequal harms being imposed on individuals. As Png (2022) argues, a disproportionate burden associated with the human costs of environmental degradation is borne by low- to middle-income countries. Furthermore, it is generally understood in terms of environmental justice that low-income and minority communities tend to suffer the greatest environmental harms, such as pollution (Bell, 2004). Vulnerable and marginalized populations, including those with disabilities, women, the elderly and Indigenous groups, are particularly vulnerable to the impacts of environmental degradation and climate change (Ngcamu, 2023). While low- to middle-income countries and their citizens are disproportionately harmed throughout this process, the profits and benefits from it are enjoyed by high-income Western countries and their economies (Muldoon & Wu, 2023). This will be explored further in Sect. 4.

3.2 Labor Extraction

AI Empire's mechanism of extractivism also extends to labor. The AI supply chain and development process is plagued by unethical labor practices both within the extraction stage and the data acquisition, curation, training, and fine-tuning stages (Tacheva & Ramasubramanian, 2023; Muldoon & Wu, 2023). Within the resource extraction phase, working conditions in mines in many developing countries, such as the Congo, are said to amount to modern slavery, while workers - which often includes children - who work in electronic dumping sites are exposed to toxins from electronic waste (Png, 2022). The Walk Free, 2023 Global Slavery Index reports that wealthy G20 countries import US$243.6 billion worth of electronic products which are at risk of involving forced labor in their production (The Global Slavery Index, 2023). There are also ethical issues around unregulated workers who do not have any formal worker or environmental protections, known as "gray-market miners", who are not employed officially and find themselves in unsafe conditions without any legal protections (Crawford, 2021; 37). While mineral extraction, poor working conditions, and electronic waste encompass much more than AI, as use of AI products increases and as AI becomes more deeply embedded in other high-use devices, such as smartphones, the negative impacts of AI in this context will only increase.

More specific to AI in particular, unethical labor practices extend to low-paid data workers who work behind the scenes of AI development. Tubaro et al. (2020) categorize these different forms of labor, referred to as micro-work, as *AI preparation*, *AI verification*, and *AI impersonation*. AI preparation involves gathering, labeling, annotating and classifying large amounts of data that are used to train AI models. AI verification involves reviewing the accuracy of AI outputs and correcting them if required. AI impersonation involves humans doing the task that is supposed to be done by an AI model to simulate AI and automation (Tubaro et al., 2020). Workers involved in AI preparation and verification play a crucial role within the AI devel-

opment process as it is the labeling, annotation, and classification of data which is ultimately fed into many AI models to train them, and verification work is used to fine-tune and develop these models. But much of this labor is invisible and hidden behind crowd sourcing platforms utilizing often poorly compensated workers from low-income countries (Tubaro et al., 2020). Similarly, workers involved in AI impersonation work play an important role as they allow AI companies to promote their models as being automated when they are not, a phenomenon referred to as "faux-tomation" (Crawford, 2021; 66). Examples of this form of labor includes X.ai's AI agent, Amy, and Facebook's personal assistant, M, where human workers worked extremely long hours to maintain the façade of perpetual automation by verifying, editing and rewriting outputs (Crawford, 2021). In this way, AI companies based in high-income countries can promote themselves as innovators in the AI field, while outsourcing much of the labor behind their technology to low- to middle-income countries behind the scenes (Newlands, 2021).

To give a prominent recent example, a *Time* investigation found that workers in Kenya received less than $2 per hour to go through and label violent and sexual abuse materials, as well as hate speech, which was then fed into OpenAI's ChatGPT to make it a safer system for users as part of fine-tuning the system (Perrigo, 2023). Muldoon et al. (2023) investigated three delivery centers operated by Sama in Kenya and Uganda, the same company reported on in the above *Time* investigation, which claims to promote an ethical AI supply chain. They found "alarming accounts of low wages, insecure work, a tightly disciplined labor management process, gender-based exploitation and harassment, and a system designed to extract value from low-paid workers to produce profits for venture capital investors" (Muldoon et al., 2023, p. 2). Furthermore, Posada (2022) found that workers undertaking AI preparation and AI verification in Latin America, particularly in Venezuela, were paid low wages and lacked social and economic protections, with platforms having complete control over their labor. This is due to both a lack of regulations to protect workers and a lack of alternative work available locally, further widening the power asymmetry of companies over workers (Posada, 2022).

These cases highlight at least three important points relevant for this analysis. First, the harms and trauma that workers in low- to middle-income countries are being subjected to within the AI supply chain and development process are significant. Second, it is the inequalities perpetuated by the global economic order which has created the conditions for such harmful labor practices to take place in low- to middle-income countries (Tubaro & Casilli, 2020). Third, these unethical labor practices exacerbate global inequalities as citizens of low- to middle-income countries are taken advantage of for their low wages, while corporations and the high-income countries they service reap the benefits that this cheap labor affords them. Additionally, the high paying software development and engineering jobs which are available within the AI development process are almost exclusively located in high-income Western countries, in stark contrast to the low paying, unregulated, and precarious work available to those in low-income countries (Muldoon & Wu, 2023).

One may object to the claim that these labor practices exacerbate inequalities, since workers are provided with a source of income that might not otherwise be available to them, which could lessen rather than exacerbate inequalities. To address

this objection, we must consider several issues. First, as we have shown in Sect. 2, the appropriation of labor from low- to middle-income countries has a clear and quantifiable impact on global inequalities. This can be exacerbated in the context of AI because, as we will show in the following sections, the majority of the benefits derived from AI will be enjoyed by citizens and corporations in high-income countries, meaning that *relative* economic inequality could deepen. Second, as AI systems develop further by building on the labor of low- to middle-income countries' citizens, there is potential that these jobs will be replaced in the future by AI systems which can complete these tasks (Mulsoon & Wu, 2023). Against this, many argue that low-level work is a structural need within the AI supply chain and unlikely to be automated away (Muldoon et al., 2023; Muldoon & Wu, 2023; Tubaro & Casilli, 2020). But if that is not the case, then low-skilled workers in low- and middle-income countries face the harms associated with unemployment from AI automation, including the inability to afford the necessities of life, as well as negative impacts on important self-attitudes, such as self-respect and self-worth (Bankins & Formosa, 2023). Thirdly, there are also concerns around deskilling, since workers undertaking low-skilled work are not exercising and developing higher-value skills and cognitive faculties, which means that workers may ultimately be at a disadvantage when or if they choose to look for other work. Given that many low-skilled AI workers report their jobs as menial, monotonous and repetitive (Le Ludec et al., 2023; Muldoon et al., 2024), there is a risk that workers will lose the more complex skills such as judgement, critical thinking, intuition, context sensitivity, and ethical deliberation (Bankins & Formosa, 2023). Finally, the rise of synthetic or artificial data may lead to less low-skilled labor being needed to train AI models (Nikolenko, 2019). If more synthetic data is used to train AI models, there may be a reduced need for data labelling labor, impacting workers currently undertaking this work. This could also have consequences for the labor market itself by reducing the number of jobs available, thereby putting downward pressure on the negotiating power of workers, which could further perpetuate inequality.

