A Qualitative Study on Cultural Hegemony and the Impacts of AI

Venetia Brown, Retno Larasati, Aisling Third, Tracie Farrell

The Open University, Walton Hall, UK
venetia.brown@open.ac.uk, retno.larasati@open.ac.uk, aisling.third@open.ac.uk, tracie.farrell@open.ac.uk

Abstract

Understanding the future consequences of artificial intelligence requires a holistic consideration of its cultural dimensions, on par with its technological intricacies and potential applications. Individuals and institutions working closely with AI, and with considerable resources, have significant influence on how impact is considered, particularly with regard to how much attention is paid to epistemic concerns (including issues of bias in datasets or potential misinterpretations of data, for example) versus normative concerns (such as societal and ecological effects of AI in the medium- and long-term). In this paper we review qualitative studies conducted with AI researchers and developers to understand how they position themselves relative to each of these two dimensions of impact, and how geographies and conditions of work influence their positions. Our findings underscore the need to gather more perspectives from low- and middle-income countries, whose notions of impact extend beyond the immediate technical concerns or impacts in the short- to medium-term. Rather, they encapsulate a broader spectrum of impact considerations, including the deleterious effects perpetrated by global corporate entities, the unwarranted influence of wealthy nations, the encroachment of philanthrocapitalism, and the adverse consequences of excluding communities affected by these phenomena from active participation in discussions surrounding impact.

Introduction

There is a long history of sociology in computing, to examine the culture of computing itself (Bloomfield 2018; Birhane et al. 2022): to understand the people involved, their motivations and their actions (Bloomfield 2018; Mindell 2015; Holton and Boyd 2021), to connect the culture around computing with its development, deployment and regulation (Kling 2000; Jasanoff and Kim 2015), and to critically address the role of power across these relationships and interactions (Kling 2000; Dignum 2019; Crawford 2021; Birhane et al. 2022). In this article, we contribute to the study of AI as a sociotechnical assemblage by reviewing in-depth, qualitative research on AI and its impacts gathered from researchers and developers (henceforth, "practitioners") within the field, highlighting several dimensions that may impact their position toward impact. In particular, we sought to understand how mainstream and alternative narratives around epistemic *versus* normative concerns are visible in how practitioners talk about the impact of AI, and explore some of the gaps in representation across the literature.

Our research questions were:

- **RQ1: How do AI practitioners attend to epistemic *versus* normative concerns, when talking about the impact of AI?**
- **RQ2: What are some of the dimensions that may shape a practitioner's view on the impacts of AI?**
- **RQ3: What can we learn from the above about the culture of Artificial Intelligence?**

Our findings suggest that wealthy, powerful countries and institutions, particularly those that currently set the trajectory and purpose of AI, also heavily influence perceptions around both epistemic and normative concerns. This is particularly visible in the consideration of long-term harms that may emerge, such as global inequality, the social and ecological implications of extractive industries, and general climate disruption, from which such actors are less likely to suffer negative consequences, directly or indirectly. The culture of Artificial Intelligence is, therefore, operating under the same global power structures as any other globally relevant field, but with the power to amplify these asymmetries. We argue that in order to fully understand and mitigate AI's impacts, we need extended consultations with AI researchers, developers and ordinary members of the public, particularly in low- to middle-income countries, to balance hegemonic views within the research literature, the media, and our sociopolitical and economic structures.

Methodology

Our study utilises Jasanoff and Kim's conceptual framework of sociotechnical imaginaries to help understand what "impact" means to those involved - the collective beliefs that give rise to shared ideas of how impact should be understood and performed, and to which outcomes. Jasanoff and Kim describe sociotechnical imaginaries as "collectively held, institutionally stabilised, and publicly performed visions of desirable futures, animated by shared understandings of forms of social life and social order attainable through, and

supportive of, advances in science and technology" (Jasanoff and Kim 2015, p. 6). We assert that the "desirable future" imagined by those with power and influence will determine future impact. Therefore, within this conceptual framework, we sought to explore mainstream and alternative narratives, government and institutional agendas, and explicitly stated goals and opinions from practitioners. We prioritised interviews and focus groups in our study, though we do include some qualitative surveys, workshops and case studies involving practitioners as researchers, as well as expert essays in our collection, particularly where extending certain points of view were helpful.

For a conceptual-mapping of the concerns around AI, we have used the categories provided by Mittelstadt et al. (2016), which include inconclusive, inscrutable, and misguided evidence, described as "epistemic concerns"; unfair outcomes and transformative effects described as "normative concerns"; and traceability (a transversal concern). These six broad categories were shown to have continued relevance over time (Tsamados et al. 2021), but even more so, the division between epistemic and normative concerns is a pragmatic way to understand the extent to which practitioners focus on the tools themselves, their quality and their specific applications (epistemic concerns), or the wider impacts they can have (normative concerns) (Morley et al. 2020).

Stage 1 – Identifying Narratives From the Literature

In the first stage of this study, we reviewed articles published between 2013-June 2023 that were specifically addressing the impact of AI. We excluded papers that had a focus other than understanding the impacts of AI more generally, such as papers addressing technology acceptance as the primary goal or papers discussing specific opportunities in AI for a given domain. As keywords, we used "qualitative research", "AI researchers", and "impacts of artificial intelligence", as well as synonyms and some sub-classes of these terms [1]. Our research focused on Google Scholar (arriving at 386 results). We then manually reviewed the abstracts of these papers (161 potentially relevant results), and then in more detail to ensure that they were qualitative studies, assessing the general impacts of AI, with AI researchers, developers and experts (27 papers). Of these 27 papers, 12 did not specifically discuss demographics of the respondents, but most were conducted with researchers, developers and other experts in technology companies, or with stakeholders in funded-projects taking place in North America and Europe. 11 papers addressed concerns of stakeholders in North America, Europe, Australia and Taiwan. 2 papers had diverse groups of participants from mixed global regions, and 2 papers focused on stakeholder concerns in the Asia-Pacific Region.

[1]The inclusion and exclusion criteria and keywords can be found at this address [removed for anonymous submission]

Stage 2 – Extending Narratives Through the Literature and Media

As this dataset was skewed toward the perspectives emerging in wealthier nations, or at large technology companies and elite Universities, we supplemented it with a narrative review, looking specifically for any qualitative work examining the perspectives of AI researchers from underrepresented countries and regions on the general impact of AI, but could only identify 4 further works that met our criteria. We discuss the limitations of our methodology in .

To represent the public and private sectors, we also include some detail on National Strategies and responses to those strategies, and discuss perspectives from large technology companies and their employees that can be found in news articles or on practitioners' own social media channels to surface some of the institutional understandings of AI, conflicts between companies and their employees, and perspectives on how best to control the impacts of AI.

Stage 3 - Categorising Narratives

In stage three of our study, we began to more clearly categorise narratives. Geography emerged as a category of interest. According to a report by Statista [2], the top 10 countries doing research on AI are the United States, China, Singapore, Switzerland, the United Kingdom, Australia, Canada, Germany, Finland, and the Netherlands (which is followed closely by Israel). There are also other indices that measure involvement in AI more directly, such as the AI Readiness Index from consultancy group Oxford Insights[3], which highlight similar regions as having the best infrastructure and capacity to build AI. Economy also emerged as a category, and we made the decision to refer to World Bank 2024 classifications of countries as low-, middle- or high-income based on Gross National Income[4]. There are disadvantages to using this method that we acknowledge, particularly in underestimating the income of countries that rely on subsistence or more informal economies or overestimating income in countries where wealth disparity is significant between different regions and communities. However, because these measures translate to real power in the global economy, including investment and monitoring, we argue that they can provide one way of looking at global exclusion in AI discourse. These categories are very much entangled. The Oxford Insights AI readiness index 2023 report, for example, found an increase in national strategies published by lower- and middle-income countries. However, in terms of data and infrastructure, lower-income countries appeared to have more barriers and dependencies on higher-income countries that have the technology to offer. High-income countries score much higher on the technology sector indicators, with the exception of Malaysia, Brazil, Russia, India and China, all middle-income countries. Taken together, it is perhaps unsurprising that having the money and infrastructure to build

[2]https://www.statista.com/statistics/1410523/top-20-ai-countries-by-research-capacity/

[3]https://oxfordinsights.com/ai-readiness/ai-readiness-index/

[4]https://datahelpdesk.worldbank.org/knowledgebase/articles/378832-what-is-the-world-bank-atlas-method

and test AI is reflected in a country's research capacity in this area.

