



**HAL**  
open science

## The (Un)bearable Whiteness of AI Ethics

Syed Mustafa Ali, Angela C. Daly, Adela Gjorgjioska, Luke Hespanhol,  
Xaroula Kerasidou, Soraya Kouadri Mostéfaoui, Oluyinka Oyeniji, Beata  
Paragi, Ana Tomičić

► **To cite this version:**

Syed Mustafa Ali, Angela C. Daly, Adela Gjorgjioska, Luke Hespanhol, Xaroula Kerasidou, et al..  
The (Un)bearable Whiteness of AI Ethics. Handbook on the Ethics of Artificial Intelligence, Edward  
Elgar Publishing, pp.217-230, 2024, 9781803926711. 10.4337/9781803926728.00020 . hal-04703344

**HAL Id: hal-04703344**

**<https://hal.science/hal-04703344v1>**

Submitted on 20 Sep 2024

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# The (Un)bearable Whiteness of AI Ethics

Ali, Syed Mustafa; Daly, Angela C.; Gjorgjioska, Adela; Hespanhol, Luke; Kerasidou, Xaroula; Kouadri Mostéfaoui, Soraya; Oyeniji, Oluyinka, Paragi, Beata and Tomičić, Ana (2024). The (Un)bearable Whiteness of AI Ethics. In: Gunkel, David J. ed. *Handbook on the Ethics of Artificial Intelligence*. Cheltenham, UK: Edward Elgar Publishing Ltd., pp. 217–230.

DOI: <https://doi.org/10.4337/9781803926728>

## 1. INTRODUCTION

Advances in the field of AI have been accompanied by the production of numerous guidelines, policy documents, and ethical frameworks with the aim of steering and regulating the development of the technology in a range of application domains. While the term AI ethics<sup>2</sup> tends to refer to collections of (usually non-binding) norms ‘issued mainly by governments and corporations’ (Daly, Devitt and Mann 2021), ‘ethics’ has a broader meaning relating to matters of morality and virtue (Bietti, 2020). That said, it is the former, ‘narrower’ meaning that has come to dominate the rhetoric of AI ethics, rather than the latter, more expansive meaning.

The numerous AI ethics initiatives launched in the past decade or so have been subject to criticism on various grounds (Kerasidou et al 2022, Mittelstadt 2019, Williams 2019). Commentators such as Wagner (2018) view the field as tantamount to a form of ‘ethics washing’ insofar as most AI ethics initiatives lack the binding force of law, divert attention from enacting binding law that would regulate AI, and do not actually prevent the companies and governments that create AI ethics frameworks from engaging in unethical conduct.<sup>3</sup> Others highlight that demanding the binding force of law for AI ethics – that is, simply making the norms and principles binding via law – is not sufficient to ensure ‘good’ processes and outcomes; as they point out, attending to the substance of the norms and principles is essential since a failure to do so can result in ‘law washing’ (Daly, Devitt and Mann, 2021).<sup>4</sup> For reasons such as these, Munn (2023) concludes that AI ethics principles are useless insofar as they are (1) meaningless (since contested and/or incoherent), (2) isolated (since articulated within contexts that largely ignore ethical concerns), and (3) toothless (because they are inconsequential). In relation to the latter, Munn points to the entanglement of AI ethics and corporate agendas, drawing on accounts setting out the corporate-sponsored genealogy of the field (Ochigame, 2019).

While endorsing such accounts, we take the view that AI ethics should be understood in relation to both governments/states and corporations/markets – that is, in political economic terms – and that the local-global relationships between these entities constitute a world system. We contend that appreciating the entanglement of whiteness with AI ethics in world-systemic terms is crucial since it allows the discourse to be interrogated – race critically, political-economically, ecologically etc. – in a manner that attends to structural asymmetries of power.

## 2. AI ETHICS: (UN)BEARABILITY

As might be apparent to some readers, the title of this contribution riffs off *The Unbearable Lightness of Being* (1984) by the late Czech writer Milan Kundera, essentially for stylistic

purposes – ‘whiteness’ rhyming with ‘lightness’ – rather than driven by any serious concern to engage with the Nietzschean sense of selfhood explored in that text. Crucially, this strategy has been used by previous commentators for purposes of doing race critical work: in this connection, consider the title of an essay by Adams (2015) exploring the relationship between feelings of superiority, White privilege, White guilt, and a denied White racial identity in a therapeutic context; and the editorial for an issue of the literary journal *Wasafiri* calling attention to ‘the unbearable lightness of whiteness’, thereby bringing the two notions into explicit contact (McIntosh 2020).

Beyond such stylistic deployments, and notwithstanding the risk of a certain essentialism, it is worthwhile briefly pausing to consider the meaning of the adjective “bearable” contained in some standard works of reference. For present purposes, it should suffice to consider the entry for this word appearing in a few dictionaries of the English language. According to the Merriam-Webster Dictionary, for example, something is *bearable* if it is ‘capable of being borne’, synonyms including *endurable*, *sufferable*, *supportable*, *sustainable*, *tolerable*. The Cambridge Dictionary refers to an unpleasant situation as being bearable if it can be accepted or dealt with. Usefully, the Oxford English Dictionary provides an entry for the derived noun “bearability” which refers to (1a) the ability to endure or tolerate something; and (1b) the quality or fact of being *endurable* or *tolerable*.

