Moral Responsibility and AI Extractivism
A Structural Analysis of AI Supply Chains

By

Deepshikha Sharma
(S3226387)

A Master's Thesis submitted for the partial fulfilment of the degree
MSc Philosophy of Science, Technology and Society - PSTS
(201300088)

To the

Faculty of Behavioural, Management, and Social Sciences,
University of Twente, Enschede, the Netherlands

August 12, 2025

Supervisor: Dr. Dominic Lenzi
Second Reader: Dr. Michel Bourban

Contents

In Memoriam

R. K. Sharma
Caring father, steadfast friend

All ink I spill is tinged with your love and support.

मेरो प्यारो बुबालाई,

मेरो गद्यको लहरमा तैरिने क्षमता
तपाईंले दिएको अक्षरको नाउँले गर्दा हो ।

Your fledgling

तपाईंकी झ्याउकिरी

Acknowledgements

Completing this thesis was one of the most challenging things I have done. It was only possible because of the support I received – whether that be through academic mentorship, intentional acts of care, or words of solace.

I am indebted to my thesis supervisor, Dr. Dominic Lenzi, for his charitable interpretations, incisive comments and extensive edits. He gave me the autonomy to explore topics and frameworks in my own way and at my own pace. And if I veered off course, he would steer me back through rigorous feedback and sincere conversations. His patience in answering my scattered questions about ethico-political philosophy, academic writing, and interdisciplinary career trajectories is unparalleled.

I would also like to thank my second reader, Dr. Michel Bourban, for his time and his feedback, as well as his reading recommendations at the start of my thesis, which significantly influenced the direction of my project.

I am so grateful for my family and the strength they provide, even when we're oceans apart. Muma, ThulPhu, and SanPhu – thank you for your reassurance, affirmations, and hopefulness. It is only by emulating your resilience that I've finished this thesis. And although he is not here anymore, I am so grateful for my father and his unrelenting support through all the choices I have made. Thank you, Buba, for instilling a deep appreciation for books and ideas in me. While I started this thesis in conversation with him about my ideas, as he emailed me articles about the topic, it's tragic that he did not get to read it.

As I've completed degrees, switched jobs, and moved cities and countries four times in the last 6 years, my partner, Karma, has been my rock. Thank you for keeping me sane through transitions and growing pains.

For someone far away from home, my friends in Enschede were crucial in providing me with care through weekly check-ins, long walks, thesis rants, home-cooked meals and shared coffee – especially Anisat, Animesh, Mert, Sebastian, and Thanuja. Overall, I feel a heartfelt appreciation for all my friends, old and new. Thank you for bearing with my grief-stricken strangeness, providing me with shoulders to cry on, opportunities to relax, and nudges to reawaken.

And here's to my favourite sociologist and sister-figure, Rama. I would not be here without you. Your brilliance, kindness and wit inspire me.

Summary

AI ethics, despite its heterogeneous voices and perspectives, can fall short due to its principlistic orientation that only fixates on the technical aspects of the technology. This has resulted in shifts of theorising that not only acknowledge the technology's sociotechnical embeddedness but also its structural aspects. With theorists hailing the structural turn in AI ethics, AI is being recognised as a global assemblage of natural resources, human labour and data from all around the world. However, the supply chains that funnel the resources for AI development result in a range of extractivist harms. And since these issues are beyond the boundaries of a single company entangled in a complex web made up of several supply chains, the problem of 'many hands' gets amplified on a global scale, yet has a disproportionate effect on the Majority World, with historical, complicating notions of moral accountability. Hence, the primary research question of this project is: how should one attribute moral responsibilities for the structural injustices involving extractivism across AI supply chains? To answer this question, this project borrows approaches from critical AI studies and philosophical literature on structural injustice, engaging in applied philosophy. After highlighting that the resources used in AI development are sourced through and uphold an extractive logic, I argue that AI ethics must acknowledge that the infrastructural harms, emerging from natural resource excavation, and exploitative harms, emerging from labour exploitation, are undivorced from the algorithmic harms as they are rooted in the same structure. I analyse these harms through the typologies of structural injustice outlined by Young (2011) and McKeown (2024b), to infer that natural resource extraction for AI development is a case of avoidable structural injustice, due to the foresseability of the harm of the mineral excavation and that labour exploitation for AI development is a case of deliberate structural injustice, as the structures that result in these harms are deliberately maintained. Furthermore, I argue that agents with power can be attributed moral responsibility for these structural injustices, as they can maintain, influence, and disrupt the structure that reproduces injustices. By tracing power through the supply chains and finding it concentrated at the downstream tier, I conclude that multinational corporations can be held morally responsible for the structural injustices involving extractivism across AI supply chains, due to the power they have in influencing and maintaining structures that result in injustice. Hence, positioning myself within the power-conscious structural turn in AI ethics, I argue for moral accountability by MNCs for the extractivist harms emerging from AI production and hint at the additional importance of the ordinary individual's political responsibility in bringing about structural change, also suggesting AI minimisation in terms of scale and necessity.

Introduction

The Artificial Intelligence (AI)[1] boom has ushered in extreme narratives of extinction and/or salvation, urging ethicists to mull over its unjust pasts, the uncertain present and alternative futures. With over 160 AI ethics Guidelines having cropped up after only 2018 (AlgorithmWatch, 2020), academia, research institutes, governance bodies, industry and civil society organisations address these issues urgently. Urging ethical considerations to go beyond technical fixes, critical discourse on AI ethics has emphasised how systemic issues are mirrored and magnified by computational systems (Benjamin, 2019; Vallor, 2024), resulting in intersectional discrimination, (Buolamwini & Gebru, 2018), automated welfare disparity (Eubanks, 2018), and racial penalisation (Benjamin, 2019). This shift in focus, from algorithmic output to systemically induced algorithmic harm[2], has run parallel to a growing literature about the production of AI technologies that require various factors and resources like environmental resources and human labour relying on planetary networks (Crawford, 2021; Muldoon, Graham and Cant, 2024). Rigorous intersectional analyses that attempt to think jointly about these parallel conversations are steadily increasing (Crawford, 2021; Peña & Varon, 2021; Ricaurte, 2022). These accounts have brought structures to the forefront, also spotlighting the grave impacts of AI technologies on the Majority World (MW)[3] (Ricaurte, 2022), from where natural resources, human labour and data are funnelled through supply chains resulting in AI technologies. These chains are plagued with instances of harm, underpinned by an extractive logic.

However, since these issues are beyond the boundaries of a single company entangled in a complex web made up of several supply chains, the problem of 'many hands' (Frank & Klincewicz, 2024) get amplified on a global scale, with historical, social and political intersections complicating notions of moral accountability. When several agents and groups collect/supply natural resources, human labour, and data

[1] By AI, this thesis refers to machine learning (ML), which uses statistics and probability to learn from large datasets, making predictions on new or unseen data (Mittelstadt et al., 2016). It also attempts to highlight the notion of scale, as larger models are the trend despite debates about their efficacy (Heikkilä, 2024). Large-scale AI models have trillions of parameters and train on vast datasets, enabling them to tackle complex, general-purpose tasks. This requires more resources, more computing power, more human labour and more data.

[2] In this thesis, I employ harm as a normative concept where setbacks to the interests of an individual are unjust or morally indefensible (Kamm, 2020). The relation between the different kinds of harms (infrastructural, social, algorithmic, extractivist, and structural) will be discussed in Chapter 2.

[3] This term is coined by Shahidul Alam (2007) as an alternative to terms like "Third World" or "Developing world", which propose a linear vision of development. Additionally, it brings to light how planetary decisions represent a tiny fraction of humankind. This term can overlap with the Global South (GS) but overcomes the problem regarding how individuals in both South/North categories can belong to the spheres of affluence and poverty. However, this also comes with its issues as it de-emphasises geography. Also, with countries that belong to the GS bloc, ergo MW, like India being categorised as the Compute North due to the availability of high-rise chips (Lehdonvirta et al., 2024), adds to the fuzziness of these categories – pointing to the fact that the MW is not a monolith or a singularity but rather people forming a bloc on historical terms or as a political compass (Amrute et. al, 2022).

all across the world, working asynchronously as per the constraints of their immediate surroundings, it becomes challenging to identify the responsible agent. Hence, the primary research question of this project is: how should one attribute moral responsibilities for the structural injustices involving extractivism across AI supply chains? This question will be broken down into sub-questions that motivate each chapter: First, it asks what the existing AI ethics literature says about the wider supply chains of AI. Second, it asks what becomes ethically pertinent when the extractivist logic conditions the AI supply chain. Third, it asks how structural injustices, involving extractivism, manifest in the production of AI technologies. Finally, it examines what this implies for the attribution of moral responsibility across the AI supply chain.

To answer this question, this project borrows approaches from critical AI studies and philosophical literature on structural injustice, engaging in applied philosophy that utilises philosophical theories and concepts to address problems in real-world contexts. Particularly, it proceeds in two steps. Firstly, it foregrounds its analysis in literature that reconceptualises AI as a global assemblage embedded in larger structures to flag the extraction of natural resources, human labour, and data as resulting in harm. While disproportionately affecting the MW, these harms are a result of trans-continental supply chains leaving a glaring gap about who ought to be deemed responsible. This leads to the second step, where these extractivist harms will be analysed through the typologies of structural injustice outlined by Young (2011) and McKeown (2024b), where individuals connected to the structure may be politically responsible. However, agents with power can be attributed moral responsibility. By tracing power through the structurally unjust supply chains and finding them concentrated at the downstream tier, this project argues that corporate actors can be held morally responsible for the extractivist harms related to natural resource extraction and labour exploitation, as the structural injustices that occur are avoidable as well as deliberate, respectively.

Chapter 1, while providing a general scope of the vast and heterogeneous field, identifies gaps in the field voiced by critics of AI ethics, in particular the limitations of general principlist frameworks, which neglect the “hidden social and ecological costs of AI” (Hagendorff, 2020, p. 8). With scholars attempting to think beyond the machine assembly line, the chapter emphasises a shift towards the ‘ structural turn’ in AI ethics (Bolte & van Wynsberghe, 2024), which focuses on social and environmental concerns with systemic ensembles becoming sites of intervention. This also highlights the infrastructural aspects of the technology, which are built through extensive supply chains. Thus, Chapter 1 positions the thesis to echo the provocation of decolonial feminist thinkers like Ricaurte (2022) who emphasise the need to gauge AI

harms globally and on every level, “transcend[ing] the dichotomy between the tree and the forest” (728) – that also includes structures and AI supply chains.

Chapter 2 expands on the notion that AI supply chains result in extractivist harms. With extractivism being the driving logic not only to accumulate, exploit and dispossess but also to redefine what is a valuable resource, this axis of operation is evident in AI production as well. The AI extractivist trifecta, first identified by Crawford (2021), being natural resources, human labour and data, unduly impacts the MW as its states and corporations are very much entangled in the production of large AI systems. However, separate analyses of infrastructural/exploitative harms and algorithmic harms neglect the fact that these harms are linked to larger structures in place, which are reproduced and reinforced by different logics sustaining the structure. Thus, Chapter 2 provides various examples and descriptions of harms to argue for an undivorced analysis of infrastructural/exploitative and algorithmic harms for the production of ethical AI.

Chapter 3 applies the structural injustice framework to the extractivist harms. With previous attempts to analyse AI technologies exclusively focusing on the examples of algorithmic harms and only through the framework of Iris Marion Young (Lin & Chen, 2022; Himmelreich & Lim, 2022; Browne, 2023), this thesis interrogates infrastructural and exploitative harms, driven by extractivism, while relying on the power-centric and pluralist framework of Maeve McKeown (2024b). Adding to growing literature that analyses AI technologies through Structural Injustice frameworks, the chapter classifies AI extractivism in two ways: where natural resource extraction is an avoidable structural injustice as it is foresseable due to years of reporting on mining, especially about green transition attempts and the questionable rapid development of data centres; and where labour exploitation is a ‘deliberate’ structural injustice where powerful entities sustain structures that bring them profit and incentivise other players to follow the same rules. Thus, Chapter 3 provides a normative analysis of the examples discussed in the previous chapter using the philosophical framework of Structural Injustice.

Chapter 4 argues for a revised understanding of moral responsibility for structural injustices created by AI, focusing on the responsibility of large MNCs. By limiting the scope of inquiry to the extractivist logic sustaining the structure and scaffolding of a framework that acknowledges the global distribution of actions and agents, this view hinges on who has the most power. Since the agents with the most power can extensively manoeuvre these widespread structures, stabilising or disrupting them for their vested

interests, tracing the power across the supply chain is essential to identify agents who are morally responsible. Moreover, although distinct and disparate agents that act in complacency and conformity to the structure, with no direct contact to the most powerful, it is the MNCs, such as Big Tech giants like Amazon or Google, that must be held morally responsible as collective agents for these structural harms in question. Hence, chapter 4 adds to growing calls to not only hold powerful agents accountable for the wider harms of AI, but also emphasises the need for a structural view of responsibility which includes past injustice, navigating the uncertain present and building alternative futures.

Chapter 1

The Structural Turn in AI Ethics

While the "AI boom" (Violino, 2024) has resulted in benefits like tackling complex decades-old scientific problems (McDonough, 2024), ominous predictions and realities are also being emphasised. This includes warnings about fatal superhuman intelligence (Bostrom, 2020), job displacement (Moradi and Levy, 2020), disinformation (Bontridder and Poullet, 2021), surveillance (Saheb, 2022), data-harvesting (Veliz and Law, 2023) and algorithmic discrimination (Albert and Delano, 2023). These concerns have led to the development of the field of AI ethics.[4] However, the dominant modes of theorising/practising AI ethics can fall short. Hence, in this chapter, I argue that structural analysis is required to assess the various processes involved in the production of AI technologies, especially supply chains. In this chapter, §1.1 provides a general outline of the scope of AI ethics, where principlism is currently the dominant mode of theorising about responsibility. However, this obscures the sociotechnical embeddedness of the system, as demonstrated in §1.2, and factors beyond the machine assembly line, as argued in §1.3. This would include the structures that these technologies are conditioned by. Hence, in §1.4, the chapter introduces the 'structural turn' in AI ethics that highlights environmental impacts and power relations as necessary when considering ethical implications of AI. Positioning the thesis in line with literature on the structural turn, this section stresses how infrastructural elements and invisibilised human labour are brought to the forefront, as well as supply chains, with §1.5 summarising the chapter.

1.1 The Scope of AI Ethics

While AI ethics is a burgeoning subfield within applied ethics, it is heterogeneous in its voices and perspectives due to the range of parties involved in co-creating this field. Many institutes and organisations, within academia, industry, civil society, governments, and international bodies, contribute to the ethical discourse on technologies under the umbrella of AI ethics. These take shape in various formats such as reports, white-papers, self-commitment documents, checklists, questionnaires, and

[4]Although the scholarly discussion is relatively recent, the origin of AI ethics has been traced back to mid-20th-century literary work, with some scholars going as far as Ancient Greece. While such stories help capture socio-political imaginaries, this thesis will only focus on the methodical frameworks created by industry, research institutes, academia, states, and civil organisations. For a brief history about AI ethics, see Borenstein et. al (2021)

scholarly papers (Heillinger, 2022), with the private and public sector organisations having (nearly) equivalent numbers of documents (Jobin et al., 2019). In a scoping review conducted by Jobin et al. (2019), it was noted that certain areas, like Central Asia, Africa, South and Central America, were underrepresented, particularly in grey literature, with policy or industry-centric debates not being conducted equally on a global scale. Nonetheless, there has been a growing interest in ethical frameworks and AI governance models that are sensitive to transcultural interests (UNESCO, 2024), with many of the nations in question developing their policies and frameworks.

With many seminal contributions already in the AI ethics discourse, the main themes that emerge in AI ethics are that of the alignment of plural and incompatible values in societies (Huang et. al, 2024); value misalignment between humans and machine (Schwerzmann and Campolo, 2025); theory and practice gap in embedding values in systems (van de Poel, 2020); algorithmic opacity and its effect on human decision-making (Peters, 2023); responsibility gaps in human-machine workings (Nyholm, 2025); privacy violations due to mass data collection (Véliz, 2021); (unintended) discrimination, biases and exclusions that exacerbate social injustice (Benjamin, 2019); and competing concerns about AI as either malicious superintelligent agents (Bostrom, 2020) or relation-bound moral patients (Danaher, 2019). Attempts to tackle these questions and put them into practice can result in differing methods depending on the disciplinary background and interests.[5]

However, despite diverging methods and recommendations, Heilinger (2022, p. 61) argues that ethical reflections across formal documents generally consist of three kinds of assessments: first, of risks and opportunities; second, rules and principles; and third, visions and ideals of AI. Hence, the dominant approach in AI ethics is observed to be principlism, where principles such as transparency, justice, non-maleficence, responsibility and privacy are enlisted to create guidelines aimed at engineers or organisations designing AI[6]. This can also take the form of frameworks that include combining high-level values with stakeholder values to be operationalised within technical systems (Ryan et al., 2024). However, there has been a growing consensus on the gaps within AI ethics due to these reductive approaches, demanding a shift in ways of theorising AI ethics.

[5] The approach to designing an LLM can differ significantly between a large corporation and a civil organisation. Commercial LLMs can use web-scraped data and incorporate content moderation and human feedback to reduce offensive outputs, focusing on scalability and profitability. Civil organisations may engage in participatory design to assess risks or prioritise data minimisation for anonymity. Ultimately, while both types of LLMs aim to incorporate ethics, they prioritise different values, goals, and user needs—commercial LLMs focusing on accuracy and scalability, and civil LLMs emphasising privacy and empathetic support.

[6] Another subfield in applied ethics that relies on principlism is bioethics, where Beauchamp and Childress' (2001) framework has been adapted and applied widely. Floridi et. al (2018) have made an explicit principlist connection between Bioethics and AI ethics, although in the latter field newer principles keep getting added to the list as technologies increase in complexity.

1.2 The Gaps in AI Ethics

The gaps within AI ethics are being increasingly flagged by thinkers dissatisfied by how industries as well as scholars instrumentalise ethics to go ahead with business-as-usual (Schultz et. al, 2024). AI ethics within the corporate landscape has also led to the acknowledgement of the "trivialization" of moral structures as reductive tool boxes for corporations fall back on as heuristics or marketing, increasingly being seen in assessments of these principles (Bietti, 2021). Words and phrases such as 'useless' (Munn, 2022), 'toothless' (Rességuier and Rodrigues, 2020) ', fig leaf function' (Hagendorff, 2020), 'invention of the AI industry (Ochigame, 2019), puppet of industry stakeholders' (Abdalla and Abdalla, 2020) make such dissatisfactions evident. Pertinent critiques regarding these gaps emphasise how the dominant approach of principlism can lead to ethics washing.[7]

A popular approach resulting in around 200 ethical guidelines (Correa et. al, 2023), principlism is adopted by many practitioners delving into AI ethics. With recurrent principles like transparency or privacy, there have been multiple critiques against these elaborate checklists, frameworks and principles. However, these principles are enforced divergently as per the practitioner's interests. Mittelstadt (2019) argues that these high-level principles fail to provide concrete action-guiding recommendations and do not address the normative tensions in these oft-cited concepts, leading to incompatible goals. Terms like "justice" can be defined in ways to suit the corporation's interests, making these principles vague (Munn, 2022). Alarmingly, thinkers who helped devise such principles are also acknowledging the gaps, specifically between principle and practice (Floridi, 2019), as empirical studies have shown how ethical considerations have a negligible impact on the decision-making of software engineers (Vakkuri et. al, 2019). Economic incentives trump other concerns with a lack of "friction" between the principles of ethics and the aims of profit (Munn, 2022). Where AI ethics is a thriving academic debate that borrows from a range of disciplines and ideas like narratology, semiotics or decolonial feminist theory, corporations that control the development of AI technologies rather have an "ethics-as-usual" approach. Hence, AI ethics with principles and value hierarchies as the dominant way of theorising and implementation can be ineffective, to the point that it can become unethical (Floridi, 2019) leading to ethics washing, where such principles are appropriated to avoid regulation (van Maanen, 2022; Frank and Klincewicz, 2024; Schultz et. al, 2024)

[7] This is a non exhaustive and incomplete list of critiques, that point towards the necessity of the structural turn (discussed further).