The Balance of Concern: Previous Studies

As a precursor to our study, we wished to establish a trajectory in how different sets of principles or guidelines around the potential impacts of AI have been understood and operationalised. In particular, we wanted to understand how different epistemic and normative concerns were expressed through these documents and how this has evolved over time. There have been many such reviews, for example, Jobin, Ienca, and Vayena (2019); Schiff et al. (2020); Hagendorff (2020); Morley et al. (2020); Fjeld et al. (2020); Kazim and Koshiyama (2021); Franzke (2022) (not exhaustive), which analyse areas of convergence and divergence, to identify areas of significance. Below we discuss three examples that were published in 2019, 2020 and 2022.

In a systematic review of a corpus of 85 ethics guidelines and principles, Jobin, Ienca, and Vayena (2019) identified 11 principles, of which 5 (transparency, justice and fairness, non-maleficence, responsibility, and privacy) appeared most often, indicating convergence around these topics. However, all 11 principles differed, semantically and conceptually, across the corpus. The authors found that justice was most often interpreted as fairness, and that fairness was most often understood as avoiding "unwanted bias" and discrimination, though discrimination was less often referenced by the private sector. Justice was also often understood in terms of having a right to appeal a decision or to access how a decision was made. This is in contrast to "social justice" concepts, like solidarity, which include (for example) medium- and long-term impacts on labour markets and the promotion of individualism, or the "need for redistributing the benefits of AI in order not to threaten social cohesion", which were identified in only 6 of the documents in the corpus. Most of the documents in their corpus were produced by private companies or government entities.

A later study by Fjeld et al. (2020) identified 8 themes in their corpus of 36 sets of principles, and found greater similarity among them, particularly among those published more recently. Again, however, one can see that the interpretation of principles varies. Principles under the category of fairness related predominantly to the avoidance of negative bias, rather than equality (for example). Principles of privacy and human agency that allow one to resist technology, such as rights to erasure or opt-out, were less represented in the corpus. Consideration of long-term effects, and access to technology was only found in 33% of the documents, even if a majority of the documents (64%) promote the idea that AI should be beneficial to society. In some ways, this may reflect a desire to rely on entities outside of technology to manage normative concerns. For example, 64% of the documents in their corpus referenced the issue of human rights and the authors suggest that existing regulations, legal statues and practices could be helpful in operationalising and adjudicating future issues. They caution that a contextual approach will be necessary to understand how principles should be actioned.

In a further, in-depth qualitative analysis of 70 ethics guidelines and principles, Franzke (2022) found that even the term ethics had a vague connotation of "goodness" or "rightness", or was conflated with other terms, in addition to the corroboration with previous studies that principles themselves are "hazy". A further finding was that most guidelines centred the technology itself, the quality of the tools and how they are built, rather than ethical applications or outcomes. Moreover, Franzke (2022) argues that there is no real consideration of the possibility that technology should not be developed, similar to what was identified by Fjeld et al. (2020) in the areas of privacy and human control. Franzke (2022) writes: "the grounding presupposition at play here is that technological progress simply constitutes the 'good"'. This sentiment reflects and extends ideas of inevitability that are common in mainstream narratives (McQuillan 2022). Franzke (2022) suggests that it is necessary to consider how power dynamics and values are embedded in the entire pipeline of AI, as a material thing based on resources, human labor and infrastructure and resulting in capital gains. She refers to AI as a "registry of power", in which perceptions of harm, and value, shift over time.

Given what is understood about the geographies and economies of those involved in both setting the trajectory of AI and discussing its future impacts, it can be proposed that the hegemonic values expressed within the culture of AI are typically those that emerge from High-income, AI-Intensive Nations.

Perspectives From High-Income, AI-Intensive Nations

In this section, we consider the perspectives of practitioners with (currently) the most decisive roles in AI impact - those based in regions of the world with considerable investment in AI and contributions of AI products and services.

Mainstream Narratives and Attention to Normative Concerns

In a 2019 study with AI users (N = 135) and developers (N = 70) in the United States (US) and Australian context around the principles of Human-centred AI, Kevin, Mark, and Bernd (2019); Bingley et al. (2023) indicated that the quality of an AI tool was important to both users and developers, but that the perceived social impact of the technology was less important for developers than for users. One potential reason for this could be the extent to which AI researchers feel that normative concerns are part of their purview. Three recent studies from institutions in high-income, AI-intensive nations have indicated that social impacts are a matter of personal and professional preference to AI practitioners (Vaast 2022; Slota et al. 2023; Bingley et al. 2023). Though normative concerns appear in principles and guidelines around the use of AI, as discussed above in , the variability of interest from practitioners may explain why concepts vary so considerably in their operationalisation. Understanding that this variability exists, one may consider different approaches for encouraging more engagement and developing a better understanding. Slota et al.

(2023), who gathered a range of stakeholder perspectives from Research and Development organisations, non-profits, government and academia, suggest that, given the significant consequences that AI can have on society, thinking about impact would need to be incentivised to encourage practitioners without an expressed interested in this area to play a more decisive role. In a study with open source researchers, Vaast (2022) suggests that the tension practitioners sense around the future of AI and their experience of its impacts will shape how practitioners view their work. Bingley et al. (2023), who contrasted the views of developers and users within their Human-Computer-Interaction (HCI) study, suggest that understanding users should drive the way we consider social impacts, particularly in understanding users' needs and how AI "helps or hinders the satisfaction of these needs".

These three different approaches, incentivising thinking about impacts, learning by experience and centring user needs are all sensible approaches for understanding normative concerns, but they still depend on the quality of information available to practitioners - what impacts are understood, what experiences the practitioner can draw from, what kind of users the practitioner engages with.

Mainstream Narratives and the Law

Another way practitioners attend to AI's impact is through legal frameworks, which set out some basic requirements regardless of personal or professional interest.

In a 2020 study based on interviews with 21 AI practitioners across public, private and academic institutions in Australia, Orr and Davis (2020) found that practitioners were more concerned with the compliance of their AI systems with established laws than with the broader ethical implications and possible future transformative effects. A similar finding was noted in a 2021 European-funded study with 42 individuals working with Smart Information Systems in different sectors (including government, health care, cybersecurity, telecommunications and insurance). The authors of that study(Stahl et al. 2022) found that all participants were aware of the concept of ethics and knew of strategies their organisations were developing to adopt and implement different frameworks for mitigating harm, but many practitioners felt they lacked the appropriate training in this area, aside from mitigation techniques related to the quality of tools (i.e. bias, fairness, etc.). So, it could be that AI practitioners are not as interested in normative concerns, or they feel they lack the training to consider them, but there are also additional challenges that are more social than technical. In a study with thirty-three AI practitioners from three technology companies (Madaio et al. 2022), researchers explored how how practitioners incorporate fairness practices into their work. The example they focused on was dis-aggregated evaluations, which can illuminate performance disparities for different demographic groups. What they found was that companies were often prioritising customers over potentially disadvantaged groups and that practitioners faced confusion and push-back over which performance metrics to use during evaluations, how to identify relevant stakeholders, and how to choose which demographic groups to focus on.