4 Harms and Benefits of AI in the Global Context

While the previous section has largely focused on the harms and burdens being imposed on low- to middle-income countries and their citizens within the AI supply chain and development process, to fully articulate the global justice concerns associated with AI we must also explore the global distribution of the potential harms and benefits that the technology itself brings. To do that, we examine here the global distribution of benefits and harms that AI use could bring, as well as explore the global justice implications of the fact that most AI is trained on largely English-language and Western-centric content.

4.1 Distribution of Benefits

Much has been said about the economic benefits that AI will bring to the global economy, with research from PwC showing that AI could lead to global GDP rising

by 14% or \$15.7 trillion by 2030 (Rao & Verweij, 2017), and research from McKinsey estimating that Generative AI could add between \$2.6 trillion and \$4.4 trillion to the economy annually (Chui et al., 2023). Both reports, however, claim that the greatest gains are likely to accrue to North America and China, as well as European and developed Asian countries, while acknowledging that the much lower rates of AI adoption in developing countries means that these countries are expected to see relatively slower economic growth (Rao & Verweij, 2017; Chui et al., 2023). It is also important to note that the research undertaken by McKinsey comprises 47 countries, which they claim represents around 80% of employment across the world (Chui et al., 2023). This exclusion of so many countries is another indication that the economic benefits derived from AI are likely to be unevenly distributed, potentially further perpetuating unjust inequalities. This reality becomes clearer when we consider some case studies from developing low- to middle-income countries. For example, research undertaken by Heng et al. (2022, 12) finds that, "a major inhibitor for the adoption of AI solutions, both in Cambodia and Senegal, is the present state of their data infrastructure". Likewise, Nadeem et al. (2023, 16) find that "the lack of adequate ICT infrastructure, lack of awareness, market challenges, monetary challenges, and lack of supporting legislative framework" are the most significant challenges facing Pakistan's digitisation attempts. Additionally, Biana and Joaquin (2024) raise concerns about the ability for low- and middle-income countries to implement AI technologies, as many of these countries do not have reliable access to enabling digital infrastructure, such as electricity and internet connectivity.

The implementation of AI can have further negative impacts on employment opportunities for those in low- to middle-income countries by either replacing certain jobs they rely on, such as data entry and customer service, or by reducing the demand for certain tasks or jobs, such as programming or creative tasks, due to automation through AI (Mannuru et al., 2023). The possibility of a significant portion of workers in low- to middle-income countries being replaced by AI may exacerbate global inequalities. However, concerns around job replacement are not exclusive to low- to middle-income countries, since workers in high-income countries will also be impacted, perhaps to an even greater extent. But an important differentiating factor is that high-income developed countries typically have significantly stronger regulatory environments, worker protections, and social safety nets, compared to their equivalents in low- to middle-income countries. Further, in theory at least, the economic gains derived from AI in high-income developed countries could be distributed to impacted domestic workers if these levers are adequately deployed, which is not an option available to countries missing out on these benefits.

Developing low- to middle-income countries do not have access to the resources and capital needed to compete with the leaders in AI development. This means they will be reliant on already powerful corporations in high-income developed countries, deepening their reliance on them and further exacerbating global power imbalances (Bremmer & Suleyman, 2023). This is clear from the increasing role of multinational corporations in the datafication process and the development of data infrastructure in low- to middle-income countries through public-private partnerships (Taylor & Broeders, 2015). This has led to Big Tech corporations having a growing influence in these countries as development actors who collect and process large amounts of data

from the population, which empowers them to dictate how development interventions should be implemented, largely in the interest of the corporations themselves who become gatekeepers of development data and social knowledge (Cinnamon, 2020). Here we see AI Empire operating through multiple mechanisms: extractivism, through the collection and control of data; essentialism, through social knowledge being processed through a set of measurable traits predetermined by corporations; surveillance, through the ability to extract data about citizens and their environments; and containment, through the use of data and surveillance to determine development interventions (Tacheva & Ramasubramanian, 2023; Cinnamon, 2020).

4.2 Harms Attributed to Malicious Use of AI

A global justice example of harm resulting from malicious AI use is Israel's use of AI-powered facial recognition cameras as part of its occupation of Palestinian territories in the West Bank and East Jerusalem (Goodfriend, 2022). Not only is the technology used as a surveillance tool against Palestinians, but companies working with the Israeli military use this as an opportunity to test and enhance these new AI technologies with minimal regulatory oversight before exporting them around the globe (Loewenstein, 2023). More disturbingly, according to one recent report, Israel has been using an AI targeting system, referred to as "Lavender", to mark and target tens of thousands of Palestinians in Gaza with minimal human oversights, insufficient regard for civilian casualties, and no requirement to examine the basis of the AI system's recommendation (Abraham, 2024). This example highlights how powerful and resourceful corporations and governments can both weaponize AI systems to target foreign civilians and exploit power imbalances to test their AI systems on foreign civilian populations for future profit. Cases such as this highlight the growing risks that AI poses to global justice when powerful corporations and governments can abuse their resource advantage to utilize AI technology to unfairly pursue their own interests to the detriment of others.

Generative AI also poses a risk in the global context where well-resourced adversarial actors or foreign governments can weaponize the technology for their own ends (Nie, 2024). Generative AI models allow for the creation and manipulation of texts, images, audio, or video based on data and prompt inputs, and the outputs of such systems are becoming increasingly difficult to identify (Janjeva et al., 2023). Generative AI can be weaponized to harm democratic processes in a number of ways including: manufacturing false perceptions of consensus around political issues through the generation of social media posts and comments, which can be dispersed and amplified by algorithms; influencing public opinion and exacerbating polarisation; the creation of false news articles and false videos and recordings of key public figures and politicians; and deceiving voters and undermining trust in the election process and democratic institutions as a whole (Wirtschafter, 2024; Formosa et al., 2024). These harms are already occurring, with a recent study showing that impersonation, amplification, and falsification are currently the most prevalent approaches to AI misuse in reported real-world cases (Marchal et al., 2024). An example of foreign agents weaponizing Generative AI to target citizens includes the case of Iran-backed hackers who targeted TV streaming services in the United Arab Emirates (UAE) to broad-

cast deepfake news which reached audiences in the UAE, UK and Canada (Milmo, 2024). Additionally, Open AI, the company developing ChatGPT, recently published a Threat Intel Report which detailed their Generative AI models being used by actors linked to Russia, China, Iran, and Israel for covert influence operations (Open AI, 2024).

When considering the distribution of such harms, it is worth noting that the most powerful tools available to detect and combat AI-generated content are trained primarily on English content, which means that harmful content targeting English speakers are more likely to be detected and harms minimised (Bontcheva et al., 2024). However, it is difficult to quantify the distribution of harms in this case given that high-income English-speaking countries may be more valuable targets for foreign adversaries. Nonetheless, specific cases, such as Facebook's role in the 2017 Rohingya genocide in Myanmar (Amnesty International, 2022), highlights the scale of harms that could be inflicted on citizens from low- and middle-income countries when powerful AI technological tools are weaponized against them, especially when those countries lack the regulatory and technological tools to protect themselves.