AI ethics predominantly hinges upon a technocentric view of the technology, which ignores its socio-technical embeddedness. Due to this myopia, even if such lists were transparently encoded and faithfully applied, it would not necessarily lead to ethical output. Most ethical propositions and practices primarily centre around fixing the algorithms, encoding values into these systems, and adding unbiased and representative data sets. While this is a necessary component of developing ethical AI, merely fine-tuning these aspects of algorithms is inadequate and leads to an incomplete diagnosis of the problem, resulting in symptomatic relief. This is because AI is not just a technical artefact but a socio-technical system (Hafez et. al, 2023; Noorman and Swierstra, 2023; Kudina and de Poel, 2024; Chen, 2024; Johnson and Verdicchio, 2024). These systems are part of broader configurations, shaping and being shaped by these aspects. In turn, the technical only becomes a singular dimension of the larger socio-technical constellation. Therefore, where the data collected, labelled and analysed as a part of the technical infrastructure of AI, is collected from larger socio-political settings that have a certain eco-ontologies, specific epistemic preferences of knowing, and certain classifications of values (Hafez et. al, 2023)

Thus, principlism in AI ethics becomes inadequate not only because of profit-driven incentives or haphazard application of such theories and guidelines, but because it suffers from "sociotechnical blindness" (Johnson and Verdicchio, 2017). Consider the case of Amazon's hiring algorithm that, when sifting through anonymised resumes for software developers or technical posts, downgrades resumes that had the words "women" in them (Dastin, 2018) as the data fed into the system had only been that of men. This case illustrates that even when privacy as a value was applied and encoded into the hiring algorithm, a high probability of gender-biased outputs exists. In this case, the machine-learning system detected the gender proxies, like "Women's Chess Captain" or an all-women's college, present in the resume, for decision-making about the suitability of the candidate. Hence, despite following the principle, the technology's situatedness in a socio-political environment where previous data reflects the prevalence of a particular gender in a specific field is unavoidable and inescapable. Thus, while principlism can be a helpful strategy in thinking about how to make these systems less harmful, it is inadequate. To think about larger structures in question where AI as a sociotechnical artefact is situated, AI ethics needs to go beyond the "ML Assembly Line."[8]

1.3 Beyond the ML Assembly Line

[8] This term is coined by and borrowed from Pasquinelli and Joler (2020)

With scholars working in aspects of race, gender, class and larger ecologies, an understanding of how socio-political patterns are mirrored in the data inputs, algorithmic outputs, or ways of design is being brought to the forefront. For instance, acknowledging how data-infrastructure may have a larger representation of light-skinned individuals can reveal specific structures in place that ultimately seep into the system (Buolamwini and Gebru, 2018), which can be addressed for the AI technology in question, but also needs to work on larger power structures in question. Such sociotechnically oriented ethical considerations focus their critique on the ML assembly line of data, algorithms and models,[9] where biased outputs reflect and reproduce patterns of discrimination. Thus, algorithmic harms can and do befall systemically disadvantaged individuals and groups, where algorithms that predict recidivism display racial bias (Purves and Davis, 2023) or credit-scoring algorithms are being used to deny loans to individuals living in specific zip codes (Garcia et. al, 2024).

Relatedly, Heillinger (2022) insists that "AI ethics should expand its focus [...] to include the structural background conditions" (p. 61). With a few attempts to interrogate these background assumptions through the framework of structural injustice,[10] where ordinary injustices are cumulative of untraceable and (seemingly) benign actions within a certain structure, ethical and political analyses of AI have scaffolded on this framework – whether that be to think about criteria of data classification in medical AI (Lin and Chen, 2022), racial politics in predictive analytics in health (Himmelreich and Lim, 2022) or the thoughtlessness it encourages due to the structure being taken for granted and the injustice, untraceable (Browne, 2023). These accounts reach similar conclusions about how debiasing algorithms cannot merely achieve AI fairness but must take into account the social structure in which the technology is embedded. This requires legitimising communications of diverse kinds, even those neglected from governance mechanisms (Browe, 2023), as structural injustice requires collective action rather than technocentric solutions from engineers or data scientists. (Lin and Chen, 2022). This requires the social and moral considerations about AI to move beyond mere compliance with the ethical codes and principles. (Himmelreich and Lim, 2022), even when considering algorithmic processes and data sets that are conditioned by these background conditions.

Increasingly, scholars have also started paying attention to how AI technologies are produced, using environmental resources, human labour and other elements in production (Crawford, 2021). This invites

[9] However, granular/grunt work of fixing the algorithmic pipeline is also necessary and can partially fix manifestations of structural issues within the AI system (Munn, 2022).

[10] This theoretical framework, detailed in Chapter 3, suggests that harm arises when social processes threaten large groups with domination or deprivation of opportunities to develop and exercise their capacities, while enabling domination by others (Young, 2011).

scrutiny of various other factors, like infrastructures (Valdivia, 2024), global supply chains (Dauvergne, 2020), geopolitical manoeuvring (Vipra and Myers-West, 2023), planetary work-force (Muldoon et. al) and product lifecycles (Falk et. al, 2024),[11] that enrich and complicate the discourse of the ethics of AI. When such details are taken into consideration, AI becomes much more than a classification of the family of complex techniques like ML or neural networks (Russsel and Norvig, 2016), to the extent that it is revealed to be "neither artificial nor intelligent" (Crawford, 2021, p.7) requiring natural resources and human labour to function.

Additionally with large-scale AI models like ChatGPT are being deployed rapidly, the computational power including the hardware (chips and GPUs), software (ML optimisations, software to use the chips, data management systems) and infrastructure (data centres that house the cloud, data sets and models) has increased massively – proportionately increasing the amount of the chips, wires and water required to maintain these technologies (Vipra and Myers-West, 2023) Because AI models trained on supercomputers are hosted in large data centers that also provide cloud services to result in user-facing applications like ChatGPT accessed by the individual consumers on their devices (Vipra and Myers-West, 2023), the software, hardware, and infrastructure all become essential in considering the costs of producing these systems. And the data work that is a result of human labour to annotate data sets or train algorithms results in the technologies that are deployed. Heilinger argues that evading these facts can lead to "misleading intuitions" (2022, 61) about what AI is and what the bounds of AI ethics are.

1.4 The 'Structural Turn' in AI Ethics

Elaborate critiques that urgently emphasise the sociotechnical aspects of AI that acknowledge its material constitution while interrogating the background conditions have led to shifts in ethical theorising, resulting in a 'Structural[12] Turn' (Bolte and van Wysenberghe, 2024). This turn in AI ethics evolves from two subsequent waves in the sub-discipline – the first being speculative and the second pragmatic.[13] Although these waves overlap and still have strains of people working in these areas, the third wave, they argue, will bring in concerns about the environmental impact of AI systems, highlighting infrastructures

[11] While acknowledging the overlaps between the discussion about the life cycle of AI infrastructure and supply chains, this project focuses on the latter while excluding logistics and disposal.

[12] The way they use structure is loose, simply pointing towards a level of analysis beyond individual objects and instead thinking about how/where/in whose favour these objects are embedded. Starting from similar intuitions, this scaffolds on a theoretical framework with a social ontology in Chapter 3.

[13] The first is a speculative wave that considers future scenarios with agenthood and uprisings centring on abstract, non-existent technology that can have a grave impact on the long-term possibilities for humankind. This gave way to critiques of impracticality and evasiveness, resulting in the second wave that wanted to focus on existing risks posed by AI technologies. Hence, the second wave was pragmatic and concrete, which attempted to tackle immediate concerns regarding bias, transparency, or explainability.

that constitute AI. It moves beyond the isolationist analyses of the second wave to bring to light the hidden social and environmental costs of such systems. Centring their analysis around the Sustainable AI movement, Bolte and van Wynsberghe (2024) argue that third-wave approaches will consider the development and deployment of such technologies as being “embedded in structures and systems, the ensemble becoming the primary locus of intervention” (p.1740). The structural turn reveals that the AI is a global assemblage materialising and functioning due to the resources from all around the world, funnelled by supply chains.

The compute infrastructure and the hardware it houses are products of trans-national cooperation, with assemblages from different corners of the planet and require a complex and layered network of supply chains. (Vipra and Myers-West, 2023) To elaborate, the GPUs that make up the hardware of AI systems are distributed by NVIDIA (among many other competitors), a multinational corporation (MNC) with its headquarters in Santa Clara, the United States of America. However, the MNC does not manufacture most of these electronics but employs suppliers for different phases of the manufacturing process (Valdavia, 2024), from water fabrication to packaging. And since GPUs are made using silicon and copper, combined with various minerals like tantalum, aluminium, tungsten, gold, boron, or palladium, the MNC relies on a global logistical chain to source these minerals from 259 mineral processing companies across the world. Further, the manufacturing processing is outsourced to the Taiwan Semiconductor Manufacturing Company (TSMC). However, these GPUs only have a life cycle of 5 years, after which they are replaced, leading the used chips to be dumped in a country in the Global South (GS) (Valdivia, 2024).

Building these infrastructures, hardware, and software requires a supply chain of labour as well. Creating and maintaining the software, from training the ML algorithms to annotating datasets, also involves a planetary workforce. Where data can be collected from proprietary sources or scraped from the internet, it will be annotated by data workers in Kenya, further shifting the training of the model to another centre or even another country. These models with the data will further be tested and evaluated by programmers in Pakistan (TechEquity, 2025). Audits of the model may be conducted by programmers based within the company itself, like OpenAI. Hence, entangling data and labour, the structural turn spotlights another supply chain where inordinate amounts of human labour are required to moderate, annotate and label data-sets to be used in models that will soon be tinkered by engineers and data-scientists, to be audited for improvement.

Therefore, the structural turn brings supply chains (Valdivia, 2024) that drive the development of AI technologies to the forefront, including natural resources and critical minerals that make up the physical infrastructure, hours of labour performed by humans, and data within the sets to power the algorithm, with extraction being the first step to the process. Valdivia (2024) puts it:

"The supply chain capitalism of AI is the orchestration of commodity chains that extract, ship, and manufacture the natural resources needed to develop AI from an infrastructural perspective, such as mines, data centres and e-waste dumps together with their human resources (miners, drivers, directors of data centres operations, e-waste dismantlers, etc.). Within this chain, digital elements such as datasets and algorithms together with human labour (data annotators, data scientists, data engineers, etc.) are also key" (p.5)

Critical scholars have looked at different infrastructural impacts, social and environmental impacts of excavation, e-disposal, data centres (Valdivia, 2024), labour value chains (Muldoon et. al, 2024) within the supply chain of AI and acknowledge the extraction it entails. Still, it is only recently that it has been considered concerning 'AI ethics.' Widder (2023) argues that AI ethics must consider AI as a supply chain problem, where engineers feel a limited notion of responsibility for the downstream impacts, yet only focuses on the technical aspects. Falk et. al (2024) consider it in tandem with life-cycle analyses regarding sustainable AI. Valdivia (2024) draws attention to the infrastructural harms brought about by the supply chain of natural resources for AI production. Insights by theorists like Crawford (2021) about how data is collected and which data sets are used can also be interpreted as signalling towards the supply chain of digital elements like datasets or algorithms. With the third wave of theorising looking at larger structures in place, an interdisciplinary perspective that borrows from these critical scholars can help think about the AI supply chain.

1.5 Summary

To summarise, principlism as the dominant mode of theorising/practising AI ethics, while necessary, is not sufficient for developing and deploying ethical AI. Increasingly, the socio-political embeddedness of these technologies and their production from natural resources and human labour have led to shifts in theorising, precipitating the structural wave. As various ensembles of AI technologies situated in larger structures become sites of ethical examination, supply chains emerge as a crucial entity, funnelling the natural resources, labour and data required for the functioning of an AI model. And while insights from the thorough accounts by critical scholars have been integrated in terms of sustainable AI, infrastructural harms, or limited responsibility of engineers, research concerns about supply chain within AI ethics (as a

discipline) are nascent yet steadily on the rise. However, to warrant a structural analysis of AI supply chains, one also needs to carefully consider the harms they bring about and the logic that conditions their emergence. Only then can we consider if and how one should attribute responsibility for these harms.

Chapter 2

Extractivist Logic and Undivorced Harms

When structures are considered in the ethical consideration of AI, the range of harms considered morally problematic also expands. Particularly, extractive harms that populate the supply chain become ethically pertinent. Since supply chains are an essential aspect of AI development, the technology is a global assemblage built by resources that have been extracted, like natural resources, human labour, and data.[14] This trifecta of extractivism, the extraction of natural resources, the exploitation of labour and the abstraction of data, can lead to infrastructural harms, exploitative harms and certain kinds of algorithmic harms. Moreover, this range of harms sustains unjust practices that become commonplace in the pursuit of innovation, which the MW unduly bears. Moreover, while there is a tendency to separate algorithmic harms as a result of deploying AI technologies from the infrastructural or exploitative harms that result from developing AI, analysing the supply chains of natural resources, human labour and data sets demonstrate that these harms cannot be divorced as they are rooted in a structure that is a result of and reproduces range of logics, like extractivism. This chapter introduces the concept of extractivism briefly in §2.1, discussing examples of the extractivist trifecta and the harms they engender when building AI systems in §2.2. §2.3 demonstrates that while harms resulting from the production seem separate from the algorithmic harms, they are undivorced on account of being rooted in structures that encourage extraction, where data extraction leads to data abstraction that can result in data bias and algorithmic discrimination. §2.5 addresses some potential criticisms, emphasising the need to complement algorithmic ethics that focus on the technical aspects of making AI ethical with a structural analysis that considers extractivism as an ethically pertinent issue in the production of AI technologies. §2.6 summarises the chapter.

2.1 The logic of Extractivism

Defined as "a mode" (Acosta, 2013, p. 62), a "structural feature of capitalist world economy" (Aráoz, 2013, 131) and a "development model" (Svampa, 2019, p.6), extractivism is a logic oriented towards

[14] This chapter will analyse data, highlighting how AI technologies differ from other technologies, such as radio devices, which also require natural resources and human labour. However, the thesis shall not apply the Structural Injustice framework to data extractivism in the next chapter. Since the data extraction pipeline is intricately woven with the deployment of AI, it is beyond the scope of this thesis, which focuses on AI development, although it argues that the two cannot be divorced, warranting further research.
.

maximising profit through instances of extraction (Durante et. al, 2021). The former abstract mode of thinking is predicated on the latter, where acts of extraction that create benefits and propagate violence entrench such mentalities. Hence, the practices and properties of extractivism can result in violence for communities and ecologies, also determining and redefining 'value' in terms of extractability. While some acts are inherently destructive, like mining, others, like labour forces, are not, with nominal stewardships and limited benefits that can nonetheless be exploitative and non-reciprocal, where, in exchange for meagre benefits, the extractivist accumulates large amounts of profit. Such realities flag extractivism as a normative issue to analyse.

When tracing the emergence of extractivism, Durante et al. (2021) clarify that while appropriation of natural resources, like ancient deforestation, originated 5000 years ago during the formation of civilisations, they were complemented by attitudes of co-existence, sustenance, and regeneration. It was only since the 16th-17th centuries that an extractivist mindset started to emerge. And while this logic originally proliferated alongside the development of the modern world system, technoscientific revolution, and European colonialism, Durante et. al (2021) maintain that it can permeate other economic systems like socialism. Hence, this makes extractivism an ontological orientation or an "onto-logic" (Durante et al., 2021) of its own that emerged out of the systems of modernity, coloniality, and technoscientificity.

However, in the past century, with economies being restructured to the neoliberal norm (Boden, 2011), extractivism has worked towards accumulation becoming a global constant (Chagnon et. al, 2022). With the increasing digitisation of the globe and worldwide data network, the conceptual expansion of the term extractivism has been proposed. Emphasis on limiting the use of the term only to describe the removal of natural resources (Gudynas, 2018) is contested through rebuttals about how what counts as a 'natural resource' itself is being redefined, with data being compared to oil or sunlight (Taffel, 2023). With concepts like "total extractivism" (Dunlap & Jakobsen, 2019) that acknowledge how extractive technologies aim to reconfigure the planet while normalising its logics on many levels, the necessity of expanding the concept becomes evident. Such conceptual expansions can highlight how central such practices and outlooks, which emerged from historical precedents, are amplified in the modern world. For instance, Mezzadra and Neilson (2017) argue that literal extraction now subsumes logistics and finances, due to the supply chains in place with the creation of new industries of managing, packaging and shipping, expanded aspects of extractivism reveal newer modes of capitalist accumulation, through data.

With the proliferation of technologies in every aspect of life, from navigation to health maintenance, user data is collected from these various applications to be monetised by corporations that maximise profit despite privacy violations and opaque data practices. Through such data mining, even human cooperation and social activities are utilised through digital mediums that produce data profiles that are being used for building predictive models to help identify specific behavioural patterns that lead to the coining of concepts like digital colonialism (Couldry & Mejias, 2019). Advocating for a nuanced take on data extractivism, Gray (2023) argues that the account of data colonialism is an incomplete one as it fails to accommodate a historicised and differentiated understanding of colonialism and its relationship to capitalism. In terms of the development of AI, acknowledging its materiality through the extractivist logic becomes evident due to the supply chains in question that create or strengthen industries and professions of their own, ranging from mining, data brokerage, and data annotation. And this is very much evident in AI technologies and their extractivist trifecta of natural resource extraction, labour exploitation and data abstraction.

2.2 The extractivist trifecta

Supply chains procuring resources like natural resources, human labour, and data for AI production are underpinned and reproduced by this logic of extractivism. The discourse on AI as an extractive technology has been growing, as scholars frame it in terms of knowledge (Pasquinelli & Joler, 2021), data (Kwet, 2019), labour (Muldoon, Graham, and Cant, 2024) or environments (Taffel et. al, 2019; Brodie, 2024; Valdavia, 2024). Scholars like Crawford (2021) further trace global patterns that despite historical contingencies that lead to specific harms are "deeply interconnected by the multiple forces of extraction." (28) And while the costs of extraction can be witnessed across the planet – whether that be in the deserts in Nevada (Crawford, 2021), the tin-mines in Indonesia (Jong, 2024), or the digital sweatshop in Kenya (Perrigo, 2020)– a disproportionate impact is rather seen in the MW i.e. regions and areas in the world that contain the majority of the population.

This does not mean that extractions do not occur in the Global North (GN) or wealthy countries. Nonetheless, an unequal juxtaposition between the GN and GS is evident in terms of who owns the digital economy and who bears the human and environmental costs (Kwet, 2019). When outsourcing labour and natural materials when building these systems, there is also an "offloading of risk" (Digital Future Labs, 2025a). Due to the prioritisation of economic growth in these nations, social and environmental concerns are often overlooked. Such practices perpetuate not-in-my-backyard politics, where production is shifted to countries with lower environmental standards, labour and energy regulations to make a profit due to

escalating demands. Regilme (2024) conceptualises this as necro-exportation, where AI systems create global life-and-death hierarchies through the supply chain, externalising the costs from the GN, where the benefits are concentrated, to the GS, which suffers environmental degradation and labour exploitation. Specifically, this follows a 'trifecta' of extractivism: where there is extraction of natural resources, exploitation of labour, and abstraction of data.

When considering the extraction of natural resources, AI production intensifies demand for rare earth elements required for constructing hardware that powers "digital technologies that enable and rely on AI: computers, smartphones, data centres, undersea cables, insulation, optical fibres and fuel cells" (Muldoon & Wu, 2023, p. 9). Data centres are physical infrastructures that compute and store data with cooling systems to regulate the networks of computing resources and also house the cloud (Digital Future Labs, 2025b). These infrastructures require land to house them, water to cool them, and minerals to build them. GPUs, chips, wires, and battery components require metals like copper, tungsten, lithium, and other minerals acquired through mining, which is an extractive practice. With ecological damage and community disruption leading to pollution, toxicity, and health risks, such mineral mining comes at a high cost. However, processing and smelting these minerals add to these costs. With China having previously dominated the market (Crawford, 2021), its recent shift towards environmental preservation has prompted other places like Indonesia to step up. (Digital Future Labs, 2025a) Additionally, these metals are excavated from conflict areas, like the Democratic Republic of Congo (DRC) (Crawford, 2021, exacerbating pre-existing issues at hand, feeding into the conflict and creating environmental migrants who are stripped of their homes.

With no transparent reports and alarming predictions about the carbon emissions and electricity usage of AI technologies, there is a rush towards finding cleaner alternatives (Menear, 2024). In attempts to "green" this infrastructure, there is also a rush for critical minerals in the name of sustainability and a shift towards building nuclear power plants by Big Tech giants. However, more than half of these minerals are procured from indigenous, peasant lands (Owen et.al 2022), displacing their communities and disrupting their ways of life. Such dynamics reveal geopolitical arrangements as the countries that make up the MW and are located in the GS serve as resource frontiers for extraction and refining. The risk of the rebound effect, where efficiency gains through greening technologies can lead to the over-production of the said technology itself (Digital Future Labs, 2025a), looms over the heads of states who want to "keep up" with the digital transformation, as larger corporations rush towards profit. Hence, one can question whether these attempts to transition to sustainable forms of AI can result in, paradoxically, more ecological harm through natural resource extraction, especially for those who come from vulnerable regions.

Further, exploitation is rampant as physical infrastructures are built through the labour of miners, technicians, and construction workers, where others help ship components as part of the logistical chain. Furthermore, when it comes to developing and deploying the algorithms, the labour required to train or fine-tune these algorithms is invisible, under the garb of "fauxtomation" (Taylor, 2018), where human labour is still powering the intelligent machine. A continuum of unpaid, poorly paid or micropaid work (Casilli, 2017), the digital labour that powers AI is extractive. With AI being trained on vast amounts of public and proprietary multimedia data, ranging from paintings to academic publications, the cultural and scholarly labour that goes behind creating and curating those works is hardly compensated for.

Additionally, a few other sources of unpaid labour include Internet users helping train AI cars through reCAPTCHA (O'Malley, 2018) or behavioural data from smart devices or environments that help produce such data (Casilli, 2017). Microwork, also known as ghostwork or clickwork, involves annotating or labelling data. This consists of flagging violent content or labelling digital archives for hours on end. However, these workers are treated and paid differently depending on where they are located. For instance, where a content moderator might be paid £2,000 a month in the UK, the pay in Colombia is found to be ten times less, amounting to £235 (McIntyre et. al, 2022). Due to lax labour policies and a country's attempt to enter the global AI chain (McKey, 2025), such exploitation is rampant, especially in the GS. A digital sweatshop culture has mushroomed with terrible working conditions with minimal pay, psychological damage and extreme surveillance in places like Kenya (Perrigo, 2023), the Philippines (Tan & Cabato, 2023) or India (Iyer, 2023). Moreover, attempts at increasing wages or tightening regulations can be risky as the industry could conveniently move to another country with fewer restrictions, resulting in exploitative harms.