Invoking “bearability” in relation to whiteness, yet complicating the former by attaching *un-*, the prefix of reversal, deprivation, negation and/or removal, in brackets, viz. (un), rather than ‘crossed through’, i.e. *un* – the latter constituting an instance of what the existential phenomenologist and philosopher Martin Heidegger referred to in his later work as “writing under erasure” – points to the ambiguity, if not outright undecidability, of the bearability of whiteness by those forced to bear it<sup>5</sup>. What is meant by this is that although whiteness is unbearable – that is, an intolerable situation – it is nonetheless bearable. That is, it is assumed to be capable of being accepted, albeit grudgingly as an existential fact, and/or dealt with as such. Yet the acceptability of the unbearable is clearly not the same as the ability to deal with the unbearable, a difference that makes all the difference when it comes to thinking about compliance with and/or opposition to ethical principles – *a fortiori* those that are racially marked in some way.

But who are those forced to bear whiteness? Are they exclusively those on its receiving end, with whiteness understood, at least at the outset, as ‘an unpleasant situation’? In short, does this category refer only to those marked, in some way, as not-white/non-white, etc.? Or might it also include those marked as white, both in some hegemonic sense but also as extended to others, viz. the near-/off-/proximally-white?

In an attempt at answering these questions, recourse is usefully made to Charles W. Mills’ (1997) distinction between *signatories* and *beneficiaries* of what he refers to as ‘the Racial Contract’, a methodological construct for thinking about the establishment of racialised polity in contractarian terms. According to Mills, those white people responsible for establishing white supremacy – which for him means a global political order in which whiteness is dominant/hegemonic – are signatories to this ‘contract’; by contrast, those white people born into the world that came into being following the establishment of the racial contract are, by default, its beneficiaries yet they can, if they so choose, disavow it by turning ‘race traitor’.

On this basis, it might be argued that while all non-white people are forced to bear whiteness, and while recognising that some non-white people might willingly accept and comply with its dictates, those white people who commit to race treachery are also forced to bear whiteness, but in a different manner, viz. as actively working against whiteness from within it and from a differential position of power.

Is it all about power then? And what does this mean for AI ethics?

### 3. AI ETHICS: POWER

Although there is clearly a need to unpack what is meant by ‘whiteness’ in the above exploration of differential bearability – something that will be attempted shortly – for the purposes of addressing the principle concerns of this collection of essays, I think we should briefly consider the implications of understanding AI ethics – and perhaps AI itself – as white. In terms of the above, it would mean that all non-white people are forced to deal with this ethical terrain, with some of them choosing to embrace it; by contrast, all white people are its de facto beneficiaries, with some of them choosing to refuse it.

In terms of power, there is a need to consider how it plays out in various registers including differential power among differentially racialised (and classed, gendered, geographically-located etc.) human actors and their differential entanglements with technological infrastructure. As a technology, AI requires large amounts of computational power and associated material and immaterial resources, including human labour, to support it. Not only that, but also a skilled labour force – with advanced mathematical and analytical skills too – at least in the design phase. Such demands lead to migration and brain-drain with the active participation of non-white individuals in strengthening white structures. This places AI development and deployment mainly in the hands of those who are already rich and powerful, i.e. large corporations and governments, and has led to the stockpiling of expertise and funding by these actors too. These resources and expertise in AI technology also enable such actors to assume the lead in developing AI ethics initiatives, i.e. crafting technology policy while shaping the development of the technology in ways that entrench their political and economic power and dominance.

That’s a good point. Many of the AI ethics initiatives have been introduced and developed by existing powerful actors such as governments and transnational corporations. Western governments, both in the US and Europe, as well as US-headquartered corporations have been extremely active in this domain. At the same time, there has been limited civil society participation, and even more limited “bottom up” ethics initiatives from people and communities throughout the world. Put simply, AI ethics tends to be an elite endeavour of, by, and for the privileged few, marked by a lack of diversity among those who are involved in shaping the field, and generally “framed by means of Western values, contexts, and concerns” (Hagendorff, 2021). In that sense, can – *should* – we understand AI ethics as a fundamentally “white” endeavour; one that hides its spatial and cultural specificities behind universalising mottos that invite us to imagine “AI for Good” or “AI for Humanity?” If so, one needs to then ask, what is this ‘humanity’ that is being invoked here? Who gets to define it? And does everyone have an equal stake in it?

Does it all boil down to matters of representation? And inclusivity? And diversity? Thinking along those lines, would a Black AI ethics, or an African AI ethics, for example, provide means by which to address the ‘problem’ of whiteness?

#### **4. AI ETHICS AND WHITENESS: REPRESENTATION, INCLUSION, DIVERSITY**

I don't think it's as simple as that. Different countries in Africa have conceptualised emerging technologies as potential tools for economic development. In South Africa, the Fourth Industrial Revolution has been identified as comprising AI, drone technology, etc. As a continent, the African Union has also established a framework for exploiting the benefits of the technology. Agenda 2063, and the Digitisation programme set to be fully achieved in 2030, are some other ongoing initiatives aimed at boosting capacity for technological advancement on the continent. In the context of AI, the 52 countries in Africa are at different stages of developing a framework for use and deployment: For example, Nigeria is in the process of framing a national policy which places a premium on the ethics of AI, and Egypt, Algeria, and other North African countries have also incorporated ethics within their AI policies and strategies as fitting for their societies. It is a Herculean, perhaps even futile, task to attempt to determine a uniform interpretation of what AI ethics might mean across the continent, let alone the global stage. Africa is known for the diversity of its political and economic positions, religious traditions, and cultures; for example, Nigeria alone has over 250 ethnic groups. What this means is that various factors may determine how different countries on the continent relate to AI ethics as being shaped largely by narratives developed in the Global North. Such factors include the differential impacts of colonialism, how AI is defined and understood, the level of technological advancement in each country and regional blocs, economic disparities and social inequalities, etc. In referring to "the whiteness of AI ethics", it is important not to understand whiteness here as something that can be reduced to a physical marker of the human body such as skin colour. This is because AI in an African context requires us to consider issues beyond embodiment including AI innovation, integration in traditional societies, the differential role and power of stakeholders, technology deployment, impact, etc.