It may be the case that entities struggling with (or against) the complexities of harm or mitigation strategies fall back on legal definitions of harm as a result. However, unless this is supported with the appropriate training and commitment on the part of the employer or institution, strategies to avoid negative impacts are difficult to follow through.

Mainstream Narratives on AI for Social Good

In addition to enhancing personal or professional interest, or aligning AI with the law, another popular approach to shape the impacts of AI is through the concept of "social good", in particular, through alignment with UN Sustainable Development Goals (SDGs) (Tomašev et al. 2020; Cowls et al. 2021; Floridi et al. 2021). This approach is not without its critics, but there is some variation in critique. For example, Floridi et al. (2021) argue that such projects have a higher burden around safeguards and incremental deployment, context driven intervention and explanation, and situated fairness, alongside privacy and consent. In contrast, Cowls et al. (2021) suggest that projects must be aligned with where the greatest needs are and not just the greatest benefits, which suggests some attention to power asymmetry and global structures. In a 2020 review of more than 1000 papers, Shi, Wang, and Fang (2020) suggested that topic coverage is ad hoc and geographic coverage is unbalanced. Among our qualitative studies, a 2021 survey conducted with a convenience sample of 1018 participants from Taiwan (all who were pre-screened for education level and access to the internet), (Yeh et al. 2021) explored local perceptions of AI with regard to SDGs. The authors found participants to be "rationally optimistic" about AI, understanding that there are very serious risks involved in the development and deployment of AI, but viewing those risks as possible to mitigate through institutional monitoring and regulation, perhaps combining legal narratives with those of social good. Participants viewed linkages with SDG 4 (quality education), SDG 9 (industry innovation), and SDG 3 (good health and well-being) as offering the most promise with the fewest negative trade-offs. Participants rated SDG 10 (reduce inequality) as having the highest trade-off rate, which the authors argue reflects the belief that wealthier countries will benefit most and have more opportunities related to AI, widening of the "gap between richer poorer countries". To support their conclusion, the authors remark that SDG 1 (no poverty) and SDG 8 (decent work and economic growth) were underneath SDG 10 in perception of trade-offs. This study highlights the wealth of nations and wealth disparity is one feature of AI development that will affect future impacts, which is less commonly found in other mainstream narratives.

National Strategies as Mainstream Narratives

National strategies in wealthier countries may suffer from some of the same challenges as those described above.

The United States' 2016 National Artificial Intelligence Research and Development Strategic Plan and its 2023 up-

date[5] suggest reliance on legal and social good frameworks to guide responsible innovation in the field, with a focus on maintaining US leadership in the field. However, critics have suggested that its lack of a "flagship legislative AI initiative" and congressional infighting over legislation leaves significant gaps that can be exploited[6]. The EU AI Act (Act 2021), in contrast, was intended as a world-leading piece of legislation to address the risks of AI, as well as its promise. Critics have suggested that it lacks any significant, mandatory guardrails (Veale and Zuiderveen Borgesius 2021), and conflates trust with acceptability (Laux, Wachter, and Mittelstadt 2024).

National strategies appear to suffer from the same challenges described in around interpretation of terms and operationalisation. In a 2020 report on Trustworthy AI from the High-Level Expert Group on Artificial Intelligence (AI HLEG), set up by the European Commission (Ala-Pietilä et al. 2020), there are no prescribed or recommended process, for example, on how ensure that stakeholders' needs and concerns are taken into account. Environmental considerations are largely around measures for energy use and carbon emissions, relating back to legal and political frameworks. The section on society and democracy also lacks in specific recommendations, asking only "did you consider the impact on society?", or "Did you take measures that ensure that the AI system does not negatively impact democracy?", immense subjects that are extremely complex and require specialist knowledge to adequately assess. As already stated, many AI practitioners do not feel they have this competency and some do not feel motivated to develop it.

Regional Comparisons

Of course, mainstream narratives may also differ slightly from region to region, but research between even two very culturally different nations still shows considerable convergence around some expectations of AI's impacts. In this extended example, we discuss a study from Hautala and Ahlqvist (2022) conducted with n=26 AI practitioners from Finnish and Singaporean contexts, both countries that are digitally competitive. The authors studied each nation's strategies toward AI, made observations of different AI related events taking place in each country, and conducted in-depth interviews with practitioners, to understand how they envision the future, their expectations of AI, and the practices that are currently undertaken or supported in recognition of that vision.

The futures imagined by the participants differed slightly between the Finish and Singaporean respondents, but the general vision reported by both sets of stakeholders was that AI will be pervasive in future society, that it will be disruptive along the way, and that it will continue to develop at a rapid pace. One expectation shared by all practitioners in the study was that AI is something that should be understood by everyone. The authors refer to this as the perspective that "AI is the new electricity", something that everyone will eventually use and appreciate, which makes educating the wider public a necessity. The second expectation is that AI will eventually become as "intelligent as a person is". Developers were less certain about the timeline to this than managers and other AI leaders in each country, and there were also differing opinions about what would have to change about people, the technology and our understanding of intelligence along the way, but the expectation that it was possible and even probable was held by all but one participant. The third expectation was that AI will replace activities that are repetitive and routine, allowing humans to focus on other types of tasks. While some participants did express concern over different impacts on social inequality, the future of work, and human agency moving forward, most expressed scepticism toward regulation due to the knowledge disparity between different groups. Some of the participants expressed concerns about the impact of such transformations on labour, exacerbating existing digital divides (between older and younger employees, for example), and the directional relationship of the human-computer interaction (machines telling humans what to do vs. humans telling machines what to do). They viewed accountability as shared between practitioners, nation states and global companies to regulate the development of AI, with some scepticism about how regulation will be implemented when the knowledge disparity is so large between different stakeholder groups.

The researchers involved in this study identified several practices which they refer to as "anticipatory", in relationship to this future vision. First, there is the practice of anticipating and "riding the wave" of AI hype. The second is "taking up challenges" that interfere with the future vision (like dealing with issues of data availability and quality of algorithms, for example, but also educating the public, correcting misconceptions around AI and challenging the narrative that AI should be applied to every problem). The authors of this work viewed the role of developers, given their knowledge of both technical requirements and potential social impacts, as deconstructing and avoiding the "costs of AI hype". When we discuss the experiences of practitioners in large technology companies in section , we will explore further how practitioners can wind up mediating the relationship between such institutions and the public, relative to the technology.

Perspectives From Low- and Middle-Income Countries

As we mentioned previously, very few papers that we reviewed addressed the in-depth points of view of AI researchers in low- and middle-income countries around the general impacts of AI. There may be several reasons for this. First, our manual review of the initial sample indicated that their perspectives are often sought as end-users, where specific applications of AI are expected to be deployed. Second, our survey was limited to English language papers, something we will expand upon in future work. Third, while AI researchers educated in low- or middle-income countries are

[5]https://www.whitehouse.gov/wp-content/uploads/2023/05/National-Artificial-Intelligence-Research-and-Development-Strategic-Plan-2023-Update.pdf

[6]https://carnegieendowment.org/2023/05/03/reconciling-u.s.-approach-to-ai-pub-89674

present within the literature, they are often living and working in wealthier countries, which could potentially complicate their perspectives. The exorbitant salaries that can be offered by large tech companies like Google, Meta and Amazon bring many promising researchers to the United States and Europe, away from local communities elsewhere. Further, the privatisation of research has been shown to negatively impact the visibility of researchers' work over time (Jurowetzki et al. 2021), making it difficult to gather perspectives once those researchers move on to industry positions.