Relatedly, the unpaid labour from cultural/scholarly work, smart devices/environments and any activities on the internet leads to the extraction of data. As mentioned before, phrases like 'data mining' conceptualisations of data as a resource (Gitelman, 2013) as a capital (Sadowski, 2019) signify an attitude of accumulation and an extractive motive. Concerns about how Big data is harnessed have been articulated using analytic categories like surveillance capitalism, where click-based attention generates meta-data to be appropriated (Zuboff, 2019) or data colonialism, where a logic of resource appropriation in the form of data by powerful actors, that reflect colonial practices, dominate (Couldry & Mejias, 2019). From copyrighted material to public data scraped off due to forms of consent for use, data is accumulated and categorised in specific ways, also leading to an abstraction of it all (Gray, 2023).

2.3 Structural Links and Undivorced Harms

Infrastructural harms and exploitative harms cannot be divorced from the algorithmic harms. While they are neither identical nor bear a causal relationship, they are both rooted in the same structures with interlocking logics.[15] Tacheva and Ramsubramanium (2023, p. 4) argue that while critical scholars have succeeded in critiquing AI through the lens of capitalism, [ableism], colonialism and patriarchy, a single-axis examination is incomplete as these logics result in an "interlocking assemblage whose parts intensify and reinforce each other in the context of AI." For instance, Muldoon and Wu (2023) argue that colonial logics underpin these AI technologies, leading to extraction and exploitation. Extractivism as a logic partially sustains the existing structures, which reproduce and are reinforced by practices of extraction.

In cases of data extraction, concepts like surveillance capitalism (Zuboff, 2019) or data colonialism (Couldry and Mejias, 2019) help identify the concerns that arise from mass data collection, including privacy concerns, attention-based surveillance, or accumulation for domination. Furthermore, in the case of data, there is an abstraction involved that can flatten and erase context for the formation of sufficiently varied aggregates that improve technical performance (Crawford, 2021), where one data point can easily substitute the other. Abstraction makes it possible to treat data as interchangeable, quantifiable inputs in each of these situations, removing the particular identities, histories, and harms from which the data originated. This makes the abstraction of data an epistemic operation as well (Gray, 2023), when data is framed as neutral, universal, and commensurable, it enforces dominant orders of knowledge and value by determining what counts as legitimate data and who gets to interpret it.

These abstractions can result in data bias that can further lead to algorithmic discrimination.[16] For instance, on a small scale, predictive policing algorithms might abstract data from neighborhoods into a single risk score, ignoring the historical and systemic biases in policing that led to higher arrest rates in certain communities. Or in large scale models, generating images of a nurse or a CEO may represent the stereotypes due to abstractions in question – or worse, attempting technical fixes may result in images of

[15] This concept will be explored in detail in the next chapter. To follow this argument, structure can be understood as being made up of mutually reinforcing, historically sedimented logics that are social and material rules enacted by relationally positioned agents who are also constrained by them.

[16] Diberardino et al. (2024) talk about different algorithmic harms – distributional, dignitary, discriminatory, exclusionary, exploitative, privacy, epistemic and democratic. This thesis only discusses discriminatory/exclusionary harms concerning data abstraction/dispossesion. However, future work could investigate to what extent the extractivist logic does or does not influence the facilitation of these harms.

Black people as the ensalvers. Hence, ethical imperatives on privacy, fairness or transparency that dominate questions of bias can be seen as connected to such extractivist orientations.

Considering the different ways in which the enlisted harms are related can further highlight the structural link between development and deployment. The previous sections have discussed four kinds of harms: algorithmic, infrastructural, exploitative, and extractivist harms. The first, algorithmic harm, is widely discussed in AI ethics as it revolves around the ML assembly line and also acknowledges its sociotechnical embeddedness. Issues around data sets that, as algorithmic inputs, result in unjust outputs in terms of decision-making or predictions in policing or healthcare have been marked as harmful (O'Neill, 2017; Noble, 2018; Eubanks, 2019; Benjamin, 2019; Buolamwini & Gebru, 2018). Hence, algorithmic harms consist of AI bias resulting in discrimination that can indirectly implicate other aspects like privacy or security as well (Debarardino et. al, 2024), concerning its use and deployment. The second and third harms, infrastructural and exploitative, come to light when considering the effects of AI supply chains. These chains funnel natural resources to build infrastructures and hardware, and they also involve labour that constructs these infrastructures, as well as the algorithms in question, through data work to develop these technologies.

While the first and/or the second harm appear to concern different phases than the third, like production versus use, these categorisations can be diluted. When it comes to the question of algorithmic harms, the issues surrounding data sets with historical bias also touch on the development of these systems. Specifically, they hint towards the data pipeline from which data is collected and abstracted, and the types of data sets fed into the system and how they have been labelled. This results in algorithmic harm, particularly in terms of data bias and algorithmic discrimination, emerging partially from an extractivist logic that flattens data. In this way, certain kinds of algorithmic harm alongside infrastructural and exploitative harm are (partially) emerging from an extractive logic that conditions the production of AI. Hence, these three harms in the context of AI supply chains correspond to natural resource extraction, labour exploitation and data abstraction. Moreover, these phenomena are empirical manifestations of the extractivist logic emerging from a structure, and this can lead to structural injustice. Since structures are dynamically enacted by different individuals yet conditioned by certain social and material rules, that is, the various logics that reinforce each other, extractivist harms, alongside other intersectional aspects, put a particular group of people at a distinct disadvantage when compared to others.

2.4 Supply Chains in AI Ethics?

Two potential objections can be made against the claims proposed so far. First, while many may concede that AI technologies are made from global supply chains, this is hardly unique to AI in the context of globalised capitalism. Instead, supply chains of a myriad of products exist – from fast fashion to electronics – and are not necessarily a flaw in AI systems to be considered the purview of normative ethical or political analysis, but are a feature of the world and the cost of innovation and international trade. Hence, this is beyond the scope of AI ethics, requiring a critique of larger forces that may or may not necessarily fall within the bounds of AI ethics.

However, what is crucial to note about the supply chain of AI is that it combines various industries: data labelling companies, big tech firms, electronic industries and mining-smelting companies. Hence, the material aspects of AI with data centres, chip factories and e-waste dumps are very much foundational to and entangled with the algorithmic makeup of AI, with the creation and curation of datasets alongside the design of AI (Valdavia, 2024). Since AI systems are embedded in societies, we are required to think about how these societies are set up, with factors like electricity usage being taken into consideration (Heillinger, 2022). Hence, acknowledging the fact that OpenAI's GPT-4 consumes much electricity and demands much water, disrupting local communities and ecologies (Zewe, 2025), is necessary in building ethical AI.

Despite these facts, the ubiquity of supply chains can be affirmed, whether they pertain to garments or technologies. However, this does not subtract from the extractive and exploitative harms they multiply as they involve practices that should not be considered the norm. Even if the human rights issues, like underage labour, were addressed, these supply chains would still involve an extractivist tendency that orients itself towards accumulation, which creates further instances of injustice. This is precisely what I am pointing towards (to be discussed in Chapter 3), where the structures that sustain such extractive practices render these acts of injustice as the norm.

The second critique would be a corollary of the first, pointing out that such a provocation is not necessarily helpful or action-guiding as it widens the responsibility gap further by locating harms and wrongs on a structural level, widening the complicity of agents, amplifying the many hands problem not only on a spatial level, by including transnational networks of production, but even temporally by referring to historical relations of inequality by involving sedimented logics and socio-material rules. Prefacing my response to the potential second critique, these analyses can be complementary to the ethical methods and frameworks mentioned in the previous chapter. Nonetheless, those principles or

frameworks commonly found in AI ethics would be incomplete without acknowledging the larger structures at play. Concerning the questions of amplification of responsibility and the inability to address them, an appropriate framework, like that of Structural Injustice, can accommodate in such distributions of responsibility.

2.4 Summary

To summarise, AI supply chains that procure resources for the production of AI technologies are driven by an extractivist logic that orients itself towards accumulation and dispossession. This becomes ethically pertinent as it leads to the trifecta of extracitvism, natural resource extraction, labour exploitation, and data abstraction, which lead to a range of harms that are rooted in the structure. While these practices implicate and affect everyone, they have a disproportionate impact on the MW. And to address these distinct moral wrongs, unfair practices and harmful abstractions, we can turn to the notion of Structural Injustice in the next chapter that will further help us understand whether with supply chains, criss-crossing the world, situated against a structure that incentivises all participants to act as per the extractivist logic, if and how moral responsibility is to be attributed.

Chapter 3

Structural Injustice and the AI Supply Chain

As AI technologies are reconceptualised to be global assemblages fabricated through supply chains, extractivist harms become ethically urgent and pertinent, especially as they have a disproportionate impact on the MW. However, these extractivist activities, resulting in ecological and social harms, take place under legal norms which hide their moral dimensions and widespread participation from a range of agents, making it a feature of the world. Diagnosing these moral issues requires normative analysis; it is the theory of Structural Injustice that can help unpack the injustices at play. This chapter introduces the frameworks of Iris Marion Young (2011) and Maeve McKeown (2024b) in §3.1, highlighting the distinct ways McKeown improves on Young's account of Structural Injustice by reconfiguring the social ontology that underpins the theory. This involves adopting a critical realist ontology to consider interplays between structure and agency as well as morphogenetic cycles, thereby highlighting the role of power. As we will see, it also involves taking a pluralist approach to structural injustice. §3.2 relies on McKeown's framework to analyse the extractivism of natural resources and labour in the production of AI technologies as cases of *avoidable* structural injustice, where injustices caused by structural processes and constraints are foreseeable yet are not mitigated; and *deliberate* structural injustice, where injustices caused by structural processes and constraints are deliberately maintained by those in power. Where its structural nature can be seen through its ontological set-up and morphogenetic cycles, the injustices manifest through oppression and domination, demonstrating how these injustices are avoidable and/or deliberate. The chapter concludes by emphasising the advantages of relying on the framework on structural injustice in §3.3.

3.1.1 Structural Injustice: Young's Conception

With its lineage being traced back to 1850 BCE Egypt, structural analyses of injustice were prominent in anti-slavery and anti-colonial thought penned by writers like DuBois (Gordon, 2024) as well as in Marxian thought (McKeown, 2021). Within liberal political philosophy, the focus was upon a select few institutional structures of the society, due to the influence of John Rawls (Forrester, 2019). However, the concept of structural injustice was popularised by Iris Marion Young through her book *Responsibility for Justice* (2011). Rejecting the false dichotomy between personal responsibility and social structural causation, Young argued that a complex interplay between individual agents and socio-material constraints creates structures that render individuals and certain groups vulnerable. However, the sources

that cause vulnerability are often “multiple, large-scale and relatively long-term” (Young, 2011, p. 47) and cannot be attributed to a single entity but instead to larger structures in place. Moreover, it is often these social, economic and political structures in place themselves that sustain, generate or reproduce injustices.

For Young (2011), structures are “socially caused conditions that position a large number of people in similar ways” (p. 18) with “accumulated outcomes [resulting from actions of individuals] enacting their projects” (p. 62). These socio-structural frameworks have four features, which can also propagate structural injustices. First, the legal and social rules of the material world can constrain as well as enable individuals, and continuing practices that have solidified over time encourage specific habitual patterns of acting. Second, these structures position different people in relation to each other, conditioning and limiting interactions among individuals. Third, these structures exist and are produced through the action of individuals, occupying social roles and following rules, having internalised this structure. Fourth and finally, these combined actions result in unintended consequences, which end up benefiting some and harming some. Hence, for Young:

'Structural injustice, then, exists when social processes put large groups of persons under systematic threat of domination or deprivation of the means to develop and exercise their capacities, at the same time that these processes enable others to dominate or to have a wide range of opportunities for developing and exercising capacities available to them. Structural injustice is a kind of moral wrong distinct from the wrongful action of an individual agent or the repressive policies of a state. Structural injustice occurs as a consequence of many individuals and institutions acting to pursue their particular goals and interests, for the most part within the limits of accepted rules and norms.' (Young, 2011, p. 52)

She explains the concept through the example of Sandy, a single mother of two who is being asked to empty the rental unit that she lives in, as the building will be converted into a condominium. Since she works at the mall, the places near her work are too expensive. While she considers the inner-city neighbourhoods, she dismisses the idea as they could be dangerous for her children. She decides to get an apartment far from work, although it requires her to buy a car. However, the apartment that she settles on renting requires her to pay a three-month deposit, which she cannot afford. Sandy, then, finds herself on the brink of homelessness. While one could hold the landlord, the developer or the local politician accountable in this specific case, it would not necessarily result in tackling similar cases and hence similar structures that create housing insecurity across the country. For Young, this requires thinking about

injustice differently and beyond cases where there are straightforward perpetrator-victim dynamics. Moreover, for Young, these cases could be global, spanning across borders.

To elaborate on global structural injustice, Young (2011) traces the production of fast fashion through sweatshop labour, where workers, usually situated in the GS, are exploited and abused with low wages, excessive working hours and dangerous working conditions. And while there are parties, like factory owners, managers, and states, that are legally responsible for the human rights violations, they claim to be operating under objective constraints. The employer cannot improve wages due to competitive markets, where economically weak states are pitted against other poor states for private investments. This makes multiple and diverse actors in different countries enabled and constrained by structural processes resulting in injustices for the garment workers in question due to the background conditions that make this social reality for them.

3.1.2 Structural Injustice: McKeown's Revisions

Young's work has sparked a rich and lively debate with many applications of and adaptations to the theory. While there are scholars who take Young's theory for granted resulting in applications of the concept (Godoy, 2017; Widdows, 2021; Browne, 2018), others expand the framework by linking it with conceptions of epistemic injustice (Jaggar & Tobin, 2024) or historic injustice (Nuti, 2019). Few even critique the framework for its flawed conception (Atenasio, 2019); its claims of decisions being made by agents as morally benign (Powers & Faben, 2019) and in the wake of post-Rawlsian liberal philosophy, its fetishisation of justice (Gordon, 2024). Others, while highlighting the limits, work within the blueprint of Young's theory. One specific account, as such, is that by Maeve McKeown.

In her book, *With Power Comes Responsibility: The Politics of Structural Injustice* (2024b), McKeown refines Young's account regarding the work of socio-structural processes. McKeown claims that Young's four-fold account (as discussed in 3.1.1) combines insights from sociological schools that are inconsistent, as Young conceives of structures as both emergent and reproduced only through actions. Grounding her analyses in Archer's (1995) critical realism, McKeown (2024b) rejects the third feature of Young's socio-structural processes, where structures only emerge through actions, as it collapses structures into agency and occludes the structural conditioning of agents over time, existing before and after the agent. It also understates the role of social positions and the vested interests in causing, maintaining or changing structures that have unjust outcomes. Instead, she turns to Archer's morphogenetic approach, according to

which a structure predates the action which transforms it, and structural elaboration necessarily post-dates those actions. Hence, a morphogenetic cycle has three stages: conditioning, interaction and elaboration. Agents inherit structures shaped by the actions of previous generations, and their actions are conditioned by the structures in place that position them in relation to one another within a world that is not of their making. Further, by their social position with vested interests, agents will interact with the structures to maintain these structures, incentivising such behaviour, or even disrupting them. These interactions result in elaboration of the structure, leading to unintended consequences and new possibilities for the structure to morph. To go back to the example, Sandy and her landlord have inherited the structure of the housing market, conditioning their behaviour as they occupy different social positions. They interact with this structure, their actions combining with those of other agents, resulting in an elaboration of the structure, which can lead to the sustenance of the housing market structure, that causes housing insecurity, or a contestation of the structure if individuals engage in collective action to change the structure. However, McKeown (2024b) emphasises that certain individuals have a larger scope of influence when interacting with and elaborating on the structure – that is, those with more power.

Drawing from Wartenberg, McKeown (2024b) conceives of power relations emerging between agents occupying social positions. For Sandy, the landlord-tenant relationship influences her behaviour towards being complacent to avoid eviction and also a black mark on her record. However, the landlord need not do anything special for Sandy to behave in this way as Sandy herself knows that any disruption could harm her chances in the future to find housing. However, this power relation is structural, where the dynamic is not reached by agreement, but where Sandy has no choice but to be in this situated power relationship with the landlord - she has little effect on the landlord, but the landlord will have a significant effecr on her. For McKeown, this reveals something important about power relations – that is, they do not exist only between the interacting parties, Sandy and her landlord, but rather these power relations also depend on “peripheral social agents”(McKeown, 2024b, p. 37). It is the system of property rights or judicial system backed by the state policies and market, and performed by the actions of other tenants and landlords that form a social alignment in generating and justifying the power relation between landlord-tenant. However, it is this situatedness that also makes the powerful/powerless distinction contextual and relational (McKeown, 2024b, p. 38). The landlord is constrained by state policies, legal regulations, and market conditions, but is also enabled by them to issue threats for rent or by having the power to ask them to leave. Hence, due to the vested interests built into the social positions, certain agents reproduce social structures that can entrench injustices, resulting in dominating and oppressive relationships, deriving benefits from the actions of the subordinate groups. Although a given due to existing social structures, power relations are also reproduced by not only the actions of the

dominant-subordinated groups but also the aligned social agents. This provides the subordinated group an opportunity to create a countering or alternative alignment for challenging the entrenched yet dynamic power relations.

By adopting a critical realist ontology and combining it with a situated notion of power, McKeown (2024b) improves upon Young's definition of structural injustice (as discussed in 3.1.1). When one accommodates the interactions between structure and agency as well as the role of power in this dynamic, structural injustices manifest more than unintended outcomes of collective actions that are benign social processes (Powers & Faden, 2019). McKeown argues that Young's conception of oppression as 'the institutional constraint on self‐development' and domination as 'the institutional constraint on self‐determination' (1990, p.37) in *Justice and the Politics of Difference* endorses a similar line of thinking about power. This leads McKeown (2024b) to introduce amendments to the definition of Structural Injustice:

"Within social structures, agents are situated in different social positions with varying degrees of power and access to resources. The social structures are unjust when they result in the oppression or domination of certain social groups. Structural injustice is 'pure' when the injustice is unintended, unforeseeable and there are no agents with the capacity to remedy the injustice; it requires wholesale social-structural change. Structural injustice is 'avoidable' when the unjust outcomes of structural processes are foreseeable and there are agents with the capacity to remedy the injustice but they fail to do so. Structural injustice is 'deliberate' when the unjust outcomes are intended because powerful agents benefit from it so they deliberately perpetuate it, and these agents have the capacity to remedy it but they fail to do so". (45)

Adopting a pluralist stance to structural injustice, McKeown (2024b) identifies three kinds of structural injustice: *pure*, *avoidable*, and *deliberate*. First, 'pure structural injustice' is the kind of structural injustice Young (2011) theorised about, involving diverse and distant agents engaging in acts to be changed only through collective action. McKeown elucidates on this through the example of climate change, where multiple, disparate individuals act to produce the unintended outcome of worsening the climate crisis. Second, 'avoidable' structural injustice takes place when the resulting harm from the structure can be avoided by actions from those with power, but fail to do so, as exemplified by global poverty. Third, 'deliberate structural injustice' is that in which agents are deliberately perpetuating unjust background conditions for their own interest, despite having the power to change them. According to McKeown, this

can very much be observed in the global garment industry, where there is a deliberate maintenance of structures to extract benefits from the said structures.[17] Such kinds of structural injustices can be traced across the several issues around the globe – whether that be about bioethics, beauty standards, colonial reparations or AI.

3.2 AI Extractivism as Structural Injustice

While there have been attempts to analyse AI through the framework of Structural Injustice (Lin & Chen, 2022; Himmelreich & Lim, 2022; Browne, 2023), these accounts mainly draw from Young and focus on questions of data bias. As argued in Chapter 2, since algorithmic harms cannot be divorced from the material harms due to their rootedness in structures, where a confluence of logics intersect, reproduce and reinforce the other – like capitalism, colonialism, patriarchy, and our case extractivism – it is essential to think about AI technologies and how it results in natural resource extraction and labour exploitation. McKeown's (2024b) framework of structural injustice helps emphasise that extractivism in supply chains for AI reproduces sites of harm that are ethically pertinent. Following the morphogenetic cycle, the steps of structural conditioning, structural interaction and structural elaboration are observed, where historical precedents for extractivist logic set the stage, current activities interact with these structures to elaborate on them to result in extractivist injustices that are dispersed yet disproportionately befall the vulnerable. Additionally, structural constraints and causes that condition social rules and social roles are navigated between agents with different vested interests and levels of power. And through these structural processes, certain individuals undergo (risk of) oppression and domination (McKeown, 2024b). This leads to natural resource extraction being a case of avoidable structural injustice, while labour exploitation is a case of deliberate structural injustice in the development of AI.

