



Artificial Intelligence in the Colonial Matrix of Power

James Muldoon¹ · Boxi A Wu²

Received: 16 February 2023 / Accepted: 2 December 2023 / Published online: 15 December 2023
© The Author(s) 2023

Abstract

Drawing on the analytic of the “colonial matrix of power” developed by Aníbal Quijano within the Latin American modernity/coloniality research program, this article theorises how a system of coloniality underpins the structuring logic of artificial intelligence (AI) systems. We develop a framework for critiquing the regimes of global labour exploitation and knowledge extraction that are rendered invisible through discourses of the purported universality and objectivity of AI. Through bringing the political economy literature on AI production into conversation with scholarly work on decolonial AI and the modernity/coloniality research program, we advance three main arguments. First, the global economic and political power imbalances in AI production are inextricably linked to the continuities of historical colonialism, constituting the *colonial supply chain of AI*. Second, this is produced through an *international division of digital labour* that extracts value from majority world labour for the benefit of Western technology companies. Third, this perpetuates *hegemonic knowledge production* through Western values and knowledge that marginalises non-Western alternatives within AI’s production and limits the possibilities for decolonising AI. By locating the production of AI systems within the colonial matrix of power, we contribute to critical and decolonial literature on the legacies of colonialism in AI and the hierarchies of power and extraction that shape the development of AI today.

Keywords Artificial intelligence · Machine learning · Digital colonialism · Coloniality · Decolonisation · Colonial matrix of power

✉ James Muldoon
James.muldoon@essex.ac.uk

Boxi A Wu
boxi.wu@jesus.ox.ac.uk

¹ Reader in Management, Essex Business School, University of Essex, Colchester, UK

² Oxford Internet Institute, University of Oxford, Oxford, UK

1 Introduction

The capability of artificial intelligence (AI) systems has undergone remarkable progress over the past decade, driven by increased computing power and algorithmic advances that enable ever-larger datasets to be fed into large-scale models. Advancements in a subset of AI research called machine learning (ML) has scaled up computational capabilities such as pattern recognition, text and image analysis and generation with broad applications for predictive analytics, automated decision-making systems and assistive technologies (Clutton-Brock et al., 2021; Kaplan & Haenlein, 2020; Vinuesa et al., 2020). These models, such as OpenAI's GPT-4 and Google's Bard are commonly referred to as large-scale AI models, although they are also referred to as "foundation models," recognising their huge underlying pre-training datasets and flexibility to be fine-tuned for a broad range of tasks (Centre for Research on Foundation Models, 2023).

Scholarship on the harms of AI systems focuses on a number of broad domains across the social and computational sciences. First, issues of bias and fairness within algorithmic inputs (data) and outputs (predictions) have been the focus of both social scientists such as Benjamin (2019) and Noble (2018) as well as computer scientists and sociotechnical scholars such as Buolamwini and Gebru (2018). Second, there is increasing attention on the social and environmental impacts of the AI supply chain; from the exploitative dynamics of AI data work (e.g. data annotation and evaluation) to the environmental consequences of AI associated with the water, energy and rare minerals required to power compute and produce the hardware underpinning AI development and its downstream applications (Bender et al., 2021; Crawford, 2021; Dauvergne, 2022; Robbins & Wynsberghe, 2022). Third, burgeoning scholarship argues that the economic and political power wielded by Western, often American, technology companies have disproportionate impacts on majority world economies that mirror and extend the extractive dynamics of historical colonialism (Adams, 2021; Birhane, 2020; Kwet, 2019a, b; Ricaurte, 2019).¹ This article adds to the emerging intersection of AI harms and theories of data and digital colonialism by drawing theoretical links between Aníbal Quijano's "colonial matrix of power" within the modernity/coloniality research program and literature on the political economy of AI production (Quijano, 2007). We locate the extractive dynamics of AI in the colonial matrix of power on two levels considering both material harms manifested through the deployment of AI technologies and the discursive harms perpetuated by the imaginaries of AI developers and the broader AI industry.

In order to interrogate AI's position in the colonial matrix of power we focus on a sub-field of AI referred to as machine learning (ML), which uses statistics and probability to learn from large datasets, making predictions on new or unseen data (Mittelstadt et al., 2016). Machine learning encompasses a variety of methodologies that address different problems. For example, deep learning, a type of ML based on multi-layer

¹ Following Ricaurte (2022) we use the term 'majority world' (coined by Bangladeshi photojournalist Shahidul Alam) to refer to countries where most of the global population resides and explicitly in acknowledgement of historical and ongoing colonisation.

neural networks has enabled breakthroughs in natural language processing and multi-modal perception capabilities which underpin digital technologies such as social media recommender algorithms, chatbots and assistants, skin cancer detection and autonomous vehicles. Contemporary deep learning can be applied to both supervised tasks (requiring labelled data, often outsourced by companies to low-paid data workers) and unsupervised tasks which require large quantities of training data (often scraped from the web without consent or generated by data workers through AI data platforms) (McQuillan, 2022). The resources required to research, develop and deploy AI technologies are largely concentrated in the hands of Western technology companies. Such companies have the financial capital to attract highly skilled labour and ensure access to both proprietary datasets and computing infrastructure such as Graphics Processing Units (GPUs) (Srnicek, 2022). Through engaging with critical and decolonial thought, we also recognise our positionality as Australian and UK institutionalised researchers whose life experiences and education have been shaped by the colonial structures that uphold Western epistemologies. We approach this research with sensitivity and humility, acknowledging that our perspectives are inevitably influenced by our own social, cultural and epistemological standpoints.

We address a number of exploratory research questions: which aspects of AI production reinforce extractive social, economic and political dynamics within the colonial matrix of power? How do historical Western imaginaries of modernity and technological progress shape the material and knowledge production of current AI development? How do these link to existing scholarship on AI harms and what are the possibilities of exploring alternative models to the current paradigm of AI development? We expand on three ways that AI can be located within the colonial matrix of power. First, we discuss the *colonial supply chain of AI*, demonstrating how AI's production reinforces the legacies of historical colonialism evidenced in global economic and political power imbalances. Second, AI is produced through an *international division of digital labour* that extracts value from the labour of workers in the majority world, generating profits for Western technology companies. The labour distribution of AI work across the globe concentrates the most stable, well-paid and desirable jobs located in key technology hubs in the West and the most precarious, low-paid and dangerous work exported to workers in the majority world. Third, AI perpetuates *hegemonic knowledge production* through the discursive and material enactments of AI's purported universality and rationality that perpetuates a frame of Western-centrism that marginalises non-Western alternatives and possibilities for decolonising AI. We conclude by highlighting the social and political consequences of AI production in the colonial matrix of power and point towards the work of scholars and activists who are exploring alternatives to the current world-system within which AI is built.

2 The Colonial Matrix of Power

Historical colonialism denotes the practice of domination over a people and its culture and the appropriation of its wealth, labour and natural environment. While many former colonies of Western nations have gained formal independence

through political decolonisation, global inequalities of wealth and power from colonialism still endure. Maldonado-Torres (2007, 243) defines “coloniality” as the “long-standing patterns of power that emerged as a result of colonialism, but that define culture, labour, intersubjective relations, and knowledge production well beyond the strict limits of colonial administrations.” The modernity/coloniality research program emerged from the writings of diasporic scholars from the Andean region of Latin America whose communities were impacted by European colonisation. Scholars were influenced by the American Subaltern Studies group in the 1990s, an intellectual community who engaged with the subalternised knowledge systems of indigenous communities and other groups oppressed by colonisation. Quijano (2000, 2007), Lugones (2016) and Mignolo (2012) among others, have contributed to the critical understanding the inextricable relationship between colonialism and modernity and the associated imposition of European cultural, social and epistemological frameworks (see *Globalisation and the Decolonial Option* edited by Walter Mignolo and Arturo Escobar for an extensive overview of the analytic of modernity/coloniality). The modernity/coloniality research program argues for a new understanding of modernity that challenges the origin of and universality of European modernity and associated social, cultural and epistemological norms; instead, focusing on modernity’s relationship with colonisation and the capitalist world-system.

Quijano (2007) conceptualised the colonial matrix of power (“*patron de poder colonial*”) to describe the current world-systems entanglement with historical colonialism. The colonial matrix of power is an organising principle involving exploitation and domination across four interrelated domains: “control of economy (land appropriation, exploitation of labour, control of natural resources); control of authority (institution, army); control of gender and sexuality (family, education) and control of subjectivity and knowledge (epistemology, education and formation of subjectivity).” (Mignolo, 2007, 156). Although we draw on all four domains in our analysis, the article focuses on the control of economy (namely labour and resources) and the control of subjectivity and knowledge. Quijano (1993) argues that within the matrix, these domains are naturalised around the colonial articulations of “objective” and “scientific” notions of race and racial classifications. These categories underpin the racial division of labour, resources and capital that constitute one of the most visible organising principles structuring modern societies.