Okay, if whiteness is not about colour, then how should it be understood?

#### **5. WHITENESS**

In terms of understanding how we think about whiteness, one useful point of departure is the sociological exploration of the phenomenon developed by Garner (2007), specifically, his (1) processual understanding of whiteness as a socially-constructed phenomenon existing in dynamic, relational tension to other racialized identities, (2) the socio-political structural manifestation of whiteness as a persistent, yet contested, globally-systemic political structure, viz. white supremacy, a position he derives from Mills (1997), and (3) the function of whiteness as a tacit invisible background standard. Regarding (1), it is crucial to appreciate that whiteness and white identities are historically-contingent formations, being the temporary (albeit currently enduring) out-

comes of racialisation processes that play out differently at different times and in different places. Although the link between whiteness and European – and here ‘European’ means “of European descent” – has come to be somewhat sedimented, this connection is, again, historically-contingent and socially-constructed rather than necessary, essential, or natural. Similarly, the centre or ‘core’ of whiteness is contingently located in ‘the West’ within the Global North – specifically in the North-Western countries within Europe, along with the US, Canada, Australia, and New Zealand – whiteness in other parts of Europe (Southern, Eastern) being somewhat more contested – hence, categories such as ‘off white’, proximal white etc. (Mills, 1997) In addition, whiteness is only contingently associated with bodily markers such as skin colour, hair texture, nose shape etc; as Hesse (2007) argues, “racialization [is] embodied in a series of onto-colonial taxonomies of land, climate, history, bodies, customs, language, all of which became sedimented metonymically, metaphorically, and normatively, as the assembled attributions of race” (2007, pp. 658–659). Regarding (2), the diffusion of whiteness around the world was largely a result of the economic and political power gained by Europeans through colonialism and capitalism, both phenomena being thoroughly entangled with Eurocentric processes of racialisation. Understanding such entanglements is crucial because although engaging with racism (Cole, 2020) and racial capitalism (Robinson, 1983) necessitates going beyond markers such as skin colour to consider how race is produced and used to justify the exploitation of certain classes by those in power, race is not reducible to class. Furthermore, attending to the *specificity* of whiteness – more specifically, white supremacy – rather than a more general *racism* is necessary in terms of identifying – that is, *naming* – historically dominant actors within a racialised world system; as Mills (2009) states, one of the “great virtue[s] of ‘white supremacy’ as a term is that it puts front and center the crucial reality of white *domination* and white *privilege* in a way that ‘racialized capitalism’ does not. Both nationally within the US (and many other nations), and globally in terms of European and Euro-settler domination of the world, ‘white supremacy’ directs us to the fact of the national and planetary privileging of whites vis-à-vis non-whites. ‘Racialized capitalism’ as a term is evasive in its failure to name the race that is (generally) created as superior, and fails to register the objectivity, the social existence, of this race, subsuming it instead to the ‘classes’ of capitalism.” (pp. 274–275). Regarding (3), it should be noted that Garner (2007) invites us to think about whiteness as a set of norms, values, and as cultural constructions which is useful since a concern with standards, norms, and values points us in the direction of ethics. In fact, many scholars argue that the process of AI formation has been led primarily by white people in the Global North, and the substance of AI ethics largely reflects Western ideas of ethics and governance (Benjamin, 2019).

## 6. AI, CULTURE, AND ‘SOFT’ COLONIALISM

Indeed. AI ethics has largely focused on bias and discrimination, yet these deep-seated cultural and political problems are often treated as technical matters with technical solutions, thereby adopting a tech solutionist stance

which overlooks the root causes of the problems with AI systems. I would say that this point underscores the problems with ethical frameworks focusing predominantly on pragmatic, operational aspects of AI. As Rizk (2020) points out, “within each of AI’s components – the data, the algorithm, and the infrastructure – lies a trigger for potential inequalities” (p. 626), and therefore biases, abuse of power, and the perpetuation of entrenched inequalities. Yet, focusing exclusively on these problems risks neglecting deeper ethical concerns related to cultural transformation prompted by widespread AI adoption and societal overreliance on it, not to mention further entrenching structural asymmetries of power.

While contemporary AI ethics discourse focuses on addressing the seemingly ‘hard’ problems of data colonialism, algorithmic bias, and a lack of infrastructural equity, it often fails to address the ‘soft’ colonisation of culture through the rhetorical normalisation of AI adoption as both desirable and useful. For example, the increasing appetite for incorporating generative AI technologies such as DALL-E or ChatGPT in the design process as creative assistants risks mixing value systems and ethical frameworks of different stakeholders including those who create AI technologies and the designers using them. The emerging professional skill of ‘prompt engineering’ – put simply, learning how to coax generative AI systems to obtain certain kinds of results – exemplifies this new power conundrum. On the one hand, when using AI as a creative partner, designers may embed algorithm-driven design inputs (and values) into the final product, uncritically orienting it toward biased outcomes informed by their cultural lens that become incorporated into the designed artefact. On the other hand, the role of prompt engineers might undergo transformation to include gatekeeping of ethical decisions informed by local practices and cultural mores, thereby minimising the risk of producing culturally inappropriate outcomes. Crucially, these developments occur independently from any measures, ethical or otherwise, adopted in relation to the sourcing of data, training of models, and rollout and deployment of the AI system. For my part, this speaks to the need for adopting culturally appropriate ethical frameworks, developed in a bottom-up fashion, used to guide AI development in accordance with local epistemologies and hermeneutics.