4.3 Missed-out Benefit from AI

Beyond the direct harms pertaining to the malicious use of AI, it is also important to consider the benefits from AI that people from low- and middle-income countries might miss out on. This leads to another global justice concern, the risks associated with AI models being trained primarily on English language and Western-centric content. Although it should be noted that there is significant development in Chinese language AI models (Biever, 2024; Bradford, 2023), Chinese language models have substantially less application globally as compared with English and Western developed AI models. Within the digitally available data used to train AI models, there are high-resource languages, which have large quantities of data available and include languages such as English, French, German, and Russian, and low-resource languages, which have significantly lower quantities of data available and include the majority of the languages spoken around the globe (Costa-jussà et al., 2022). To give context to the domination of English content online, there are over 7000 languages spoken around the globe, but over 50% of the content online is from one language, English (Shwartz, 2024; Ta & Lee, 2023). Furthermore, it is reported that approximately two thirds (between 50 and 90%) of the leading Natural Language Processing research is targeted towards developing the technology for English speakers (Søgaard, 2022).

This disparity gives rise to a "digital language divide", whereby the speakers of a majority of global languages are at a significant disadvantage due to not being speakers of a dominant online language, particularly English (Ta & Lee, 2023). Researchers have found that when testing the multilingual language generation ability of ChatGPT, high-resource languages such as French and Chinese had higher quality translations and generated fewer hallucinations compared to low-resource languages such as Javanese and Sudanese (Bang et al., 2023). They also found that ChatGPT is stronger in generating sentences of Latin script languages compared to non-Latin script languages, even if the non-Latin script languages are high-resourced

(Bang et al., 2023). Additionally, researchers testing the ability of Large Language Models (LLMs) to generate text in languages from Southeast Asia, such as Indonesian, Malay and Tamil, found that ChatGPT and other publicly available multilingual instruction-tuned models produced higher rates of grammatically incorrect or semantically meaningless texts in these languages (Yong et al., 2023). These findings show that there is a clear global inequality when it comes to the benefits and opportunities afforded by LLMs, since global users whose native language is not English, or any other high-resource language, are at a significant disadvantage.

The dominance of English and other high-resource languages in the training data of LLMs also raises concerns around cultural bias, which can make the technology harmful or at least less useful for people from low- and middle-income countries (Shwartz, 2024). For example, researchers assessing the underlying cultural background of ChatGPT found that it aligns strongly with American culture, with English prompts producing consistent responses biased towards American culture while also flattening out cultural differences, and prompts from other cultural contexts produced responses that were less effective and less relevant (Cao et al., 2023). Other research has found that LLMs tend to produce responses more aligned with the opinions found in the USA, as well as some European and South American countries, when asked about global issues, highlighting potential embedded biases that favour the views of WEIRD populations (Durmus et al., 2024). They also found that harmful cultural stereotypes may be reflected in responses when prompted to consider a particular country's perspective (Durmus et al., 2024). AI systems that perpetuate harmful stereotypes, for example about women, immigrants or minorities, can also bring about material harms, such as lesser job prospects for members of those groups.

Researchers found that in over 200 Explainable AI studies that were reviewed, cultural variations were not considered, and explanations were found to be tailored to individualistic, western populations, with non-western, collectivist cultures overlooked (Peters & Carman, 2024), which raises justice concerns around the explainability and inclusivity of AI models. It is important to note that other forms of biases exist within different AI models. For example, AI-powered facial recognition algorithms are also found to be biased due to representational imbalances in their training datasets (Leslie, 2020), with researchers finding that these systems perform significantly better when identifying light-skinned and male faces, while darker-skinned female faces are the worst identified (Buolamwini & Gebru, 2018). These biases have resulted in direct harms, including cases where individuals have been wrongfully imprisoned (Thanawala, 2023). These cases highlight the inequalities that exists within the global AI context, and the significant risks cultural bias in AI systems have for minorities across the globe.

5 Global Justice and the Capabilities

Before we move on to assess the global justice implications of the above, we first need to acknowledge that we have only covered a relatively small number of benefits and harms associated with AI. But rather than attempting to be comprehensive in our coverage, we have instead focused on cases where the global justice implications are

clearest and most acute. We should also note that increased adoption and use of AI in high-income countries means that certain harms will be more prevalent in higher use countries due to the sheer number of users. For example, cumulative harms from AI hallucinations will probably be highest in countries that use the technology the most, which is likely to be higher income countries that have the infrastructure and resources to access the technology. But this should be understood primarily as an AI *safety* issue (Millière, 2023), rather than a global *justice* issue, as it results from a privilege that allows for higher use of the technology rather than any form of active discrimination or unfairness in the technology. Further, concerns around AI safety are already being addressed in high-income countries, whether via market solutions or through regulatory and legislative means. Furthermore, high-income countries are largely responsible for the development and adoption of AI, and they therefore have significantly more control over its progress and regulation. It is important to differentiate between harms stemming from lack of AI access and those resulting from AI saturation. The former, such as the digital language divide and cultural biases within training datasets, is primarily a global justice issue affecting low- and middle-income countries due to unequal resource distribution. The latter, including cumulative effects of AI hallucinations, is mainly a safety issue caused by higher AI use in high-income countries. While high-income countries have the means and regulatory powers to address these safety concerns, low- and middle-income countries risk being left behind in AI development and largely lack the powers needed to address these justice concerns themselves.

What is clear from our analysis in Sects. 3 and 4 is that, on the global scale, many of the benefits of AI tend to flow to high-income countries and their citizens and many of the harms tend to flow to lower-income countries and their citizens. In this section we will assess the justice implications of this by applying the Capabilities Approach, an approach we justified in Sect. 2. We will first examine Nussbaum's list of the central human Capabilities to assess whether the global impacts of AI are large enough to put downward pressure on individuals to the extent that justice concerns are raised about their abilities to meet the threshold levels of each capability needed to live a dignified life. We acknowledge that there is difficulty in determining what the basic threshold should be in different contexts, but we argue that there is enough downward pressure on people who are *already* marginalized to raise significant justice concerns. This is especially so given that those concerned are low-income and marginalized citizens and, as we have already shown in previous sections, global inequality precedes widespread AI use, and the concerns we have raised here thus perpetuate those existing inequalities. We then turn to Nussbaum's list of global justice principles to further articulate the justice concerns pertaining to AI. We note that our analysis will be covering the Capabilities and principles in order of relevance and importance in the context of AI.

5.1 AI and the Ten Central Human Capabilities

We first consider the Capabilities of *Life* and *Bodily Health*. *Life* constitutes one's ability to live a human life worth living, while *Bodily Health* considers one's ability to achieve good health, adequate nourishment and shelter. Resource depletion

and environmental degradation from the AI pipeline, as outlined in detail in Sect. 3, have a significant negative impact on these Capabilities. For example, the destructive impact of mineral extraction on local farms, forests and fish supplies (Crawford, 2021) depletes food and water supplies in vulnerable communities, putting downward pressure on the ability of people there to live a full human life and obtain the necessities for nourishment given that these are the basic thresholds for survival and bodily health. Additionally, the depletion of resources makes it harder for people to find other means (e.g. through work or trade) of accessing the goods they need, including shelter. Further, it is the most under-privileged communities, particularly those in low-income countries, that tend to suffer the greatest from environmental harms, such as pollution, degraded water or air quality, and the impact of toxic metals on the health of communities (Bell, 2004). This also threatens the ability to find safe and adequate shelter.