The exercise of economic dominion over both labour and resources is closely intertwined with the ideological construct of European technological preeminence used to justify European colonialism. European colonisers imposed authority in science, culture and societal administration while communities from Africa and Asia were expected to acculturate and obey while providing labour to the colonial project (Adas, 1989). Quijano (2016) links the manipulation and control of technological resources by colonial powers to the analytic of modernity/coloniality. He argues that technologies of communication and transport were used to instrumentalise the superiority of European modernity, which were mutually constitutive with the violence of colonialism. This highlights the important fact that the colonial legacy of Western dominance, in terms of both modernity and technological progress, continues

the legacy of colonial domination despite the elimination of political colonialism in many areas of the world (Quijano, 2007).

The modernity/coloniality analytic emphasises that the imaginaries and epistemologies of AI development are inextricably predicated upon rationality and universality and are not divorced from coloniality. Indeed scholars of sociotechnical imaginaries and feminist science and technology studies remind us that technology is co-created by the social, cultural and political environment around it (Jasanoff, 2004). In the case of algorithmic technologies, these are modelled on visions of the social world, with their outcomes influenced by the agendas of corporate and political elites who stand to gain from the economic and political changes engendered by the deployment of these technologies (Beer, 2017).

The modernity/coloniality analytic seeks to critically examine the logics of colonialism with respect to both the economy, but also knowledge and cultural production. This was closely tied to the hierarchical paradigm of European rationality as a universal form of knowledge – rendering Europeans as rational “subjects” while others were merely “objects of knowledge of domination practices” (Quijano, 2007, 174). This served the epistemological authority for the destruction of language and cultural practices during colonial rule. The epistemologies of technology are also influenced by the colonial matrix of power; local technologies were viewed as primitive and traditional (Quijano, 2007). The continuity of this logic can be traced to the post-war discourses about international development, led by Western institutions such as the World Bank, that infantilised non-Western and local knowledge and technology in favour of unilaterally imposing Western development practices (Escobar, 2007).

This dynamic is also evidenced in the epistemology of academic disciplines such as anthropology that have perpetuated the subjugation of non-Western knowledge and culture, long after former colonies gained independence. We argue that these logics can be traced throughout the development and deployment of AI technologies, from Big Tech’s concentrated ownership of AI’s intellectual property and computing infrastructures, to the export of technologies and their embedded values into the majority world. It is through these exercises of power that a Western mode of knowing has been constituted and reinforced as the dominant imaginary about AI which is tightly bound to a story of modern growth and progress. However, as the colonial matrix of power analytic elucidates, the vestiges of colonialism persist and necessitate a de-linking from narrow ways of thinking about colonial relations in order to accomplish the unfinished and incomplete dream of decolonisation.

3 Decolonial Perspectives on Data and Technology

Emerging scholarship from the social sciences, across platform studies, critical data studies and media studies, trace the legacies of colonialism in digital technologies and their data infrastructures. These can broadly be parsed into digital colonialism, data colonialism, and decolonial approaches to technology and AI development. The study of “digital colonialism” emerged as a theoretical lens for analysing how the flow of data, technology and value, usually from wealthy Western economies to the

majority world, reinforce colonial legacies and new forms of economic and political domination. Theorists trace how technology companies exercise new forms of control over the majority world through manufactured dependence on their digital tools and platforms, subjecting populations to new forms of surveillance and data extraction (Avila, 2020; Casati, 2013; Coleman, 2018; Kwet, 2019b). Avila (2018, 2020) shows that major technology companies in the US seek to obtain control over critical infrastructure in order to gain access to data produced by the majority world. She defines digital colonialism as “the deployment of imperial power over a vast number of people, which takes the form of rules, designs, languages, cultures and belief systems serving the interests of dominant powers” (Avila, 2020, 1). Avila (2020) asserts it is not states but “technology empires” that hold the dominant positions through their control of “digital infrastructures, data and the ownership of computational power” in addition to solidifying their power through favourable global trade agreements.

Kwet (2019b) has described digital colonialism as a “structural form of domination” which technology companies, and particularly Big Tech, exercise through “three core pillars of the digital ecosystem: software, hardware, and network connectivity.” Kwet argues this is a neo-colonial form of “imperial control,” particularly due to the integration of the surveillance apparatuses of corporations into the intelligence services of the US state and the dominance of a Silicon Valley ideology across the globe. Kwet (2019a) highlights the relationship between Government elites and Big Tech in South Africa, citing the example of Operation Phakisa Education, a failed education reform attempt from the South African government aiming to “leverage ICTs to improve basic education.” He argues that Big Tech investment in majority world classrooms, through programs like Microsoft Partners in Learning and Google Classroom, create infrastructure dependencies amongst youth while biasing local tech ecosystems towards Big Tech. This is also evidenced through Big Tech’s provision of free internet via platforms like Facebook which centralise Facebook’s control over communications and news, and drains local advertising revenues. In both examples, the historical mechanisms of surveillance capitalism are perpetuated, extending the use of colonial patterns of racialised extraction and domination through to the ongoing policing of Black bodies. Abeba (2020) illustrates how the ‘algorithmic colonisation of Africa’ imposes Western values onto locales while impoverishing local tech development, citing that Nigeria imports 90% of all software used in the country. In a similar vein, but from a legal perspective, Coleman (2018) analyses how data protection laws and regulations are often absent in African countries which has given rise to a modern day “scramble for Africa” in which tech companies offer purportedly “free” and altruistic services in order to harness the power of users’ data for profit. In each of these cases, the authors argue that issues of power and technology cannot be discussed without attention to local context, history and the values embedded in technologies. This research agenda is wide reaching and lays the groundwork for this article to further systematise these ideas and to integrate a political economy perspective of the production of AI.

Scholars have focused on the extractive dynamics of data collection in digital platforms, forming discourses around “data colonialism”, which Couldry and Mejias (2019) define as human life being captured and controlled through the appropriation

of data extracted by social media platforms for profit. Contrary to theorists who describe the business models of surveillance tech companies as a radical break from pre-existing capitalism (Zuboff, 2019), Couldry and Mejias (2019) argue these latest developments are an extension and intensification of colonialism and capitalism's drive to commodify human life for profit. What is distinctive about their theory is not only that they explicitly frame their critique within capitalism's long standing relationship to colonialism, but they understand contemporary data practices as a "new form of colonialism" for the twenty-first century, represented by new industries devoted to the capture and storage of personal data. Unlike previous uses of the term by Thatcher et al. (2016), Couldry and Mejias (2019) do not focus on the Western imposition of free digital infrastructure to the majority world. Rather, "social life all over the globe becomes an 'open' resource for extraction that is somehow 'just there' for capital" (Couldry & Mejias, 2019, 337). Tech elites in the US and China are the main beneficiaries of this new regime, while "North–South, East–West divisions no longer matter in the same way." For Couldry and Mejias (2019), this emerging framework sets the stage for a new phase of capitalism characterised by the capitalisation of life without limits.

Mumford (2022) raises the important concern that some of the key decolonial insights raised by decolonial scholars such as Anibal Quijano, Walter D. Mignolo, Nelson Maldonado-Torres and others do not appear as central concerns within theories of data colonialism. This includes their critique of the false universality of European knowledge and the ways in which the colonial project marginalises non-Western ways of understanding the world. Mumford (2022) asserts that data colonialism therefore shies away from addressing some of the modernity/coloniality research program's primary concerns, around the imposed centrality and objectivity of European superiority used to legitimise the conquest and destruction of other ways of knowing and being. The focus of data colonialism is on describing an extractive rationality inherent in new business models and showing how new forms of value are produced through data relations. In this research, we seek to centre critiques of Western centrality in social, cultural and epistemological frameworks used to analyse AI, by drawing from the modernity/coloniality school to locate AI within the colonial matrix of power.

The paradigms of digital colonialism and data colonialism have a shared concern for the continuing effects of colonialism on asymmetries of power within global capitalism. These dynamics are magnified by the technical properties of AI technologies that rely on processes of extraction. Here we turn to theories of 'extraction' and 'extractivism' that have typically focused on the exploitative logics of resource extraction in global capitalism such as raw mineral mining and plantations. However, the extractive dimensions of global digital capital, enabled by AI technologies, extend beyond the material and territorial (Mezzadro & Nielson, 2017; Gago & Mezzadra, 2017). For example, Ricaurte (2019) argues that AI technologies rely on the extraction of data, on which information and knowledge are created. Colonial hierarchies are perpetuated through asymmetric flows of data extractivism which concentrate data and therefore value in wealthy Western countries (Thatcher et al., 2016). Pasquinelli and Joler (2021) argue that the technical components of machine learning are instruments of knowledge extractivism derived from social, cultural and

scientific labour. Unlike other digital technologies, the current paradigm of machine learning focuses on improved performance through scale, which has increased demand for human generated images, text and videos. They argue that processes of pattern extraction, recognition and generation that characterise machine learning are designed to extract “analytical intelligence.” Data is extracted through trawling the web, collecting user data from the digital platforms in which algorithms are embedded and through the often low-paid and precarious labour of AI data annotation (Bender et al., 2021). These examples of extractive regimes imposed by the developers of and technical capabilities of AI can be considered sites of ‘algorithmic coloniality,’ which researchers suggest are also discursively wielded in technology companies, international development organisations and within the broader AI governance discussions (Mohamed et al., 2020; Png, 2022). Scholars have explicitly argued for the need to apply decolonial perspectives to ensure that AI harms are discussed within the historical context of colonialism and its continuities. These examples highlight the plurality of extractive dynamics in AI production; from the extraction of data and therefore value from globally distributed AI data works to the algorithmic tracking and surveillance of daily life as described by Couldry and Mejias (2019). Returning to the lens of the colonial matrix of power, the extractive dynamics of AI production reinforce global power asymmetries.

