MALMÖ UNIVERSITY

Decolonising the hidden labour of global 'AI Empire'

A qualitative content analysis of the EU Artificial Intelligence Act

Carsten Skov O'Donnell

One-year Political Science MA programme in Global Politics and Societal Change
Dept. of Global Political Studies
Course: Political Science Thesis ST632L (15 credits)
[Spring semester, August 2024]
Supervisor: [Bo Petersson]

Abstract

Billed as the 'world's first comprehensive AI law', the EU AI Act (AIA) sets a precedent for the regulation of Artificial Intelligence (AI) – one that could influence policy-making globally. Referencing the whole 'AI value chain', AIA employs a risk-based legislative approach. Yet, *prima facie,* this constitutes an internal contradiction of the regulation. Given that risk is a fundamentally forward-looking concept predicated on the 'unrealised potentialities' of future events, any risk-based regulation offers a limited scope for regulating retrospective harms along the value chain. Certainly, by drawing on the theory of 'AI empire', this thesis explores how the AI value chain is saturated by the exploitation of 'AI labour' – that is, human labour which produces metadata for AI systems. In determining *to what extent AIA addresses the exploitation of human-performed 'AI labour'* – my research problem – I undertake a qualitative content analysis of AIA using both data-driven and theory-driven coding. I conclude that AIA contains a wholly inadequate address of AI labour since AIA 1) contains no conceptualisation of AI labour; 2) fails to enshrine AI labourers' rights by only protecting rights deemed '*at risk'* by functional AIs; and 3) legislates solely against prospective risks late in the AI value chain.

Word count of thesis: 16500

Table of contents

Introduction

In December 2023, the EU Council and European Parliament reached a provisional agreement concerning the EU Artificial Intelligence Act (henceforth, all references to 'AIA' and its textual components refer to this version of the AI Act)[1]. Branded as the "world's first comprehensive AI law", AIA was initially proposed by the European Commission in April 2021 as part of the European AI Strategy (European Parliament, 2023a). Subject to subsequent amendments – including a major overhaul in June 2023 to address the advent of generative AI models like OpenAI's ChatGPT (European Parliament, 2023b) – the regulation was adopted by Parliament in March 2024 and approved by the EU Council in May 2024 (Council of the EU, 2024). While appreciating that Artificial Intelligence (AI) represents the "functional combination" of data, algorithms, and hardware (Calderaro and Blumfelde, 2022:416), this thesis also agrees with Article 3(1) which defines AI as a "machine-based system designed to operate with varying levels of autonomy and that may exhibit adaptiveness after deployment and that, for explicit or implicit objectives, infers, from the input it receives, how to generate outputs […] that can influence physical or virtual environments". Certainly, in many different ways – from credit checks to job screening – AI systems have begun producing outputs and "making decisions that affect human lives" (Pavlidis, 2024:294). Recognising as much, the ambition of AIA is "to address risks to health, safety and fundamental rights" (European Commission, 2023:1). While these hopes for AIA are great and grand, the lack of a similar legislative precedent warrants scrutiny of just how 'comprehensive' and effectual AIA will be.

AIA's claim to being a world-first AI law is a justifiable one. While, for example, a US blueprint for an "AI Bill of Rights" has been proposed (White House, 2022), the EU remains the sole frontrunner in terms of regulating AI. As such, the global relevance of AIA and its potentially international impact cannot be overstated. Recognising AI's "strategic importance" at the crossroads of geopolitics, national security, and industry, the European Commission makes it clear that AIA represents an attempt to become the "global leader" in AI-regulation. More specifically, the Commission claims that "the EU will take actions to

[1] The provisional interinstitutional Parliament-Council agreement, published by the European Parliament 2/2/2024, can be accessed here: https://www.europarl.europa.eu/meetdocs/2014_2019/plmrep/COMMITTEES/CJ40/AG/2024/02-13/1296003EN.pdf (see also List of references).

foster the setting of global AI standards" (European Commission, 2023:9). Commonly referred to as the 'Brussels effect' – that is, the externalisation of EU legislation and the functional diffusion of EU standards beyond the Union itself – numerous scholars have highlighted AIA's potential to spread throughout "the rest of the world" (Bas et al, 2024:56). Leveraging its market power to incentivise the development of EU-compliant AI systems, AIA may constitute a vehicle for propagating EU standards abroad (Siegmann and Anderljung, 2022; Stix, 2021:29). Furthermore, AIA could act as a "springboard" for "global regulatory efforts" (Friedl and Gasiola, 2024:para.15). As Bas et al (2024) suggest, AIA "will constitute a relevant experience to learn from and possibly serve as an archetype to inspire future legislation elsewhere" (p.56). Indeed, within the academic literature on 'policy transfer', it is no secret that policy-makers frequently "learn from abroad" (Dolowitz and Marsh, 2000; see also Minkman et al, 2018; Stone, 2012, 2017). In short, then, AIA could influence industry standards for AI development within and beyond Europe whilst also having its regulatory efforts serve as a global template for AI legislation. Depending on the success of AIA, such a template could serve as a model for emulation or active rejection by foreign policy-makers. Either way, AIA is of global significance.