Unsafe working conditions, lack of legal protections, low wages and insecure work also put downward pressure on achieving the *Life* and *Bodily Health* Capabilities. This includes the dangerous conditions which local miners and their communities are subjected to, with deaths and injuries common among workers, including children (Frankel, 2016). Local mining communities are also exposed to toxic metals which cause health problems and birth defects (Frankel, 2016). The trauma, abuse and suffering that is reported by low-income data workers in the AI pipeline, as well as the lack of mental health support provided by corporations who benefit from their work (Perrigo, 2023; Muldoon et al., 2023), also puts downward pressure on the ability of people to achieve the threshold to fulfil these Capabilities. Furthermore, the extremely long hours that can be associated with AI work (Crawford, 2021) is another factor which threatens the ability to achieve these key Capabilities. Whether this is to complete enough micro-work in AI Preparation, AI Verification and AI Impersonation tasks, extended working hours can not only be damaging to one's health, but it can also take away time from the pursuit of the things needed for *Life* and *Bodily Health*, such as adequate nourishment or physical exercise, putting downward pressure on people's abilities to meet the thresholds for these Capabilities. In addition to this, cases such as those outlined in Sect. 4.2, including Israel's use of AI-powered systems against Palestinians (Abraham, 2024), highlight the threat that the weaponization of AI systems can pose to these Capabilities, by leading to death, destruction and an inability to pursue a life worth living.

Other Capabilities which are directly threatened within the AI supply chain are: *Bodily Integrity*, which constitutes one's ability for unrestricted movement, free of violence and assault; *Control Over One's Environment*, which has a political dimension pertaining to the political choices governing one's life, and a material dimension which situates rights related to property and employment; and *Affiliation*, which relates to one's self-respect and ability to live with others with reciprocal concern and dignity. These Capabilities are all at risk of harm in areas where resource extraction is leading to geopolitical conflict and violence, which can be especially detrimental to women and children. As Pogge (2002) argues, resource extraction has the propensity to lead to greater civil conflict because of internal disputes to capture political power and gain control over those resources, a process that is enabled and encouraged by the international borrowing privilege maintained as part of the global institutional

structure by high-income countries. Unsafe working conditions, lack of control over working conditions, being subjected to abuse, trauma and gender-based exploitation and harassment, whether in the mineral extraction context or the low-skilled AI data work context, also threatens one's ability to meet the basic threshold for these Capabilities, especially for women. For example, workers may struggle to engage fully with one another or their families and loved ones due to trauma or reduced self-respect resulting from poor working conditions and engaging with toxic online material as part of AI Preparation and Verification roles. Another important mechanism of the AI Empire which puts significant downward pressure on people's abilities to meet the threshold for these Capabilities is surveillance. For example, AI-powered surveillance using facial recognition technology and other AI-powered technologies, whether it is deployed by powerful corporations or foreign governments, has a potentially significant negative impact on some vulnerable people's ability to move freely and avoid violence. As a result, they may find themselves constantly watched, followed and influenced by the constant gaze of a powerful external agent, which can undermine the Capability of *Control Over One's Environment*, as well as impacting their self-respect and dignity. This could lead to communities feeling as though they are not entitled to equal worth or self-determination, a core requirement for meeting the achievement of *Affiliation.*

Additionally, environmental damage and degradation also harms the capability of *Other Species*, i.e. being able to live alongside plants, animals and the natural world. It is harder to enjoy the benefits of sharing the world with other animals and plants when the habitats and ecosystems those animals and plants need to survive are depleted, destroyed, or poisoned through resource extraction. Environmental degradation and destruction can also be extremely harmful to the *Emotions* capability, especially for Indigenous communities, who perceive the natural world as a living entity which ought to be looked after and respected, and not exploited or destroyed for financial gain (see Tapia & Peña, 2020).

There are three Capabilities which are less directly impacted in the context of AI, but are still worth considering here. These Capabilities are: *Senses*, *Imagination*, *and Thought*, which constitutes one's ability to use their senses, imagination and cognitive faculties; *Emotions*, one's ability to achieve the full range of emotional development; *and Practical Reason*, which pertains to one's ability to reflect critically and form a conception of the good. For example, tedious menial work, deskilling, and unfair labor management processes may undermine the capabilities of *Senses*, *Imagination*, *and Thought*, as well as *Emotions* and *Practical reason.* This is especially evident when we consider some of the important dimensions of meaningful work, which have a direct impact on the achievement of these capabilities. Bankins and Formosa (2023) outline five dimensions for meaningful work which are relevant for this current analysis. The first dimension is *task integrity* which affords a worker the opportunity to undertake a piece of work to completion as opposed to undertaking fragmented tasks. This is clearly lacking in the context of AI Preparation, AI Validation, and AI Imitation roles, where workers are undertaking fragmented micro-tasks. The second dimension is *skill cultivation and use* which entails workers utilizing and developing a variety of skills. Again, this is not the case within much of the low-skill work being done in the AI supply chain, such as AI Preparation, which does not

require or cultivate complex skills. The third dimension involves *task significance* which gives meaning to the work in its connection to the wider world. As Le Ludec (2023) highlights, workers in AI preparation report that they perceive their work as invisible, menial and lacking in significance, compared to their higher paid and more visible colleagues such as software engineers, who are largely based in high-income countries. This also negatively impacts their self-worth (Le Ludec, 2023). The fourth dimension, *autonomy*, is clearly lacking for these workers as they have no control over their work approaches and are subject to intrusive labor management processes, including surveillance and monitoring (Muldoon et al., 2023). Finally, the dimension of *belongingness* is also likely absent given that there is a lack of collaboration in the micro-work they are undertaking.

Furthermore, the weaponization of Generative AI also poses threats to these capabilities, particularly those of *Senses*, *Imagination*, *and Thought*, as well as *Practical Reason*, given that, as we have outlined in Sect. 4, the technology can be used to manipulate texts, images, audio, and video at mass scale to influence public opinion, create false narratives, impersonate key political, civic, and business leaders, and deceive citizens. These capabilities can also be undermined by the biases that are reported to be present in AI models, including LLMs, as well as the weaker models available to majority of the non-English speaking world, in particular those who speak low-resources languages primarily in low- to middle-income countries. The perpetuation of harmful stereotypes, whether indirectly through biases within AI models or directly through the weaponization of Generative AI content, can also pose as a barrier to meeting the thresholds for the Capabilities of *Emotions* and *Affiliation*, given that it may lead to hostilities towards minority groups, or even oneself, and influence the emotional reactions expressed towards the self and others.