4 The Colonial Supply Chain of AI

Technology companies frame AI technologies as tools for increasing human and technical productivity, with promises of creating an environmentally friendly and efficient global economy (Dauvergne, 2022; Nost & Colven, 2022). Crawford (2021, 48) reminds us that corporate narratives of AI are often completely abstracted from the physical reality of its operation, “[l]ike running an image search of ‘AI,’ which returns dozens of pictures of glowing brains and blue-tinted binary code floating in space, there is a powerful resistance to engaging with the materialities of these technologies.” Crawford and Joler’s (2018) project *Anatomy of an AI system* captures the planetary reach of AI’s vast supply chain of human labour, data, algorithmic processing and resource extraction. The decolonial lens looks beyond this immaterial depiction of AI and draws attention to its material infrastructure, ongoing energy consumption, complex geopolitics and the long histories that underpin them (Png, 2022). Through analysis of the connections within multi-layered supply chains, we uncover how the global relationships of labour exploitation and knowledge extraction in the AI supply chain are actively enabled by the continuities of historical colonialism.

There is a growing critical literature on AI production that analyses the provenance of its datasets and the human work involved in its production (Brevini, 2020; Newlands, 2021; Tubaro et al., 2020). Kemper and Kolkman (2019) speak of a “data value chain,” while Newlands (2021) offers the framing of “dataset supply chains.” Many of the current discussions of the value chain and production of AI begin with the process of data collection and gathering as the first site of inquiry (Bechmann & Bowker, 2019; Miceli & Posada, 2022; Tubaro et al., 2020). Tubaro et al. (2020)

describe the necessary processes of data collection, data cleaning, model training and evaluation, required to train and deploy machine learning models. We expand this framework through the concept of “the colonial supply chain of AI” to interrogate the physical infrastructure of AI and the human and material resources necessary to power machine learning processes. In doing so, we respond to calls from the modernity/coloniality research program to engage with alternative narratives “geared towards the search for a different logic” than uncritical celebrations of European rationality and progress (Mignolo, 2012, 22). From this perspective, the development of AI technologies should be understood as constituted by a colonial supply chain that relies on an unjust international division of digital labour and the longstanding material and epistemological subordination of countries in the majority world from which resources are extracted and labour is exploited.

The dominant discourse on AI technologies consciously limits the visibility of machine learning models and applications are portrayed as environmentally friendly and contributing towards a cleaner future (Clutton-Brock et al., 2021; Giuliano, 2020). Critics have questioned this narrative with environmental concerns receiving increasing mainstream attention within AI policy circles – such as the Global Partnership on AI’s “A responsible AI Strategy for the Environment” and the OECD AI’s Working Group on Compute and Climate (OECD, 2022; OECD.AI, 2023). These discussions should also foreground the colonial history of these extractive practices which have important consequences for how they operate today (Clutton-Brock et al., 2021; Png, 2022). When it comes to the environmental costs of AI – both in terms of costs of production and waste outputs – the harms of these processes are disproportionately shouldered by countries in the majority world while profits flow to wealthy Western economies (Dauvergne, 2022).

These environmental concerns begin with the observation that machine learning algorithms require a large amount of computational capacity to both store and process data which require servers and infrastructure. AI production also intensifies demand for rare earth elements which are used in the hardware necessary for powering digital technologies that enable and rely on AI: computers, smart phones, data centres, undersea cables, insulation, optical fibres and fuel cells (Abraham, 2017; Dauvergne, 2022; Fei et al., 2019). This is driving a global mining boom for cobalt, lithium and coltan, among others (Bird et al., 2020; Kiggins, 2015). Dauvergne (2022) notes that in the case of the metal tantalum, 60% of the world’s supply is extracted from Africa. Rising demand in Democratic Republic of Congo has had devastating environmental and social consequences from the pollution of water systems to the degradation of ecosystems. Open pit mining, such as in China’s Bayan-Obo district, creates enormous lakes of toxic waste and can pollute groundwater and contaminate local surroundings (Ali, 2014). The environmental costs of widespread environmental degradation caused by AI’s supply chain are first and foremost carried by communities in the majority world, including increases in mining industry violence against indigenous communities. These examples point towards the role of AI’s supply chain in the perpetuation of environmental degradation with human and ecological impacts on climate justice. Reflecting on Png’s (2022) calls for attention to the coloniality of power in AI governance, we aim to further elicit the social and ecological impacts of AI’s supply chain which arguably have not been sufficiently

interrogated within the modernity/coloniality research program (Coronil, 1997; Escobar, 2007).

From the 1990s onwards, demand from wealthy Western countries for technology-related minerals created an enormous spike in mining in the Democratic Republic of Congo, Ethiopia, Mozambique, Rwanda, South Africa, and Zimbabwe (Fuchs, 2014, 180). However, the climate crisis, the war in Ukraine and disruptions to global supply chains have led to a rise of resource nationalism in which European and North American governments have attempted to secure strategic resources for technological development by maximising what they can extract on their own soil and reducing their dependence on foreign suppliers (Valdivia, 2023). Some countries which have previously been sites of extraction have also made moves to nationalise their materials. For example, Chile has announced that it would nationalise all mines and Mexico has prevented private companies from mining its lithium (BNamericas, 2023; Hurtado, 2022). This demonstrates how the geopolitics of mineral extraction continue to be shaped by historical and ongoing colonial patterns of exploitation and dependencies, even as “AI nationalism” has altered the way in which these resources are mined and traded (Hogarth, 2018).

Furthermore, even once established, the physical infrastructure of AI requires an enormous amount of energy to train machine learning models, including large quantities of electricity and water usage (Brevini, 2020; Crawford, 2021). Some of the dramatic recent advances in machine learning have been the result of using more computational power in training, which requires greater levels of energy consumption. The amount of compute used to train AI models necessarily increases every year as companies want to develop more powerful systems (Schwartz et al., 2020). However, this technical increase comes at a large environmental cost. It is difficult to estimate the total carbon footprint of the field of machine learning, and no existing articles provide a specific number (OECD, 2022). One of the most cited studies by Strubell et al. (2019) found that a single natural language processing model produced 660,000 pounds of emissions, amounting to as much as five cars over their lifetime. At an organisational level, Google has released figures which show that machine learning accounts for 15% of the company’s total energy consumption (Patterson et al., 2022). The environmental costs of this technology are not distributed equally, with countries vulnerable to climate change related catastrophes most at risk (Bender et al., 2021; Westra & Lawson, 2001). Technological growth also increases the amount of e-waste the world produces, which is another cost disproportionately shouldered by the majority world. E-waste increased to 6.8 kg per capita in 2021, with long-term estimates predicting over 120 million metric tons of e-waste per year by 2050 (Dauvergne, 2022).

In summary, the production of AI systems requires a material and environmentally costly infrastructure that is often ignored in mainstream accounts of its social and economic benefits. The AI supply chain relies on rare earth minerals from the majority of the world; the process of mining and assembling these minerals follows colonial patterns of uneven global trade in which those closest to the source of the minerals often obtain the least value from their extraction. The process of mining can be destructive, release pollutants and give rise to a host of other social and political issues related to mining and selling resources. As we have demonstrated through

the emerging research on these harms, the human and ecological harms of these extractive processes fall disproportionately on communities in the majority world who are also excluded from both global climate and AI governance (Png, 2022). An analysis of the coloniality of the AI supply chain, which begins with its material infrastructure and concomitant environmental and human costs, must then turn to questions of data collection, cleaning and analysis, which is another key site where colonial legacies can be traced through the lens of the colonial matrix of power.

5 An International Division of Digital Labour

Theorists from the modernity/coloniality school have analysed how Europe's colonisation of non-European geographies, particularly Latin America, resulted in a horrific and massive extermination of indigenous and local populations through the imposition of forced labour as an expendable labour force. This laid the foundations for the development of a world market through the creation of a new structure of control over labour and chains of commodity production (Mignolo, 2007; Quijano, 2000, 535). An international division emerged between a core and periphery with workers in each region assigned different forms of labour with different modes of social, cultural and epistemic control (Mignolo, 2011, 18). Race was a central organising principle of this new division of labour as supposed differences in biological structures were used to justify the subordinate status of different racial groups and their respective social roles within colonial capitalism. Whiteness was associated with wages and powerful positions in colonial administrations, while other forms of labour such as slavery and serfdom were confined to non-White races (Quijano, 2000, 535). This new "social geography of capitalism," as Quijano (2000, 539) refers to it, ensures "the entire production of such a division of labour was articulated in a chain of transference of value and profits whose control corresponded to Western Europe." A system of wage labour eventually spread across many parts of the colonised world, but the division of labour and a hierarchy of roles in a global production chain remained in place (Quijano, 2000, 565). Quijano (2001) emphasises that the contemporary global distribution of resources and the hierarchies of labour and exploitation continue to follow relations of coloniality, despite the elimination of political colonialism in many geographies.