LMICs as Recipients of AI

Low- and middle-income countries (LMICs)are often viewed as potential loci for AI projects that address the UN Sustainable Development Goals (SDGs) (Wakunuma, Jiya, and Aliyu 2020; Ciecierski-Holmes et al. 2022; Tomašev et al. 2020; Cowls et al. 2021; Floridi et al. 2021), or as a part of industrialisation strategies with a view on being "competitive" with wealthier nations (Fejerskov 2017; Heng et al. 2022). Research also indicates that the negative impacts of AI are likely to be more keenly felt in regions of the world that are globally minoritised and peripheralised, a *global* mirroring of the kinds of impacts AI can have on marginalised populations from within AI advanced countries as well (Hagerty and Rubinov 2019; Petermann et al. 2022). Fejerskov (2017) refers to LMICs as a "laboratory for technical experimentation" fuelled by the machinations of global corporations, private foundations and what Fejerskov (2017) refers to as "philanthrocapitalists", that could have serious consequences for local development, democracy and populations themselves. However, there have been a few pockets of qualitative research conducted with a view of understanding how AI researchers living and working in low- and middle-income countries view their local context of AI development and innovation.

In a study based on 16 in-depth interviews with stakeholders from the public and private sectors in Cambodia and Senegal, for example, Heng et al. (2022) aimed to map out the ecosystem of AI in each country. In both cases, foreign entities, such as consultants or even national governments of other countries were identified as key actors. Participants from Senegal also discussed the lack of available open datasets, which were often held by private companies. This points to the challenge of interference in the agency of low- and middle-income countries to control the impact of AI and the use of their own data. In a short qualitative analysis of public statements by the co-founders of the African Institute for Mathematical Sciences and its African Masters of Machine Intelligence (in partnership with Google and Facebook, now Meta), Hassan (2023) argues that the partnership with Western countries is a contested issue even within individuals themselves. He cites statements from the programs' founders and key academics that express a desire to develop local expertise, not only for the purposes of job creation and global competitiveness, but primarily to direct the trajectory of AI research toward the issues and challenges that are viewed as important by those with experience of that context.

National policy narratives are also often intertwined with the global supply chain of AI, meaning that countries that are relying on wealthier nations to help them implement their AI road-maps, are also likely to be heavily influenced by the interests of those countries (Kak 2020; Heng et al. 2022).

LMICs and Alternative Moral Frameworks

In pre-aligning the development of AI in LMICs with SDGs, there is a kind of digital colonialism evident. In the report, ESCAP et al. (2020) publish a selection of essays prepared by a multi-disciplinary team of researchers who explore the unique perspectives from nations and communities in Asia and the Pacific, alongside their policy recommendations for AI development in the region. The findings in this report showed that researchers in Asia and the Pacific relied on different moral frameworks to determine what should and should not happen with AI, like altruism over individualism, collectivity, and communal approaches to data. Contributed essays discuss developing capacity to manage risk (rather than avoid risk altogether), question centralised approaches and monopolies over data in favour of universal access, and the necessity to provide multi-stakeholder oversight, specifically, including the perspectives of younger generations. They also highlight the need for experts to be more "sensitive to the concerns of ordinary people" around AI, in addition to business leaders, politicians, technology experts and other stakeholders that are typically acknowledged.

Similarly, a 2021 discourse analysis based on 36 qualitative interviews conducted with AI researchers in India found that local conditions significantly impact both the conceptualisation and operationalisation of algorithmic fairness (Sambasivan et al. 2021). The authors argue that the western orientation of fairness, including the legal framing often based on the regulatory environment of the US (for example) and philosophical underpinnings that prioritise consequentialism and deontic notions of justice, seemingly exclude other "moral foundations", such as purity/sanctity or restorative justice that may be more common in older, more traditional societies.

Many African nations, for example, Kenya, Tunisia, South Africa, Ghana and Uganda are also developing strategies around data protection and ethics, but researchers active in the region have argued that regional diversity means that multiple moral frameworks and religions could (and should) inform and interact with this process (Goffi 2023; Okengwu 2023). In fact, (Goffi 2023) proposes the need for a regionally relevant, African Ethical Framework, to decentre the West in determining what should and should not happen with AI. A recent survey from (Nakatumba-Nabende, Suuna, and Bainomugisha 2023) on ethics education and African Universities indicated that African students in computer science and AI are largely aware of global concepts of ethics, but without local examples to learn from, the subject remains largely theoretical. (Nakatumba-Nabende, Suuna, and Bainomugisha 2023) recommends growing local capacity for engaging in the subject of ethical AI, and a "glocalisation" (a combination of globalisation and localisation) of ethical frameworks, customised to the African context. Hassan (2023) notes, however, that calling for the decentring of

White European values and norms may not go far enough to question the notions of ethics and intelligence themselves, or decentre corporate interests. The extraction of both materials and human resources from poorer countries by wealthier nations (Monasterio Astobiza et al. 2022; Goffi 2023), feeding the interests of wealthier nations and supporting their investments abroad, also ignore what (Stilgoe 2020) argues is a central ethical question in artificial intelligence: who ultimately benefits from AI?

Wakunuma et al. (2021) argue that Responsible AI, as it is envisioned by wealthier, industrialised nations, places too much emphasis on linear progress, technical innovation, profitability and the creation of new markets, which is not fit for purpose in regions where more emphasis is placed on local communities and "informal knowledge systems". The authors champion the need to reconfigure "responsible research and innovation" (RRI) that is not only focused on "capital-oriented elements of RRI but livelihood-oriented RRI", that takes into account the heterogeneous landscape of customs and culture. This is perhaps exactly where the reliance on legal frameworks imported from other regions like the US and Europe will fall short in circumscribing that could be viewed as responsible in any global sense.

Perspectives From Large Tech Companies

While the perspectives of large technology companies are visible across the studies we have discussed above, it is worth discussing these organisations as spaces where some competing values (and narratives) emerge.

Diversity and Values

The lack of diversity at large tech companies[7] can increase the potential for transformative harm, when mainstream narratives crystallise (Mittelstadt et al. 2016), and personal values cannot break through institutional practices (Ryan et al. 2022). AI researchers can be the instruments of change in organisations, if organisations embrace it. Meyerson's proposition of "tempered radicalism" (Meyerson and Scully 1995) describes how employees can shift values and inspire change within an organisation over time.

AI practitioners have already impacted areas of AI research into automated weapons, in setting out principles of Fair, Accountable, Transparent and Ethical research, in whistle-blowing and organising within their companies (Belfield 2020), and have founded many internationally recognised institutes addressing the impact of AI technology, such as the Algorithmic Justice League[8] and the Distributed AI Research Institute (DAIR)[9].

However, organisations do not always welcome the new perspectives of the diverse talent they recruit. In the spring of 2023, Microsoft fired its entire team for responsible and ethical AI development as part of a larger lay-off[10]. The former team believes that one reason for this may have been internal critiques of the speed of innovation, in which the company was keen to become more competitive with Google on AI products[11]. Amazon's platform Twitch also fired its ethics and safety teams in the past two years[12] amidst complaints that the platform was biased against women and people of color[13]. Elon Musk, in his takeover of Twitter (now "X") fired one-third of the company's workforce, including most of its ethics and safety team[14] despite many different concerns about lacking representation within the company[15] and biases in its recommendation and amplification algorithms (Huszár et al. 2022).

These events highlight how the maintenance of power asymmetry, particularly in the private sector, is accomplished through a lack of regulation, where it is permitted to revoke power or backslide on social gains.