3.2.1 Natural Resource Extraction in AI as Avoidable Structural Injustice

While humans are recorded to have extracted minerals since prehistoric times, it is only in the recent 70 years that an increased demand for strategic and rare earth minerals has been noticeable (Brown et. al, 2021), alongside dependencies on fossil fuels that have roughly doubled since 1980 (Ritchie & Rosado, 2024). Moreover, while it is individuals in average and high-income countries that consume these resources more (Alvaredo et. al, 2018), due to geopolitical arrangements, low-income nations also depend

[17] See McKeown (2024b)

on extracting and exporting these resources. This paradigm has been traced back to the European colonial era, where colonised localities in the Americas, Africa and Asia became sources of natural and human resources to meet the demands of the world economy, largely defined by capitalism (Acosta, 2013). These regions were forced to specialise in the extraction of raw materials, resulting in the violent appropriation of minerals, metals, oil and gas (Gudynas, 2015) alongside other commodities like sugar, rubber or timber. Such dynamics, coupled with industrialisation, resulted in unequal labour relations and limited access to means of production (Acosta, 2013), structurally conditioning how contemporary events unfold. Hence, the social and material rules that knit the structure together as natural resource extraction has become "modus operandi for economic growth and diplomatic and military forms of geopolitics" (Sörlin, 2022, p. 4), as companies and countries compete to build larger AI models. McKeown's (2024b) observation about social relations creating structural constraints unfold here as the very nations that suffered at the hands of an extractivist paradigm, sovereign and liberated, replicate it for catching up at the development race, as observed in countries located in the Latin American region (Acosta, 2013) due to the said rules and conditioning in place. With the modern state being historically responsible for economic development, these extractivist policies form a core part of the developmental legacies normalising such mass accumulations (Peša & Ross, 2021). And these practices resulted in structural interaction where tools have been used to extract minerals for building technologies like AI, leading to a range of environmental harms and public health crises for populations who live near the mining sites.

These structures are further elaborated on (McKeown, 2024b) with large-scale AI technologies like LLMs being rolled out at an alarming speed that require large data centres, which further require large amounts of minerals and materials to construct and sustain these infrastructures. (Stacciarini and Gonçalves, 2025). Current estimates suggest that there are 12000 data servers in the world (Fleck, 2024), with 992 classified as hyperscale (Taylor, 2024), containing at least 5000 servers and spanning more than 10,000 square feet (Powell & Smalley, 2024). Typically, data centre facilities house thousands of servers organised in racks, storage systems, networking devices, processing units, power supply and backup systems, connection cables, cooling mechanisms, environmental monitoring sensors, and other essential components (Stacciarini & Gonçalves, 2025) – and all these components are built using metals. To name a few, aluminium and steel are used to construct the larger grids to accommodate parts of the computer. Copper, gallium, germanium, indium, lithium, tantalum and tellurium are essential for building microchips. Hazardous metals like lead and mercury are also used in specialised AI hardware as well as toxic elements like gold, platinum and neodymium (Falk et. al, 2024).

Such interactions with the structure lead to elaborations that can start another cycle of conditioning the actions taken by the agents in their different roles (McKeown, 2024b). For instance, with projects like the Meta's hyperscale data center with 350,000 NVIDIA H100 GPUs (Lee et. al, 2024) or OpenAI's Stargate, an arms race has begun for AI chips (Falk et. al, 2024) – and for metals and minerals required to build these chips, that make up the infrastructures and other systems that also try to green the AI. As investments pour into the data centre business with projected increases for copper (Chance and Hammersley, 2025), the extractivist logic sustains these production chains through practices like mining, which is further endorsed by poorer states who need to continue business-as-usual.

Hence, different agents participate in this complex structure where extractivism becomes the norm, yet unduly affects disadvantaged people, due to structural constraints (McKeown, 2024b), destroying their environments. Historically, mining has transformed landscapes and polluted environments, with differently interpreted notions of environmental damage and economic benefits (Pesa & Ross, 2021). And in the rush to build compute infrastructures, economic benefits of critical and rare earth minerals dominate the conversation, neglecting the disproportionate environmental costs of extracting minerals.[18] Moreover, these practices are not a one-time affair, as data centres require repair. As GPU AI chips are reported to last only for three years (Whitman, 2024), coupled with the fact that AI innovations are moving at a breakneck pace, such extractions are an ongoing reality amplified by the fact that data centres are projected to grow by 20% in 2025 alone (Taylor, 2025b).

As the differently socially positioned agents have different vested interests (McKeown, 2024b), different agents bear different costs and benefits, whether that be local companies, larger mining corporations, semiconductor manufacturers, chip distributors or the AI development company – and through the structural processes, injustice takes the form of domination and oppression. Domination or constraint on self-determination is experienced on different fronts. Despite depleting water tables and open pits, local communities cannot directly challenge these projects, as they also generate employment for the individuals working in the sector for generations. Where lands can have various meanings attached to them, beyond notions of accumulation as seen in Indigenous traditions, the structures in place due to the extractivist logic reconfigure land as a commodity to profit from. Additionally, resource-rich nations and mining communities are locked into extractive economies due to historical precedents, international debt, and trade dependency, preventing them from regulating prices or diversifying their economy (Pesa and Ross, 2021). Hence, it is unjust for such groups to have little self-determination over their circumstances.

[18] In China, mining for dysprosium yields only 0.2% of the element from mined clay, with the rest discarded, causing pollutants like ammonium. Refining also produces significant waste—75,000 litres of acidic water and one ton of radioactive waste. (Crawford and Joler, 2018)

And while historically, coal mining communities have unionised and leaned towards the left, they are increasingly seen leaning towards the populist right in wanting to maintain the existing extractivist social order (Sörlin, 2022). Natural resource extraction also leads to oppression, where the capacities of individuals and communities for self-development are hindered. As landscapes change, toxins diffuse, settlements jostle, and ecosystems weaken, mining fosters exploitation of the communities. Workers in Chile, Bolivia or the DRC who extract these minerals are poorly paid and the value of their labour is disrupted, while disrupting marginal communities (Matanzima & Loginova, 2024). Communities are being displaced as marginalised, having to move due to the excavation of the Earth, and are excluded from the benefits of AI as well. In the specific instances of the extraction of minerals for the construction of AI infrastructures, these injustices are avoidable.

As extractive practices pollute, destroy, and disrupt human settlements and natural ecosystems, certain social groups are even more at risk of domination. Vulnerabilities engendered by natural resource extraction for AI production can be observed in many cases: lithium mining in the Atacama Desert at Chile, that is depleting freshwater sources essential to wildlife and Indigenous people, who in turn are being discriminated against (Hao, 2025); copper and cobalt artisanal miners in the Democratic Republic of DRC, who are exploited and suffering fatal accidents and critical ailments (Lemma, 2025); nickel mining in Indonesia that destroys sea beds and leads to water contamination, disrupting communities and leading to the social domination of indigenous groups (Jong, 2024). Unsurprisingly, the costs of extraction are borne by the MW, while Western Countries primarily enjoy the benefits. As different countries can fund AI research differently, while simultaneously shaping governance frameworks, such advancements are rarely enjoyed by those who engage in dangerous extractive practices from mines all over the world. Additionally, the safety measures and environmental standards followed by high-income countries are not prevalent in the GS, where the sector is informal. (Falks et. al, 2024). Hence, the production of AI infrastructure produces injustices that are enabled and reproduced by structural factors.

However, these structural injustices are avoidable. As McKeown argues, certain individuals in this structural network have more power, can foresee these harms despite not knowing the exact consequences and can fail to prevent these harms. As scholars who encourage looking at the entire AI product life-cycle have pointed out (Crawford & Joler, 2018; Brevini, 2020; Ricaurte, 2022; Falk & Wynsberghe, 2023; Valdivia, 2024), the dangers linked with the material production of AI technologies can be predicted. With the acquisition and processing of minerals and metals being flagged as a significant environmental indicator of impact at the European Union life cycle assessments of information and communication technologies (Bordage et al, 2021), information about such harms is widespread. Although structural

factors constrain states, they are aware of the environmental harms of mining, as evidenced by local protests and increasing ecological fragility. Companies are aware of the massive reduction in the production of electric devices, from raw material to device, where the manufacturing of a 2 kg laptop requires approximately 600 kg of raw material (Magnier and Mugge, 2022), leading one to speculate on the amount of minerals required to build hyperscale data centres. Despite these factors, data centres are built at a rapid pace with an inexhaustible demand for natural resources that are depletable and polluted and local communities protest, with states continuing with these practices.

Additionally, an agent's ability to act is relative to the influence of other agents and larger constraints (McKeown, 2024b). Afraid of being left behind, developing nations feed into the hype by entering the production chain at any cost. More powerful states, while participating themselves, can also offload the risks onto other states. Although companies are well aware of these costs, they continue to build such large-scale infrastructures to integrate AI technologies into various platforms. Despite the information and ability to act in limiting such unrestrained growth, companies and states move forward, where people on the margins become collateral damage in the pursuit of innovation and profit. While states, corporations and tech-entrepreneurs may not actively want to harm communities and destroy ecologies, when choosing to invest in and carry out rapid development of AI through natural resource extraction, these powerful agents fail to prevent harm.

Hence, natural resource extraction for building AI technologies supported by infrastructure leads to avoidable structural injustice. With critical minerals being sourced for clean transitions as well as green AI, these questions become even more urgent. As NGOs, academic studies and media reports expose the harms of mineral extraction and technological companies publishing ESG reports acknowledging such unjust working conditions, environmental degradation, and geopolitical grabs for minerals, there is awareness about the harms emerging from such practices. And these companies with power have the capacity to intervene, with resources to audit, regulate or reform supply chains and even recycle materials properly. States can also enforce mandatory due diligence laws in the MW. Nevertheless, they fail to eliminate exploitative conditions and go ahead with business-as-usual building data centres left and right, making this an avoidable[19] case of structural injustice. If companies and states acknowledge that the current developmental trajectory of large-scale AI is neither necessary nor sustainable, leading to a reduction of the number of data centres, there would be a significant reduction in natural resource extraction as well, leading to a decrease in harms and injustices to the local communities.

[19] However, if companies *knowingly* partner with suppliers that engage in child labour, unbridled ecological destruction and union suppression, they actively maintain such practices as it benefits them, making these cases deliberate

3.2.2 Labour Exploitation in AI as Deliberate Structural Injustice [20]

Inherited structures conditioning the current practices (McKeown, 2024b) in handling human labour behind AI, go back to debates in 17th century where sceptics like Karl Marx when in conversation with the ideas articulated by Ure and Babbage, wanting to mechanise human skill, cognition and knowledge, cautioned how the automaton requires continuous amounts of energy fed by "mechanical and intellectual organs" (Marx and Engels, 2010, p.82) that can exploit as well as appropriate intellectual resources, privatised for the sake of capital. As inventions like the Mechanical Turk invisibilised the human behind the machine (Crawford, 2021), European colonialism that set up relations in the global economy that gave birth to management techniques broke down manufacturing for efficiency with little room for unionising (Muldoon et. al, 2024), ITC work being outsourced to the post-liberal nations in the MW through BPOs (Muldoon et al., 2024) set the stage for labour behind AI.

The seamless interface of an AI system is maintained by an army of individuals who belong to different professions, located in different parts of the planet. However, the narrative centres around Silicon Valley and Big Tech with its CEOs and programmers, who emphasise automated intelligence, advancing narratives around intelligence, automation, and technoscientific development. Such conditioning narratives elude the fact that these machines are run and powered by individuals who not only invest in products, write only code or fix algorithmic bugs but also work in highly-surveilled Amazon warehouses, manage BPOs and engage in arduous data work. AI infrastructures as global assemblages require a large workforce dispersed around the globe, to build, maintain and repair the entire system – from the infrastructural to the logistical to the algorithmic. Even when focusing on individuals who contribute their labour in making the algorithms work, there is a range of individuals who are positioned differently, with different interests and different benefits. (McKeown, 2024b) – from the content moderator to the engineer who works on making the systems work, but are incentivised differently.

The structural processes are also held together by material and social norms that help agents interact and elaborate these structures in question (McKeown, 2024b). Since the late 20th century, globalisation and neoliberal policies set the stage for large-scale business process outsourcing (BPO) to the GS, resulting in

[20] Debates about unpaid cultural labour have become even more urgent with the emergence of Generative AI that churns out texts, images, and videos having been trained on the publicly available internet commons, exploiting cultural and scholarly labour. However, since these discussions are deeply implicated in the aspects of data extraction, AI deployment, and algorithmic output, these aspects will not be explored in this thesis. Instead, this research project will only focus on precarious data work.

strategies of global labour arbitrage where multinational companies moved their production to a cheaper place to accumulate profit (Delgado Wise & Martin, 2015). This was further enabled by decreasing logistical costs of communication and transport and supported by deregulation and tax incentives in host countries. This structured how the AI labour chain is proliferating, as the Internet connectivity opened newer avenues for connection, trade and outsourcing. (Tyson et. al, 2023) Moreover, unlike an earlier era where moving production facilities required significant capital, say in the automobile industry, this was not the same as shifting data annotation contracts, where clients could transfer their data and lucrative contracts to another BPO in another country (Muldoon et al., 2024, p. 39).

These structural processes nonetheless lead to injustices, as "the extraction machine" (Muldoon et al., 2024) ends up dominating and oppressing specific agents. This is specifically because when these machines are deployed in the workplace, they start another cycle of extraction, following the morphogenetic cycle as laid out by McKeown (2024b). However, due to the AI hype and myths of automated intelligence, the human labour required to sort, label and annotate vast amounts of data and train the algorithms is invisibilised. Nonetheless, a global force of individuals in different positions powers the machine, exploiting the marginalised and incentivising those who help maintain it. The discourse on digital labour, which is the "value-adding activities performed by humans on Internet platforms" (Casilli, 2017), centres around those who are employed precariously, rather than high-tech professionals or data scientists, as it is usually these individuals who are vulnerable to exploitation, manifesting as domination and oppression. It dominates as it inhibits the self-determination of data workers with surveilled workspaces and extreme work hours. As these workplaces actively discourage unionising, as seen in cases of Amazon warehouses (Sainato, 2024), it does not provide the people who offer their labour any means to negotiate. Data workers in the GS are not being compensated adequately, further compounding oppression. Exploitation is foundational as these digital sweatshops create unsafe environments where labour and value are transferred up the supply chain through sub-contractors and BPOs. Precariously employed, the data workers are powerless due to their expendability, in countries like India or Kenya with high unemployment rates, constraining them to such jobs. This traps people in dangerous, exhausting and underpaid jobs. What emerges as a "planetary market" becomes a race to the bottom in terms of wages and contracts, where BPOs are pitted against each other, and workers are exploited. At the same time, investors and managers enjoy the spoils, where different individuals with different amounts of power have more freedom in the structure. (Muldoon et al., 2024) Hence, nations, groups and individuals are constrained by the structure that compels them to transfer their productive labour, in exploitative conditions, for the growth of other economies, limiting their chance for self-development.

As these tasks are outsourced to third parties in the MW in countries like India, Kenya, Venezuela and the Philippines, the wages are reduced to as little as $1-3 per hour, depending on the region and subcontractor. Moreover, while data workers may take up such jobs as an extra source of income in the GN, for those in the MW, this serves as their primary source of income (Muldoon et al., 2024). Whether one works from home or reports to a tightly managed BPO, temporary contracts via third-party platforms expose precarious working conditions wth no benefits and protections. With recorded instances of children taking over their parents' work in Venezuela (Posada, 2022), Kenyan content-moderators having to sift to disturbing multimedia leading to PTSD yet being fired when trying to unionise (Periggo, 2022), and annotators being denied their payment if clients are not satisfied (Suri & Gray, 2019), these data workers are heavily exploited and further managed by the algorithms they help create.

Built from the exploitation of labour, AI technologies can enable further extraction of labour, and during their production, they come to be through a deliberate case of structural injustice. Deliberate structural injustice takes place when people who are relatively more powerful, intentionally maintain these structures, despite foreseeing the harms and having other alternatives, as it works in their interests to stabilise these processes to maintain their power and social position. Technological companies actively resist structural change, instead claiming plausible deniability due to the nodes of supply chains that form the labour value chain, washing their hands off the blame by pointing to BPOs or subcontractors (Mutemi, 2025). However, such shifting of blame and/or inaction is made possible by the discursive and institutional power these companies hold, deliberately maintaining these structures that outsource labour to other parties who then, in turn, look for the cheapest alternative for inflating the profit margins. The fact that alternative practices can be envisioned and deployed demonstrates that this is a case of deliberate structural injustice. The possibility of fair compensation models and ethical labour standards, as exemplified by a not-for-profit organisation like Karya,[21] shows that if flagged as an issue by regulatory mechanisms and structures are navigated to create incentives for alternative alignments within existing practices, alternatives are possible (McKeown, 2024b). Hence, by working with these alternatives, companies can ensure that these supply chains are fair. However, these harms become strategically and deliberately maintained by powerful groups or individuals due to their vested interests in generating profits. It is specifically the data-workers who collect, annotate and verify datasets, which comprises 80% of the AI training (Muldoon et al., 2024). However, despite being an indispensable and the most lucrative

[21] Karya in Bengaluru, India, collaborates with NGOs to help data workers record their mother tongue via an app, creating new datasets and promoting collective ownership for ongoing benefits from sales. Karya pays significantly above the minimum wage and limits individual working hours (Muldoon et al., 2024)

portion of the AI infrastructure, the workers who power it suffer through dehumanising and precarious conditions, with companies maintaining these structures of subcontracting.

3.3 Interplays of structure and agency

As the structural injustice framework tries to capture the ordinary injustices that are cumulative of (seemingly) benign actions within a certain structure, ethical and political analyses of AI have scaffolded on this framework – whether that be to think about criteria of data classification in medical AI (Lin & Chen, 2022), racial politics in predictive analytics in health (Himmelreich & Lim, 2022) or the thoughtlessness it encourages due to the structure being taken for granted and the injustice, untraceable (Browne, 2023). These accounts reach similar conclusions to the claims made in Chapter 1 about how debiasing algorithms cannot merely achieve AI fairness but must take into account the social structure in which the technology is embedded. This requires legitimising communications of diverse kinds, even those neglected from governance mechanisms (Browe, 2023), as structural injustice requires collective action rather than technocentric solutions from engineers or data scientists. (Lin and Chen, 2022). This requires the social and moral considerations about AI to move beyond mere compliance with the ethical codes and principles (Himmelreich and Lim, 2022). However, these accounts rely on the Youngian (2011) framework, mainly focusing on the question of data bias AI generated from a structurally unjust society (Lin & Chen, 2022; Himmelreich & Lim, 2022) that is untraceable and requires larger deliberation and collective responsibilisation (Browne, 2023). Instead, this Chapter has applied McKeown's (2024b) framework of structural injustice to the material production of AI. This would result in a transition of theorising from second-wave concerns of the sociotechnical to the third-wave concerns of the power-infused structural.

However, scholars worry that McKeown's revisions to Young's framework may make it redundant altogether as the structural element may be missing in the pluralist account by the blurring between the structural/international boundaries or the injection of intentionality through the notion of power (Marin, 2024; Bufacchi, 2024; De Bernardi, 2024). McKeown (2024a) defends her account against these critiques by emphasising two aspects of her concept of deliberate structural injustice. First, everyone is constrained within the system, from the powerful to the powerless, either to make profits or to survive. However, powerful agents can wield their power through various means (discussed in Chapter 4), like lobbying or setting industry standards to either maintain the status quo or to change it, as they do in cases of intense scrutiny. However, since such processes have several layers, like fast fashion production, which has dense

layers of subcontracting, the powerful agents may not have total control within these larger systems. Nonetheless, just by possessing power, there is immense pressure on agents on the downstream tier of the supply chain to perform on a particular scale and at a rapid pace to fend off competition. Moreover, this pressure is partly shaped and even manipulated by the activities of the powerful. Second, the fact that such structural conditions arise from several larger factors, like decolonisation, globalisation, changes in trades, etc., does not change the reality that there is manipulation of the system by powerful agents through lobbying, law-making or policy-insertions. Hence, McKeown emphasises that both aspects can be true where injustices emerge from structural processes with powerful actors manipulating situations for maximum benefit. Rather than prioritising one over the other, this demonstrates the "interplay of structure and agency" (McKeown, 2024a), reflecting the reality of how structures work in practice. Also, several issues work together, impacting the same individual, where poverty and sweatshop labour can work to create vulnerability and exploitation, resulting in avoidable and deliberate forms of structural injustices to work in tandem. This explanation of McKeown (2024a) can also help understand the cases laid above of avoidable and deliberate forms of structural injustice, where structure and agency can have interplays and avoidable and deliberate injustices can co-exist.

An advantage of relying on this conceptual framework is that it starts from identifying and comprehending injustice, rather than making a preemptive normative proposal, making it sensitive to the contingencies of the particular case. On a broader scale, this affords the theory what Himmelreich and Lim (2022) call explanatory power that can unify disparate phenomena that "show up among causes and consequences" (p. 8). And this can very much help us understand the pre-technical structural foundation that undergirds the production of AI. Additionally, relying on McKeown's (2024b) framework, it also explains why vulnerable agents endure (and sometimes propagate) forms of oppression, as it works in their favour, albeit in a limited scope, as they need to play along the rules of the structure due to their powerlessness. Also, as articulated by McKeown, with detailed analysis that integrates powerful agents, possibilities of structural change emerge as it provides a framework to consider what alternatives are available to subordinate agents, who, by their participation, reproduce these structures. And this is what leads to questions of responsibility for structural injustice – political for the ordinary individual agent (which is beyond this thesis yet important for its future research trajectory) and moral for the powerful ones.