This division of labour is particularly pronounced in the technology and AI research and development industries with the majority of highly-paid software development jobs located in the wealthy Western countries while low paid and often precarious clerical, data entry, assembly and mining work is located in the majority world (Casilli, 2017; Fuchs, 2014; Irani, 2019). As we have seen, when it comes to resource extraction, there are certain components of AI systems that are mined under exploitative conditions of forced labour in countries including the Democratic Republic of Congo and China (Niarchos, 2021). These workers in extraction zones around the world often work in highly volatile conditions of exploitation and threats of violence, in which union-busting, poor working conditions and widespread pollution are the norm (Global Witness, 2022). While the extraction of minerals tends to take place in parts of Africa, China and Latin America, the refinement and assembly

of electronic components is more likely to occur in Asian countries such as Japan, Taiwan, and South Korea (Brodzicki, 2021). The final products are then shipped to consumer markets throughout Europe, North America and the rest of the world.

Aside from the physical infrastructure required to support AI systems, computational approaches such as machine learning are reliant on human work in the development of datasets and in the ongoing training and execution of machine learning algorithms (Bechmann & Geoffrey, 2019). As Jones (2021) notes, “the magic of machine learning is the grind of data labelling.” Developing the datasets that are needed to train algorithms consists of a complex process involving both humans and machines working in close connection (Raisch & Krakowski, 2021). This is often referred to as microwork, “a form of work on digital platforms in which short tasks are assigned to workers, who are paid piece wages for completing them” (Jones & Muldoon, 2022: 5). Workers performing this type of work on platforms such as Amazon Mechanical Turk, Appen and Clickworker undertake a variety of tasks which can include consumer surveys, identifying photos, and coding data. Previous research on microwork has focussed predominantly on the position of workers rather than the connection between this work and the production of datasets for AI (Newlands, 2021).

Tubaro and Casilli (2019) argue there is a structural demand for microwork in AI, which is unlikely to be merely a temporary phenomenon. It could be argued that as algorithms improve in quality they will no longer need human involvement in preparing and training datasets and that they will grow better at labelling, tagging and categorising their own datafiles. Although AI can now solve tasks considered challenging five years ago, the kinds of problems that industry demands are solved by AI have grown more complicated. Companies purchasing AI services now demand customised resources and have a wide variety of new tasks to perform from the development of autonomous vehicles, to medical imaging and advanced analytics. Tubaro et al. (2020) show that “data preparation tasks represent over 80% of the time consumed in most AI and machine learning projects, and that the market for third-party data labelling solutions is \$150 M in 2018, growing to over \$1B by 2023.” Moreover, human labour will also still be needed to verify and correct algorithms to ensure they are accurately performing tasks and that issues of bias and fairness, or factuality of outputs are not diminishing overall performance. As Lilly Irani notes, “human labour is necessary to configure, calibrate, and adjust automation technologies to adapt to a changing world, whether those changes are a differently shaped product or a bird that flies into the factory” (Irani, 2015).

Microworkers undertaking data labelling and preparation are usually based in low-income countries in the majority world, with extensive ethnographic and social sciences research demonstrating the poor compensation and working conditions experienced by workers (Gray & Suri, 2019; Miceli & Posada, 2022; Posada, 2022; Crawford, 2021). Aside from workers inside the United States, the countries with the largest concentrations of microworkers are India, Pakistan, Bangladesh, Indonesia and the Philippines (Berg et al., 2018; Kuek et al., 2015; Stephany et al., 2021). Research and journalistic work has revealed that many workers have no employment contracts and work for piece meal rates as independent contractors on digital labour platforms that broker agreements for clients to post their tasks on the platform. In

addition to the mode of employment having been traditionally racialised and therefore the purview of people of colour within a colonial system, workers also experience dangerous and psychologically harmful work that multinational companies would rather keep out of the spotlight (Gray & Siddharth, 2019; Perrigo, 2023). Tasks such as moderating social media for racist and toxic content or identifying child pornography and other graphic images is also work that is usually outsourced to workers in the majority world (Elliott & Parmar, 2020).

The work is structured in a way to render invisible the contributions of workers in the majority world from the public imaginary of AI's production, preferencing the so-called highly skilled work of engineers. Gray and Siddharth (2019) argue that the design of microwork platforms is engineered to anonymise workers, provide them limited avenues for communicating with clients and make their individual work contributions invisible within the overall operation of the platform. This mirrors analysis from Hatton (2017) who argues that work can be invisibilised through three intersecting sociological mechanisms identified as cultural, legal and spatial processes that devalue certain types of labour which both separates the worker from the observer and contributes to the work's economic devaluation. The not-for-profit coalition, Partnership on AI (2021) has raised concerns that although AI systems are dependent on clean and labelled datasets, companies' marketing campaigns may result in "efforts to hide AI's dependence on this large labour force when celebrating the efficiency gains of technology. Out of sight is also out of mind, which can have deleterious consequences for those being ignored."

One prominent example of this is OpenAI's engagement with outsourced Kenyan labourers to help reduce harmful content on its chatbot ChatGPT (Perrigo, 2023). OpenAI worked with an outsourcing partner, the San Francisco-based firm Sama, which employs workers in a range of majority world countries including Uganda and Kenya to prepare and label data for Silicon Valley clients. Sama frames the company as an "ethical AI" company claiming to have lifted 50,000 people out of poverty in East Africa through their business model. While most microworkers on larger platforms such as Amazon Mechanical Turk are independent contractors with no employment agreement, these workers were employed by the company on contracts through which they could expect to take home between US\$1.32 per hour after tax and US\$1.44 per hour, depending on their performance (Perrigo, 2023). In contrast, the minimum wage for a receptionist in Nairobi is US\$1.52 an hour. Sama workers who performed data labelling tasks would work nine-hour shifts reviewing violent, toxic and abusive content and attempting to filter it out before it reached end users. This content was taken from some of the darkest corners of the Internet and described horrific details of child sexual abuse, bestiality, murder, suicide, self harm and torture, leading workers to report being mentally scarred from the experience.

6 Hegemonic Knowledge Production

AI reinforces the hegemony of Western values and epistemologies that marginalise non-Western alternatives. The modernity/coloniality research program offers a critical lens for analysing these epistemological concerns, which they refer to as the

“coloniality of knowledge” (Grosfoguel, 2007; Quijano, 2000). They contend that the framework of Western scientific knowledge with its claims to objective truth and universal validity was part of a global project to enforce European hegemony over Latin America and other colonised regions. The knowledge projects that began in Europe in the seventeenth century, including the so-called anthropological study of colonised communities, engaged in de-humanising and extractive data collection practices with unidirectional flows of information and data towards Europe. As Mignolo and Walsh (2018, 2) have argued, “all theories and conceptual frames, including those that originate in Europe and the Anglo United States, can aim at and describe the global but cannot be other than local.” One example of this was early Enlightenment cosmographers who sought to collect all knowledge about the universe as a rational and ordered set of universal truths. As Graham and Dittus (2022, 10) argue, “everything in the universe could be described in predetermined ways and placed into predetermined systems. It extended to all corners of the globe and did not tolerate alternate epistemologies.”

Recent scholarship has argued that Western knowledge epistemologies are embedded in AI development. From this perspective, the dominant epistemological paradigm that underpins technology is a direct result of the development of European classificatory systems and the broader scientific and cultural project that grew out of it. McQuillan (2022) describes how the statistical logics underpinning artificial intelligence reveal continuities with “racial imperialist views of national progress.” He argues that the political positions emerging around AI can be traced to the eugenicist conceptualisation of a racialised hierarchy of intelligence that was used to justify European colonial expansion. These historical biases of racism and colonialism are inseparable from the context of AI research and development in elite universities and technology companies in the West. Following the emerging critical scholarship on values embedded in machine learning and the broader AI industry, we argue that imaginaries of AI’s generality and neutrality constitute a reproduction of hegemonic Western knowledge and epistemology (Birhane, 2021).

Large scale ML is necessarily reliant on data at all stages of its development pipeline – from the ingestion of training data, to fine-tuning for a particular use case, to the evaluation of performance on golden data. AI datasets and common benchmarks such as ImageNet and GLUE are inherently political and value-laden. Raji et al. (2021, 8) argue that those without the power to define themselves are viewed by the model through a “distorted data lens.” The issue of dataset bias has been analysed by Buolamwini and Gebru (2018), who found that commercial facial analysis algorithms and datasets had error rates of up to 34.7% for darker skinned females, compared to lighter skinned males. Dataset biases occur when the data ingested by models is encoded with social prejudices and inequalities through the source data or through the decisions of individuals who transform data through annotation, sorting and analysis. The larger the dataset, the greater the risk of bias: the Stanford University AI Index Report states that “a 280 billion parameter model developed in 2021 shows a 29% increase in elicited toxicity over a 117 million parameter model considered the state of the art as of 2018” (Zhang et al., 2022).