## **7. THE WHITENESS OF AI ETHICS**

I would agree with this. The discourse of AI ethics embodies and embeds whiteness in both its form and substance, since the process of AI ethics formation has largely (although not exclusively) been led by white people in the Global North. The substance of AI ethics norms also largely reflects Western ideas of morality and governance. In fact, what is meant by ‘AI’ and ‘ethics’ can also be viewed as embodying and embedding whiteness since other, non-white worldviews and ways of being may not separate ‘ethics’ from other concerns such as innovation, creation etc. In the present context, this means that AI ethics might not be so easily separated or rendered separable from the conception and development of AI itself; consider, for example, how the idea of AI is often conceptualised in relation to human intelligence in contrast to alternative schemes engaging more expansive relational worldviews incorporating non-

/other than human intelligences; in short, as Kalulé and Joque (2019) and others have suggested, ‘intelligence’ needs to be subjected to critique. Initiatives such as Ubuntu-AI, Indigenous-AI, queer-AI are arguably responding to this line of argument.

## **8. AI ELSEWHERE AND OTHERWISE**

Interesting and important as such initiatives are, they remain unconvincing, at least for me. Paradigms for use and dissemination of technology have overwhelmingly been skewed towards favouring big American tech corporations, although this position has been contested by Couldry and Mejias who argue that “data colonialism involves not one pole of colonial power (‘the West’), but at least two: the United States and China” (Couldry and Mejias, 2019, p. 337), and Gravett (2020) who raises concerns about the imposition by default of the ‘China model’ of digital governance in Africa as a form of digital neo-colonialism. However, Kwet (2021) dismisses such ‘New Cold War’ narratives, pointing out that China’s technological dominance is largely localised, and still trails behind that of Japan, South Korea, and Taiwan in certain areas of technology development. In this connection, it is crucial to note that China’s technology products remain substantially dependent on American intellectual property (computer chips being one example), which suggests the US continue to occupy the position of hegemon in the Big Tech arena, thereby dominating how the discourse around AI and AI Ethics is shaped and dictated. However, that could all change since China’s large technology companies, which are increasingly transnational, and the Chinese government have led on local AI ethics initiatives, and in some instances, have participated in the formation of global AI ethics policy (Daly et al 2021). Crucially, there has been a significant emphasis on cases involving the use of Chinese AI in trials conducted in Xinjiang/East Turkestan. Attention to these events has stimulated discussion surrounding the potential ramifications of such practices and their implications in terms of (digital) colonialist dynamics (Anand 2019; Daly, 2019). Developments such as these have generated concerns about the export of Chinese AI alongside (neo)colonial Western digital technology to the African continent (Birhane 2020).

Approaches to AI outside the West vary widely. In East Asia, generally, AI policies are determined in a top-down manner, as a vehicle for societal prosperity (Gal, 2019). In China, AI is perceived as part of nation building incorporating “four unique Chinese characteristics in comparison with Western engineering ethical guidelines: responsibility precedes freedom, obligation precedes rights, the group precedes the individual, and harmony precedes conflict” (Gal, 2019). Japan and South Korea focus on AI as tools to assist with providing services in society. Japanese contemporary culture, however, has long entertained the possibility of AI as companions, relationships and ethical boundaries being negotiated in bottom-up fashion with little discussion around the potential real ethical consequences of urban social robots if effectively rolled out (Jecker and Nakazawa, 2022).



Top-down approaches have also been adopted in the Middle East and North Africa (MENA) region, as well as in Latin America. The MENA region is very heterogeneous, with some countries resource rich, yet having a small local labour force, while others are resource poor but with a large labour force. For the former, AI is considered strategic in terms of nation building, offering an opportunity to reduce reliance on foreign workers. For the latter, the prospect of implementing AI would require upskilling of the local population. The MENA region is also marred by a high degree of inequality, both socio-economically and gender-based (Rizk, 2020). The situation is similar in Latin America, where strategies for AI rollout, and ensuing ethical approaches or lack thereof, vary widely according to the political orientation of countries. For example, in countries with more pluralist governments such as Chile, Argentina, and Uruguay, bargaining and negotiation is favoured across sectors within civil society, with a greater commitment to institutions. By contrast, more populist and often authoritarian governments such as those in Brazil, Colombia, and Mexico have tended to “distrust procedures and produce conflicts that aggravate the disintegration of the social fabric” (Filgueiras, 2023, p.10).