Chapter 4

Moral Responsibility and AI Extractivism

This chapter argues that, based on McKeown's account, moral responsibility can be attributed to certain agents in the AI supply chain by virtue of having more power over it. Building on Chapter 3, which analysed mineral extraction and labour exploitation as types of structural injustices, where it was shown that certain agents have more influence within these structures, this chapter will identify these powerful agents across the AI supply chains. As a result, I argue that moral responsibility ought to be attributed to the most powerful agents, which are the MNCs or the BigTech Giants investing in computer infrastructure and outsourcing data work. To do so, §4.1 and §4.2 compare Young's Social Connection Model, which favours a forward-looking notion of responsibility, to McKeown's notion of responsibility, which argues for allocating moral accountability to agents with the most power. This view is then contrasted with Miller's notions of outcome and remedial responsibility to highlight the centrality of power in attributing moral responsibility, as on a structural level, an agent requires resources, positioning, and capacity to produce outcomes and rectify injustices. §4.3 traces power in the AI supply chain, where power concentrates around the Big Tech giants. This supports McKeown's claim that MNCs are, in fact, the most powerful agents in the global structure, especially in relation to nations in the MW. To specify what forms of power are exercised by these corporations, §4.4 provides a descriptive analysis of the MNCs' disposition, episodic and systemic power in the extractive processes. §4.5 applies McKeown's theory to cases of natural resource extraction and labour exploitation and argues for the conditions that attribute moral responsibility to MNCs involved in avoidable structural injustice, consisting of structural harms that are foreseeable and deliberate structural injustice, consisting of structural harms that are maintained. §4.6 contrasts the dominant discourse in AI ethics, introduced in chapter 1, to the account proposed in the thesis, to argue for the importance of collective responsibility for the whole supply chain of AI to account for extracivist harms and concludes the thesis hinting on the notion of structural responsibility, which requires the interplay of both moral responsibility and political responsibility for a structural overhaul.

4.1 Moral Responsibility: Young's Social Connection Model

For Young (2011), the liability model that attributes moral responsibility does not work in cases of structural injustice. Under this model, an agent can only be held responsible for a harm if they have directly and intentionally acted in a way to cause harm, with the knowledge of the likely consequences of their actions. However, since structural injustice is cumulative with many agents with varying intentions acting within structures without knowing how the consequences unfold, the liability model falls short.. Most importantly, the liability model ignores the background conditions that make up structures which reproduce injustices and perpetuate harm. (Young, 2011, pp. 100-106). To address structural issues, Young (2011) proposes the "social connection model" that requires political responsibility from individuals connected to the structure and reproducing injustices unintentionally, indirectly or negligently. It does not mean having to do something for or by oneself, but is to be shared among everyone, even the victims of the injustices. Political responsibility, being forward-looking, also aims to overcome structural injustice, without attaching blame to an individual or a group, making it a non-blameworthy form of responsibility. Hence, it entails "an obligation to join with others who share that responsibility to transform the structural processes to make their outcomes less unjust" (Young, 2011, p. 96).

Nuanced critiques have been levelled against Young's Social Connection Model and notion of political responsibility. While some scholars point out that the forward-looking idea of political responsibility may grant a free pass to wrongdoers (Nussbaum, 2011) while indulging in victim blaming (Gould, 2009), others question what type of connection to the structure generates political responsibility (Aragon & Jaggar, 2018; McKeown, 2018). However, the most prominent objections have been regarding non-attributability (Zheng, 2018). Many scholars ask whether this model or form of responsibility fails due to its refusal to attribute moral blame on any agent whatsoever (Barry & Macdonald, 2016; Hahn, 2009; Neuhäuser, 2014; Riley, 2017).

4.2 McKeown: Power and Responsibility

For McKeown (2024b), the fine line between political responsibility and moral responsibility can be navigated by considering the role of power in structural injustice. While she defends Young's conclusion about the ordinary individual being excused from moral responsibility, McKeown leans towards an Arendtian conception of political responsibility to hold individuals accountable. However, she distinguishes her position from that of Young by arguing that people with power within these structures should be held morally responsible. If powerful agents can alleviate injustices yet fail to do so or deliberately maintain structures that cause such harms, then they can be held liable for injustices. While these instances are still cases of structural injustice that are characterised by structural causation and

structural constraints, she accommodates moral responsibility within the framework where one ascribes blame to an agent by their position in the structure that offers them influence. Hence, she argues that collective moral responsibility should also integrate the notion of power.

4.2.1 From the Individual to the Collective

McKeown (2024b) subscribes to philosopher R.J. Wallace's (1998) reasons-responsive conception of moral responsibility, where an individual can be held morally accountable if they fulfil these two conditions: first, they can grasp moral reasons, and second, they can control their behaviour or reactive attitudes accordingly. However, they can be excused from these conditions if the act was performed inadvertently, was a result of unintentional bodily movements, or if the individual was physically constrained or coerced under duress. McKeown points out that in the context of, say, deliberate structural injustice, where injustices resulting from complex structural causation and constraints are deliberately maintained by powerful agents, if an ordinary individual inadvertently contributes to the structural maintenance, without intending to cause harm, due to external pressures in question, they can be excused from moral responsibility. For instance, a low-income individual purchasing a fast-fashion item for their interview that requires them to dress formally can be excused from moral responsibility. This is due to the structural nature of sweatshops, with their global network, multiple processes and several agents and institutions entangled in the production of garments, where a single purchase makes an imperceptible difference that cumulatively adds up.

However, unlike consequentialist claims about cumulative harms, where the harm is the consequence of several agents working together, there is no direct linear connection in cases of structural injustice. When one purchases a fast-fashion item, due to the complexity of supply chains with disparate agents with different degrees of power alongside intermediary agents, the consumer cannot fully gauge the impact of their actions on a macro scale. Additionally, notions of triggering the production of clothes do not necessarily translate into negative utility, as they also employ women in the GS through chains of impact sourcing, making an overall evaluation more complicated. Hence, not consuming these products can also lead to job losses for the agents involved in the production. McKeown calls the consequentialist focus on individuals' causal connection or triggering contribution to the structure a "red herring" (2024b, p. 137), as it takes away the focus on the structures and the agents with power to change these structures. McKeown also argues against the complicity principle, according to which one can be made accountable if one intentionally participates in structures or acts that result in wrong or harm, independent of one's

actions. However, these accounts fail to account for power, says McKeown (2024b), as companies or states can influence these structures of distribution. If companies did not produce fast fashion, consumers would not purchase it. Additionally, with little capacity to make moral choices, the individual's capacity to create larger structural changes is limited due to their lack of power. Instead, the attention must be turned to the most powerful agents in global structures, who are usually larger corporate agents like institutions, states or companies.

To emphasise grave moral accountability owed by corporate agents in particular, McKeown (2024b) starts from the conditions of moral responsibility that must be adhered to identify morally responsible agents. She prefaces her arguments by specifying that it is only conglomerates that can be considered as institutional moral agents if they have maintained an identity over time and if they are more than the sum of their parts, being self-asserting and defining with an internal decision-making structure. (McKeown, 2024b, 182). If a corporation meets this criterion, the corporate agent itself, and not its members, can be held responsible for its behaviour and actions. First, she extends the conception of individual moral responsibility to that of collective responsibility, starting from Strawson's claim that moral responsibility is a social practice. She argues that despite the accounts that reject the possibility of collective responsibility, which very much align with her conception of individual moral responsibility based around intention and causation, the reasons-responsive account is adequate to argue for the possibility of collective responsibility. Against the claim that moral responsibility requires *actus reus*, that is a body to carry out acts, she argues that in the case of corporations, a principal-agent relationship applies. Hence, when a principal orders the agent to carry out the deed, both the agent and the principal are responsible; however, the arrangement can unfold as such where a corporate agent, as a member, acts like the principal who authorises the corporate agent as an entity. This can be inconsistent with the fact that the internal decision-making committee also becomes the principal to authorise employees/functionaries further to act as agents on its behalf. Hence, sceptics would claim that the responsibility would devolve into individual members.

This also leads to the objection of a lack of *mens rea* (intent), which is seemingly absent from corporate agents. McKeown (2024b) posits that through decision-making structures within corporate arrangements, corporate entities can have deliberative processes between persons, with the social capacity for moral reasoning by drawing on the reasoning capacities and moral emotions of members who talk to each other about what the corporate entity ought to do. Additionally, the intent is not reducible to that of the individual acting but rather is authorised by internal decision-making procedure while also being a product of corporate culture and a set of policies and rules. And while corporate agents are constrained by

the structures as discussed in Chapter 3, to secure the interests of their shareholders, these entities do have the capacity to think about moral reasons and act on them. And these agents especially have more of a capacity due to their greater power and resources in comparison to individuals acting alone.

McKeown (2024b) also contextualises her arguments regarding how moral responsibility ought to be carried out by corporate agents in the structural ontology that she upholds – on the two fronts of the entity being morally responsible and aspects of holding the entity responsible.[22] She argues that although corporate agents are constrained by structural factors, like capitalist corporations primarily conforming to the profit motive, McKeown (2024b) insists that corporations should bear moral responsibility in cases they deliberately perpetuate or fail to redress structural injustices and fulfil the conditions listed above, arguing against the notion of structural constraint. Hence, the blameworthiness emerges from the fact that these agents maintain structures through and despite their powers, which result in injustices through the exercise or the possesion of such capacities. A corporate entity becomes morally responsible because it upholds, contributes to, and continues structures that result in injustices, despite no direct linkages or causal traces by virtue of its power.

And it is vital to hold these entities responsible as well. If not, this can lead to responsibility deficits, leading to the incentivisation of wrongs and encouraging individuals to perform as corporations to mask their wrongdoings, which can reinforce structural rules that further perpetuate the injustice. Additionally, she emphasises that some responsibilities, like climate agreements, ought to be imputed to corporate agents, or else individuals, who have limited influence over the structure, may be the only ones recognised as the bearers of responsibility. Hence, this can lead to deficits for the past where the corporate agent was in the wrong, also leading to prospective responsibility deficits. Further, public blaming of such corporations can spur actions and generate efforts to change behaviour and restore reputation – this in turn can lead to changes in laws and a sense of legal liability. Corporations, due to their size and the resources at their disposal, have the capacity to implement changes in their industries, which, despite being small-scale in the face of global power asymmetries, can still have the potential to improve lives through partnerships with civil society organisations or NGOs. Further, in common practice, corporations do perform acts that show them shouldering responsibility – they can formally apologise for past wrongdoings, write codes of conduct, engage in philanthropy and private-public partnerships or address

[22] The distinction between entities being responsible and holding these entities responsible has been widely discussed collective responsible discourse. While collective agents' being morally responsible pertains to what they have done, holding them responsible pertains to what should be done if they are morally responsible. Hence, while the former is a moral fact about the corporate agent, the latter pertains to legal or social factors surrounding the same. Defenders of collective responsibility do not necessarily distinguish between the two (Smiley, 2022) and, although McKeown does not explicitly outline this distinction, I interpret that she addresses both aspects.

human rights violations.[23] And what provides the capacity for corporations to shoulder responsibility is the extent of their power in larger structures.

4.2.2 Forms of Power

Power is central to McKeown's account of moral responsibility. She deepens her analysis[24] on the situated notion of power (as discussed in 3.1.2), by combining it with other accounts of power. For the forms of power, McKeown draws from political scientist Mark Haugaard (2010), who conceives of power as a family resemblance concept, where the meaning of what it means depends on the language game – whether that is describing how agents exercise their power or how they ought to navigate such relations. She enlists the three 'forms' of power: first, episodic power being linked to exercising one's agency; second, dispositional power signifying the inherent capacities of an agent, whether or not they exercise this capacity; and third, systemic power referring to the way social systems confer dispositional power on agents, structuring possibilities for action.

These aspects of power can interlock to unfold in various ways, as seen in the cases of larger corporate agents like states or MNCs. McKeown (2024b) argues that the role of MNCs is very much neglected in the larger debates on global justice despite having the most influence over the structure. She argues that MNCs have immense systemic power as the international political economy is structured around them, especially those that are geographically concentrated in an economically powerful state. Since MNCs exist at the discretion of the states, which invest corporations with personhood, they are not just shaped by their embedded contexts but also can restructure their settings. Hence, they are also conferred with the dispositional power to make changes to the structure itself actively. Due to their immense economic power, they can constrain the options of other actors without exercising their power, as seen in the cases of poorer countries that depend on the investments of MNCs, acting in anticipation of their wants. NGOs partner up with MNCs despite their fundamental disagreements, as MNCs can make a significant environmental and social impact. Additionally, due to their relationship to their states where corporate headquarters are located in only a few countries like the US, EU, China, and Japan, where advancing the interests of the MNC also advances the interests of the home country, as discussed above, and where MNCs are also largely shaped by the culture of the home countries.

---

[23] While this thesis is not focusing on states due to pragmatic constraints, McKeown does consider and affirm the moral responsibility of the state. Despite the challenges in holding states liable due to their evolving composition and burden on citizens, failing to do so can dilute international cooperation, intergenerational responsibilities and historical reparations.

[24] McKeown also mentions that there are five dimensions of power that interlock that manifest in the forms of power, which shall not be considered in this thesis.

Dispositional power of MNCs is strengthened by the fact that governments invite MNCs to discuss policy, where bureaucrats depend on their inputs due to the complexity of regulation, with politicians also being involved in businesses. MNCs have a seat at the policy-making table, both at the domestic and international levels. Hence, due to their structural position, MNCs can get other actors to do what they would not otherwise do, even if they do not act. They affect agenda setting without actively getting involved. They shape the needs of other actors in a way that is perceived as normal and inevitable, and MNCs produce and discipline other actors within the global political economy. Also, they not only constrain options but can also exercise episodic power to make changes that benefit them. This can be observed as these actors exercise their power through lobbying, engaging in regulation setting, Corporate Social Responsibility (CSR)[25], and instances of manipulation through advertisements and social media. (McKeown, 2024b).

4.2.3 Power and Moral Responsibility

McKeown's (2024b) equation of power and moral responsibility could be questioned, especially when it comes to the pragmatic reasons of addressing and compensating for harms. For instance, Miller's (2007) account, which also extends individual responsibility to collective responsibility, seems to provide a straightforward answer to holding corporate agents like MNCs responsible without having to account for power. However, Miller primarily focuses on nations as collective/corporate agents, unlike McKeown, who considers both nations and MNCs. For Miller (2007), nations, being more than groups of individuals, are cohesive moral agents with shared institutions, values, and a continuous identity over time. This continuity allows for attributing collective responsibility to agents across generations. Additionally, since citizens participate in national decisions and benefit from collective outcomes, moral responsibility can be reasonably assigned to the nation as a whole, justifying shifting responsibility from the individual to the collective.

In the context of global justice, Miller (2007) distinguishes between "outcome responsibility" and "remedial responsibility." Outcome responsibility pertains to an agent's accountability for their actions and the resulting consequences, emphasising agency and foreseeability. Miller points out that while moral responsibility requires outcome responsibility, the reverse is not true, and also that outcome responsibility is distinct from causal responsibility. In contrast, remedial responsibility refers to the obligation to

[25] CSR refers to a business's commitment to integrate social, environmental, and ethical considerations into business operations and strategy, going beyond legal obligations. However, these face a range of critiques where they are seen as greenwashing attempts or moral licenses.

alleviate a negative situation or provide aid to those in need. Its criteria for assignment include moral responsibility, outcome responsibility, benefit, capacity and community ties.

In this way, agent-focused outcome responsibility can serve as a basis for identifying who bears patient-focused remedial responsibility. Hence, for Miller (2007), moral responsibility would support outcome responsibility and remedial responsibility. And Miller specifies that for an agent to be at moral fault, they must have either deprived people deliberately or recklessly or failed to provide for them despite a pre-existing obligation (p. 100). This would be because the agent would have acted in the way that resulted in an outcome that was blameworthy, for which they ought to be held liable, so that the people affected are provided with aid. Hence, Miller's account could account for moral responsibilities in cases without causal connection, like structural injustices that are avoidable or deliberate, without having to resort to power, offering both conditions for identification and assignment that are liability and pragmatic reasons.

However, McKeown's approach is better suited to examples of structural injustice as it more directly identifies the role of power in maintaining unjust structures (discussed in Chapter 3). Despite acknowledging the immense power states and nations have as corporate agents, McKeown argues that the global justice debate neglects considering how MNCs are implicated in the power of the state itself, and how MNCs situated in the geographical GN have more resources, capabilities and capital than lower-income countries. Hence, her account can align with Miller's framework yet emphasise the fact that in a globalised world, states themselves contribute to and draw from the power of MNCs. Although Miller acknowledges power asymmetries in his book, his notion of collective responsibility through outcome or remedial responsibility does not explicitly mention power. Nonetheless, through Miller's criteria for outcome/moral responsibility, these injustices can be seen as the byproducts of deliberate in/action and foreseeability by specific agents, who, on a structural level with diffused actions of peripheral social actors, make use of their resources to maintain their position and the status quo. And moral/remedial responsibility would entail for the outcome-identified agents that can rectify harms that work in their benefit, while being conscious of the fact that several other agents are involved in the said structure. Hence, on a structural level, when dealing with ideas of poverty or climate change, outcome-responsible agents who have the resources or the position to deliberately cause or fail to address deprivation and remedy-responsible agents who can rectify this have power. Both conditions for outcome/moral responsibility and moral/remedial responsibility, when addressing large-scale, structurally created and constraining issues, rely on the assumption of an agent possessing power.

4.3 Tracing Power in the Development of AI Infrastructure

As discussed in the previous chapters, AI technologies materialise and function due to a global supply chain of natural resources, data, and labour, with different agents working within structures. While differing geographies and political systems have distinct modes of operations, they follow the same path, partially shaped by the systemic forces of extractivism. For natural resource extraction, the chain starts from the mining of elements by multinational mining corporations to extract valuable elements. These processes are regulated and supported by national governments, state-owned enterprises or international mining companies. Following the extraction, the materials are handled by refineries, which can be privately owned or state-funded. The processed minerals are manufactured as hardware for manufacturers to fabricate components or build data centres that are owned mainly by Big Tech. Labour chains are also deeply stratified, with the base layers being the AI data workers, often gig workers or precariously employed annotators from the GS who are hired through business process outsourcing companies or annotation platforms that act as intermediaries to extract value from human labour. The chain is further maintained by subcontractors who are also large outsourcing firms. All these units together report to BigTech and AI firms like Open AI, Google and Meta, who design the models but outsource the labour-intensive steps of data curation and feedback tuning.

In both these supply chains, distinct agents acting within structures, resulting from a confluence of systems, are globally positioned with differing vested interests. And their situated power relation (discussed in chapter 2) depends on whether they form the upstream or the downstream component of the supply chain. The upstream of a supply chain consists of those agents that source, supply, and refine materials for manufacturing, and the downstream involves agents who engage in activities post-manufacturing for distribution and consumption (Johnson-Wood, 2023). This involves different ventures and processes for natural resource extraction and labour exploitation. For natural resource extraction, upstream activities involve extraction, processing, refining and smelting, with downstream activities involving marketing and distribution. (Bingham Chee, 2022) When thinking about material AI infrastructures, this combines two industries and supply chains, of mining and electronic manufacturing, resulting in a longer chain. For labour exploitation, upstream activities centre around data collection, data annotation, and model training and downstream activities include deploying AI applications and model evaluation (TechEquity, 2025). With multiple agents being a part of webs of supply chains, power mediates the relations among all agents involved. With upstream tiers usually associated with critical

social issues (Bingham Chee, 2022), having to supply materials/labour out of necessity and precarity, power flows downstream.

This can be analysed through McKeown's situated notion of power and collective responsibility. As discussed in the previous chapter regarding the situated notion of power, social positions in the structure often influence agents' behaviour towards complacency and/or dominance (McKeown, 2024b). This can be observed in both cases of natural resource extraction and labour exploitation, where agents working in upstream activities are positioned in ways that make them vulnerable due to their expendability. Whether it be the miners, small-scale mining firms, data labellers, platform workers, BPOs or sub-contractors, their employment and agreements with actors who can influence the chain become imperative to their employment. This is especially true for agents from the MW, where struggling economies and rampant unemployment make them complacent towards altering their behaviour to align with the expectations of the agents on the downstream tier, whether it be states or multinational companies. However, the geological availability of specific resources in resource-rich countries remains an important factor in how poorer states can strategically position themselves in the AI development pipeline and at what costs. This geographical advantage is not provided to states in the GS when it comes to the structural dynamics of labour exploitation, as the risk of moving locations leads to precarious and underpaid employment of highly educated youth (ILO, 2024).

The power dispersed across the supply chain also depends on peripheral social agents like workers, intermediaries, labour ministries, where coordinated practices create alignments to legitimise power (McKeown, 2024b) While poorer states like the DRC or Indonesia do regulate and enable the flow of these resources, they grant mining concessions to MNCs as these governments are incentivised to prioritise debt relief or short-term economic growth over enviornmental or social protection. Hence, the need for investment leads the states to roll out policies that allow tax exemptions, weak environmental oversight or land dispossesion (Ali et. al, 2017; Manberger & Stenqvist, 2018) when catering to the interests of more powerful corporations like MNCs, especially BigTech like Amazon or Microsoft, who are wealthier than the powerful states themselves (Wallach, 2021). Further, GN states can enable such structures by participating in trade agreements, export controls and investment protections that secure mineral supply chains for their tech industries. Policies like the EU's Critical Raw Materials Act or China's Belt and Road initiative (García-Herrero, 2024) also hint at how richer states have control over critical minerals in attempts to transition their economies, secure their futures and bolster economies, while also letting them have control over AI-relevant materials. These states also offer tax subsidies

(Tortorelli, 2025) and diplomatic support (Höne, 2022) to homegrown BigTech firms, with legal systems being amended to regulate yet advance their efforts.

Additionally, the states in the MW can passively regulate precarious labour, as countries like Kenya, India, or the Philippines have welcomed annotation platforms like Sama or Appen to attract BPO investment as a part of their digital development strategy. With lax regulatory mechanisms creating a disciplined workforce and lower costs, states can passively, if not reluctantly, legitimise exploitation (Muldoon et. al, 2024). On the other hand, GN states can uphold structures that allow tech companies to outsource labour while avoiding employer responsibility, as there are no binding obligations on transnational companies to ensure labour rights in their supply chains, despite their capacity to do so via trade or corporate law (Fairwork, 2024).[26] Hence, different parties, whether that be the workers or the venture capitalists. Several agents agree with the fiscal, legal, and socio-political rules of the structure, stemming from the extractivist logic, to maintain their power.