This has been documented through the work of AI researchers who audit datasets used to train ML algorithms such as DALL-E 2 and ChatGPT. Birhane et al.

(2021) audited the multimodal training dataset LAION-400 M which is commonly used to train large-scale ML models. The audit returned images and text that hypersexualised “overwhelmingly sexualize Black women, and fetishize Asian women” (Birhane, 2021, 11). In later work on the upsized LAION-2B model, now with 2 billion samples, researchers found a “hate-scaling phenomena” whereby the scaled up dataset increased harmful content constituted 12% of content (Birhane et al., 2023, 2). Additionally, when tested on racial bias compared to the smaller model, LAION-2B was found to associate Black individuals with criminality five times more than the smaller model LAION-400 M.

The toxicity in the datasets is driven by the social world represented on the Internet. Participants on platforms such as Wikipedia, Reddit and YouTube who skew young, white, male and American; which means that overtly racist, sexist and ageist perspectives are overrepresented in the data and many other viewpoints are excluded altogether (Pew Research, 2021; World Bank, 2023). ImageNet, a foundational and globally utilised computer vision dataset consisted of 45% US-sourced images, and over 60% from a selection of Western countries. In comparison, only 1 and 1.2% of images were sourced from China and India, despite their respective populations (Raji et al., 2021, 8). These examples may seem amenable to technical fixes such as the tools of technical fairness such as data filtering, classifiers and evaluations. However, they represent the algorithmic reinforcement of hegemonic values and knowledge that is deeply entwined with the Western centric rationality of both science and colonisation, illustrating how AI technologies perpetuate the colonial matrix of power.

The reproduction of hegemonic knowledge begins with data but flows through to the outputs of downstream use cases, manifesting in the algorithmic reproduction of societal biases and harms. Ruha Benjamin’s *Race After Technology* (2019) calls attention to the magnified social biases codified into machines through learnt patterns in the training data. Benjamin (2019) points towards numerous examples of algorithmic bias including the algorithmic codification of Whiteness into technical systems. In one example, she investigates The Beauty AI initiative which attempted to use an algorithmic system to judge a neutral and objective measure of beauty across 100 countries, resulting in 86% of winners being white. In another example, Google Photo was found in 2015 to auto-tag two Black friends as gorillas in a photo, an egregious racist depiction formalised through the history of scientific racism. She elucidates the implicit and explicit anti-blackness built into predictive policing algorithms that used data reflected ongoing surveillance and targeting of predominantly Black neighbours. In illustrating these examples, she points towards the inherent invisibility of whiteness, as the default within both society and algorithmic systems.

These harms persist despite the developers of large-scale AI systems claiming to prioritise safety and responsibility in the design and deployment of consumer facing systems. For example, OpenAI’s DALL-E 2 System Card, intended to document the risks and limitations of the generative model, states that “the default behaviour of the DALL-E 2 Preview produces images that tend to overrepresent people who are White-passing and Western concepts generally” (OpenAI, 2022). Melissa Heikkilä’s (2022) reporting on her experience as an asian woman with the Lensa digital portrait app revealed that the underlying model, Stable Diffusion, generated

hyper-sexualised and pornographic representations of the author. This model, which is open-source and thus available for developers to create downstream applications, was built using the open-source dataset LAION-5B, a larger version of the LAION datasets audited by Prabhu and Birhane (2020) and Birhane et al. (2023). Even when technical safety measures have been put in place, the hegemony of algorithmic neutrality and objectivity still subjugates marginalised subjectivities (Bergman et al., 2023). Indeed evidence from ChatGPT's release showed that even when technical safety features are built into systems, they are easily circumvented, allowing the chatbot to produce alarming racist, sexist and derogatory commentary derived from its underlying training datasets (Asare, 2022; Vock, 2022). As Beer (2017) has argued, algorithms have a "social power" through the ability to reorder the social world, which is particularly prescient when applied to the decisions developers of large-scale AI make during AI production.

AI technologies are produced within particular social contexts, including the way that algorithms are socially and organisationally constructed (Beer, 2017). Anthropological studies of engineering teams including AI researchers demonstrate that researchers enact their cultural discourses into technical objects through their work (Forsythe, 2001; Seaver, 2017; Star, 1999). This is necessarily influenced by the majority white, male and educated demographics of the global technology elite who contribute to the design, deployment and regulation of AI (Bender et al., 2021; Chowdhury, 2023). Tech companies often portray their services as "colour-blind," which as Safiya Noble (2018) has argued represents a myth of post-racialism which inhibits critical engagement with racialised social inequalities. Silicon Valley elites maintain power and control through this myth which is used to suppress concerns about race and diversity issues while justifying a so-called meritocracy.

Ideologies of race are also central to how AI is imagined and sold to investors and companies seeking to purchase AI products. Cave and Dihal (2020) argue that AI systems themselves are imagined as White: "to imagine an intelligent (autonomous, agential, powerful) machine is to imagine a White machine because the White racial frame ascribes these attributes predominantly to White people." They show how AI is deployed in narratives around three core categories of intelligence, professionalism and power and in each case notions of whiteness predominate. These examples show that through the Western values and knowledge embedded in the social world of AI production, the machines themselves embody Whiteness in popular culture and the commercial world. In perpetuating hegemonic knowledge production, we locate AI in the colonial matrix of power.

7 Conclusion

This article seeks to understand how AI production reinforces the extractive social, economic and political dynamics within the 'colonial matrix of power'. In doing so, we make three contributions to the critical literature on AI. First, we systematise an analytic framework through which to understand the material and ideological implications of the AI supply chain, through the lens of the colonial matrix of power. We map the organising principles of the colonial matrix of power to the

interwoven levels of extraction throughout the AI supply chain (Crawford & Joler, 2018). Second, we combine theories of labour exploitation from the modernity/coloniality school with contemporary work on platform labour, to illustrate how value is extracted from the data and labour generated by workers. This labour is a key mechanism in the valorisation process behind downstream AI use cases, despite its relative invisibility amongst the hype of recent advances in large scale algorithms such as ChatGPT. Third, we interrogate the values and knowledge epistemologies within the AI industry, tracking the continuity of colonial frames of Western modernity and Whiteness. Following the critical, socio-technical work of Crawford and Joler (2018) and Joler and Pasquinelli (2020), we specifically engage with the technical aspects of ML systems from data collection, model training and algorithmic outputs, to excavate the link between racist ideologies and grounded AI harms.

Throughout this article, the frameworks of coloniality and decolonial perspectives need to be understood in a nuanced and reflexive manner. Firstly, the framework of modernity/coloniality has its own tradition from which it has emerged in the American Subaltern Studies group, drawing from world systems theory, underdevelopment theory and the critical theory of the Frankfurt School (Bhambra, 2014). While the colonial matrix of power is a powerful theoretical framework from which to understand how AI perpetuates coloniality, it should also be situated within the broader developments in decolonial and postcolonial theory and other critical perspectives on technology and AI (Bhambra, 2014; Ricaurte, 2019). The interpretive frameworks of modernity/coloniality should also not be reduced to a series of over-simplified binaries which portray everything “Western” as bad and everything coming from the majority world as an emancipatory force for good. Decolonial thought aims to interrogate and disrupt Western systems of knowledge and ways of being, but this does not involve ascribing simplistic categories of “oppressor” and “oppressed” over complex issues and debates (Mohamed et al., 2020). Here we also acknowledge the need to contextualise Western and non-Western binaries within geopolitical realities emerging from the increasing dominance of both the US and China in technology development. China has both been exploited by European powers and has engaged in its own form of expansionism, including its African Policy and ambitions to become a leading superpower in AI and digital technology (Robles, 2018). Further research should interrogate the diversification of power within AI development, questioning how history and geopolitics shapes the way we interpret decolonial frameworks.

For Quijano, coloniality is not a model or object of study but a framework for subverting hegemonic and Western ways of knowing. Therefore, we end this article with a reflection on critical and decolonial scholarship that is working towards the dismantling of Western-centricity AI production. Scholars and activists are looking to non-Western and relational schools of thought such as Ubuntu ethics (a Sub-Saharan African philosophy) (Birhane, 2021; Mhlambi, 2020). Mhlambi (2020) argues that AI production is shaped by Western preconceptions of personhood based on rationality, which Ubuntu philosophy is diametrically opposed to. Introducing a critique of AI production through the lens of Ubuntu, he highlights: the exclusion of marginalised communities from design, the codification of biases within data, the misguided conception of technology as neutral (and therefore privileging Whiteness), the construction of human value through the individualised lens of data and

the concentration of power in the hands of a powerful elite. Mhlambi (2020) suggests a way forward through the lens of Ubuntu, encouraging the decentralisation of technology and acknowledging our human and ecological connectedness. Mhlambi and other scholars and activists have put forward a “Manifesto on AI,” (2021) calling for a future for AI that is decolonial and open to contributions from marginalised communities. Additionally, indigenous approaches to AI have included participatory initiatives to develop community owned language AI tools and mechanisms to protect the use and ownership of Māori data (Birhane et al., 2022; Hao, 2022). Other efforts have brought Indigenous groups together to articulate guidelines for Indigenous-Centred AI design and to work towards a multiplicity of Indigenous protocols across diverse community groups (Lewis, 2020). These efforts suggest that the project of decolonisation involves both “top-down” approaches of reflecting on different histories and epistemologies but most importantly, through “bottom up” approaches led by marginalised and subaltern groups (Cruz, 2021). Future attempts to resist and/or develop AI technologies require both reflecting on the inheritance of colonial legacies and experimenting with practices drawn from inter-cultural and existing critical praxis that break free from the “distorting mirror” of the colonial matrix of power (Quijano, 2001).