What this shows is that within the AI supply chain and development process, as well as its implementation and use, there are significant harms and burdens which are being carried by low- to middle-income countries and their citizens. We have also shown that these harms and burdens pose significant barriers to leading a decent life with dignity, given that they put downward pressure on people's abilities to meet the basic thresholds needed to develop the Capabilities required to live such a life. This is especially concerning as we are dealing with negative impacts on people who, due to existing inequalities, may already be near to or below threshold levels for various central human Capabilities. While it is difficult to determine what the thresholds are in different contexts, there is enough evidence to show that the downward pressure that is imposed on low- to middle-income countries and their citizens is sufficient to raise significant global justice concerns.

5.2 AI and the Ten Global Justice Principles

Before we expand our analysis by examining Nussbaum's (2004) ten principles for global justice in the context of AI, it is important that we acknowledge that some of these principles will be more relevant for the current analysis, whereas others are more general in nature and not necessarily tied directly to the AI context. However, it is important to note that the development and implementation of AI is a continuation of the technological and economic dominance of high-income countries and the

corporations they empower on a global scale. Therefore, while some of the global principles of justice are not specific to the AI context, the AI supply chain and its use and implementation still plays an important role in helping to maintain unjust global structures and patterns. We explore these principles in order of relevance and importance in the context of AI.

The evidence we have outlined highlights that the principle that *multinational corporations have responsibilities for promoting human capabilities in the regions in which they operate* is not being adequately adhered to. This is possibly the most important principle in the context of resource and labor extraction because it applies directly to multinational corporations, such as large AI technology companies, and because the burdens and harms which arise in this context directly threaten the achievement of the central human Capabilities needed to live a decent life. This is made worse by the fact that these corporations are making significant profits off the back of these harms (Muldoon et al., 2023). While much effort has been taken by multinational corporations to keep the supply chain as opaque as possible (Crawford, 2021), we can see that corporations integral to the AI development process, such as NVIDIA, are sourcing materials from regions and countries where the harms outlined here are perpetuated (NVIDIA Corporation, 2024), potentially pushing citizens in poorer countries below the threshold needed to achieve the central human Capabilities. This is also clear when we consider the lack of worker protections and the trauma and suffering that is reported by low-skilled AI workers such as those engaged in AI Preparation, as well as the lack of mental health support provided by corporations (Perrigo, 2023). Furthermore, as we outlined in Sect. 4, while low- to middle-income countries are reliant on powerful corporations for their infrastructure development needs, corporations are taking few steps to promote human Capabilities and are instead exploiting public-private partnerships to extract further resources, including data, for their own benefit. Additionally, as our evidence suggests, significantly more needs to be done by corporations to address the biases and shortcomings of their models so that citizens in low- to middle-income countries can benefit from them and for this principle to be adequately implemented.

Next, consider the colonial and capitalistic logics of extractivism (Sultana, 2022; Tacheva & Ramasubramanian, 2023), whereby the labor and natural resources of low- to middle-income countries are appropriated for the benefit of high-income countries and their corporations, and the associated lack of respect for Indigenous sovereignty as it pertains to resource extraction and the environment. This implies that the principle that *national sovereignty should be respected*, *within the constraints of promoting human capabilities* has, arguably, at best been undermined and at worst completely ignored in the pursuit of materials for the AI pipeline. This principle is also undermined, and is already being threatened in many instances, when AI is weaponized by foreign agents and governments against other countries and their citizens, such as in the case of Israel's use of AI against the Palestinian people. Additionally, national sovereignty is undermined given the procedural justice concerns raised by the fact that governments in high-income countries, primarily the EU and the US, and China to a lesser extent, are dominating the regulation of the global digital economy and AI (Bradford, 2023). This gives a few high-income countries control over global norms of AI regulation that impact everyone in the world, without fair procedures

that require input from all impacted nations. Another principle which is clearly being violated is that *all institutions and individuals should focus on the problems of the disadvantaged in each nation and region.* As is clear from the growing inequality, resource extraction and depletion, and environmental destruction, this principle is also being neglected by the global community. This is further supported by workers in low- to middle-income countries being subject to abuse, harassment and violence. The examples and cases we have outlined also highlight that this principle is not being adhered to, with biases within AI systems clearly disadvantaging minorities of different kinds including races, genders and cultures. In fact, it could be argued that the disadvantaged are the most harmed by AI, whether through a lack of access to benefits or through the malicious use of the technology.

There are also multiple principles of global justice which are more general and not specific to AI, but it is still useful to consider these briefly here. For example, two principles which are clearly being breached are that *the main structure of the global economic system must be designed to be fair to poor and developing countries* and that *prosperous nations have a responsibility to give a substantial portion of their GDP to poorer nations.* This is evident by Hickel et al.'s (2022) findings that the losses incurred by low- to middle-income countries in the global south due to the appropriation of embodied resources and labor is 30 times greater than their total aid receipts over the same period. There is no evidence to suggest this trend is changing or likely to change. While this is a more general trend that is not directly related to AI, nonetheless AI development is a continuation of the ongoing technological and economic dominance of prosperous high-income countries, and thus helps to maintain this unjust system. Additionally, in Sect. 4 we outlined the economic benefits attributable to the implementation of AI and showed that the majority of those benefits will likely go to high-income countries, while low- to middle-income countries are likely to fall behind as a result of the lack of resources, adequate infrastructure, and technical expertise. This trend is likely to not only perpetuate global inequality, but it may exacerbate it on a relative basis if steps are not taken to address this.

Another principle which must be addressed if global justice is to be achieved in the context of AI is that *we should cultivate a thin*, *decentralized*, *yet forceful global public sphere.* The importance of this principle becomes clear when we consider that AI, particularly Generative AI, could be weaponized to harm democratic citizenship, political deliberation, and decision making (Formosa et al., 2024). Given the growing importance and influence of digital media within the information landscape, such as social media mediated by AI algorithms, it may prove more difficult to have civil discourse if there is a flood of AI-generated misinformation. Additionally, AI may be weaponized to fuel polarization in order to increase online engagement and thus profits for Big Tech platforms within the attention economy (Formosa et al., 2024). Overall, this will make it harder for a fair global public sphere to emerge that will promote civility, justice, and human rights for all. These justice concerns will only exacerbate if governments, corporations and adversarial actors are allowed to weaponize this powerful technology without any accountability measures in place to hold those who violate democratic rights responsible. Finally, when considering the principle that *care for the ill*, *the elderly and the disabled should be a prominent focus of the world community*, there has been a lot of discussion around the healthcare ben-

efits of AI. However, AI health technologies will be expensive and tailored for high-income patients in resource-rich settings (Gerke et al., 2020), posing a risk of patients in low- to middle-income countries missing out on key health benefits given they will not be a priority in this context and may not be able to afford to access the technology.

6 Conclusion, Limitations and Future Research

In this paper we have drawn from numerous case studies and examples to highlight the global justice concerns that are raised by both the AI supply chain and development process, as well as the uses and inadequacies of the AI technology itself. By utilizing the Capabilities Approach we were able to articulate these justice concerns with AI on the individual level, in terms of a downward pressure on people's abilities to meet the central human Capabilities needed for a dignified life, and on the global level, in terms of an inadequate adherence to ten principles of global justice. It is worth noting again that at least some of these global justice concerns, especially those around mineral extraction, apply to many other forms of technology, such as smartphones, televisions, or even fridges. Other justice concerns, however, are specific to the AI industry, such as the use of AI micro-workers to label traumatic content, the inherent bias in LLMs, weaponization of AI to target vulnerable populations, the use of AI for surveillance, and the potentially chilling impacts of mass impersonation and falsification on the quality of discourse comprising the global public sphere.