This situatedness also makes the powerful/powerless distinction contextual and relational (McKeown, 2024b). States and local mining bodies are constrained by historical-colonial patterns of resource exploitation, leading to strategies to relieve debt and bolster economic growth, yet have the power to deem lands as fit for mining and decide what to pay individuals. BPOs or subcontractors who are constrained by the financial contracts they acquire through tech companies to perform and deliver efficiently are also enabled by them to discipline, surveil and exploit individual data workers. This makes power distributed across the chain shift its axis, with a higher concentration of it as one traces the supply chain downstream. Hence, while poorer states have more power than their citizens, they are also constrained by larger forces downstream. And the most powerful[27] among all agents, individual and corporate, are the MNCs – especially BigTech, who ought to be held morally responsible.

4.4 Powerful MNCs

As discussed above, although states are complicit, the most influential agents in how much they can stabilise or disrupt the structure are usually MNCs. This becomes especially evident when we consider the

[26] However, states can intervene, as seen in Kenya, where judicial bodies ilke the court have become sites for data workers to sue platforms like Sama and Meta over harmful working conditions. Additionally, nascent EU legislation on AI and supply chain due diligence could expand the state's regulator role.

[27] While larger states can also bear immense power and hence, responsibility especially considering the historical structures that give rise to the current structures that were largely shaped by these nation states and forgoing the state versus market dichotomy, for the purposes of this thesis and its constraints, we shall only be focusing on MNCs. However, this is something that would be an interesting angle to be explored in future research.

kind of power the MNCs wield over these structures, ranging from episodic to systemic. When it comes to AI infrastructures, MNCs can exercise power on all three levels for natural resource extraction. And with a longer supply chain, this combines corporations working not only in mining and electronic manufacturing but also Big Tech, who build such data centre facilities. On a systemic level, the global economy is structured to normalise and sustain the extractive relations between the GN and the MW, with the resources serving Eurocentric and Anglocentric, and recently, Sinocentric interests. With unequal terms of conditions and commodity dependence that ensure raw materials are undervalued compared to finished good from the North (Ricci, 2019), the AI sector can widen these inequalities as there is an intensification of extraction from the DRC and Chile (Pereira, 2025) The rampant development of AI infrastructure that depends on elements sourced from postcolonial states with weak regulatory capacities a systemic power dynamic is sustained as global trade regimes, investment treaties and development institutions continue to prioritise profit over human rights or ecological stability. These structures are the legacy of colonial economic systems that have been sustained through neoliberal reforms and free trade agreements that lock resource-rich countries into roles as raw material suppliers. MNCs such as Google, Amazon, Apple or NVIDIA benefit from such systemic power that further confers them with dispositional leverage.

With the capacity to shape outcomes due to their structural position, these firms exert their power through their procurement dominance (Ølykke, 2023), where buyers have a strong position in the market, and they wield significant influence over suppliers. Hence, BigTech being clients of the Mining MNCs can also exhibit dispositional power by setting sourcing terms, technical requirements and sustainability thresholds. Due to the ubiquity of AI and rising demand for AI infrastructures, demands for mining or smelting companies to supply these data centres have increased. Hence, due to their wealth and technological primacy, MNCs can dictate terms across continents without taking direct ownership of extractive operations due to their entanglement in dense supply chains. MNCs (mining or BigTech) can exercise at an episodic level by engaging in interventions like CSR campaigns and sustainability certifications with initiatives that often lack meaningful enforcement (Brevini, 2022).

The systemic, dispositional and episodic faces of power can also be observed in labour exploitation for building AI. The data labour divide (Casilli, 2017) is being sustained by platform capitalism emerging from a deregulated labour market through neoliberal reforms. States in the GS influenced by structural adjustment programs and digital economy incentives have created policy environments where informal, underpaid digital labour is normalised as a path to development (Graham et. al, 2017). Big Tech, like Meta or Google, also wield a lot of dispositional power in this labour structure. While they do not employ the data workers directly, they exert immense influence as clients of platforms like Sama, Appen or

iMerit, which manage labour under metrics, quotas and opaque performance systems designed to meet the standards and timelines of tech MNCs. Such control over the value and time of the worker enables MNCs to shape the working conditions of thousands of workers without bearing legal responsibility. Their wealth, market dominance and data dependency give them disproportionate influence over how labour is harnessed across continents. These firms also exercise episodic power through participating in governance and regulation settings by investing in AI ethics boards, signing human rights declarations, and funding social impact programs that offer symbolic gestures of reform, as seen with Google. Yet, these actions often work to obscure their reliance on precarious, invisible labour. For instance, Meta has repeatedly claimed it is committed to safe working conditions, even as its subcontracted moderators in Kenya have taken legal actions over exposure to graphic content and psychological trauma (Roberts, 2019). Hence, the episodic power reinforces the systemic and dispositional advantages of corporate actors.

4.5 Moral Responsibility, AI Extractivism, and Structural Injustices

Technology MNCs, consisting of the giants of BigTech corporate agents, can be classified as conglomerates. These firms are demonstrably more than the sum of their parts, as despite personnel changes, CEOs or COOs, Amazon or Google remain the same. They also maintain a distinct identity over time. Despite economic diversification or strategic "rebrands," Amazon maintains its customer-first principle. Meta and its merged social media platforms, Microsoft, as a software enterprise and cloud provider, have access to their history to chart projects for future initiatives. These MNCs also possess sophisticated internal decision-making structures, with investors, major shareholders, C-Suite executives and managers allotted roles and procedures to follow. Big Tech companies are profoundly self-asserting; with their charters, mission statements, and products not as a collection of designers, engineers or marketeers but as the firm itself, as a corporate entity.

As discussed in the previous chapters, natural resource extraction reflects and reproduces extractivist systems rooted in larger structures, resulting in avoidable structural injustice. These are avoidable as they are foreseeable due to rampant yet unnecessary construction of AI infrastructures that unintentionally lead to the domination and oppression posed upon local communities due to environmental pollution, disruption of communities and toxic working environments. And this can be mitigated or avoided by powerful people so that people who are harmed due to the structural processes of normalisation of resource extraction for economic progress do not suffer collaterally. Additionally, they fulfil the conditions of collective responsibility for these entities to be held morally accountable.

The decision-making committees, which are instantiated to discuss issues surrounding the sustainability of these systems, whether that be mining or land acquisition, or the outsourcing of labour to the MW, with protests, reports and legal warnings providing evidence, provide MNCs the means to grasp and apply moral reasons. Even if these larger decision-making committees do not directly make decisions about this, as principals, they endorse the action of agents like regional managers to decide which mining companies for mineral extraction or which nations to outsource to, whose priorities and motives are shaped by the ethos of these MNCs, centring around innovation and profit. And the fact that these institutes have an ethics board that devises charters with principles displays their intentional efforts at engaging in moral reasoning to regulate their behaviour. However, despite these commitments, avoidable and deliberate structural injustices in the production of AI technologies continue. Where companies like Google promise AI for sustainability (Google, 2025) yet keep building data centres (Harkansas, 2024) that can set back sustainability goals, which must be thought beyond energy efficiency, van Wynsberghe's contrast between AI for sustainability and sustainability of AI becomes stark (van Wynsberghe, 2024). In the case of labour chains, despite labour union protests and national jurisdiction involved in action against exploitative BPOs with harmful work environments like Sama, MNCs like Meta have infamously denied responsibility, discontinuing their association with the BPO, with Open AI outsourcing their annotations to the very same BPO (Rowe, 2023). With terminated contracts and set sustainable goals, BigTech giants also closely guard these details as they have increasingly come under scrutiny. Hence, despite moral reasons at hand and capacities to reflect, apply and learn, they fail to regulate their behaviour.

These MNCs also cannot be excused on the grounds of inadvertence or constraint. While structural injustices are a result of socio-structural processes that have unintended consequences, having more resources, research teams, and ethics boards enables them to reasonably foresee the harms that will result from the rampant construction of data centres and also choose to maintain supply chains that result in more profit. These Big Tech giants have immense influence over the structures, which can be complex and knotted, due to the various forms of power they possess to maintain or facilitate the structures that reproduce extractivist harms and structurally unjust outcomes. They also have the power to make the production chains of AI technologies less opaque and the capacity to change things. Hence, with much influence to continue injustice or remedy it, the power endows them with moral responsibility (McKeown, 2024b).

And despite being constrained by the rules of the market, corporations that deliberately perpetuate or fail to redress structural injustices should be held accountable. Pragmatically, if this does not happen, it may

lead to responsibility deficits, allowing for the replication of such injustices as seen in the case of data workers and the BPO Sama, first used by Meta then Open AI (Perrigo, 2023) – especially, as such responsibility can only be applied to corporate agents who can bear it and act on it. Corporations, due to their size and the resources at their disposal, can implement changes in their industries, where Microsoft actively maintains – or claims to – its supply chain integrity (Microsoft, 2025). Since these companies can support private-public partnerships and engage in philanthropy, holding them morally responsible for these structural harms can lead to small-scale changes. Hence, it is the MNCs in question who ought to assume moral responsibility in this case.

4.6 Towards Structural Responsibility in AI Ethics

The structural injustices pertaining to natural resource extraction, labour exploitation and data abstraction, despite being ethically pertinent, have rarely entered the dominant discourse in AI ethics. While questions surrounding embedding values and negotiating responsibility gaps are crucial, they are inadequate due to the current trajectory of AI development and require a structural analysis. The general themes of value alignment and theory-practice gaps, while proving to be technically and ethically challenging, merely focusing on these complex challenges in a controlled environment can flatten the socio-political embeddedness of the technology. While such ahistorical and depoliticised simplicity can help articulate ethical and technical baselines, moral considerations should also accommodate the texture and friction of non-ideal environments in which these technologies are developed and deployed. Relatedly, principlist notions, even when tackling themes of harmful discrimination due to biases, also fall short due to their lack of consideration of power relations. Recall the Amazon hiring algorithm that downgraded the anonymised resumes based on gendered proxies as discussed in Chapter 1. While completely anonymised resumes or audited algorithms could fix such issues to deploy these algorithms in a particular context, such approaches do not acknowledge the fact that aspects like race or gender are not “isolatable attributes [....but] broader systems of power” (Drage and Mackereth, 2022) with historical legacies. Hence, the additional layer of ethico-political analyses that the third wave of AI ethicists offers is crucial. A partial yet important reason as to why specific ways of theorising/practising AI ethics are dominant, resulting in the various types of ethics washing, is also because they are published and ratified by technology companies, who, despite collaborating with other stakeholders, have the most power. Additionally, AI ethics must also attempt to accommodate the aspects of production of AI systems that require human labour and natural resource extraction, as they are entangled in the very same structural processes, integrating insights from other AI researchers who work on issues of digital labour, sustainable AI or political economy of AI. Specifically with regards to this research project, identifying structural

injustices, using McKeown's (2024b) framework, can add to the efforts of the Structural Turn by attributing moral responsibility to corporate agents for supply chains required to produce the AI technologies.

This is crucial as, with structural processes being complex, these corporations are not held accountable for such extractivist practices in the case of AI, despite the moral scrutiny by a few agents. This is because local populations who are situated near sites of excavation or construction sites and data-workers from single-income households (largely) go along with these practices, as it offers employment, constraining their choices. And while these MNCs have sustainability pledges for sourcing elements and labour, these considerations are treated separately and not integral to the development of AI. Such vocal expectations of accountability may result in efforts from these MNCs to amend their workings to restore their reputation, and can also lead to changes in laws. Meta's Cambridge-Analytica scandal, which led to a #DeleteFacebook movement (Bright et. al, 2021) led to CEO Mark Zuckerberg apologising, with Meta also changing its terms and conditions, alongside other promises (Ivanova, 2018) which also resulted in a $5 billion fine from the US Federal Trade Commission (FTC, 2019) and a push towards formulating GDPR laws (Simberkoff, 2018). While there have been discussions around cobalt mining in the DRC, with many Big Tech companies under moral scrutiny about the harms emerging from the extraction (Amnesty International, 2021), these companies have faced no legal liability due to the fuzzy opaqueness of supply chains. (Library of Congress, 2024). This necessitates a structural analysis that identifies moral wrongdoings despite complex networks, to emphasise the power they possess within structures to maintain the status quo, avoid responsibility deficit and aid in legal liability, especially when it comes to thinking about large-scale AI technologies beyond the algorithmic pipeline. Although ethics should not be used as regulatory mechanisms (as done in cases of AI ethics charters for self-regulation), such accounts can help formulate arguments for such mechanisms.

Failing to hold such powerful entities would result in more harm, both structural and interactional. All the cases mentioned previously surrounding mineral extraction, resulting in environmental degradation, habitat loss, community disruption, public health crises, or labour exploitation, resulting in financial insecurity, toxic work environments, and psychological harm, would continue, strengthening the material and social rules that stabilise structures that further normalise and incentivise such actions. Thus, moral responsibility would entail not only being accountable for past harms but also regulating one's behaviour for future partnerships. Hence, they ought to only partner with mining companies that pay their fiscal and legal obligations. Due to a lack of enforceable global rules, private actors self-regulate, leading to a range of soft and non-binding initiatives dressed up in CSR lingo and engaging in avoidance tactics like tax

evasion or concessions or abandoning mines without recourse. (Elbra, 2024). Hence, BigTech that decides to build infrastructures for AI applications must ensure that they are a part of sourcing processes from mining companies, to minimise and compensate for extractions, say by participating in a circular economy that recycles metals. The BigTech company must, hence, attempt to redefine its scope of responsibility in the supply chain (Bingham Chee, 2022) by demanding transparency from the mining companies and also, investigating the supply chain themselves to adopt risk-based due diligence frameworks that can identify, prevent and mitigate potential negative impacts (IEA, 2023) – as AI data centres are rapidly being constructed.

Additionally, it is the MNCs who should primarily assume responsibility for the deliberate structural injustices that take place across the labour chain to prevent harm. Due to their systemic disposition and episodic efforts, they must also work with governments to address governance gaps and ensure policy/legal measures that are not only enshrined but also acted upon. Upstream regulations require them to partner with platforms that provide workers' rights without pretences about impact sourcing, as claimed by Sama, who was sued alongside META by data workers. This would mean working with not-for-profit organisations like Karya (discussed in Chapter 3), which ensure workers' rights and decent wages, and was opted for by Microsoft. Hence, like the case of natural extraction and labour exploitation, MNCs should invest in supply chain checks and regulations.

However, fulfilling these conditions of moral responsibility and harbouring a significant amount of power does not amount to structural change. In fact, following the ontology laid out by McKeown, as endorsed by this thesis, BigTech MNCs with power and vested interests in the maintenance of these structures would, in all likelihood, shirk accountability or file it under company policies of ethical sourcing and logistics, rather than an issue of AI ethics. This is what makes the role of the other agents and ordinary individual actors crucial, as they can exercise political responsibility. While this is not within the scope of this thesis, political responsibility has been interpreted differently by different theorists, where for McKeown it is the cultivation of political solidarity (McKeown, 2024b). A common theme in these various interpretations follows Young's heuristic: non-liable, yet forward-looking action that supports the voices of the most vulnerable in the complex structural processes. As demonstrated by McKeown (2024b), in her book, the Bangladesh Accord on Fire and Building Safety was an agreement rolled out after the tragic aftermath of fires in the garment sweatshops in Bangladesh. It was signed by 190 corporations from 20 countries, demanding inspections to remedy faults and provide fire safety training to staff. This was possible through the moral accountability assumed by the garment-centric MNCs to bring about such laws due to their systemic position that conferred upon them dispositional power. However, as

pointed out by Marin (2024), this was also due to the pressure put on the MNCs by the other agents involved – unions, NGOs, lawmakers and the victims of this structural injustice. This demonstrated how political responsibility, whether that be solidarity among different agents to create a counter alignment or the performance of one's role responsibility (Zheng, 2018), resulted in the recognition of a moral responsibility to be carried out, which can further lead agents to assume similar responsibility. Hence, structural responsibilities to bring about structural change require both moral responsibility and political responsibility to mutually sustain each other to improve lives, mitigate harms and trigger long-term changes. And in the context of AI production, this would also mean unsubscribing to extractivism as a system of operation, thus dismantling the structural ideologies at play about profit, technoscientific innovation and perpetual growth.

Hence, a structural analysis that acknowledges the differential extractivist impacts of supply chains that drive the production of AI technologies as ethically pertinent can also advocate for AI minimisation in terms of scale and necessity. With rampant construction and billion-dollar projects competing with and combining attempts towards green transitions, an extractivist urgency looms with critical decisions about mining projects to be made. As MNCs have enough power to switch continents for profitable labour, leaving a massive trail of traumatised data annotators, content moderators and click-workers, data is being collected and dispossessed from as many avenues as possible. Hence, the issue that emerges pertains to the ethics of developing large-scale models that are computationally resource-intensive, requiring large amounts of data and arduous data work that is exploitative and specifying in which contexts are necessary. Since AI models can be small and task-specific with their training data contained and knowable (Hao, 2025), they do not require large-scale mineral extraction or harmful labour practices, making these structural injustices avoidable and/or deliberately maintained. This pre-empts measures that are cautious about which tasks require what scale of computation, and additionally, requires us to echo the question raised by Kalluri (2019):

"It is not uncommon now for AI experts to ask whether an AI is 'fair' and 'for good'. But 'fair' and 'good' are infinitely spacious words that any AI system can be squeezed into. The question to pose is a deeper one: how is AI shifting power?"

Conclusion

The research questions raised in the introduction can now be addressed. Although supply chains that lead to the production of AI technologies are globally dispersed and structurally complex with no linear causal ties, moral responsibility for the extractivist harms and structural injustices can be attributed to agents with power.

As the dominant discourse in AI ethics focuses on the outcomes of deploying AI systems and principlistic frameworks to analyse their ethical impacts, it can prove to be inadequate considering the current developmental trajectory of AI technologies. Such dissatisfications have led to shifts in theorising, where the third wave of AI ethicists increasingly focus on larger structures, dabbling with questions of power or the environmental impacts of these technologies. This also expands the geographical ambit of ethically pertinent issues beyond a single algorithm in a single company to consider the planetary consequences of developing these technologies. Hence, the 'structural turn' highlights the socio-technically embedded, infrastructural materiality of AI, assembled by agents and resources from all around the globe and funnelled through supply chains.

Such relatively recent shifts in theorising encourage the interrogation of background conditions and the logic that sustains them. AI technologies as global assemblages require resources from all around the world for their development, whether that be natural resources, human labour or data. However, the mode of acquisition that is common in sourcing these three aspects results in injustices. Hence, once structures are acknowledged, extractivism and the harms it engenders become most ethically pertinent, especially since the MW disproportionately bears them. This also demonstrates why algorithmic harms should not be divorced from infrastructural/material, as they are rooted in the same structures with the same logic conditioning how the development and deployment of AI takes place.

These extractivist injustices are reproduced by agents working according to social norms, where the aggregate consequences result in harms, making this a case of structural injustice. McKeown's (2024b) framework helps articulate how different kinds of structural injustices manifest across the AI supply chain, especially concerning natural resource extraction and labour exploitation. On one hand, natural resource extraction is a case of avoidable structural injustice, as despite readily available information about violent practices of mining companies, articulate concerns about life-cycle impacts of technologies and growing green transition concerns, the rapid construction and indiscriminate deployment of AI technologies continue. And this is avoidable if some discernment is exercised in considering what scale is

required for which tasks in AI. On the other hand, labour exploitation is a case of deliberate structural injustice where companies can rely on a culture of subcontracting to low-income nations that have exploitative work environments as a cheaper alternative, triggering a race to the bottom among developing countries. And this is deliberately maintained as there can be alternatives to ensure fair pay and better working environments.

While it may seem difficult to hold any agent morally accountable for such distributed harms due to structural causation, due to the many (million) hands contributing to this problem, moral responsibility can be attributed to agents with power. As avoidable structural injustices and deliberate structural injustices demonstrate a greater range of interplay between structure and agency, where powerful agents have more influence in the structure, they can be held responsible if they meet the criteria for reasons' responsiveness. And in the case of AI development, the power is concentrated in the hands of the MNCs. However, this is not adequate for structural change itself – for that, we need to consider political responsibility.

I contribute to the structural theorising of AI ethics, positioning myself within the third wave, responding to the call by Ricuarte (2022, p. 727) that urgently stresses the articulation of ethics not only on an individual or institutional level but on a structural level. This thesis, specifically, maps the relations between extraction and labour through the framework of Structural Injustice. While previous analyses employing the Structural Injustice framework focused on the accumulation of data bias in deployed AI systems, this research project focuses on the development or production of AI technologies, focusing on natural resource extraction and labour exploitation funnelled through global supply chains. Thereby, I add to the budding discourse on structural injustice and AI by drawing attention to the infrastructural/exploitative harms that arise from the production of AI technologies. Additionally, I advance an argument that emphasises how the production of these technologies cannot be divorced from their deployment, explicitly connecting the two separate analyses through the notion of the structure.

On a theoretical level, this research project bridges the area of political philosophy, specifically structural injustice, with AI Ethics while also highlighting the ethical importance of supply chains in not only the deployment or technical development of the technologies but also their material development in terms of natural resources and human labour. Indirectly, this thesis also contributes to and hints towards future work on supply chain ethics and structural injustice, and how implements frameworks that accommodate power revealing the interplay between (benign/malicious) agency and structure, can helps understand how businesses can and do maintain background conditions that are interpreted as morally uncontroversial

through supply chains – further opening avenues to explore and develop ontologies that show how interactional and structural harms reinforce each other.