Acknowledgements We would like to thank Lola Alaska and SJ Zhang for their generous feedback and input into the article.

Author Contributions All authors contributed to the study conception and design. The first draft of the manuscript was written by JM and was then significantly revised by BW. All authors read and approved the final manuscript.

Funding The authors declare that no funds, grants, or other support were received during the preparation of this manuscript.

Data Availability Not applicable.

Declarations

Ethics Approval and Consent to Participate Not applicable. This is an observational study that does not involve research with human or animal participants. Ethical approval is not required.

Consent for Publication Not applicable.

Competing Interests The authors have no relevant financial or non-financial interests to disclose.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Abraham, D. (2017). *The elements of power: Gadgets, guns, and the struggle for a sustainable future in the rare metal age*. Yale University Press.
- Adams, R. (2021). Can artificial intelligence be decolonized? *Interdisciplinary Science Reviews*, 46(1–2), 176–197. <https://doi.org/10.1080/03080188.2020.1840225>
- Adas, M. (1989). *Machines as the measure of men: Science, technology, and ideologies of Western dominance*. Cornell University Press.
- Ali, S. H. (2014). Social and environmental impact of the rare earth industries. *Resources*, 3(1), 123–134. <https://doi.org/10.3390/resources3010123>
- Asare, J. G. (2022). The Dark Side Of ChatGPT. *Forbes*. Retrieved 6 February 2023, from <https://www.forbes.com/sites/janicegassam/2023/01/28/the-dark-side-of-chatgpt/>. Accessed 12 Nov 2023
- Avila, R. (2018). *Digital Sovereignty or Digital Colonialism.* Sur - International Journal on Human Rights. Retrieved February 6, 2023, from <https://sur.conectas.org/en/digital-sovereignty-or-digital-colonialism/>. Accessed 12 Nov 2023
- Avila, R. (2020). Against Digital Colonialism. In J. Muldoon & W. Stronge. *Platforming Equality: policy challenges for the digital economy*. Autonomy. Retrieved February 6, 2023, from <https://autonomy.work/wp-content/uploads/2020/09/Avila.pdf>. Accessed 12 Nov 2023
- Bechmann, A., & Bowker, G. C. (2019). Unsupervised by any other name: Hidden layers of knowledge production in artificial intelligence on social media. *Big Data & Society*, 6(1). <https://doi.org/10.1177/2053951718819569>
- Bechmann, A., & Geoffrey, C. B. (2019). Unsupervised by Any Other Name: Hidden Layers of Knowledge Production in Artificial Intelligence on Social Media. *Big Data & Society*, 6(1). <https://doi.org/10.1177/20539517188195>
- Beer, D. (2017). The social power of algorithms. *Information, Communication & Society*, 20(1), 1–13. <https://doi.org/10.1080/1369118X.2016.1216147>
- Bender, E. M., Timnit, G., Angelina, M., & Mitchell, M. (2021). On the dangers of stochastic parrots: Can language models be too big?. *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*. <https://doi.org/10.1145/3442188.3445922>
- Benjamin, R. (2019). *Race after technology: Abolitionist tools for the new Jim code*. Polity.
- Berg, J., Marianne, F., Ellie, H., Uma, R., & Silberman, M. S. (2018). *Digital Labour Platforms and the Future of Work: Towards Decent Work in the Online World. Report*. International Labour Organisation. Retrieved 5th February, 2023, from https://www.ilo.org/global/publications/books/WCMS_645337/lang--en/index.htm. Accessed 12 Nov 2023
- Bergman, A. S., Hendricks, L. A., Rauh, M., Wu, B., Agnew, W., Kunesch, M., Duan, I., Gabriel, I., & Isaac, W. (2023). Representation in AI Evaluations. In *Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency*.
- Bhambra, G. K. (2014). Postcolonial and decolonial dialogues. *Postcolonial Studies*, 17(2), 115–121. <https://doi.org/10.1080/13688790.2014.966414>
- Bird, E., Fox-Skelly, J., Jenner, N., Larbey, R., Weitkamp, E., & Weitkamp, A. (2020). *The Ethics of Artificial Intelligence: Issues and Initiatives*. European Parliament Scientific Foresight Unit. Retrieved 5th February, 2023, from [https://www.europarl.europa.eu/RegData/etudes/STUD/2020/634452/EPRS_STU\(2020\)634452_EN.pdf](https://www.europarl.europa.eu/RegData/etudes/STUD/2020/634452/EPRS_STU(2020)634452_EN.pdf). Accessed 12 Nov 2023
- Birhane, A. (2020). Algorithmic colonization of Africa. *SCRIPTed: A Journal of Law Technology & Society*, 17(2), 389–409. <https://doi.org/10.2966/scrip.170220.389>
- Birhane, A. (2021). Algorithmic injustice: a relational ethics approach. *Patterns*, 2(2). <https://doi.org/10.1016/j.patter.2021.100205>
- Birhane, A., Isaac, W., Prabhakaran, V., Diaz, M., Elish, M. C., Gabriel, I., & Mohamed, S. (2022). Power to the people? Opportunities and challenges for participatory AI. *Proceedings of the 2022 ACM conference on Equity & Access in Algorithms, Mechanisms, and Optimization*, 6, 1–8. <https://doi.org/10.1145/3551624.3555290>
- Birhane, A., Prabhu, V. U., & Kahembwe, E. (2021). *Multimodal datasets: misogyny, pornography, and malignant stereotypes*. ArXiv. Retrieved August 28, 2023, from <https://arxiv.org/abs/2110.01963>. Accessed 12 Nov 2023
- Birhane, A., Prabhu, V., Han, S., & Boddeti, V. N. (2023). *On Hate Scaling Laws For Data-Swamps*. ArXiv. Retrieved August 28, 2023, from <https://arxiv.org/abs/2306.13141>. Accessed 12 Nov 2023

- BNamericas. (2023). *Propuesta para nacionalizar minas de Chile supera primer escollo* BNamericas. Retrieved February 2, 2023, from <https://www.bnamericas.com/es/noticias/propuesta-para-nacionalizar-minas-de-chile-supera-primer-escollo>. Accessed 12 Nov 2023
- Brevini, B. (2020). Black Boxes, Not Green: Mythologizing artificial intelligence and omitting the environment. *Big Data & Society*, 7(2). <https://doi.org/10.1177/205395172093514>
- Brodzicki, T. (2021). *The Role of East and Southeast Asia in the Global Value Chain in Electronics*. S&P Global. Retrieved February 2, 2023, from <https://www.spglobal.com/marketintelligence/en/mi/research-analysis/the-role-of-east-and-southeast-asia-in-the-global-value-chain-.html>. Accessed 12 Nov 2023
- Buolamwini, J., & Gebru, T. (2018). Gender shades: Intersectional accuracy disparities in commercial gender classification. *Proceedings of the 2022 ACM Conference on Fairness, Accountability and Transparency*, 81, 77–91.
- Casati, R. (2013). *Contro Il Colonialismo Digitale*. I Robinson / Letture.
- Casilli, A. (2017). Digital labour studies go global: Toward a digital decolonial turn. *International Journal of Communication*, 11, 21.
- Cave, S., & Dihal, K. (2020). The whiteness of AI. *Philosophy & Technology*, 33(4), 685–703. <https://doi.org/10.1007/s13347-020-00415-6>
- Centre for Research on Foundation Models. (2023). *Our Mission*. Retrieved August 28, 2023, from <https://crfm.stanford.edu/>. Accessed 12 Nov 2023
- Chowdhury, H. (2023). Sam Altman's Big Problem? ChatGPT Needs to Get "woke" If He Wants Cash from Corporate America. *Business Insider India*. Retrieved February 6 2023, from https://www.businessinsider.in/tech/news/sam-altmans-big-problem-chatgpt-needs-to-get-woke-if-he-wants-cash-from-corporate-america/amp_articleshow/97586935.cms. Accessed 12 Nov 2023
- Clutton-Brock, P., Rolnick, D., Donti, P. L., & Kaack, L. H. (2021). Climate Change and AI. *Global Partnership on Artificial Intelligence*. Retrieved February 2, 2023, from <https://www.gpai.ai/projects/climate-change-and-ai.pdf>. Accessed 12 Nov 2023
- Coleman, D. (2018). Digital colonialism: The 21st century scramble for Africa through the extraction and control of user data and the limitations of data protection laws note. *Michigan Journal of Race & Law*, 24(2), 417–440.
- Coronil, F. (1997). *The magical state: Nature, money, and modernity in Venezuela*. University of Chicago Press.
- Couldry, N., & Mejias, U. A. (2019). *The costs of connection: how data is colonizing human life and appropriating it for capitalism*. Stanford University Press.
- Crawford, K. (2021). *Atlas of AI: Power*. Yale University Press.
- Crawford, K., & Joler, V. (2018) *Anatomy of an AI System*. Retrieved February 3, 2023, from <https://anatomyof.ai/>. Accessed 12 Nov 2023
- Cruz, C. C. (2021). Decolonizing philosophy of technology: Learning from bottom-up and top-down approaches to decolonial technical design. *Philosophy & Technology*, 34(4), 1847–1881. <https://doi.org/10.1007/s13347-021-00489-w>
- Dauvergne, P. (2022). Is artificial intelligence greening global supply chains? Exposing the political economy of environmental costs. *Review of International Political Economy*, 29(3), 696–718. <https://doi.org/10.1080/09692290.2020.1814381>
- Elliott, V., & Parmar, T. (2020). *The Despair and Darkness of People Will Get to You*. Rest of World. Retrieved February 3, 2023, from <https://restofworld.org/2020/facebook-international-content-moderators/>. Accessed 12 Nov 2023
- Escobar, A. (2007). Worlds and knowledges otherwise. *Cultural Studies*, 21(2–3), 179–210. <https://doi.org/10.1080/09502380601162506>
- Fei, F., Tambe, T., Dilkina, B., & Plumptre, A. J. (2019). *Artificial intelligence and conservation*. Cambridge University Press.
- Forsythe, D. (2001). *Studying those who study us: An anthropologist in the world of artificial intelligence*. Stanford University Press.
- Fuchs, C. (2014). *Digital labour and Karl Marx*. Routledge.
- Gago, V., & Mezzadra, S. (2017). A Critique of the Extractive Operations of Capital: Toward an Expanded Concept of Extractivism. *Rethinking Marxism*, 29(4), 574–591. <https://doi.org/10.1080/08935696.2017.1417087>
- Giuliano, R. (2020). Echoes of myth and magic in the language of artificial intelligence. *AI & Society*, 35(4), 1009–1024. <https://doi.org/10.1007/s00146-020-00966-4>