Big Tech and the Singularity

Of course, other critiques from within large tech organisations are less about the current harm that AI is causing and more about the future harm that ever-more-powerful AI and language models could produce. The Centre for AI Safety has an open statement on the risk of extinction from advanced AI[16] that has been signed by hundreds of AI researchers, developers and public personalities. Many signatories are the heads of large technology companies, like OpenAI and DeepMind, professors and other senior academics at influential academic institutions, such as Berkley and MIT. Another open letter from the Future of Life Institute[17] has more than 30,000 signatories to date and calls for "all AI labs to immediately pause for at least 6 months the training of AI systems more powerful than GPT-4". This letter has also been signed by many influential people in the field of technology, including the co-founder of Apple, Steve Wozniak, co-founder of Skype, Jaan Tallinn, co-founder of Pinterest, Evan Sharp and Space-X CEO Elon Musk. The subject of co-called advanced, artificial general intelligence (AGI) has also been the subject of multiple articles from large publication outlets like the New York Times[18], Finan-

[7] https://www.globaldata.com/data-insights/technology–media-and-telecom/ethnic-representation-in-big-tech-companies-in-2091383/

[8] https://www.ajl.org/

[9] https://www.dair-institute.org/

[10] https://www.theverge.com/2023/3/13/23638823/microsoft-ethics-society-team-responsible-ai-layoffs

[11] https://www.platformer.news/p/microsoft-kickstarts-the-ai-arms

[12] https://www.bloomberg.com/news/articles/2023-04-11/twitch-cuts-in-safety-ai-ethics-raise-concerns-among-ex-workers

[13] https://senseient.com/ride-the-lightning/tech-firms-are-laying-off-their-ai-ethicists-sigh/

[14] https://www.wired.com/story/twitter-ethical-ai-team/

[15] https://peopleofcolorintech.com/articles/musks-twitter-layoffs-takes-a-hammer-to-diversity-and-inclusion-efforts/

[16] https://www.safe.ai/statement-on-ai-risk

[17] https://futureoflife.org/open-letter/pause-giant-ai-experiments/

[18] https://www.nytimes.com/2023/03/29/technology/ai-artificial-intelligence-musk-risks.htm

cial Times[19], and the BBC[20].

However, researchers at the Distributed AI Research Institute argue that the risks already posed now to human privacy, security, and well-being should remain in sharp focus for researches and regulators, and that we need to carve out space for more imaginative and meaningful applications of this very powerful technology to build a better future[21]. Moreover, some critics have pointed out that a large majority of signatories to such open letters as discussed above are also those currently involved in generating the hype around artificial intelligence to facilitate a "utopia" within which we can cure all diseases, colonise space, and transcend all of our bodily limitations[22]. Some have even argued that both the "utopias" proposed and the prioritisation of risks related to AGI are a reflection of societal power dynamics that centre Whiteness and the priorities of racialised industrial capitalism[23].

At large tech companies, the number of stakeholders, the power asymmetries between them, and the variable decision-making ability means that the proverbial waters can become quite muddied. In a study based on 26 interviews conducted with stakeholders from research, law, and policy, (Slota et al. 2021) argue that "distributed agency", and the knock-on effects for accountability, responsibility and liability create an environment where the design, evaluation and regulation of AI systems is incredibly inconsistent and difficult to grasp. This perhaps mimics the wider global context.

Resistance Narratives

There are also several key narratives around the development of AI that are sceptical of its potential for anything *other* than harm. While this research is not like the empirical, qualitative studies we have discussed in previous sections, it serves to motivate the need for increased collective understanding around this technology and its future impacts.

Yarden Katz's 2020 book, "Artificial Whiteness: Politics and Ideology in Artificial Intelligence" (Katz 2020), argues that the very organisation of AI technology is predicated on power, specifically power associated with racialised, industrial capitalism, which is undermined by what he describes as an inauthentic claim of neutrality (AI as a "tool"). He explores historical connections between AI and the military-industrial complex, and demonstrates how AI reproduces the logic of White supremacy by presenting the performance of Whiteness as the professional and moral standard. This point of view is echoed by other researchers and activists working in the field of decolonial AI. Dr Kanta Dihal, who heads the Decolonising AI initiative at the University of Cambridge's Leverhulme Centre for the Future of Intelligence (CFI), has said "Given that society has, for centuries, promoted the association of intelligence with White Europeans, it is to be expected that when this culture is asked to imagine an intelligent machine it imagines a White machine"[24]. The idea that machines will be better or smarter than humans at making decisions, combined with the reality that machines are more likely to perform Whiteness, creates the kind of impact that ethical or responsible AI principles are not likely to capture.

Likewise, Dan McQuillan's 2022 book, "Resisting AI: An Anti-fascist Approach to Artificial Intelligence" (McQuillan 2022), explores a future trajectory of AI that is likely to increase authoritarianism and austerity, and impact distribution of power, as a political technology. McQuillan points to precarious labour markets and discriminatory AI, as do other scholars, but he goes further in suggesting a necropolitics served by AI that will help to determine through its modelling, optimising, and course-suggesting/correcting the life and death of human beings. This is perhaps most apparent in the use of AI in warfare. Reports on the Israeli Defense Force's (IDF) use of an AI system in rapidly identifying new targets such that it produces twice the number of targets each day than human intelligence officers could predict in one year, is exactly one such example[25][26][27]. Indicators also suggest that Russia is investing in the development of advanced autonomous weapons to replace human infantry in combat settings[28][29]. Other examples include the use of lethal robots in the context of policing[30] and defence, for example in Ukraine[31]. According to the news articles footnoted above, several smaller manufacturers, many of them US-based, are even entering directly into negotiations with nation states who wish to use their software, in lieu of more traditional nation-to-nation channels associated with other types of combat technology and equipment. Dahab (2019) has suggested that AI may threaten society in ways similar to the arms race and nuclear proliferation, because of its ability to impact military operations. Ricaurte (2022)argued more generally that Hegemonic AI, within which the processes of datafication, algorithmisation and automation occur, is currently consolidated within a few wealthy countries and organisations with little oversight and is therefore likely

[19]https://www.ft.com/content/03895dc4-a3b7-481e-95cc-336a524f2ac2

[20]https://www.bbc.co.uk/news/technology-65110030

[21]https://www.theguardian.com/technology/2023/sep/25/experts-disagree-over-threat-posed-but-artificial-intelligence-cannot-be-ignored-ai

[22]https://www.salon.com/2023/06/11/ai-and-the-of-human-extinction-what-are-the-tech-bros-worried-about-its-not-you-and-me/

[23]https://davidgolumbia.medium.com/the-great-white-robot-god-bea8e23943da

[24]https://www.cam.ac.uk/research/news/whiteness-of-ai-erases-people-of-colour-from-our-imagined-futures-researchers-argue

[25]https://www.theguardian.com/world/2023/dec/01/the-gospel-how-israel-uses-ai-to-select-bombing-targets

[26]https://www.politico.eu/article/israel-drones-high-tech-weapons-united-states-ai/

[27]https://www.jpost.com/israel-news/article-771419

[28]https://www.russiamatters.org/analysis/roles-and-implications-ai-russian-ukrainian-conflict

[29]https://www.cna.org/our-media/newsletters/ai-and-autonomy-in-russia

[30]https://medium.com/whats-at-stake-in-a-fourth-industrial-revolution/race-necropolitics-and-the-human-in-an-age-of-intelligent-machines

[31]https://www.nationaldefensemagazine.org/articles/2023/3/24/ukraine-a-living-lab-for-ai-warfare

to become an additional instrument of subjugation.

McQuillan also points to the participation of the academy in promoting AI hype with little introspection on these issues. He suggests that nothing short of a new structure to our sociopolitical and economic lives will be enough to prevent the harms we live and die with today from becoming exacerbated by the tool of AI. To facilitate this, he suggests a stronger community of resistance to explore alternative uses of this technology.

Discussion

In the following subsections, we address each of our research questions and the conclusions we have drawn from the literature. To help illustrate some of our points, we include some reflections from specific domains that were not included in our main study, but which address some of the points raised in this section.