The practical implications of our analysis for AI regulation and governance are significant and warrant further exploration. Our capability-based approach suggests that AI regulation should prioritize protecting and promoting central human capabilities globally, with particular attention paid to vulnerable populations in low- to middle-income countries. This might involve measures such as: (1) mandating capability impact assessments for AI systems before deployment, especially in critical domains such as healthcare, education, and employment; (2) establishing international cooperation mechanisms to ensure a more equitable distribution of AI benefits and mitigation of harms across nations, such as by creating global programs to support capability enhancement in areas where AI might cause capability deprivation and by preventing harms such as AI weaponization; (3) implementing strict regulations on AI-driven resource extraction to better protect environmental-related capabilities; and (4) setting and enforcing global standards for fair wages, safe working conditions, and mental health support to better protect workers in the AI pipeline, particularly in countries with weaker labor laws. In addressing these concerns, efforts should be made to address the language and digital divides identified here, such as by investing in AI development for low-resource languages and improving digital infrastructure where needed. These potential interventions aim to align AI development and deployment with the goal of ensuring all individuals can achieve a threshold level of central capabilities, regardless of their global position, and will also help to address the procedural justice concerns raised by the current dominance of a few high-income countries in setting global AI regulatory norms. While some of these regulatory measures have broader relevance beyond the AI industry, such as those

related to resource extraction, others are more AI-specific, such as investing in AI development for low-resource languages.

In terms of limitations and future research, as we noted above, our analysis can – and should – be complimented by other approaches to global justice, as well as more context specific analyses. Furthermore, our analysis utilizes Nussbaum's lists of Capabilities and principles, and while we argue that her framework provides us with a powerful starting point for this project, future work could explore whether, in the context of AI, this list could be extended or altered to be more domain specific. Future work could do this by exploring whether a list of capabilities which are applied to the domain of technology and AI more specifically, such as AI literacy, would be helpful in this context (Santoni de Sio et al., 2024). This would further help us to better understand how we can globally reap the benefits of AI in ways that respect the dignity of all persons and lessen rather than exacerbate existing global justice concerns.

References

Abraham, Y. (2024, April 3). *'Lavender'*. +972 Magazine. https://www.972mag.com/lavender-ai-israeli-army-gaza/

Allen, D., & Weyl, E. G. (2024). The real dangers of generative AI. *Journal of Democracy*, *35*(1), 147–162. https://doi.org/10.1353/jod.2024.a915355

Amnesty International (2022). *Myanmar*. Amnesty International. https://www.amnesty.org/en/documents/asa16/5933/2022/en/

Arsenault, A. C., & Kreps, S. E. (2024). AI and International politics. In J. B. Bullock, Y. C. Chen, J. Himmelreich, et al. (Eds.), *The Oxford Handbook of AI Governance*. Oxford University Press. https://doi.org/10.1093/oxfordhb/9780197579329.013.49

Bang, Y., Cahyawijaya, S., Lee, N. (2023). *A Multitask, Multilingual, Multimodal Evaluation of ChatGPT on Reasoning, Hallucination, and Interactivity* (No. arXiv:2302.04023). arXiv. http://arxiv.org/abs/2302.04023

Bankins, S., & Formosa, P. (2023). The ethical implications of Artificial Intelligence (AI) for meaningful work. *Journal of Business Ethics*, *185*(4), 725–740. https://doi.org/10.1007/s10551-023-05339-7

Bell, D. (2004). Environmental Justice and Rawls' Difference Principle. *Environmental Ethics*, *26*(3), 287–306. https://doi.org/10.5840/enviroethics200426317

Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the Dangers of Stochastic Parrots. *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, 610–623. https://doi.org/10.1145/3442188.3445922

Biana, H. T., & Joaquin, J. J. (2024). The irony of AI in a low-to-middle-income country. *AI & SOCIETY*, s00146-023-01855–2. https://doi.org/10.1007/s00146-023-01855-2

Biever, C. (2024). China's ChatGPT. *Nature*, *629*(8014), 977–978. https://doi.org/10.1038/d41586-024-01495-6

Bird, E., Fox-Skelly, J., Jenner, N., et al. (2020). The ethics of artificial intelligence. *European Parliamentary Research Service*. https://doi.org/10.2861/6644

Bontcheva, K., Papadopoulous, S., Tsalakanidou, F. (2024). Generative AI and Disinformation. https://edmo.eu/wp-content/uploads/2023/12/Generative-AI-and-Disinformation_-White-Paper-v8.pdf

Borenstein, J., & Howard, A. (2021). Emerging challenges in AI and the need for AI ethics education. *AI and Ethics*, *1*(1), 61–65. https://doi.org/10.1007/s43681-020-00002-7

Bradford, A. (2023). *Digital empires*. Oxford University Press.

Bremmer, I., & Suleyman, M. (2023). *Can States learn to govern Artificial Intelligence before it's too*. Late? Foreign Affairs.

Buolamwini, J., & Gebru, T. (2018). Gender Shades. *Proceedings of the 1st Conference on Fairness, Accountability and Transparency*, 77–91. https://proceedings.mlr.press/v81/buolamwini18a.html

Cao, Y., Zhou, L., Lee, S. (2023). *Assessing Cross-Cultural Alignment between ChatGPT and Human Societies* (No. arXiv:2303.17466). arXiv. http://arxiv.org/abs/2303.17466

Chui, M., Hazan, E., Roberts, R. (2023). The economic potential of generative AI. *McKinsey*.

Cinnamon, J. (2020). Data inequalities and why they matter for development. *Information Technology for Development*, *26*(2), 214–233. https://doi.org/10.1080/02681102.2019.1650244

Costa-jussà, M. R., Cross, J., Çelebi, O., et al. (2022). No Language Left behind. *arXiv*. https://doi.org/10.48550/arXiv.2207.04672

Crawford, K. (2021). *The Atlas of AI*. Yale University Press.

Dauvergne, P. (2022). Is artificial intelligence greening global supply chains? *Review of International Political Economy*, *29*(3), 696–718. https://doi.org/10.1080/09692290.2020.1814381

De Sio, S., Almeida, F., T., & Van Den Hoven, J. (2024). The future of work. *Critical Review of International Social and Political Philosophy*, *27*(5), 659–683. https://doi.org/10.1080/13698230.2021.2008204

de Vries, A. (2023). The growing energy footprint of artificial intelligence. *Joule*, *7*(10), 2191–2194. https://doi.org/10.1016/j.joule.2023.09.004

Diemel, J. A., & Hilhorst, D. J. M. (2019). Unintended consequences or ambivalent policy objectives? *Development Policy Review*, *37*(4), 453–469. https://doi.org/10.1111/dpr.12372

Durmus, E., Nguyen, K., Liao, T. I. (2024). Towards Measuring the Representation of Subjective Global Opinions in Language Models (No. arXiv:2306.16388). *arXiv*. http://arxiv.org/abs/2306.16388

Eichstaedt, P. (2011). *Consuming the Congo*. Chicago Review.