- Global Witness. (2022). *Myanmar's Poisoned Mountains*. Global Witness. Retrieved February 2, 2023, from <https://www.globalwitness.org/en/campaigns/natural-resource-governance/myanmars-poisoned-mountains/>. Accessed 12 Nov 2023
- Graham, M., & Dittus, M. (2022). *Geographies of Digital Exclusion: Data and Inequality*. Pluto Press.
- Gray, M., & Suri, S. (2019). *Ghost work: How to stop Silicon Valley from building a new global underclass*. Houghton Mifflin Harcourt.
- Grosfoguel, R. (2007). The epistemic decolonial turn. *Cultural Studies*, 21(2–3), 211–223. <https://doi.org/10.1080/09502380601162514>
- Hao, K. (2022). *A New Vision of AI for the People*. MIT Technology Review. Retrieved August 29, 2022, from <https://www.technologyreview.com/2022/04/22/1050394/artificial-intelligence-for-the-people/>. Accessed 12 Nov 2023
- Hatton, E. (2017). Mechanisms of invisibility: Rethinking the concept of invisible work. *Work, Employment & Society*, 31(2), 336–351. <https://doi.org/10.1177/0950017016674894>
- Heikkilä, M. (2022). *How it feels to be sexually objectified by an AI*. Technology Review. Retrieved August 27, 2023, from <https://www.technologyreview.com/2022/12/13/1064810/how-it-feels-to-be-sexually-objectified-by-an-ai/>. Accessed 12 Nov 2023
- Hogarth, I. (2018). *AI Nationalism*. Retrieved February 13, 2023, from <https://www.ianhogarth.com/blog/2018/6/13/ai-nationalism>. Accessed 12 Nov 2023
- Joler, V., & Pasquinelli, M. (2020). *AI as Instrument of Knowledge Extraction*. Noosope.ai. Retrieved August 27, 2023, from <https://noosope.ai/>
- Hurtado, J. (2022). *Economía - Senado de México aprueba reforma para nacionalizar la explotación de litio*. France 24. Retrieved February 2, 2023, from <https://www.france24.com/es/programas/econom%C3%ADa/20220420-mexico-nacionalizacion-explotacion-litio>. Accessed 12 Nov 2023
- Irani, L. (2019). *Chasing innovation*. Princeton University Press.
- Irani, L. (2015). *Justice for "Data Janitors"*. Public Books. <https://www.publicbooks.org/justice-for-data-janitors/>. Accessed 3 Feb 2023
- Jasanoff, S. (2004). The idiom of co-production. In S. Jasanoff (Ed.), *States of knowledge: The co-production of science and the social order* (pp. 12–23). Routledge.
- Jones, P. (2021). *Work without the worker*. Verso.
- Jones, P., & Muldoon, J. (2022). *Rise and grind - microwork and hustle culture in the UK*. Autonomy.
- Kaplan, A., & Haenlein, M. (2020). Rulers of the World, Unite! The challenges and opportunities of artificial intelligence. *Business Horizons*, 63(1), 37–50.
- Kemper, J., & Kolkman, D. (2019). Transparent to whom? No algorithmic accountability without a critical audience. *Information, Communication & Society*, 22(14), 2081–2096. <https://doi.org/10.1016/j.bushor.2019.09.003>
- Kiggins, R. D. (2015). *The political economy of rare earth elements: Rising powers and technological change*. Palgrave Macmillan.
- Kuek, S. C., Paradi-Guilford, C., Fayomi, Toks., Imaizumi, S., & Ipeiritos, P. (2015). *The Global Opportunity in Online Outsourcing*. World Bank. Retrieved February 2, 2023, from <https://thedocs.worldbank.org/en/doc/212201433273511482-0190022015/The-Global-Opportunity-in-Online-Outsourcing>. Accessed 12 Nov 2023
- Kwet, M. (2019a). *Digital Colonialism: South Africa's Education Transformation in the Shadow of Silicon Valley*. PhD dissertation, Rhodes University.
- Kwet, M. (2019b). Digital colonialism: US empire and the new imperialism in the Global South. *Race & Class*, 60(4), 3–26. <https://doi.org/10.1177/0306396818823172>
- Lewis, J. E. (2020). *Position Paper: Indigenous Protocol and Artificial Intelligence*. Honolulu, Hawai'i: The Initiative for Indigenous Futures and the Canadian Institute for Advanced Research (CIFAR). Retrieved August 28, 2023, from <https://www.indigenous-ai.net/position-paper>. Accessed 12 Nov 2023
- Lugones, M. (2016). The coloniality of gender. In W. Harcourt (Ed.), *The Palgrave Handbook of Gender and Development: Critical Engagements in Feminist Theory and Practice*. Palgrave Macmillan UK.
- Maldonado-Torres, N. (2007). On the coloniality of being: Contributions to the development of a concept. *Cultural Studies*, 21(2–3), 240–270. <https://doi.org/10.1080/09502380601162548>
- Manyfesto. (2021). *AI Decolonial Manifesto*. Retrieved August 28, from <https://manyfesto.ai/>. Accessed 12 Nov 2023
- McQuillan, D. (2022). *Resisting AI: an anti-fascist approach to artificial intelligence*. Policy Press.
- Mezzadra, S., & Neilson, B. (2017). On the multiple frontiers of extraction: Excavating contemporary capitalism. *Cultural Studies*, 13(2–3), 1–20. <https://doi.org/10.1080/09502386.2017.1303425>