Connecting the three areas with the notion of power-endowed responsibility also lends itself to many practical implications. Holding these organisations responsible would involve compensation and reparation for the retrospective structural processes that led to harm. This would include providing reparations to affected workers in the labour chain, or funding the remediation of ecosystems damaged by natural resource excavation. MNCs ought to invest in transparent supply chains, ensuring living wages, conscious subcontracting, and regular audits, moving beyond profit to also disincentivise other peripheral actors from causing harm. However, this would not be enforceable without regulatory mechanisms. Thus this requires interdisciplinary collaboration between AI ethics, legislation and policymaking, necessitating structural responsbility. Such mechanisms could de-incentivise constructing hyperscalers that despite pushing the bounds of technical innovations, can normalise harm through hype cycles, making AI unethical.

The various limits of the thesis also provide directions for future research. While a significant advantage of using this theory is its emphasis on not only moral responsibility but also political responsibility in bringing out structural change, this thesis, due to its defined scope of inquiry, focuses primarily on the question of moral responsibility. Hence, future work could develop this line of argumentation further, as hinted in the concluding section of Chapter 4, to think about structural change using both dimensions of responsibility. Since large-scale AI technologies involve data-driven techniques like ML, the responsibility question looms large regarding data extractivism. However, since data abstraction is rooted in extractivism that not only entails mass collection of data but also the conversion of these datasets into objects of value as per epistemic hierarchies and influencing how algorithmic discrimination takes place, further conjoining processes of development and deployment, it requires an intricate analysis which requires another thesis to tackle. However, these processes of collection and dispossession would blur the lines between avoidable, deliberate, and pure structural injustice. Lastly, for pragmatic reasons, this research project theorises around a unimodal understanding of power-endowed responsibility, despite relying on a social ontology that acknowledges how all agents in the network contribute to the structure in question, like national governments or states. Since these corporate agents also have considerable power in the structure, future research work could develop a multimodal approach that questions if, when, how and to what extent responsibility ought to be shared among different powerful agents.

Bibliography

Abdalla, M., & Abdalla, M. (2021). The Grey Hoodie Project: Big Tobacco, big tech, and the threat on academic integrity. *Proceedings of the 2021 AAAI/ACM Conference on AI, Ethics, and Society*, 287–297. https://doi.org/10.1145/3461702.3462563

Acosta, A. (2013). Extractivism and neoextractivism: two sides of the same curse. *Beyond development: alternative visions from Latin America*, *1*, 61-86.

Alam, S. (2008). Majority world: Challenging the West's rhetoric of democracy. *Amerasia Journal*, *34*(1), 88–98. https://doi.org/10.17953/amer.34.1.l3176027k4q614v5

Albert, K., & Delano, M. (2023). Algorithmic exclusion. *Handbook of Critical Studies of Artificial Intelligence*, 538–548. https://doi.org/10.4337/9781803928562.00056

AlgorithmWatch (2020*) In the realm of Paper Tigers – exploring the failings of AI ethics guidelines*. https://algorithmwatch.org/en/ai-ethics-guidelines-inventory-upgrade-2020/

Alvaredo, F., Chancel, L., Piketty, T., Saez, E., & Zucman, G. (Eds.). (2018). *World inequality report 2018*. Belknap Press.

Amnesty International. (2021, August 16). *Exposed: Child labour behind smart phone and electric car batteries*. https://www.amnesty.org/en/latest/news/2016/01/child-labour-behind-smart-phone-and-electric-car-batteries/

Amrute, S., Singh, R., & Guzmán, R. L. (2022). A primer on AI in/from the majority world: An empirical site and a standpoint. *Available at SSRN 4199467*.

Aragon, C., & Jaggar, A. M. (2018). Agency, complicity, and the responsibility to resist structural injustice. *Journal of Social Philosophy*, 49(3), 439–460. https://doi.org/10.1111/josp.12251

Archer, M. (1995). *Realist Social Theory: The Morphogenetic Approach*. Cambridge University Press, 1995.

Aráoz, H. M. (2013). Crisis ecológica, conflictos socioambientales y orden neocolonial: Las paradojas de NuestrAmérica en las fronteras del extractivismo. *REBELA-Revista Brasileira de Estudos Latino-Americanos*, *3*(2).

Barratt, L. (2025, April 9). *Revealed: Big tech's new datacentres will take water from the world's driest areas*. The Guardian. https://www.theguardian.com/environment/2025/apr/09/big-tech-datacentres-water

Barry, C., & Macdonald, K. (2016). How should we conceive of individual consumer responsibility to address labour injustices? In F. Milman-Sivan, Y. Dahan, & H. Lerner (Eds.), *Global justice and international labour rights*. (pp. 92–118). Cambridge University Press.

Benjamin, R. (2019). Assessing risk, automating racism. *Science*, *366*(6464), 421–422. https://doi.org/10.1126/science.aaz3873

Bietti, E. (2021). From ethics washing to ethics bashing: A view on tech ethics from within Moral Philosophy. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3914119

Boden, M. (2011). Neoliberalism and counter-hegemony in the Global South: Reimagining the State. *Social Movements in the Global South*, 83–103. https://doi.org/10.1057/9780230302044_4

Bolte, L., & van Wynsberghe, A. (2024a). Sustainable AI and the third wave of AI Ethics: A structural turn. *AI and Ethics*, *5*(2), 1733–1742. https://doi.org/10.1007/s43681-024-00522-6

Bolte, L., & van Wynsberghe, A. (2024b). Sustainable AI and the third wave of AI Ethics: A structural turn. *AI and Ethics*, *5*(2), 1733–1742. https://doi.org/10.1007/s43681-024-00522-6

Bontridder, N., & Poullet, Y. (2021). The role of Artificial Intelligence in disinformation. *Data & Policy*, *3*. https://doi.org/10.1017/dap.2021.20

Bordage, F., de Montenay, L., Benqassem, S., Delmas-Orgelet, J., Domon, F., Prunel, D., ... & Lees Perasso, E. (2021). Digital technologies in Europe: an environmental life cycle approach. *The Greens/EFA, Brussels*

Borenstein, J., Grodzinsky, F. S., Howard, A., Miller, K. W., & Wolf, M. J. (2021). Ai ethics: A long history and a recent burst of attention. *Computer*, *54*(1), 96–102. https://doi.org/10.1109/mc.2020.3034950

Bostrom, N. (2020). Ethical issues in advanced artificial intelligence. *Machine Ethics and Robot Ethics*, 69–75. https://doi.org/10.4324/9781003074991-7

Bowker, G. C., & Star, S. L. (2000). *Sorting things out: Classification and its consequences*. MIT press.

Bright, L. F., Sussman, K. L., & Wilcox, G. B. (2021). Reaching the tipping point: A critical analysis of the #deletefacebook movement. *Journal of Data Protection & Privacy*, *4*(4), 424. https://doi.org/10.69554/gtih8748

Brodie, P. (2024). Smarter, Greener Extractivism: Digital Infrastructures and the Harnessing of New Resources. *Information, Communication & Society*, *28*(6), 1061–1080. https://doi.org/10.1080/1369118x.2024.2341013

Brown, T. J., Idoine, N. E., Wrighton, C. E., Raycraft, E. R., Hobbs, S. F., Shaw, R. A., Everett, P., Deady, E. A., & Kresse, C. (2021, March 22). *World mineral production 2015-2019*. NERC Open Research Archive. https://nora.nerc.ac.uk/id/eprint/529923/

Browne, J. (2023). AI and Structural Injustice: A Feminist Perspective. *Feminist AI*, 328–346. https://doi.org/10.1093/oso/9780192889898.003.0019

Browne, J., & McKeown, M. (2024). *What is structural injustice?* (p. 305). Oxford University Press.

Bufacchi, V. (2024). Where is the injustice in structural injustice? *Filosofia e Questioni Pubbliche*, 221. https://doi.org/10.17473/2240-7987-2024-2-2

Buolamwini, J., & Gebru, T. (2018, January). Gender shades: Intersectional accuracy disparities in commercial gender classification. In *Conference on fairness, accountability and transparency* (pp. 77-91). PMLR.

Calderon, J. L. (2021, October 31). *How critical minerals became so critical*. Milken Institute Review. https://www.milkenreview.org/articles/how-critical-minerals-became-so-critical

Casilli, A. A. (2017). Global digital culture| Digital labor studies go global: Toward a digital decolonial turn. *International Journal of Communication*, *11*, 21.

Chagnon, C. W., Durante, F., Gills, B. K., Hagolani-Albov, S. E., Hokkanen, S., Kangasluoma, S. M., ... & Vuola, M. P. (2022). From extractivism to global extractivism: the evolution of an organizing concept. *The Journal of Peasant Studies*, *49*(4), 760-792.

Chance, H., & Hammersley, M. M. (2025). *AI is mining more than data*. Vermont Journal of Environmental Law. https://vjel.vermontlaw.edu/top-ten/2025-top-ten/2024/12/ai-is-mining-more-than-data/

Chandrashekar, A. (2017, August 31). *Can ai-powered chatbots create a safe space while helping deconstruct rape myth acceptance?*. Medium. https://medium.com/compnanion-chat/can-ai-powered-chatbots-create-a-safe-space-while-helping-deconstruct-rape-myth-acceptance-d548cc55f66a

Chee, M. E. B. (2022). Ethical supply chains of metals and minerals. In *Routledge Handbook of the Extractive Industries and Sustainable Development* (pp. 480-502). Routledge.

Chen, B. J. (2024, May 29). *Why AI policy needs a sociotechnical perspective*. Tech Policy Press. https://www.techpolicy.press/why-ai-policy-needs-a-sociotechnical-perspective/

Corrêa, N., Galvão, C., Santos, J., Del Pino, C., Pinto, E., Barbosa, C., Massmann, D., Mambrini, R., Galvão, L., & Terem, E. (2023). *Worldwide AI Ethics: A Review of 200 Guidelines and Recommendations for AI Governance*. https://doi.org/10.2139/ssrn.4381684

Couldry, N., & 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). *Atlas of AI*. Yale University Press.

Crawford, K., & Joler, V. (n.d.). *Anatomy of an AI system*. Anatomy of an AI System. https://anatomyof.ai/

Creamer, E. (2023, July 5). *Authors file a lawsuit against openai for unlawfully "ingesting" their books*. The Guardian. https://www.theguardian.com/books/2023/jul/05/authors-file-a-lawsuit-against-openai-for-unlawfully-ingesting-their-books

Danaher, J. (2017). The rise of the robots and the crisis of moral patiency. *AI & SOCIETY*, *34*(1), 129–136. https://doi.org/10.1007/s00146-017-0773-9

Dastin, J. (2018). *Insight - Amazon scraps secret AI recruiting tool that showed bias against women | reuters*. Reuters. https://www.reuters.com/article/world/insight-amazon-scraps-secret-ai-recruiting-tool-that-showed-bias-against-women-idUSKCN1MK0AG/

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

De Bernardi, R. (2024). From climate change to Sweatshop Labor: Do "structural" injustices exist, after all? *Filosofia e Questioni Pubbliche*, 241. https://doi.org/10.17473/2240-7987-2024-2-4

Delgado Wise, R., & Martin, D. (2015). The Political Economy of Global Labor Arbitration. *Problemas del desarrollo*, *46*(183), 13-32.

Diberardino, N., Baleshta, C., & Stark, L. (2024). Algorithmic harms and algorithmic wrongs. *The 2024 ACM Conference on Fairness, Accountability, and Transparency*, 1725–1732. https://doi.org/10.1145/3630106.3659001

Digital Future Labs . (2025, February). *Issue 05: Glutton AI*. Code Green Asia. https://cdn.sanity.io/files/urj6jz1n/production/f1ebfc07585b9bd614cb13370c7a7604e41806a3.pdf

Digital Future Labs. (2025, March 6). *Issue 06: Critical panic*. Code Green Asia. https://codegreenasia.substack.com/p/issue-06-critical-panic

Drage, E., & Mackereth, K. (2022). Does AI debias Recruitment? Race, Gender, and AI's "Eradication of Difference." *Philosophy & Technology*, *35*(4). https://doi.org/10.1007/s13347-022-00543-1

Dunlap, A., & Jakobsen, J. (2019). Introduction: Consuming everything—capitalism and the imperative of total extractivism. In *The Violent Technologies of Extraction: Political ecology, critical agrarian studies and the capitalist worldeater* (pp. 1-12). Cham: Springer International Publishing.

Elbra, A. (2024). *Arc of Avoidance: An Analytical Framework for Analysing Mining Companies' Actions in the Global South*. https://doi.org/10.2139/ssrn.4685236

Eubanks, V. (2019). *Automating Inequality: How high-tech tools profile, police, and punish the poor*. Picador, St. Martin's Press.

Falk, S., van Wynsberghe, A., & Biber-Freudenberger, L. (2024). The attribution problem of a seemingly intangible industry. *Environmental Challenges*, *16*, 101003. https://doi.org/10.1016/j.envc.2024.101003

Federal Trade Commission, USA. (2024, August 20). *FTC imposes $5 billion penalty and sweeping new privacy restrictions on Facebook*. Federal Trade Commission. https://www.ftc.gov/news-events/news/press-releases/2019/07/ftc-imposes-5-billion-penalty-sweeping-new-privacy-restrictions-facebook

Fleck, A. (2024, September 20). *Infographic: Which countries have the most data centers?*. Statista Daily Data. https://www.statista.com/chart/24149/data-centers-per-country/

Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., ... & Vayena, E. (2018). AI4People—An ethical framework for a good AI society: Opportunities, risks, principles, and recommendations. *Minds and machines*, *28*(4), 689-707.

Floridi, L. (2019). Translating principles into practices of Digital Ethics: Five risks of being unethical. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3835010

Forrester, K. (2019). *In the shadow of justice: Postwar liberalism and the remaking of political philosophy*. Princeton University Press.

Frank, L. E., & Klincewicz, M. (2024). Uses and abuses of AI ethics. *Handbook on the Ethics of Artificial Intelligence*, 205–217. https://doi.org/10.4337/9781803926728.00019

Gambacorta, L., & Shreeti, V. (2025). The AI supply chain. *BIS Papers*.

Ganesh, M. I. (2020). The Ironies of Autonomy. *Humanities and Social Sciences Communications*, *7*(1). https://doi.org/10.1057/s41599-020-00646-0

Garcia, A. C., Garcia, M. G., & Rigobon, R. (2023). Algorithmic discrimination in the credit domain: What do we know about it? *AI & SOCIETY*, *39*(4), 2059–2098. https://doi.org/10.1007/s00146-023-01676-3

García-Herrero Alicia García Herrero is a Senior Fellow at Bruegel.She is the Chief Economist for Asia Pacific at French investment bank Natixis, A. (2024, December 11). *David and Goliath: The EU's Global Gateway versus China's belt and road initiative*. Bruegel. https://www.bruegel.org/newsletter/david-and-goliath-eus-global-gateway-versus-chinas-belt-and-road-initiative

Gitelman, L. (Ed.). (2013). *Raw data is an oxymoron*. MIT press.

Godoy, E. S. (2017). What's the harm in climate change. Ethics, Policy & Environment, 20(1), 103–117. https://doi.org/10. 1080/21550085.2017.1291828

Goetze, T. S. (2024). AI art is theft: Labour, Extraction, and exploitation: Or, on the dangers of Stochastic Pollocks. *The 2024 ACM Conference on Fairness, Accountability, and Transparency*, 186–196. https://doi.org/10.1145/3630106.3658898

Google (2025) *AI for Sustainability*. Google AI - Sustainability AI. (2025). https://ai.google/sustainability/

Gordon, L. (2024). Decolonizing Structural Justice and Political Responsibility. *What Is Structural Injustice*, 187-200.

Gould, C. C. (2009). Varieties of global responsibility: Social connection, human rights, and transnational solidarity. In A. Ferguson & M. Nagel (Eds.), *Dancing with Iris: The philosophy of Iris Marion Young* (pp. 199–213). Oxford University Press.

Graham, M., Hjorth, I., & Lehdonvirta, V. (2017). Digital labour and development: impacts of global digital labour platforms and the gig economy on worker livelihoods. *Transfer: European review of labour and research*, *23*(2), 135-162.

Gray, M. L., & Suri, S. (2019). *Ghost work: How to stop Silicon Valley from building a new global underclass*. Harper Business

Gray, C. (2023). More than extraction: Rethinking Data's colonial political economy. *International Political Sociology*, *17*(2). https://doi.org/10.1093/ips/olad007

Gudynas, E. (2018). Extractivisms: Tendencies and consequences. In *Reframing Latin American Development* (pp. 61-76). Routledge.

Hafez , V., Wazir, R., & Karimi, F. (n.d.). *AI as complex sociotechnical systems: Problems, ...* Women In AI: Austria. https://www.womeninai.at/wp-content/uploads/2023/11/WhitePaper_AI_as_SociotechnicalSystems_final.pdf

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

Hagendorff, T. (2022). AI Ethics and its pitfalls: Not living up to its own standards? *AI and Ethics*, *3*(1), 329–336. https://doi.org/10.1007/s43681-022-00173-5

Hahn, H. (2009). The global consequence of participatory responsibility. *Journal of Global Ethics*, 5(1), 43–56. https://doi.org/10.1080/17449620902765336

Hao, K. (2025). *Empire of AI: Dreams and nightmares in Sam Altman's OpenAI*. Penguin Press.

Harkanas, M. (2024). *Google pours billions into new U.S. data centers: Here's where*. CRN. https://www.crn.com/news/cloud/2024/google-pours-billions-into-new-u-s-data-centers-here-s-where

Haugaard, M. (2010). Power: A 'family resemblance'concept. *European journal of cultural studies*, *13*(4), 419-438.

He, M. (2025a). Investment in R&D of AI: Evidence from the Global South. *Future of Business and Finance*, 61–85. https://doi.org/10.1007/978-3-031-73639-1_4

He, M. (2025b). Investment in R&D of AI: Evidence from the Global South. *Future of Business and Finance*, 61–85. https://doi.org/10.1007/978-3-031-73639-1_4

Heikkilä, M. (2024, October 1). *Why bigger is not always better in AI*. MIT Technology Review. https://www.technologyreview.com/2024/10/01/1104744/why-bigger-is-not-always-better-in-ai/#:~:text=In%20AI%20research%2C%20everyone%20seems,concentration%20of%20power%2C%20says%20Luccioni.

Heilinger, J.-C. (2022). The ethics of AI ethics. A constructive critique. *Philosophy & Technology*, *35*(3). https://doi.org/10.1007/s13347-022-00557-9

Himmelreich, J., & Lim, D. (2022). AI and structural injustice. *The Oxford Handbook of AI Governance*, 210–231. https://doi.org/10.1093/oxfordhb/9780197579329.013.13

Hogan, M. (2015). Data flows and water woes: The Utah Data Center. *Big Data & Society*, *2*(2). https://doi.org/10.1177/2053951715592429

Huang, L. T.-L., Papyshev, G., & Wong, J. K. (2024). Democratizing value alignment: From authoritarian to Democratic AI ethics. *AI and Ethics*, *5*(1), 11–18. https://doi.org/10.1007/s43681-024-00624-1

Höne, K., & By. (2022, March 30). *What is tech diplomacy?: Heinrich-Böll-Stiftung: Tel Aviv - israel*. Heinrich-Böll-Stiftung | Tel Aviv - Israel. https://il.boell.org/en/2023/03/30/what-tech-diplomacy#:~:text=Thus%2C%20according%20to%20an%20article,deepening%20existing%20socio%2Deconomic%20gaps

Ivanova, I. (2018, April 10). *8 promises from Facebook after Cambridge Analytica*. CBS News. https://www.cbsnews.com/news/facebooks-promises-for-protecting-your-information-after-data-breach-scandal/

Iyer, K. (2024.). *"Time to hold big-tech accountable for using cheap labour in Artificial Intelligence Data-labelling work."* Article 14. https://article-14.com/post/-time-to-hold-big-tech-accountable-for-using-cheap-labour-in-artificial-intelligence-data-labelling-work--6607926d7f9fe

Jaggar, A. M., & Tobin, T. W. (2024). Moral justification and structural epistemic injustice. *What is Structural Injustice*.

Johnson, D. G., & Verdicchio, M. (2017). Reframing AI discourse. *Minds and Machines*, *27*(4), 575–590. https://doi.org/10.1007/s11023-017-9417-6

Johnson, D. G., & Verdicchio, M. (2024). The sociotechnical entanglement of AI and values. *AI & SOCIETY*, *40*(1), 67–76. https://doi.org/10.1007/s00146-023-01852-5

Johnson-Woods, C. (2023, March 28). *Upstream, downstream, extended, and ethical: The Growing Vocabulary of Supply Chain Management*. Resonance Global. https://www.resonanceglobal.com/blog/the-growing-vocabulary-of-supply-chain-management

Jonas, H. (1985). *The imperative of responsibility: In search of an ethics for the technological age*. University of Chicago Press.