EPRI. (2024). Powering intelligence – analyzing Artificial Intelligence and Data Center Energy Consumption. *Electric Power Research Institute*. https://www.epri.com/research/products/3002028905

Formosa, P., & Mackenzie, C. (2014). Nussbaum, Kant, and the capabilities Approach to Dignity. *Ethical Theory and Moral Practice*, *17*(5), 875–892.

Formosa, P., Kashyap, B., & Sahebi, S. (2024). Generative AI and the future of democratic citizenship. *Digital Government: Research and Practice*. https://doi.org/10.1145/3674844

Frankel, T. C. (2016). Cobalt mining for lithium ion batteries has a high human cost. *Washington Post*. https://www.washingtonpost.com/graphics/business/batteries/congo-cobalt-mining-for-lithium-ion-battery/

Fui-Hoon Nah, F., Zheng, R., Cai, J., et al. (2023). Generative AI and ChatGPT. *Journal of Information Technology Case and Application Research*, *25*(3), 277–304. https://doi.org/10.1080/15228053.2023.2233814

Gabriel, I. (2022). Toward a theory of Justice for Artificial Intelligence. *Daedalus*, *151*(2), 218–231. https://doi.org/10.1162/daed_a_01911

Gerke, S., Minssen, T., & Cohen, G. (2020). Ethical and legal challenges of artificial intelligence-driven healthcare. In *Artificial Intelligence in Healthcare* (pp. 295–336). Elsevier. https://doi.org/10.1016/B978-0-12-818438-7.00012-5

Goodfriend, S. (2024). June 28. How the Occupation Fuels Tel Aviv's booming AI Sector. *Foreign Policy*. https://foreignpolicy.com/2022/02/21/palestine-israel-ai-surveillance-tech-hebron-occupation-privacy/

Guo, S., Lin, X., Coicaud, J. M., et al. (2019). Conceptualizing and measuring global justice. *Fudan Journal of the Humanities and Social Sciences*, *12*(4), 511–546. https://doi.org/10.1007/s40647-019-00267-1

Hagendorff, T. (2020). The Ethics of AI Ethics. *Minds and Machines*, *30*(1), 99–120. https://doi.org/10.1007/s11023-020-09517-8

Heng, S., Tsilionis, K., Scharff, C., & Wautelet, Y. (2022). Understanding AI ecosystems in the Global South. *International Journal of Information Management*, *64*, 102454. https://doi.org/10.1016/j.ijinfomgt.2021.102454

Hickel, J., Sullivan, D., & Zoomkawala, H. (2021). Plunder in the post-colonial era. *New Political Economy*, *26*(6), 1030–1047. https://doi.org/10.1080/13563467.2021.1899153

Hickel, J., Dorninger, C., Wieland, H., & Suwandi, I. (2022). Imperialist appropriation in the world economy. *Global Environmental Change*, *73*, 102467. https://doi.org/10.1016/j.gloenvcha.2022.102467

Hickok, M. (2022). Public procurement of artificial intelligence systems. *AI & SOCIETY*. https://doi.org/10.1007/s00146-022-01572-2

Hodgson, C., & Morris, S. (2024, July 2). Google's greenhouse gas emissions jump 48% in five years. *Ars Technica*. https://arstechnica.com/gadgets/2024/07/googles-greenhouse-gas-emissions-jump-48-in-five-years/

IEA (2024). Electricity 2024 - Analysis and forecast to 2026. *International Energy Agency*. https://iea.blob.core.windows.net/assets/6b2fd954-2017-408e-bf08-952fdd62118a/Electricity2024-Analysisandforecastto2026.pdf

Janjeva, A., Harris, A., Mercer, S. (2023). The Rapid rise of generative AI. *Centre for Emerging Technology and Security*. The Alan Turing Institute.

Kak, A. (2020). The Global South is everywhere, but also always somewhere. *Proceedings of the AAAI/ACM Conference on AI Ethics and Society*, 307–312. https://doi.org/10.1145/3375627.3375859

Lauer, D. (2021). You cannot have AI ethics without ethics. *AI and Ethics*, *1*(1), 21–25. https://doi.org/10.1007/s43681-020-00013-4

Le Ludec, C., Cornet, M., & Casilli, A. A. (2023). The problem with annotation. *Big Data & Society*, *10*(2), 20539517231188724. https://doi.org/10.1177/20539517231188723

Leslie, D. (2020). Understanding bias in facial recognition technologies. *The Alan Turing Institute*. https://doi.org/10.5281/zenodo.4050457

Loewenstein, A. (2023). *The Palestine Laboratory*. Verso Books.

Mannuru, N. R., Shahriar, S., Teel, Z. A., et al. (2023). Artificial intelligence in developing countries. *Information Development*, 02666669231200628. https://doi.org/10.1177/02666669231200628

Marchal, N., Xu, R., Elasmar, R. (2024). *Generative AI Misuse* (No. arXiv:2406.13843). arXiv. https://doi.org/10.48550/arXiv.2406.13843

McQuillan, D. (2022). *Resisting AI*. Policy.

Millière, R. (2023). The Alignment Problem in Context. *arXiv*. arXiv:2311.02147. http://arxiv.org/abs/2311.02147

Milmo, D. (2024, February 8). Iran-backed hackers interrupt UAE TV streaming services with deepfake news. *The Guardian*. https://www.theguardian.com/technology/2024/feb/08/iran-backed-hackers-interrupt-uae-tv-streaming-services-with-deepfake-news

Morley, J., Kinsey, L., Elhalal, A., et al. (2023). Operationalising AI ethics. *AI & SOCIETY*, *38*(1), 411–423. https://doi.org/10.1007/s00146-021-01308-8

Muldoon, J., & Wu, B. A. (2023). Artificial Intelligence in the Colonial Matrix of Power. *Philosophy & Technology*, *36*(4), 80. https://doi.org/10.1007/s13347-023-00687-8

Muldoon, J., Cant, C., Graham, M., et al. (2023). The poverty of ethical AI. *AI & SOCIETY*. https://doi.org/10.1007/s00146-023-01824-9

Muldoon, J., Cant, C., Wu, B., & Graham, M. (2024). A typology of artificial intelligence data work. *Big Data & Society*, *11*(1), 20539517241232630. https://doi.org/10.1177/20539517241232632

Nadeem, M., Ali, Y., Rehman, O., et al. (2023). Barriers and strategies for Digitalisation of Economy in developing countries. *Journal of the Knowledge Economy*. https://doi.org/10.1007/s13132-023-01158-3

Newlands, G. (2021). Lifting the curtain. *Big Data & Society*, *8*(1), 20539517211016024. https://doi.org/10.1177/20539517211016026

Ngcamu, B. S. (2023). Climate change effects on vulnerable populations in the Global South. *Natural Hazards*, *118*(2), 977–991. https://doi.org/10.1007/s11069-023-06070-2

Nie, M. (2024). Artificial Intelligence. *Proceedings of the AAAI Symposium Series*, *3*(1), 376–379. https://doi.org/10.1609/aaaiss.v3i1.31239

Nikolenko, S. I. (2019). *Synthetic Data for Deep Learning* (No. arXiv:1909.11512). arXiv. http://arxiv.org/abs/1909.11512

Nussbaum, M. (2000). Women's capabilities and Social Justice. *Journal of Human Development*, *1*(2), 219–247. https://doi.org/10.1080/713678045

Nussbaum, M. (2003). Capabilities as fundamental entitlements. *Feminist Economics*, *9*(2–3), 33–59. https://doi.org/10.1080/1354570022000077926

Nussbaum, M. (2004). Beyond the Social Contract. *Oxford Development Studies*, *32*(1), 3–18.