- Mhlambi, S. (2020). *From rationality to relationality: Ubuntu as an Ethical & Human Rights Framework for Artificial Intelligence Governance*. Carr Center Discussion Paper.
- Miceli, M., & Posada, J. (2022). The data-production Dispositif. *Proceedings of the ACM on Human-Computer Interaction*, 6. <https://doi.org/10.48550/arXiv.2205.11963>
- Mignolo, W. D. (2007). Introduction: Coloniality of power and de-colonial thinking. *Cultural Studies*, 21(2–3), 155–167. <https://doi.org/10.1080/09502380601162498>
- Mignolo, W. D. (2011). *The darker side of western modernity: Global futures*. Duke University Press.
- Mignolo, W. D. (2012). *Local histories/global designs: Coloniality, Subaltern knowledges, and border thinking*. Princeton University Press.
- Mignolo, W. D., & Walsh, C. E. (2018). *On decoloniality: Concepts, analytics*. Duke University Press.
- Mittelstadt, B., Allo, P., Taddeo, M., Wachter, S., & Floridi, L. (2016). The ethics of algorithms: Mapping the debate. *Big Data & Society*, 3(2), 205395171667967. <https://doi.org/10.1177/2053951716679679>
- Mohamed, S., Png, M.-T., & Isaac, W. (2020). Decolonial AI: Decolonial theory as sociotechnical foresight in artificial intelligence. *Philosophy & Technology*, 33, 659–684. <https://doi.org/10.1007/s13347-020-00405-8>
- Mumford, D. (2022). Data colonialism: Compelling and useful, but whither epistemes? *Information, Communication & Society*, 25(10), 1511–1516. <https://doi.org/10.1080/1369118X.2021.1986103>
- Newlands, G. (2021). Lifting the curtain: Strategic visibility of human labour in AI-as-a-Service. *Big Data & Society*, 8(1), 205395172110160. <https://doi.org/10.1177/20539517211016026>
- Niarchos, N. (2021). *The Dark Side of Congo's Cobalt Rush*. *The New Yorker*, Retrieved August 30, 2023, from <https://www.newyorker.com/magazine/2021/05/31/the-dark-side-of-congos-cobalt-rush>. Accessed 12 Nov 2023
- Noble, S. (2018). *Algorithms of oppression: How search engines reinforce racism*. New York University Press.
- Nost, E., & Colven, E. (2022). Earth for AI: A political ecology of data-driven climate initiatives. *Geoforum*, 130, 23–34.
- OECD. (2022). *Measuring the Environmental Impacts of Artificial Intelligence Compute and Applications: The AI Footprint*. OECD. Retrieved August 30, 2023 from <https://www.oecd.org/publications/measuring-the-environmental-impacts-of-artificial-intelligence-compute-and-applications-7baf571-en.htm>. Accessed 12 Nov 2023
- OECD.AI. (2023). *OECD Working Party and Network of Experts on AI*. OECD.AI. Retrieved August 24, 2023, from <https://oecd.ai/en/network-of-experts/working-group/1136>. Accessed 12 Nov 2023
- OpenAI. (2022). *DALL-E 2 Preview - Risks and Limitations*. GitHub. Retrieved August 17, 2023, from <https://github.com/openai/dalle-2-preview/blob/main/system-card.md#explicit-content>. Accessed 12 Nov 2023
- Partnership on Artificial Intelligence (PAI). (2021). *Responsible sourcing of data enrichment services*. Partnership on AI.
- Pasquinelli, M., & Joler, V. (2021). The Noosphere manifested: AI as instrument of knowledge extractivism. *AI & Society*, 36, 1263–1280. <https://doi.org/10.1007/s00146-020-01097-6>
- Patterson, D., Gonzalez, J., Hölzle, U., Le, Q., Liang, C., Munguia, L.-M., Rothchild, D., So, D., Texier, M., & Dean, J. (2022). The carbon footprint of machine learning training will plateau, then shrink. *Computer*, 55(7), 18–28. <https://doi.org/10.1109/MC.2022.3148714>
- Perrigo, B. (2023). Exclusive: OpenAI Used Kenyan Workers on Less Than \$2 Per Hour to Make ChatGPT Less Toxic. *Time*. Retrieved February 3, 2023, from <https://time.com/6247678/openai-chatgpt-kenya-workers/>. Accessed 12 Nov 2023
- Pew Research. (2021). Internet/Broadband Fact Sheet. *Pew Research Center: Internet, Science & Tech*. Retrieved February 6, 2023, from <https://www.pewresearch.org/internet/fact-sheet/internet-broadband/>. Accessed 12 Nov 2023
- Png, M.-T. (2022). At the tensions of South and North: Critical roles of global south stakeholders in AI governance. *Proceedings of the 2022 ACM Conference on Fairness, Accountability and Transparency*. <https://doi.org/10.1145/3531146.3533200>
- Prabhu, V. U., & Birhane, A. (2020). Large image datasets: A pyrrhic win for computer vision?. *arXiv*. Retrieved August 27, 2023, <https://arxiv.org/abs/2006.16923>. Accessed 12 Nov 2023
- Quijano, A. (2000). Coloniality of power and eurocentrism in Latin America. *International Sociology*, 15(2), 215–232. <https://doi.org/10.1177/0268580900015002005>
- Quijano, A. (2001). Colonialidad del poder, globalización y democracia. *Utopías, nuestra bandera: revista de debate político*, 188, 97–123.

- Quijano, A. (2007). Coloniality and modernity/rationality. *Cultural Studies*, 21(2–3), 168–178. <https://doi.org/10.1080/09502380601164353>
- Quijano, A. (2016). “‘Bien Vivir’ –Between ‘Development’ and the De/Coloniality of Power”. *Alternautas* 3(1). <https://journals.warwick.ac.uk/index.php/alternautas/article/view/1023/893>. Accessed 12 Nov 2023
- Raisch, S., & Krakowski, S. (2021). Artificial intelligence and management: The automation–augmentation paradox. *Academy of Management Review*, 46(1), 192–210.
- Raji, I. D., Bender, E. M., Paullada, A., Denton, E., & Hanna, A. (2021). *AI and the everything in the whole wide world benchmark*. arXiv. Retrieved August 28th, 2023, from arXiv:2111.15366.
- Ricourte, P. (2019). Data epistemologies, the coloniality of power, and resistance. *Television & New Media*, 20(4), 350–365. <https://doi.org/10.1177/1527476419831640>
- Ricourte, P. (2022). Ethics for the majority world: AI and the question of violence at scale. *Media, Culture & Society*, 44(4), 726–745. <https://doi.org/10.1177/01634437221099612>
- Robbins, S., & van Wynsberghe, A. (2022). Our new artificial intelligence infrastructure: Becoming locked into an unsustainable future. *Sustainability*, 14(8), 4829. <https://doi.org/10.3390/su14084829>
- Robles, P. (2018). *China Plans to Be a World Leader in Artificial Intelligence by 2030*. South China Morning Post. Retrieved February 14, 2023, from <https://multimedia.scmp.com/news/china/article/2166148/china-2025-artificial-intelligence/index.html>. Accessed 12 Nov 2023
- Schwartz, R., Dodge, J., Smith, N., & Etzioni, O. (2020). Green AI. *Communications of the ACM*, 63(12), 54–63. <https://doi.org/10.1145/3381831>
- Seaver, N. (2017). Algorithms as culture: Some tactics for the ethnography of algorithmic systems. *Big Data & Society*, 4(2). <https://doi.org/10.1177/2053951717738104>
- Srnicek, N. (2022). Data, compute, labor. In M. Graham & F. Ferrari (Eds.), *Digital work in the planetary market*. Oxford University Press.
- Star, S. L. (1999). The ethnography of infrastructure. *American Behavioural Scientist*, 43(3), 377–391. <https://doi.org/10.1177/00027649921955326>
- Stephany, F., Kässi, O., Rani, U., & Lehdonvirta, V. (2021). Online Labour Index 2020: New ways to measure the world’s remote freelancing market. *Big Data & Society*, 8(2). <https://doi.org/10.1177/20539517211104324>
- Strubell, E., Ganesh, A., & Mccallum, A. (2019). Energy and Policy Considerations for Deep Learning in NLP. *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*. Association for Computational Linguistics.
- Thatcher, J., O’Sullivan, D., & Mahmoudi, D. (2016). Data colonialism through accumulation by dispossession: New metaphors for daily data. *Environment and Planning D: Society and Space*, 34(6), 990–1006.
- Tubaro, P., & Casilli, A. (2019). Micro-work, artificial intelligence and the automotive industry. *Journal of Industrial and Business Economics*, 46(3), 333–345. <https://doi.org/10.1007/s40812-019-00121-1>
- Tubaro, P., Casilli, A., & Coville, M., (2020). The trainer, the verifier, the imitator: Three ways in which human platform workers support artificial intelligence. *Big Data & Society*, 7(1). <https://doi.org/10.1177/2053951720919776>
- Valdivia, A. (2023). *Silicon Valley and the Environmental Costs of AI*. Political Economy Research Centre. Retrieved February 2, 2023, from https://www.perc.org.uk/project_posts/silicon-valley-and-the-environmental-costs-of-ai/. Accessed 12 Nov 2023
- Vinuesa, R., Azizpour, H., Leite, I., Balaam, M., Dignum, V., Domisch, S., Felländer, A., Langhans, S. D., Tegmark, M., & Nerini, F. F. (2020). The role of artificial intelligence in achieving the sustainable development goals. *Nature Communications*, 11(1), 233. <https://doi.org/10.1038/s41467-019-14108-y>
- Vock, I. (2022). *ChatGPT Proves That AI Still Has a Racism Problem*. New Statesman. Retrieved February 8, 2023, from <https://www.newstatesman.com/quickfire/2022/12/chatgpt-shows-ai-racism-problem>. Accessed 12 Nov 2023
- Westra, L., & Lawson, B. (2001). *Faces of environmental racism: Confronting issues of global justice* (2nd ed.). Rowman & Littlefield Publishers.
- World Bank. (2023). *Individuals Using the Internet (% of Population)*. World Bank. Retrieved February 6, 2023, from <https://data.worldbank.org/indicator/IT.NET.USER.ZS>. Accessed 12 Nov 2023
- Zhang, D., Maslej, N., Brynjolfsson, E., Etchemendy, J., Lyons, T., Manyika, J., Ngo, H., Niebles, J. C., Sellitto, M., Sakhaee, E., Shoham, Y., Clark, J., & Perrault, R. (2022). *The AI Index 2022 Annual Report*. Stanford University Human-Centered Artificial Intelligence. Retrieved February 2, 2023, from <https://aiindex.stanford.edu/ai-index-report-2022/>. Accessed 12 Nov 2023

Zuboff, S. (2019). *The age of surveillance capitalism: The fight for a human future at the new frontier of power*. Public Affairs.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.