## 9. THE HEGEMONY OF (RACIALISED) NEOLIBERALISM

Okay. But do any of these cases escape the hegemony of neoliberalism and its racialised inflection by whiteness? As a hegemonic system whose moral imperative is to reshape the social world in the image of a narrowly construed market (Zamora and Behrent, 2015), neo-liberalism has resulted in the privatization of public assets and human services, and policies that prioritize corporate interests over societal well-being. In his *Critique of Black Reason*, Cameroonian philosopher Achille Mbembe (2017, p. 3) understands neoliberalism as more narrowly referring to a “phase in the history of humanity dominated by the industries of the Silicon Valley and digital technology.” Yet neoliberalism is not merely a worldview according to which everything in the human life-world can be assigned a market value, but should also be understood in terms of the production of indifference, that is, “the codification of social life according to norms, categories, and numbers” (Mbembe (2017, p. 3). It is crucial to appreciate here that such codification plays out differently among differently marked groups, differently located across the world; bodies with closer proximity to centres of whiteness located in the Global North tend to be subjected to regimes of biopolitical governance, while those at the peripheries of whiteness and located in the Global South tend to be subjected to necropolitics, considered ‘surplus’ and on occasion rendered ‘killable’ (Mbembe 2003). Capital, Mbembe argues, no longer needs workers to function, only data and artificial and digital memory wherein the human is merely “one animal among others, lacking an essence of his own to protect or safeguard” and lacking “any limits placed on the modification of his genetic, biological structure” (Mbembe 2017, pp. 3–4). In the context of AI and AI ethics, such developments can foster a culture where researchers and practitioners are pressured to align their work with the interests of private corporations rather than with human values, broader societal goals and/or public interest

(Moulaert et al., 2013). Consider, in this connection, how neoliberalism has promoted the idea of ‘innovation’ as a key driver of economic growth and development, leading to an emphasis on the development of new AI technologies and a focus on commercial applications of AI at the expense of social costs (Brown, 2015). Neoliberalism is premised on a set of beliefs, values, and ideas that support and maintain the existing social hierarchy, specifically the power and interests of those in positions of control. It is a dominant force that takes precedence over and subordinates ethical considerations, including AI ethics. Or perhaps AI ethics discourses should be seen as generated by and subservient to neoliberalism given their non-binding ‘window dressing’ status perpetuating the laissez faire orientation of unregulated capitalism?

Expanding on some lines of argumentation sketched earlier, I think the points you make take us back to Garner’s exploration of whiteness as a concept – more specifically, to his point about its socio-political structural manifestation. Drawing on Anibal Quijano’s extension to Immanuel Wallerstein’s world-systems theory in which the modern/colonial world system is identified as the market-based system of global capitalism emerging in the long durée of the 16th century CE following the onset of European colonialism (Quijano and Wallerstein, 1992), I would argue that whiteness needs to be framed relative to the systemic and structural backdrop of racialised global political economy, more specifically, to racial capitalism, which in its current form operates as a racialised neoliberalism that continues to be inflected by global white supremacy (Goldberg 2011) (Kundnani 2021). Thinking about whiteness and racial capitalism in terms of the modern/colonial world system requires us to think about centres peripheries (and possibly semi-peripheries), with whiteness positioning itself as and at the centre and non-white others as and at the periphery.

But what does this mean for AI ethics?

## **10. HYPHENATION AND THE WHITENESS OF AI ETHICS**

For a start, I suggest that we should reject the view that AI ethics can be approached as the mere conjunction of AI and ethics with each considered as separate disciplines and/or phenomena. Although AI ethics needs to be understood as a singular phenomenon in its own right, to better understand the relationship between AI ethics and whiteness, it is useful to think about the relationship between whiteness and AI itself. According to Katz (2020), AI serves a dual purpose: politically, it functions as a prosthetic tool in the maintenance of racial social order and the advance of imperialist and capitalist projects; and ideologically, it functions as a site for reinscribing the invisibility of whiteness, invisibility being interpreted along the lines suggested by Garner and others. Building on what has been said above, I would argue analogously to Katz in relation to AI ethics, viz. politically, it is a rhetorically-charged disciplinary discourse with a corporate-sponsored genealogy (Ochigame, 2019) that has emerged within racialised neoliberalism for the purpose of biopolitical (and sometimes necropolitical) governance, maintaining – through principles, values, standards, and regulatory

frameworks – racial social order and the advance of racial capitalism. On this basis, I would suggest that AI ethics, at least in its hegemonic (that is, dominant) form, is white; further, that it functions as both a discursive facilitator and liberal regulatory distractor, enabling yet masking the accelerating rollout of AI as a colonising racialised political-economic socio-technical infrastructure.

If this line of argument is sound, it means that efforts suggesting an alternative bottom-up and ‘pluriversal’ approach to AI ethics forged through the emergence of a shared consensus of subaltern and what might be described as ‘hyphenated’ perspectives (e.g. Ubuntu-AI, Indigenous-AI, queer-AI etc.) must be found wanting. Moreover, I am concerned that such hyphenated forms of AI ethics might amount to little more than regional dialects within a global language (game) established, maintained, expanded, and refined by an unhyphenated white hegemonic power.