RQ1: How Do AI Practitioners Attend to Epistemic *versus* Normative Concerns, When Talking About the Impact of AI?

The results from our study suggest that most AI practitioners, globally, will wind up focusing on issues such as the legality of their tools, reducing bias in their tools and seeking higher quality data, because national strategies are too inconclusive and lacking in legislative or regulatory will, or, in the case of low-middle income countries, too heavily influenced by powerful institutions and AI advanced countries. AI researchers also lack specific education and guidance around more normative harms that could seriously impact society in the long-term.

This does not mean that there are no AI researchers operating in high-income countries that have their eye on normative concerns. Virginia Dignum's manuscript on Responsible Artificial Intelligence (Dignum 2019) centres the responsibility to view AI technology within the framework of power, to "ensure human flourishing and well-being in a sustainable world' (p. 119). Pratyusha Kalluri of the Radical AI network underlined this sentiment in her 2020 article on Nature online[32] that we should be asking questions around how AI shifts power, rather than asking what is "good" or "bad" to do with AI. AI researchers coordinated by the AI startup Hugging Face, released an open-source large-language-model (LLM) BLOOM, through the efforts of 1,000 volunteer researchers[33] to ensure the transparency and accessibility that other companies like OpenAI and Google will not. Academics from well known institutions have criticised facial recognition software (Buolamwini 2017) and software for predicting recidivism (Benjamin 2023), which are both deployed within the criminal justice system have been shown to disproportionately impact people of non-White ethnicities. FinTech also struggles with biases against both women and other peripheralised ethnic groups through implied characteristics related to issues like housing or finance, where such groups have been historically excluded and/or disadvantaged in the places where they live(Kelley et al. 2022).

However, we argue that these efforts are not mainstream when we explore the entirety of viewpoints, from different nation states, from the experiences of low- and middle-income countries, and from the experiences of diverse employees at large technology companies. Rather, they constitute the efforts of pockets of researchers who deeply understand this technology and the motivations of the powerful actors around them.

Moreover, perceptions from the public, which are more likely to reflect what they encounter in the cultural artefacts of our times (such as media, books and films), appear to respond to the mainstream narratives. Cave and Dihal (2019) analysed 300 fictional and non-fictional works that reflect both the utopian and dystopian futures suggested by (predominantly White, wealthy) entrepreneurs, of an AI that will become ubiquitous, free us from the necessity of work, and potentially become more powerful than us, alongside the inevitability of the technology criticised by McQuillan (2022) and others. Cave and Dihal (2019) explores how these perceptions become reciprocal along the pipeline of AI from development to deployment and regulation, as sociotechnical imaginaries (Jasanoff and Kim 2015). This is why it is so important when works like "Coded Bias" from Joy Buolamwini appear on Netflix[34], or books like Safiya Umoja Noble's "Algorithms of Oppression" become part of mainstream discussion in the media. These are points of view that enter the public consciousness in countries most responsible for the negative impacts of AI and create spaces for resistance.

RQ2: What Are Some of the Dimensions That May Shape a Practitioner's View on the Impacts of AI?

In our study, those from, or working in, low- and middle-income countries appear more aware of specific transformative effects, namely the reciprocal relationship between the impact of wealth and political power in determining the impacts of AI, and the downstream effects of that AI from a global perspective. Our research also indicates that, though AI practitioners in wealthier nations can have a transformative impact on their organisations, their hands are often tied by decision-makers who prioritise business agendas over deep exploration of impact and harm mitigation. Even when researchers with unique perspectives on the consequences of AI are recruited from low- and middle-income countries to attend elite Universities or work at large technology companies in the US, Europe and Australia, for example, their perspectives are often undermined, and the nature of private research means that their perspectives often become less visible, unless the individual makes a very public stand against their company. This type of activity can come at great personal and professional cost (Kwarteng et al. 2021).

[32] https://www.nature.com/articles/d41586-020-02003-2

[33] https://www.technologyreview.com/2022/07/12/1055817/inside-a-radical-new-project-to-democratize-ai/

[34] https://www.netflix.com/gb/title/81328723

RQ3: What Can We Learn From the Above About the Culture of Artificial Intelligence?

When practitioners live and work in a low- to middle-income country where wealthier countries are meddling in national strategies and foreign companies are the providers of so much technology that countries do not have the capacity to develop themselves, it becomes difficult to disentangle the local and regional viewpoint from that of its international partners. This makes it difficult to ascertain what Jasanoff and Kim (2015) describe as "collectively held, institutionally stabilised, and publicly performed visions of desirable futures". Much like any cultural hegemony, the mainstream narrative around the impacts of AI reflects the point(s) of view held by the most powerful and influential. However, it is clear from our research that the wider conversation around issues such as responsible and ethical AI, or AI for social good, would benefit from decentring mainstream narratives, if the purpose of such frameworks is to limit harm. As low- and middle-income countries are more likely to carry the burden for the most serious consequences of AI technology over time, it is important for their local and regional stakeholders to contribute meaningfully to this debate and shape such frameworks for their own contexts.

Limitations and Future Research

The limitations of this study relate to three methodological choices. First, focusing on general impact of Artificial Intelligence may obfuscate perspectives of AI researchers in low- and middle-income countries. If a qualitative study was addressing the impacts of AI within a specific field - education, healthcare, etc. - it was not included in our study. This may reflect the focus of AI practitioners in these countries on resolving specific needs, rather than focusing on more general topics in AI. For example, we didn't identify any qualitative studies on general impact from many LMICs in several regions including Central and South America or many Island Nations. That does not mean there are none, but we did not identify them with our methodology. Second, our requirement of English language papers prevents us from accessing research that is conducted and disseminated nationally or regionally. This is something to consider for future research. Finally, geographic regions are not cultural monoliths. The focus in this paper on hegemony as it relates to geographic and economic position, without including other intersectional variables such as race, gender or class, reduces the dimensions of what is meant by cultural hegemony and their addition would enhance future research.

Conclusion

This study suggests that the culture of AI has the potential to exclude, and more must be done to understand and include the perspectives from those more likely to receive downstream effects from both bad AI technology and transformative effects. This is particularly true given that such a large majority of the globe are not living in high-income, AI-intensive nations. Normative theorising and deterministic claims that dominate much of our AI research will not explain the interactions between the majority of the world and AI.

In the interim, slowing or resisting AI is a potential mitigation strategy. In particular, one should be able to question the mainstream narratives around inevitability, as was demonstrated in the background studies by Fjeld et al. (2020) and Franzke (2022), as well as the qualitative study by Hautala and Ahlqvist (2022). This has been critiqued by McQuillan (2022) and others as a serious oversight in the potential to limit harm. While we do not have many examples from low- and middle-income countries, the studies in this paper that highlighted the interference of wealthier countries, like Fejerskov (2017), Kak (2020) and Heng et al. (2022) illustrate how the inability to resist (or the lack of attention to possibilities of resistance) may influence countries that are already marginalised on the global scale. Our societies and our technologies are intertwined through complex matrices of power, politics, and other interactions (Kling 2000; Jasanoff and Kim 2015; Birhane et al. 2022).

In the current environment, AI practitioners working on the front lines of our most intoxicating technology should apply a more critical lens. National and international institutions are not always inclined to produce strong regulatory environments that would make legal alignment a good strategy. AI for social good requires greater effort for topic and geographic coverage, as well as stakeholder involvement, to avoid reflecting global power structures. To understand AI's impacts, including both epistemic and normative concerns, we require greater visibility of those most marginalised in the discourse and a reconsideration of how our understanding of AI's impact can be an extension of cultural hegemony.