Jong, H. N. (2024, February 26). *Indonesian nickel project harms environment and human rights, report says*. Mongabay Environmental News. https://news.mongabay.com/2024/02/indonesian-nickel-project-harms-environment-and-human-rights-report-says/#:~:text=JAKARTA%20%E2%80%94%20A%20massive%20nickel,process%2C%20a%20new%20report%20says

Kamm, F. M. (2020). *Harms, wrongs, and meaning in a pandemic - the philosophers' magazine*. The Philosophers' Magazine -. https://philosophersmag.com/harms-wrongs-and-meaning-in-a-pandemic/

Kalluri, P. (2020). Don't ask if artificial intelligence is good or fair, ask how it shifts power. *Nature*, *583*(7815), 169–169. https://doi.org/10.1038/d41586-020-02003-2

Kudina, O., & van de Poel, I. (2024). A Sociotechnical System perspective on ai. *Minds and Machines*, *34*(3). https://doi.org/10.1007/s11023-024-09680-2

Lee, K., Lee, K., Gangidi, A., & Oldham, M. (2024, May 14). *Building Meta's genai infrastructure*. Engineering at Meta. https://engineering.fb.com/2024/03/12/data-center-engineering/building-metas-genai-infrastructure/

Lehdonvirta, V., Wu, B., & Hawkins, Z. (2024). *Compute North vs. Compute South: The Uneven Possibilities of Compute-Based AI Governance around the Globe*. https://doi.org/10.31235/osf.io/8yp7z

Lemma, A. (n.d.). *Critical Minerals, critical moment: Africa's role in the AI Revolution | ODI: Think change*. ODI Global. https://odi.org/en/insights/critical-minerals-critical-moment-africas-role-in-the-ai-revolution/

Lin, T.-A., & Chen, P.-H. C. (2022). Artificial Intelligence in a structurally unjust society. *Feminist Philosophy Quarterly*, *8*(3/4). https://doi.org/10.5206/fpq/2022.3/4.14191

List, C., & Pettit, P. (2011). *Group agency: The possibility, design, and status of corporate agents*. Oxford University Press.

Magnier, L. B. M., & Mugge, R. (2022). Are short product lifetimes ineluctable? An exploration of consumers' perceptions of lifetime extension strategies. In D. Lockton, S. L. Lenzi, P. P. M. Hekkert, A. Oak,

Malik, S. (2022, May 13). *Series on digital colonialism: Western tech companies saving billions in taxes in the Global South*. netzpolitik.org. https://netzpolitik.org/2022/series-on-digital-colonialism-western-tech-companies-saving-billions-in-taxes-in-the-global-south/

Marin, M. (2024). Reflections on power and structural change. commentary on Maeve McKeown's with power comes responsibility. The Politics of Structural Injustice. *Filosofia e Questioni Pubbliche*. https://doi.org/10.17473/2240-7987-2024-2-3

Marx, K., & Engels, F. (1996). Collected works. *(Vol 35)*. Lawrence and Wishart.

Matanzima, J., & Loginova, J. (2024). Sociocultural risks of resource extraction for the low-carbon energy transition: Evidence from the Global South. *The Extractive Industries and Society*, *18*, 101478. https://doi.org/10.1016/j.exis.2024.101478

McConnon, A. (2025, August 8). *Are bigger language models always better?*. IBM. https://www.ibm.com/think/insights/are-bigger-language-models-better#:~:text=In%20the%20race%20to%20dominate%20AI%2C%20bigger,seizing%20the%20news%20headlines%20reinforce%20this%20trend.

McDonough, M. (2025, June 30). *Did ai solve the protein-folding problem?*. Harvard Medicine Magazine. https://magazine.hms.harvard.edu/articles/did-ai-solve-protein-folding-problem

McIntyre, N., Bradbury, R., & Perrigo, B. (2022, October 20). *Behind TikTok's boom: A legion of traumatised, $10-a-day content...* TBIJ. https://www.thebureauinvestigates.com/stories/2022-10-20/behind-tiktoks-boom-a-legion-of-traumatised-10-a-day-content-moderators

McKeown, M. (2018). Iris Marion Young's "social connection model" of responsibility: Clarifying the meaning of connection. *Journal of Social Philosophy*, 49(3), 484–502. https://doi.org/10.1111/josp.12253

McKeown, M. (2021). Structural Injustice. *Philosophy Compass*, *16*(7). https://doi.org/10.1111/phc3.12757

McKeown, M. (2024a). Response to critics. *Filosofia e Questioni Pubbliche*, 261. https://doi.org/10.17473/2240-7987-2024-2-6

McKeown, M. (2024b). *With Power comes Responsibility: The Politics of Structural Injustice*. Bloomsbury Academic.

McKey, J. (2025, May 6). *Ai Hustle: Human Capital, sovereignty, and the Global South*. New America. https://www.newamerica.org/planetary-politics/blog/ai-hustle-human-capital-sovereignty-and-the-global-south/

Menear, H. (2024, April 19). *Data Centres could consume energy "equivalent to Japan" thanks to Ai* . Interface. https://interface.media/blog/2024/04/19/data-centres-could-consume-energy-equivalent-to-japan-thanks-to-ai/

Mezzadra, S., & Neilson, B. (2017). On the multiple frontiers of extraction: Excavating contemporary capitalism. *Cultural studies*, *31*(2-3), 185-204.

Miller, D. L. (2007). *National Responsibility and Global Justice* . Oxford University Press.

Minneti, J. J. (2018). Environmental governance and the Global South. *Wm. & Mary Envtl. L. & Pol'y Rev.*, *43*, 83.

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

Mittelstadt, B. (2019). Principles alone cannot guarantee ethical AI. *Nature Machine Intelligence*, *1*(11), 501–507. https://doi.org/10.1038/s42256-019-0114-4

Moradi, P., & Levy, K. (2020). The future of work in the age of ai. *The Oxford Handbook of Ethics of AI*, 270–288. https://doi.org/10.1093/oxfordhb/9780190067397.013.17

Morozov, E. (2014). *To save everything, click here: The folly of technological solutionism*. PublicAffairs.

Muldoon, J., & Wu, B. A. (2023). Artificial intelligence in the colonial matrix of power. *Philosophy & Technology*, *36*(4), 80.

Muldoon, J., Graham, M., & Cant, C. (2024). *Feeding the Machine: The Hidden Human Labour Powering AI*. Canongate.

Munn, L. (2022). The uselessness of AI Ethics. *AI and Ethics*, *3*(3), 869–877. https://doi.org/10.1007/s43681-022-00209-w

Mutemi, M. (2025, April 1). *African workers are taking on Meta and the world should pay attention*. Al Jazeera. https://www.aljazeera.com/opinions/2025/4/1/african-workers-are-taking-on-meta-and-the-world-should-pay-attention

Neuhäuser, C. (2014). Structural injustice and the distribution of forward-looking responsibility. *Midwest Studies in Philosophy*, 38(1), 232–251. https://doi.org/10.1111/misp.12026

Nicoletti, L., Ma, M., & Bass, D. (n.d.-a). *The AI Boom is draining water from the areas that need it most*. Bloomberg.com. https://www.bloomberg.com/graphics/2025-ai-impacts-data-centers-water-data/

Nicoletti, L., Ma, M., & Bass, D. (n.d.-b). *The AI Boom is draining water from the areas that need it most*. Bloomberg.com. https://www.bloomberg.com/graphics/2025-ai-impacts-data-centers-water-data/

Noble, S. U. (2018). Algorithms of oppression: How search engines reinforce racism. In *Algorithms of oppression*. New York University Press.

Noorman, M., & Swierstra, T. (2023). Democratizing AI from a sociotechnical perspective. *Minds and Machines*, *33*(4), 563–586. https://doi.org/10.1007/s11023-023-09651-z

Nussbaum, M. C. (2011). Foreword. In I. M. Young (Ed.), *Responsibility for justice* (pp. ix–xxv). Oxford University Press.

Nuti, A. (2019). Injustice and the reproduction of history: Structural inequalities, gender and redress. Cambridge University Press.

Nyholm, S. (2025). The Future of Human Responsibility. *A Companion to Applied Philosophy of AI*, 399–414. https://doi.org/10.1002/9781394238651.ch28

Ochigame, R. (2023, July 25). *How big tech manipulates academia to avoid regulation*. The Intercept. https://theintercept.com/2019/12/20/mit-ethical-ai-artificial-intelligence/

OpenAI. (2025). *Announcing the stargate project | openai*. https://openai.com/index/announcing-the-stargate-project/

Owen, J. R., Kemp, D., Lechner, A. M., Harris, J., Zhang, R., & Lèbre, É. (2022). Energy transition minerals and their intersection with land-connected peoples. *Nature Sustainability*, *6*(2), 203–211. https://doi.org/10.1038/s41893-022-00994-6

O'Malley, J. (2018, January 12). *Captcha if you can: How you've been training AI for years without realising it*. TechRadar. https://www.techradar.com/news/captcha-if-you-can-how-youve-been-training-ai-for-years-without-realising-it

O'Neil, C. (2017). *Weapons of math destruction: How big data increases inequality and threatens democracy*. Crown.

Pasquinelli, M., & Joler, V. (2020). The nooscope manifested: AI as an instrument of knowledge extractivism. *AI & SOCIETY*, *36*(4), 1263–1280. https://doi.org/10.1007/s00146-020-01097-6

Pereira, D. (2025, June 17). *Critical minerals and the AI supply chain: Geopolitics, extraction, and global power shifts*. OODAloop. https://oodaloop.com/analysis/security-and-resiliency/critical-minerals-and-the-ai-supply-chain-geopolitics-extraction-and-global-power-shifts/

Pereira, N. (2025, April 17). *China launches world's first AI-Powered Underwater Data Centre!* Sify. https://www.sify.com/data-centers/china-launches-worlds-first-ai-powered-underwater-data-centre/

Perrigo, B. (2022, February 17). *Inside Facebook's African sweatshop*. Time. https://time.com/6147458/facebook-africa-content-moderation-employee-treatment/

Perrigo, B. (2023, January 18). *OpenAI used Kenyan workers on less than $2 per hour: Exclusive*. Time. https://time.com/6247678/openai-chatgpt-kenya-workers/

Peters, U. (2022). Explainable AI lacks regulative reasons: Why ai and human decision-making are not equally opaque. *AI and Ethics*, *3*(3), 963–974. https://doi.org/10.1007/s43681-022-00217-w

Peša, I., & Ross, C. (2021). Extractive Industries and the environment: Production, pollution, and protest in global history. *The Extractive Industries and Society*, *8*(4), 100933. https://doi.org/10.1016/j.exis.2021.100933

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

Powell, P., & Smalley, I. (2025, July 22). *What is a hyperscale data center?*. IBM. https://www.ibm.com/think/topics/hyperscale-data-center

Powers, M., & Faden, R. (2019). Structural injustice: Power, advantage, and human rights. Oxford University Press. Rawls, J. (1971). A theory of justice (Revised Ed). The Belknap Press of Harvard University Press.

Purves, D., & Davis, J. (2023). Should algorithms that predict recidivism have access to race? *American Philosophical Quarterly*, *60*(2), 205–220. https://doi.org/10.5406/21521123.60.2.07

Rani, U., & Dhir, R. K. (2024, December 16). *The artificial intelligence illusion: How invisible workers fuel the "automated" economy*. International Labour Organization. https://www.ilo.org/resource/article/artificial-intelligence-illusion-how-invisible-workers-fuel-automated

Regilme, S. S. F. (2024). Artificial intelligence colonialism: Environmental damage, labor exploitation, and human rights crises in the Global South. *SAIS Review of International Affairs*, *44*(2), 75-92.

Rességuier, A., & Rodrigues, R. (2020). AI Ethics should not remain toothless! A call to bring back the teeth of Ethics. *Big Data & Society*, *7*(2). https://doi.org/10.1177/2053951720942541

Ricaurte, 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

Ricci, A. (2019). Unequal exchange in the age of globalization. *Review of Radical Political Economics*, *51*(2), 225-245.

Ritchie, H., & Rosado, P. (2017, October 2). *Fossil fuels*. Our World in Data. https://ourworldindata.org/fossil-fuels

Roach, J. P. S. (2020, September 14). *Microsoft finds underwater datacenters are reliable, practical and use energy sustainably*. Microsoft. https://news.microsoft.com/source/features/sustainability/project-natick-underwater-datacenter/

Rowe, N. (2023, August 2). *"it's destroyed me completely": Kenyan moderators decry toll of training of AI models*. The Guardian. https://www.theguardian.com/technology/2023/aug/02/ai-chatbot-training-human-toll-content-moderator-meta-openai

Russell, S. J., & Norvig, P. (2016). *Artificial Intelligence: A modern approach*. Pearson.

Ryan, M., de Roo, N., Wang, H., Blok, V., & Atik, C. (2024). AI through the looking Glass: An empirical study of structural social and ethical challenges in ai. *AI & SOCIETY*, *40*(5), 3891–3907. https://doi.org/10.1007/s00146-024-02146-0

Sadowski, J. (2019). When data is capital: Datafication, accumulation, and extraction. *Big data & society*, *6*(1), 2053951718820549.

Saheb, T. (2022). "ethically contentious aspects of Artificial Intelligence Surveillance: A social science perspective." *AI and Ethics*, *3*(2), 369–379. https://doi.org/10.1007/s43681-022-00196-y

Sainato, M. (2024, April 3). *"they are breaking the law": Inside Amazon's bid to stall a Union Drive*. The Guardian. https://www.theguardian.com/technology/2024/apr/03/amazon-union-warehouse-california

Schultz, M. D., Conti, L. G., & Seele, P. (2024). Digital Ethicswashing: A systematic review and a process-perception-outcome framework. *AI and Ethics*, *5*(2), 805–818. https://doi.org/10.1007/s43681-024-00430-9

Schwerzmann, K., & Campolo, A. (2025). "Desired Behaviors": Alignment and the emergence of a machine learning ethics. *AI & SOCIETY*. https://doi.org/10.1007/s00146-025-02272-3

Seele, P., & Schultz, M. D. (2022). From greenwashing to machinewashing: A model and future directions derived from reasoning by analogy. *Journal of Business Ethics*, *178*(4), 1063–1089. https://doi.org/10.1007/s10551-022-05054-9

Simberkoff, D. (2018, August 30). *How facebook's Cambridge Analytica scandal impacted the intersection of privacy and regulation*. CMSWire.com. https://www.cmswire.com/information-management/how-facebooks-cambridge-analytica-scandal-impacted-the-intersection-of-privacy-and-regulation/

Skovgaard Ølykke, G. (2023). Public procurement and abuse of dominance. *Research Handbook on Abuse of Dominance and Monopolization*, 420–437. https://doi.org/10.4337/9781839108723.00032

Smiley, M. (2022, December 19). *Collective Responsibility*. Stanford Encyclopedia of Philosophy. https://plato.stanford.edu/entries/collective-responsibility/

Stacciarini, J. H., & Gonçalves, R. J. (2025). *Data Centers, Critical Minerals, Energy, and Geopolitics: The Foundations of Artificial Intelligence*. https://doi.org/10.31235/osf.io/2zvkt_v1

Supply Chain Integrity: Microsoft CSR. Microsoft. (2025). https://www.microsoft.com/en-us/corporate-responsibility/supply-chain-integrity

Sörlin, S. (2022). The extractivist paradigm. *Resource Extraction and Arctic Communities*, 3–32. https://doi.org/10.1017/9781009110044.003

Svampa, M. (2019). *Neo-extractivism in Latin America: Socio-environmental conflicts, the territorial turn, and new political narratives*. Cambridge University Press.

Tacheva, J., & Ramasubramanian, S. (2023). AI Empire: Unraveling the interlocking systems of oppression in generative AI's global order. *Big Data & Society*, *10*(2), 20539517231219241.

Taffel, S. (2023). Data and oil: Metaphor, materiality and metabolic rifts. *New media & society*, *25*(5), 980-998.

Tan, R., & Cabato, R. (2023, August 28). Scale AI's remotasks workers in the Philippines cry foul over low pay - The Washington Post. https://www.washingtonpost.com/world/2023/08/28/scale-ai-remotasks-philippines-artificial-intelligence/

Taylor, A. (2023, April 26). *The Automation Charade*. Logic(s) Magazine. https://logicmag.io/failure/the-automation-charade/

Taylor, P. (2025, June 30). *Hyperscale Data Centers Worldwide 2023*. Statista. https://www.statista.com/statistics/633826/worldwide-hyperscale-data-center-numbers/

Taylor, P. (n.d.). *Topic: Data Centers*. Statista. https://www.statista.com/topics/6165/data-centers/#:~:text=However%2C%20with%20global%20data%20center,strain%20on%20local%20power%20grids.

Taylor, R. M. (2023). Methodological Reflections on Normative Case Studies: What they are and why we need better quality criteria to inform their use. *Educational Theory*, *74*(3), 301–311. https://doi.org/10.1111/edth.12555

TechEquity. (2025, June 19). *The Ai Supply Chain, explained*. TechEquity Collaborative. https://techequity.us/2025/01/31/ai-supply-chains-explained/

Tortorelli, P., Stevens, P., & Tolockaite, A. (2025, June 20). *In race to attract data centers, states can forfeit hundreds of millions of dollars in tax revenue to tech companies*. CNBC. https://www.cnbc.com/2025/06/20/tax-breaks-for-tech-giants-data-centers-mean-less-income-for-states.html

Tyson, L. D. A., Zysman, J., & Judge, B. (2023). *The new logic of globalization: uncertainty, volatility, and the digital economy*. Berkeley, University of California, Berkeley Roundtable on the International Economy

UNESCO. (2024). *UNESCO launches Global Ai Ethics and governance observatory at the 2024 global forum on the Ethics of Artificial Intelligence*. Digital Skills and Jobs Platform. https://digital-skills-jobs.europa.eu/en/latest/news/unesco-launches-global-ai-ethics-and-governance-observatory-2024-global-forum-ethics

United States: Appeals Court holds tech companies are not liable for forced labor used by Cobalt Suppliers. The Library of Congress. (n.d.). https://www.loc.gov/item/global-legal-monitor/2024-04-02/united-states-appeals-court-holds-tech-companies-are-not-liable-for-forced-labor-used-by-cobalt-suppliers/

Vakkuri, V., Kemell, K. K., & Abrahamsson, P. (2019). AI ethics in industry: a research framework. *arXiv preprint arXiv:1910.12695*.

Valdivia, A. (2024). The supply chain capitalism of AI: A call to (re)think algorithmic harms and resistance through environmental lens. *Information, Communication & Society*, 1–17. https://doi.org/10.1080/1369118x.2024.2420021

Vallor, S. (2024). *The AI Mirror: How To Reclaim Our Humanity in an age of machine thinking*. Oxford University Press.

van de Poel, I. (2020). Embedding values in artificial intelligence (AI) systems. *Minds and Machines*, *30*(3), 385–409. https://doi.org/10.1007/s11023-020-09537-4

van Wynsberghe, A. (2021). Sustainable AI: AI for Sustainability and the sustainability of ai. *AI and Ethics*, *1*(3), 213–218. https://doi.org/10.1007/s43681-021-00043-6

Varon, J., & Peña, P. (2021). Artificial Intelligence and consent: A feminist anti-colonial critique. *Internet Policy Review*, *10*(4). https://doi.org/10.14763/2021.4.1602

Veliz, C. (2021). *Privacy is power*. Melville House.

Vipra, J., & West, S. M. (2023). Computational power and AI. *AI Now Institute*, *27*.

Violino, B. (2024, September 8). *Amid ai boom, tech can't afford to neglect spending in these it areas*. CNBC. https://www.cnbc.com/2024/09/08/amid-ai-boom-tech-cant-afford-to-neglect-spending-in-these-it-areas.html

Véliz, C., & Law, S. (2023). Artificial Intelligence and data harvesting: An interview with Carissa Véliz. *Think*, *22*(63), 59–62. https://doi.org/10.1017/s1477175622000215

Wagner, B. (2019). Ethics as an escape from regulation. from “ethics-washing” to ethics-shopping? *BEING PROFILED*, 84–89. https://doi.org/10.1515/9789048550180-016

Wallace, R. J. (1998). *Responsibility and moral sentiments*. Harvard University Press.

Wallach, O. (2021). *The World’s tech giants, compared to the size of economies*. Visual Capitalist . https://www.visualcapitalist.com/the-tech-giants-worth-compared-economies-countries/

Whitman, R. (n.d.). *AI is mining more than data – Vermont Journal of Environmental Law*. ExtremeTech. https://vjel.vermontlaw.edu/top-ten/2025-top-ten/2024/12/ai-is-mining-more-than-data/

Widder, D. (2023). *Ethics down the AI supply chain: playing with power* (Doctoral dissertation, Carnegie Mellon University)

Widdows, H. (2021). Structural injustice and the requirements of beauty. Journal of Social Philosophy. https://doi.org/10. 1111/josp.12369

Young, I. M. (1990). *Justice and the Politics of Difference* . Princeton Univ. Pr.

Young, I. M. (2011). *Responsibility for Justice* . Oxford University Press.

Zalnieriute, M. (2021). “transparency washing” in the digital age: A corporate agenda of procedural fetishism. *Critical Analysis of Law*, *8*(1), 139–153. https://doi.org/10.33137/cal.v8i1.36284

Zewe, A. (2025, January 17). *Explained: Generative AI’s environmental impact*. MIT News | Massachusetts Institute of Technology. https://news.mit.edu/2025/explained-generative-ai-environmental-impact-0117

Zheng, R. (2018). What is my role in changing the system? A new model of responsibility for structural injustice. *Ethical Theory and Moral Practice*, *21*(4), 869–885. https://doi.org/10.1007/s10677-018-9892-8

Zuboff, S. (2019, January). Surveillance capitalism and the challenge of collective action. In *New labor forum* (Vol. 28, No. 1, pp. 10-29). Sage CA: Los Angeles, CA: Sage Publications.