Unfortunately, I am inclined to agree. The search for an ethical AI has laid the ground for ‘Other’-ed endeavours such as those seeking a good AI / just AI / trustworthy AI / Indigenous AI / decolonial AI<sup>6</sup> etc. Yet even if taken in good faith, such initiatives also serve an identity-political distractor function under racial neoliberalism insofar as state, corporate, and civil society actors are kept busy. It should be noted that I am not here maintaining the reality of class against the socially-constructed fiction of political identity since I take the view that *all* identities – including class – are constructed; rather, what I am pointing to is how focusing on political identity without considering its entanglement with economy obscures the operation of capital and its relation to hegemony. Going further, I suggest that the mobilisation of the various AI ethics initiatives needs to be understood in relation to other emerging fields within science and technology including synthetic biology and genomics, where a desire to “anticipate and manage societal issues, reassure publics, and maintain political legitimacy” has prompted the institutionalisation and instrumentalization of ‘ethics’ as a subsidiary component of emergent technoscientific programmes (Hilgartner et al, 2017, p. 2). In this way, subaltern AI ethics and other such hyphenated perspectives, while perhaps not sharing much in terms of political orientation or motivation, are collectively put to work in similar ways relative to a tacitly instrumentalist understanding of AI as technology. (I maintain this while recognising that various ‘Other’-ed AI – and AI ethics – initiatives frame AI as a sociotechnical assemblage.) This instrumentalist approach to the technology tends to direct the gaze to its ‘downstream’ uses, implications, and applications rather than focusing critical attention on more ‘upstream’ questions concerning its formation – or rather, ‘figuration’ (Suchman, 2012). A re-focusing on more upstream concerns would challenge not only the power of AI (Kalluri, 2020) and its attendant whiteness, but also the persistent commitment to technoscientific innovation and its identification with social progress and the public good, by foregrounding questions such as whose interests are best served by invoking instrumentalist AI imaginaries, who benefits the most from their deployment, and who might be most harmed (McQuillan, 2022; Crawford, 2021; Benjamin, 2019; Eubanks, 2018; Irani, 2016), while still refusing to take for granted who/what constitutes human/nonhuman/intelligent/machinic/subhuman.

## 11. CONCLUSION

By way of a conclusion, and instead of attempting to summarise the various twists and turns in the above discussion, we return to the issue of (un)bearability with a view to sketching a response to this condition. In preparing the ground for this conclusion, it is important to state that although we have committed to the view that AI ethics is entangled with whiteness and concur with Katz (2020) that AI itself can – perhaps *should* – be seen as a tool of colonialism, imperialism, and capitalism and an amorphous ideology analogous to whiteness, this does not exhaust the possibilities for thinking about AI. Put simply, we maintain that while it might be the case that AI in the form of Big Data-driven ML (Machine Learning) cum DL (Deep Learning) inherently manifests what Dourish and Mainwaring (2012) refer to as a ‘colonising impulse’, and while earlier types of AI have historically (and geographically) originated in highly militarised, imperialist, colonial, and capitalist contexts, this does not preclude the possibility of AI being done ‘otherwise’.

One useful way to begin answering the question ‘what is to be done?’ in terms of doing AI otherwise is by returning to (racialised) political economy and engaging the matter in terms of the Marxist distinction between *forces* of production and *relations* of production. On this basis, it might be argued that reconfiguring the latter – that is, transforming (re-organising, re-aligning) the social relations within the AI sociotechnical assemblage along more equitable class-based lines (deferring to workers) – provides the necessary means by which to overcome the exploitative and extractivist logics associated with AI under industrialised racial capitalism (as neoliberalism). However, we are inclined to hold that while *necessary*, such a shift would be *insufficient* insofar as not engaging the question of which affordances follow from which material configurations of AI technology. It simply does not follow that a transformation in the relations of production entails a transformation in the forces of production including forces marked by a colonising impulse such as those associated with Big Data-driven ML/DL.

Workers of the world can unite – and historically *have* united – in pursuit of utopian projects marked by a technocratic orientation.

If a transformation in social relations is necessary yet insufficient for doing AI otherwise, what might be needed for sufficiency? Put simply, a disavowal of Big Data-driven ML/DL and a commitment to the limited development of localised, small data-driven systems for specific, community-agreed purposes. In fact, more than merely disavowing, we argue for the necessity of preventing the ongoing rollout of AI as a colonising infrastructure, and where such rollout has already taken place, commit to its abolition (that is, rollback). In arguing along these lines, we draw inspiration from the historical actions – specifically, machine breaking – of the Luddite movement that emerged in the early decades of the Industrial Revolution in the North of England (Ali 2019). Contrary to mainstream disinformation and caricature, this movement was not anti-technology *per se*, but opposed to specific sociotechnical assemblages considered oppressive, viz. ‘machinery hurtful to the commonality’. If AI is to be done otherwise via a scaling-down to the local, this is going to necessitate a breaking up of the large-scale Big Data-driven AI assemblage in its contemporary hegemonic form; and insofar as AI ethics is a mask for the whiteness driving this assemblage with an accelerating technological momentum, this mask must be removed. Irrespective of how (un)bearable a task, it must be undertaken.

Resistance might be futile, yet it must be attempted regardless.

## NOTES

<sup>1</sup> The chapter assumes the form of a collective conversation and co-authors are listed alphabetically rather than according to the extent of their individual contributions. It is

worth noting that the first-person plural pronoun 'we' is used in the Introduction to establish a collective voice representing the collaborative effort of the authors in arguing the various positions set out in the chapter. This inclusive 'we' signifies a shared purpose and intention to bring together diverse positionalities in exploring an attempt at decolonising AI ethics. However, in subsequent sections of the chapter, which are intentionally formatted as an articulated dialogue with a view to highlighting different positionalities, the authors make use of both formatting and the first-person singular pronoun 'I' to indicate these different positionalities. It is suggested that this intentional use of 'I' contributes to the richness of the dialogue by explicitly signalling shifts in viewpoints and fostering a deeper understanding of the various stances presented.

<sup>2</sup> The 'turn' to ethics in AI to some extent reflects the problems that AI may (potentially) cause and a need to govern AI to ensure these problems do not arise or are addressed when they do arise. This can be contrasted with what transpired in relation to the (sometimes) predecessor technology of the Internet, which developed in a neoliberal deregulatory environment in the US and other Western countries during the 1980s, only experiencing a strong turn to institutions and governance from the 2000s onwards.