Acknowledgments

This work was supported by a UKRI Future Leaders Fellowship [grant number MR/W011336/1].

Research Ethics and Social Impact Statement

Ethical Considerations

There are ethical considerations in citing research where the details of methodologies were not always available. We were unsure of how to handle this, as qualitative research in the computing field often omits practices that are more common in the Humanities and Social Sciences, such as providing details on ethical approvals and consent forms. We hope that principles of academic integrity and institutional requirements were sufficient to address this concern, but it is not certain. This study is limited by the inclusion of studies in English only and the number of papers that were included, from which a complete picture of the landscape cannot be distilled. Our results provide further context and detail to statements that have already been expressed across the literature, and we attempt not to exceed this in our analysis.

Researcher Positionality

Author 1's research positionality aligns with non-positivist frameworks, reflecting how I design research, analyse and conceptualise data.

Author 2's positionality is impacted by western imperialism and economic hardship, fostering an anti-war stance and awareness of minority struggles.
Author 3 is a White, queer, neurodiverse, transgender woman, with a middle-class background from a colonising country. My research perspectives are shaped by my experiences and reflections on privilege and marginalisation, and their perceptions, as well as the interdisciplinary nature of my development as an academic.
Author 4 is a White, queer, cisgender woman, with a middle-class background, who was raised in the United States (a colonising country operating on stolen land). My privileged experiences as a White immigrant, and the solidarity and anti-capitalist movements I encountered in my second country of residence, impact how I look at research, my responsibilities as an academic, and my theoretical and analytical choices.

Adverse Impacts

This research, in setting out a conclusion that there is a representational gap, creates a responsibility to support efforts to re-define, shape and fill that gap. However, we should be cautious to think that we, as researchers operating from wealthy, AI advanced countries, have the competency to fully understand and interpret perspectives that are coming from contexts we do not understand. When engaging in such research in the future, we must ensure that our participants have full autonomy over their data, our interpretations and conclusions.

References

Act, A. I. 2021. Proposal for a regulation of the European Parliament and the Council laying down harmonised rules on Artificial Intelligence (Artificial Intelligence Act) and amending certain Union legislative acts. *EUR-Lex-52021PC0206*.

Ala-Pietilä, P.; Bonnet, Y.; Bergmann, U.; Bielikova, M.; Bonefeld-Dahl, C.; Bauer, W.; Bouarfa, L.; Chatila, R.; Coeckelbergh, M.; Dignum, V.; et al. 2020. *The assessment list for trustworthy artificial intelligence (ALTAI)*. European Commission.

Belfield, H. 2020. Activism by the AI community: Analysing recent achievements and future prospects. In *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society*, 15–21.

Benjamin, R. 2023. Race after technology. In *Social Theory Re-Wired*, 405–415. Routledge.

Bingley, W. J.; Curtis, C.; Lockey, S.; Bialkowski, A.; Gillespie, N.; Haslam, S. A.; Ko, R. K.; Steffens, N.; Wiles, J.; and Worthy, P. 2023. Where is the human in human-centered AI? Insights from developer priorities and user experiences. *Computers in Human Behavior*, 141: 107617.

Birhane, A.; Kalluri, P.; Card, D.; Agnew, W.; Dotan, R.; and Bao, M. 2022. The values encoded in machine learning research. In *Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency*, 173–184.

Bloomfield, B. P. 2018. *The question of artificial intelligence: Philosophical and sociological perspectives*. Routledge.

Buolamwini, J. A. 2017. *Gender shades: intersectional phenotypic and demographic evaluation of face datasets and gender classifiers*. Ph.D. thesis, Massachusetts Institute of Technology.

Cave, S.; and Dihal, K. 2019. Hopes and fears for intelligent machines in fiction and reality. *Nature machine intelligence*, 1(2): 74–78.

Ciecierski-Holmes, T.; Singh, R.; Axt, M.; Brenner, S.; and Barteit, S. 2022. Artificial intelligence for strengthening healthcare systems in low-and middle-income countries: a systematic scoping review. *npj Digital Medicine*, 5(1): 162.

Cowls, J.; Tsamados, A.; Taddeo, M.; and Floridi, L. 2021. A definition, benchmark and database of AI for social good initiatives. *Nature Machine Intelligence*, 3(2): 111–115.

Crawford, K. 2021. *The atlas of AI: Power, politics, and the planetary costs of artificial intelligence*. Yale University Press.

Dahab, G. O. 2019. *he weaponization of artificial intelligence (AI) and its implications on the security dilemma between states: could it create a situation similar to "mutually assured destruction" (MAD)*. Ph.D. thesis, American University in Cairo, AUC Knowledge Fountain.

Dignum, V. 2019. *Responsible artificial intelligence: how to develop and use AI in a responsible way*, volume 2156. Springer.

ESCAP, U.; et al. 2020. The AI for Social Good Summit. Technical report, Association of Pacific Rim Universities.

Fejerskov, A. M. 2017. The new technopolitics of development and the global south as a laboratory of technological experimentation. *Science, Technology, & Human Values*, 42(5): 947–968.

Fjeld, J.; Achten, N.; Hilligoss, H.; Nagy, A.; and Srikumar, M. 2020. Principled artificial intelligence: Mapping consensus in ethical and rights-based approaches to principles for AI. *Berkman Klein Center Research Publication*, 3518482(2020-1).

Floridi, L.; Cowls, J.; King, T. C.; and Taddeo, M. 2021. How to design AI for social good: seven essential factors. *Ethics, Governance, and Policies in Artificial Intelligence*, 125–151.

Franzke, A. S. 2022. An exploratory qualitative analysis of AI ethics guidelines. *Journal of Information, Communication and Ethics in Society*, 20(4): 401–423.

Goffi, E. R. 2023. Teaching Ethics Applied to AI from a Cultural Standpoint: What African "AI Ethics" for Africa? In *AI Ethics in Higher Education: Insights from Africa and Beyond*, 13–26. Springer International Publishing Cham.

Hagendorff, T. 2020. The ethics of AI ethics: An evaluation of guidelines. *Minds and machines*, 30(1): 99–120.

Hagerty, A.; and Rubinov, I. 2019. Global AI ethics: a review of the social impacts and ethical implications of artificial intelligence. *arXiv preprint arXiv:1907.07892*.

Hassan, Y. 2023. Governing algorithms from the South: a case study of AI development in Africa. *AI & SOCIETY*, 38(4): 1429–1442.

Hautala, J.; and Ahlqvist, T. 2022. Integrating futures imaginaries, expectations and anticipatory practices: practitioners of artificial intelligence between now and future. *Technology Analysis & Strategic Management*, 1–13.

Heng, S.; Tsilionis, K.; Scharff, C.; and Wautelet, Y. 2022. Understanding AI ecosystems in the Global South: The cases of Senegal and Cambodia. *International Journal of Information Management*, 64: 102454.

Holton, R.; and Boyd, R. 2021. 'Where are the people? What are they doing? Why are they doing it?'(Mindell) Situating artificial intelligence within a socio-technical framework. *Journal of Sociology*, 57(2): 179–195.

Huszár, F.; Ktena, S. I.; O'Brien, C.; Belli, L.; Schlaikjer, A.; and Hardt, M. 2022. Algorithmic amplification of politics on Twitter. *Proceedings of the National Academy of Sciences*, 119(1): e2025334119.

Jasanoff, S.; and Kim, S.-H. 2015. *Dreamscapes of modernity: Sociotechnical imaginaries and the fabrication of power*. University of Chicago Press.

Jobin, A.; Ienca, M.; and Vayena, E. 2019. The global landscape of AI ethics guidelines. *Nature machine intelligence*, 1(9): 389–399.