NVIDIA Corporation (2024). *Conflict Minerals Report as required by Items 1.01 and 1.02 of this Form - EX-1.01 - May 23, 2024*. https://fintel.io/doc/sec-nvidia-corp-1045810-ex101-2024-may-23-19866-8971

Odeku, K. O. (2022). Climate injustices due to the unequal and disproportionate impacts of climate change. *Perspectives of Law and Public Administration*, *11*(1), 103–110.

Okin, S. M. (2003). Poverty, Well-Being, and gender. *Philosophy & Public Affairs*, *31*(3), 280–316. https://doi.org/10.1111/j.1088-4963.2003.00280.x

OpenAI (2024). AI and Covert Influence operations. *OpenAI*.

Perrigo, B. (2023). *OpenAI Used Kenyan Workers on Less Than $2 Per Hour*. Time. https://time.com/6247678/openai-chatgpt-kenya-workers/

Peters, U., & Carman, M. (2024). Cultural Bias in explainable AI research. *Journal of Artificial Intelligence Research*, *79*, 971–1000. https://doi.org/10.1613/jair.1.14888

Png, M. T. (2024). The critical roles of Global South stakeholders in AI Governance. In J. B. Bullock, Y. C. Chen, J. Himmelreich, et al. (Eds.), *The Oxford Handbook of AI Governance*. Oxford University Press.

Pogge, T. (1988). Rawls and Global Justice. *Canadian Journal of Philosophy*, *18*(2), 227–256. https://doi.org/10.1080/00455091.1988.10717175

Pogge, T. (2001). Priorities of global justice. *Metaphilosophy*, *32*(1–2), 6–24. https://doi.org/10.1111/1467-9973.00172

Pogge, T. W. M. (2002). *World poverty and human rights*. Polity.

Posada, J. (2022). Embedded reproduction in platform data work. *Information Communication & Society*, *25*(6), 816–834. https://doi.org/10.1080/1369118X.2022.2049849

Qizilbash, M. (2002). Development, Common foes and Shared values. *Review of Political Economy*, *14*(4), 463–480.

Rafanelli, L. M. (2022). Justice, injustice, and artificial intelligence. *Big Data & Society*, *9*(1), 20539517221080676. https://doi.org/10.1177/20539517221080676

Rao, D. A. S., & Verweij, G. (2017). *Sizing the prize*. PwC Publication.

Rawls, J. (1971). *A theory of Justice*. Harvard University Press.

Ricaurte, P. (2022). Ethics for the majority world. *Media Culture & Society*, *44*(4), 726–745. https://doi.org/10.1177/01634437221099612

Sen, A. (1979). Equality of what? The Tanner lecture on Human Values. *Stanford University, May, 22*, 1979.

Sen, A. (1993). Capability and well-being. In M. Nussbaum, & A. Sen (Eds.), *The quality of life* (pp. 30–53). Oxford University Press.

Sen, A. (2005). Human rights and capabilities. *Journal of Human Development*, *6*(2), 151–166. https://doi.org/10.1080/14649880500120491

Shwartz, V. (2024, February 13). Artificial intelligence needs to be trained on culturally diverse datasets to avoid bias. *The Conversation*. http://theconversation.com/artificial-intelligence-needs-to-be-trained-on-culturally-diverse-datasets-to-avoid-bias-222811

Sloan, R. H., & Warner, R. (2020). Beyond Bias. *Virginia Journal of Law & Technology*, *24*, 1.

Søgaard, A. (2022). Should We Ban English NLP for a Year? In Y. Goldberg, Z. Kozareva, & Y. Zhang (Eds.), *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing* (pp. 5254–5260). https://doi.org/10.18653/v1/2022.emnlp-main.351

Sultana, F. (2022). The unbearable heaviness of climate coloniality. *Political Geography*, *99*, 102638. https://doi.org/10.1016/j.polgeo.2022.102638

Ta, R., & Lee, N. (2023). How Language gaps constrain generative AI development. *International Journal of Comparative Studies in International Relations and Development*, *9*, 48–52. https://doi.org/10.48028/iiprds/ijcsird/ijcsird.v9.i1.03

Tacheva, J., & Ramasubramanian, S. (2023). AI empire. *Big Data & Society*, *10*(2), 20539517231219240. https://doi.org/10.1177/20539517231219241

Tapia, D., & Peña, P. (2020). White gold, digital destruction. *Technology, the Environment and a Sustainable World*, 160–164.

Taylor, L., & Broeders, D. (2015). In the name of development. *Geoforum*, *64*, 229–237. https://doi.org/10.1016/j.geoforum.2015.07.002

Thanawala, S. (2023, September 25). Facial recognition technology jailed a man for days. *AP News*. https://apnews.com/article/mistaken-arrests-facial-recognition-technology-lawsuits-b613161c56472459df683f54320d08a7

Tubaro, P., & Casilli, A. A. (2020). Portraits of micro-workers. *2nd Crowdworking Symposium 2020*. https://hal.science/hal-02960775

Tubaro, P., Casilli, A. A., & Coville, M. (2020). The trainer, the verifier, the imitator. *Big Data & Society*, *7*(1), 2053951720919776. https://doi.org/10.1177/2053951720919776

Valentini, L. (2012). Ideal vs. non-ideal theory. *Philosophy Compass*, *7*(9), 654–664. https://doi.org/10.1111/j.1747-9991.2012.00500.x

Vallor, S. (2016). *Technology and the virtues*. Oxford University Press.

Veneziani, R., & Yoshihara, N. (2024). Unequal exchange and International Justice. In B. Ferguson, & M. Zwolinski (Eds.), *Exploitation*. Oxford University Press.

Walk Free (2023). The Global Slavery Index 2023. https://cdn.walkfree.org/content/uploads/2023/05/17114737/Global-Slavery-Index-2023.pdf

Wirtschafter, V. (2024). The impact of generative AI in a global election year. *Brookings*. https://www.brookings.edu/articles/the-impact-of-generative-ai-in-a-global-election-year/

Yong, Z. X., Zhang, R., Forde, J. (2023). Prompting Multilingual Large Language Models to Generate Code-Mixed Texts. *Proceedings of the 6th Workshop on Computational Approaches to Linguistic Code-Switching* (pp. 43–63).