<sup>3</sup> For Tafani (2022), "the enormous financial resources invested in AI ethics research actually fund smokescreens and mirrors and produce a distortion of ethics, reducing it to an empty shell and generating distrust of ethics itself." Insofar as the discourse of AI ethics is "a form of 'tech-nomoral' revolution aimed at exploitation, oppression and centralization of power in the hands of a few", she suggests that "a realistic counternarrative should present AI ethics as a political issue, as a matter of democracy." (p. 15) On her view, "genuine AI ethics are political in nature, and this means calling things by their proper names, remunerating work, recognizing the environmental costs, not over-collecting individual data on the basis of extorted consents, not treating humans as things and therefore not making decisions about their lives based on opaque automated statistics." (p. 16)

<sup>4</sup> Furthermore, and perhaps most problematically for thinking about AI ethics in relation to law and regulation, there is the thorny issue of hegemony to consider, viz. the possibility of powerful state actors within the international system exercising veto power and/or enacting a 'state of exception', suspending the binding force of the law in relation to its own actions; in this connection, see McQuillan (2022) and others.

<sup>5</sup> This strategy was previously deployed by one of the contributors to the present essay in a brief exploration of the entanglements of race, religion, and informational phenomena (Ali 2018).

<sup>6</sup> For early framings of 'decolonial AI', see Mohamed et al. (2020) and Adams (2021).

## REFERENCES

- Adams, D.M. (2015). The Unbearable Lightness of Being White. *Women & Therapy* 38(3–4): 327–340. DOI: 10.1080/02703149.2015.1059215.
- Adams, R. (2021). Can Artificial Intelligence be Decolonized? *Interdisciplinary Science Reviews* 46(1–2): 176–197. DOI: 10.1080/03080188.2020.1840225.
- Ali, S.M. (2019). Fugitive Decolonial Luddism – A Hauntology. *Intercultural Digital Ethics Symposium*, Digital Ethics Lab, Oxford Internet Institute (OII), 10 December 2019, University of Oxford, Oxford. Ali, S.M. (2018). The (Un)bearable Whiteness of Informational Phenomena. *Contemporary Religion in Historical Perspective*, 29 May 2018. Available at: <https://www.open.ac.uk/blogs/religious-studies/?p=729> (accessed 1 September 2023).
- Anand, D. (2019). Colonization with Chinese Characteristics: Politics of (in)security in Xinjiang and Tibet. *Central Asian Survey* 38(1): 129–147. DOI: 10.1080/02634937.2018.1534801.