Jurowetzki, R.; Hain, D.; Mateos-Garcia, J.; and Stathoulopoulos, K. 2021. The Privatization of AI Research (-ers): Causes and Potential Consequences–From university-industry interaction to public research brain-drain? *arXiv preprint arXiv:2102.01648*.

Kak, A. 2020. " The Global South is everywhere, but also always somewhere" National Policy Narratives and AI Justice. In *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society*, 307–312.

Katz, Y. 2020. *Artificial whiteness: Politics and ideology in artificial intelligence*. Columbia University Press.

Kazim, E.; and Koshiyama, A. S. 2021. A high-level overview of AI ethics. *Patterns*, 2(9).

Kelley, S.; Ovchinnikov, A.; Hardoon, D. R.; and Heinrich, A. 2022. Antidiscrimination laws, artificial intelligence, and gender bias: A case study in nonmortgage Fintech lending. *Manufacturing & Service Operations Management*, 24(6): 3039–3059.

Kevin, M.; Mark, R.; and Bernd, S. 2019. Understanding Ethics and Human Rights in Smart Information Systems: A Multi Case Study Approach. *The ORBIT Journal*, 2(2): 1–34.

Kling, R. 2000. Learning about information technologies and social change: The contribution of social informatics. *The information society*, 16(3): 217–232.

Kwarteng, J.; Perfumi, S. C.; Farrell, T.; and Fernandez, M. 2021. Misogynoir: public online response towards self-reported misogynoir. In *Proceedings of the 2021 IEEE/ACM international conference on advances in social networks analysis and mining*, 228–235.

Laux, J.; Wachter, S.; and Mittelstadt, B. 2024. Trustworthy artificial intelligence and the European Union AI act: On the conflation of trustworthiness and acceptability of risk. *Regulation & Governance*, 18(1): 3–32.

Madaio, M.; Egede, L.; Subramonyam, H.; Wortman Vaughan, J.; and Wallach, H. 2022. Assessing the Fairness of AI Systems: AI Practitioners' Processes, Challenges, and Needs for Support. *Proceedings of the ACM on Human-Computer Interaction*, 6(CSCW1): 1–26.

McQuillan, D. 2022. *Resisting AI: an anti-fascist approach to artificial intelligence*. Policy Press.

Meyerson, D. E.; and Scully, M. A. 1995. Crossroads tempered radicalism and the politics of ambivalence and change. *Organization Science*, 6(5): 585–600.

Mindell, D. A. 2015. *Our robots, ourselves: Robotics and the myths of autonomy*. Viking.

Mittelstadt, B. D.; Allo, P.; Taddeo, M.; Wachter, S.; and Floridi, L. 2016. The ethics of algorithms: Mapping the debate. *Big Data & Society*, 3(2): 2053951716679679.

Monasterio Astobiza, A.; Ausín, T.; Liedo, B.; Toboso, M.; Aparicio, M.; and López, D. 2022. Ethical Governance of AI in the Global South: A Human Rights Approach to Responsible Use of AI. *Proceedings*, 81(1): 136.

Morley, J.; Floridi, L.; Kinsey, L.; and Elhalal, A. 2020. From what to how: an initial review of publicly available AI ethics tools, methods and research to translate principles into practices. *Science and engineering ethics*, 26(4): 2141–2168.

Nakatumba-Nabende, J.; Suuna, C.; and Bainomugisha, E. 2023. AI Ethics in Higher Education: Research Experiences from Practical Development and Deployment of AI Systems. In *AI Ethics in Higher Education: Insights from Africa and Beyond*, 39–55. Springer International Publishing Cham.

Okengwu, U. A. 2023. Practical Implications of Different Theoretical Approaches to AI Ethics. In *AI Ethics in Higher Education: Insights from Africa and Beyond*, 27–35. Springer International Publishing Cham.

Orr, W.; and Davis, J. L. 2020. Attributions of ethical responsibility by Artificial Intelligence practitioners. *Information, Communication & Society*, 23(5): 719–735.

Petermann, M.; Tempini, N.; Kherroubi Garcia, I.; Whitaker, K.; and Strait, A. 2022. Looking before we leap: Expanding ethical review processes for AI and data science research.

Ricaurte, P. 2022. Ethics for the majority world: AI and the question of violence at scale. *Media, Culture & Society*, 44(4): 726–745.

Ryan, M.; Christodoulou, E.; Antoniou, J.; and Iordanou, K. 2022. An AI ethics 'David and Goliath': value conflicts between large tech companies and their employees. *AI & SOCIETY*, 1–16.

Sambasivan, N.; Arnesen, E.; Hutchinson, B.; Doshi, T.; and Prabhakaran, V. 2021. Re-imagining algorithmic fairness in india and beyond. In *Proceedings of the 2021 ACM conference on fairness, accountability, and transparency*, 315–328.

Schiff, D.; Biddle, J.; Borenstein, J.; and Laas, K. 2020. What's next for ai ethics, policy, and governance? a global overview. In *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society*, 153–158.

Shi, Z. R.; Wang, C.; and Fang, F. 2020. Artificial intelligence for social good: A survey. *arXiv preprint arXiv:2001.01818*.

Slota, S. C.; Fleischmann, K. R.; Greenberg, S.; Verma, N.; Cummings, B.; Li, L.; and Shenefiel, C. 2021. Many hands make many fingers to point: challenges in creating accountable AI. *AI & SOCIETY*, 1–13.

Slota, S. C.; Fleischmann, K. R.; Greenberg, S.; Verma, N.; Cummings, B.; Li, L.; and Shenefiel, C. 2023. Locating the work of artificial intelligence ethics. *Journal of the Association for Information Science and Technology*, 74(3): 311–322.

Stahl, B. C.; Antoniou, J.; Ryan, M.; Macnish, K.; and Jiya, T. 2022. Organisational responses to the ethical issues of artificial intelligence. *AI & SOCIETY*, 37(1): 23–37.

Stilgoe, J. 2020. Who's driving innovation. *New Technologies and the Collaborative State. Cham, Switzerland: Palgrave Macmillan*.

Tomašev, N.; Cornebise, J.; Hutter, F.; Mohamed, S.; Picciariello, A.; Connelly, B.; Belgrave, D. C.; Ezer, D.; Haert, F. C. v. d.; Mugisha, F.; et al. 2020. AI for social good: unlocking the opportunity for positive impact. *Nature Communications*, 11(1): 2468.

Tsamados, A.; Aggarwal, N.; Cowls, J.; Morley, J.; Roberts, H.; Taddeo, M.; and Floridi, L. 2021. The ethics of algorithms: key problems and solutions. *Ethics, Governance, and Policies in Artificial Intelligence*, 97–123.

Vaast, E. 2022. Future imperfect: How AI developers imagine the future. In *Proceedings of the International Conference on Information Systems (ICIS)*. Copenhagen, Denmark.

Veale, M.; and Zuiderveen Borgesius, F. 2021. Demystifying the Draft EU Artificial Intelligence Act—Analysing the good, the bad, and the unclear elements of the proposed approach. *Computer Law Review International*, 22(4): 97–112.

Wakunuma, K.; Castro, F. d.; Jiya, T.; Inigo, E. A.; Blok, V.; and Bryce, V. 2021. Reconceptualising responsible research and innovation from a Global South perspective. *Journal of Responsible Innovation*, 8(2): 267–291.

Wakunuma, K.; Jiya, T.; and Aliyu, S. 2020. Socio-ethical implications of using AI in accelerating SDG3 in Least Developed Countries. *Journal of Responsible Technology*, 4: 100006.

Yeh, S.-C.; Wu, A.-W.; Yu, H.-C.; Wu, H. C.; Kuo, Y.-P.; and Chen, P.-X. 2021. Public perception of artificial intelligence and its connections to the sustainable development goals. *Sustainability*, 13(16): 9165.