- Benjamin, R. (2019). Assessing risk, automating racism. *Science* 366(6464): 421–422.
- Bietti, E. (2020). From Ethics Washing to Ethics Bashing: A View on Tech Ethics from Within Moral Philosophy. *Proceedings of ACM FAT\* Conference (FAT\*2020)*. ACM, New York. DOI: <https://doi.org/10.1145/3351095.337286>.
- Birhane, A. (2020). Algorithmic Colonization of Africa. *SCRIPTed* 17(2), 389. Available at: <https://script-ed.org/?p=3888>. DOI: 10.2966/scrip.170220.389 (accessed 1 September 2023).
- Brown, W. (2015). *Undoing the Demos: Neoliberalism's Stealth Revolution*. Cambridge, MA: MIT Press.
- Chakrabarty, D. (2008). *Provincializing Europe: Postcolonial Thought and Historical Difference (New Edition)*. Princeton: Princeton University Press.
- Cole, M. (2020). A Marxist Critique of Sean Walton's Defence of the Critical Race Theory Concept of 'White Supremacy' as Explaining all Forms of Racism, and Some Comments on Critical Race Theory, Black Radical and Socialist futures. *Power and Education* 12(1): 95–109.
- Couldry, N., and Mejias, U. A. (2019). Data Colonialism: Rethinking Big Data's Relation to the Contemporary Subject. *Television & New Media* 20(4): 336–349.
- Crawford, K. (2021). *The Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence*. New Haven: Yale University Press.
- Daly, A. (2019). Algorithmic Oppression with Chinese Characteristics: AI against Xinjiang's Uyghurs. In: *Artificial Intelligence: Human Rights, Social Justice and Development*. APC, pp.108–112. ISBN 9789295113138
- Daly, A., Devitt, S. K., and Mann, M. (2021). AI Ethics Needs Good Data. In P. Verdegem (ed.), *AI for Everyone? Critical Perspectives*. University of Westminster Press, pp.103–121.
- Dourish, P. and Mainwaring, S. D. (2012). Ubicomp's Colonial Impulse. In *Proceedings of the 2012 ACM Conference on Ubiquitous Computing - UbiComp '12*. ACM, New York, 133.
- Eubanks, V. (2018). *Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor*. New York: St. Martin's Press.
- Filgueiras, F. (2023). Designing AI Policy: Comparing Design Spaces in Latin America. *Latin American Policy* 14(1): 5–21.
- Gal, D. (2020). Perspectives and Approaches in AI Ethics: East Asia. In M. D. Dubber, F. Pasquale and S. Das (eds.), *The Oxford Handbook of Ethics of AI*. Oxford: Oxford University Press, pp. 607–624.
- Garner, S. (2007). *Whiteness: an Introduction*. New York: Routledge.
- Goldberg, D. T. (2011). *The Threat of Race: Reflections on Racial Neoliberalism*. New Jersey: Wiley-Blackwell.
- Gravett, W. (2020). Digital Neo-Colonialism: The Chinese Model of Internet Sovereignty in Africa. *African Human Rights Law Journal* 20(1): 125–146. Available at: <https://dx.doi.org/10.17159/1996-2096/2020/v20n1a5> (accessed 1 September 2023).
- Harvey, D. (2005). *The New Imperialism*. Oxford: Oxford University Press.
- Hesse, B. (2007). Racialized Modernity: an Analytics of White Mythologies. *Ethnic and Racial Studies* 30(4): 643–663.
- Hilgartner, S., Prainsack, B., and Hurlbut, J. B. (2017). Ethics as Governance in Genomics and Beyond. In U. Felt, R. Fouché, C. A. Miller, and L. S.-D. (eds.), *The Handbook of Science and Technology Studies*, 823–851. MIT Press.
- Irani, L. (2016). The Hidden Faces of Automation. *XRDS: Crossroads, The ACM Magazine for Students* 23(2): 34–37.
- Jecker, N. S., and Nakazawa, E. (2022). Bridging East-West Differences in Ethics Guidance for AI and Robotics. *AI* 3(3): 764–777.
- Kalluri, P. (2020). Don't ask if Artificial Intelligence is Good or Fair, ask How it Shifts Power. *Nature* 583(7815): 169. DOI: 10.1038/d41586-020-02003-2. PMID: 32636520
- Kalulé, P., and Joque, J. (2019). Law & Critique: Technology Elsewhere, (yet) Phantasmically Present. *Critical Legal Thinking*, 16 August 2019. Available at: <https://criticallegalthinking.com/2019/08/16/law-critique-technology-elsewhere-yet-phantasmically-present/> (accessed 1 September 2023).
- Katz, Y. (2020). *Artificial Whiteness: Politics and Ideology in Artificial Intelligence*. New York: Columbia University Press.
- Kerasidou, C. X., Kerasidou, A., Buscher, M., and Wilkinson, S. (2022). Before and Beyond Trust: Reliance in Medical AI. *Journal of Medical Ethics* 48(11): 852–856.
- Kundnani, A. (2021). The Racial Constitution of Neoliberalism. *Race & Class* 63(1): 51–69.
- Kwet, M. (2021). Digital Colonialism: The Evolution of US Empire. *TNI Longreads*, 4 March 2021. Available at: <https://longreads.tni.org/digital-colonialism-the-evolution-of-us-empire> (accessed 1 September 2023).
- McIntosh, M. (2020). The Unbearable Lightness of Whiteness. *Wasafiri* 35(3): 1–2. DOI: 10.1080/02690055.2020.1760451
- McQuillan, D. (2022). *Resisting AI: an Anti-Fascist Approach to Artificial Intelligence*. Bristol: Bristol University Press.
- Mbembe, A. (2017). *Critique of Black Reason*. Durham: Duke University Press.
- Mbembe, A. (2003). Necropolitics. *Public Culture* 15: 11–40.
- Mills, C. W. (2009). Critical Race Theory: A Reply to Mike Cole. *Ethnicities* 9(2): 270–281.
- Mills, C. (1997). *The Racial Contract*. Ithaca: Cornell University Press.

- Mittelstadt, B. (2019). Principles Alone Cannot Guarantee Ethical AI. *Nature Machine Intelligence* 1(11): 501–507.
- Mohamed, S., Png, M. T., and Isaac, W. (2020). Decolonial AI: Decolonial Theory as Sociotechnical Foresight in Artificial Intelligence. *Philosophy and Technology* 33, 659–684. DOI: 10.1007/s13347-020-00405-8.
- Moulaert, F. (ed.) (2013). *The International Handbook on Social Innovation: Collective Action, Social Learning and Transdisciplinary Research*. Cheltenham, UK and Northampton, MA, USA: Edward Elgar Publishing.
- Munn, L. (2023). The Uselessness of AI Ethics. *AI Ethics* 3, 869–877. DOI: 10.1007/s43681-022-00209-w
- Ochigame, R. (2019). The Invention of ‘Ethical AI’: How Big Tech Manipulates Academia to Avoid Regulation. *The Intercept*, 20 December 2019. Available at: <https://theintercept.com/2019/12/20/mit-ethical-ai-artificial-intelligence/> (accessed 3 November 2023).
- Quijano, A., and Wallerstein, I. (1992). Americanity as a Concept, or the Americas in the Modern World. *International Social Science Journal* 44(4): 549–557.
- Robinson, C. J. (1983). *Black Marxism: The Making of the Black Radical Tradition*. New York: Chapel Hill.
- Suchman, L. (2012). *Human-Machine Reconfigurations: Plans and Situated Actions*. Cambridge: Cambridge University Press.
- Tafari, D. (2022). What’s wrong with ‘AI ethics’ narratives. DOI: 10.5281/zenodo.6977134.
- Wagner, B. (2018). Ethics as an escape from regulation. From ‘ethics-washing’ to ethics-shopping? In I. E. Bayamlioglu, I. Baraliuc, L. Janssens, and M. Hildebrandt (eds.), *Being Profiled: Cogitas Ergo Sum: 10 years of “Profiling the European Citizen”*, pp. 84–88. Amsterdam: Amsterdam University Press.
- Williams, O. (2019). How Big Tech Funds the Debate on AI Ethics. *New Statesman*, 2019. Available at: <https://www.newstatesman.com/science-tech/2019/06/how-big-tech-funds-debate-ai-ethics> (accessed 22 August 2023).
- Zamora, D., and Behrent, M. C. (eds.) (2016). *Foucault and Neoliberalism*. New York: John Wiley & Sons.