The risks and potential harms of AI are many. While the benefits of AI range from energy optimisation to scientific breakthroughs (Bas et al, 2024:56), the potential harms include cybercrime, discrimination, privacy breaches, and even the development of biochemical weapons (Hacker, 2023:para.5-8; Pavlidis, 2024:296). Arguably, there is a broad societal need to regulate AI in order to reap its benefits and contain its dangers. In attempting to do so, AIA adopts a specific risk-based regulatory approach. It features a "consequentialist vision of regulation" whereby it regulates only those AI systems deemed most 'risky' – in effect, exempting less risky technologies from (over)regulation (Salgado-Criado and Fernandez-Aller, 2021:61). At the "core" of the EU's AI governance, however, are fundamental rights (Stix, 2021:6). Since both the Coordinated Plan on AI (European Commission, 2018a) and the EU AI Strategy (European Commission, 2018b), respect for fundamental rights has remained firmly situated as the guiding principle of the EU's approach to AI regulation. In Recital 1, it is made clear that AIA serves to protect the "health, safety, [and] fundamental rights enshrined in the Charter [of Fundamental Rights of the European Union]". Indeed, as Ho-Dac (2024) comments, fundamental rights are given a "major role" which, alongside 'health' and 'safety', are considered "public interests" to be protected through a comprehensive risk-based approach (pp.2-5).

Fundamentally, 'risk' is a forward-looking concept. As Lupton (2023) puts it, the very nature of risk is transitory: it exists in a state of "always becoming", a potentiality of the future (p.8). Accordingly, risk-based regulation is concerned with "the *likelihood* of certain consequences occurring" (van Dijk et al, 2016:292, original emphasis). AIA is no different. It adopts a risk-based approach as "the basis for a proportionate and effective set of binding rules" (Recital 14a) – categorising AI systems according to the risks posed by their "intended purpose[s]" (European Commission, 2023). Yet, AIA's risk-based approach may lead to regulatory pitfalls. If AIA is primarily concerned with mitigating the risks of AI – i.e. the potentialities of future events – then it may neglect to address injustices, harms, and rights violations occurring earlier in the AI value chain.

Sketching the contours of "AI empire" – a "networked and distributed global order" premised on "interlocking systems of oppression" – Tacheva and Ramasubramanian (2023:1) reveal how the AI value chain begins much earlier than with the development or release of the AI system itself. From the mining of critical minerals to hardware assembly, AI fundamentally relies on the "exploitation of labor" (*ibid.*:p.7). In particular, human-performed labour is required for the production of data – an essential and defining element of AI. The production of data may occur via the digitally-mediated abstraction of individuals' data – a phenomenon Couldry and Mejias (2019) dub "data colonialism" – or through the annotation and correction of datasets. Regardless, this thesis will henceforth refer to 'AI labour' in a specific, restricted sense: as the labour which directly contributes to the production of data for AI systems. This definition is not to neglect the myriad other forms of labour along the AI value chain, but rather an acknowledgement of the limited scope of this thesis and the impossibility of sufficiently addressing all forms of labour. The term 'AI labour' will be expanded upon, as will the theories of data colonialism and AI empire. Not least because they reflect the premise of this thesis and offer an important research puzzle. For how can AIA's forward-looking risk-based approach adequately address the exploitation of labour occurring earlier along AI empire's value chain?

There appears to be a puzzling contradiction within AIA. It consistently references the 'AI value chain' but fails to indicate when (or where) the chain begins. Article 28, for example, is titled 'Responsibilities along the AI value chain' yet only discusses the providers of AI systems or third-party suppliers of AI components. The article features no provisions for practices occurring prior to the final, development stages of AI models. As such, without a clear conceptualisation of the full AI value chain, AIA risks having its risk-based approach rendered inadequate. To the extent that AIA is only concerned with the risks and rights

attached to the final 'links' of the AI value 'chain', AIA may be vulnerable to overlooking earlier misdeeds. In terms of AI empire, this means possibly disregarding the "ghost work" (Tacheva and Ramasubramanian, 2023:8) of "largely invisible workers" (Arora et al, 2023:1) whose "hidden human labor" (Perrigo, 2023) appears systematically exploited. Still, the merit of AIA – given its risk-based approach with fundamental rights at its 'core' – would seem to be a form of dual protection: while risk management seeks to pre-empt the potentiality of unwanted harms, rights enshrine one's entitlement to protection from such harms. Indeed, Watson (2021) notes that the function of rights is to "protect right-holders" (p.7). As such, to uncover whether AIA is truly concerned with the *whole* AI value chain and the apparent exploitation of labour embedded throughout it, we must not only ask how AIA conceptualises 'labour', but also how it operationalises different types of 'risks' and 'rights'. Fundamentally, this thesis aims to reveal how AIA's approach to 'labour', 'risk', and 'rights' is integral to understanding *who* AIA considers worth protecting from *which* dangers occurring at *what* stages of the AI value chain – and *how* this leads to a more or less adequate regulatory address of labour under AI empire. In sum, guided by the theory of AI empire, and in acknowledgement of AIA's internal contradiction – that is, claiming to address the AI value chain but never fully conceptualising it – this thesis takes the following as its research problem:

To what extent does AIA address the exploitation of human-performed 'AI labour'?

In order to solve this research question, I strive to answer three research questions:

1) *How does AIA conceptualise 'AI labour'?*
2) *Which rights does AIA aim to protect?*
3) *What risks does AIA endeavour to legislate against?*

These research questions reflect a tripartite hypothesis:

AIA fails to adequately address 'AI labour' because:

H1) ...AIA includes an insufficient conceptualisation of AI labour
H2) ...few, if any, of the rights appearing in AIA pertain to 'AI labour' or 'AI labourers'
H3) ...AIA legislates against prospective risks occurring late in the AI value chain

Following this introduction, the thesis offers a brief review of the literature on rights, risks, and data colonialism. In the Theory section, I then relate data colonialism to the wider theory of AI empire. Here, I also outline the global AI value chain. Next, I offer theoretical elaborations of the Marxist approach to 'labour' as well as the relationship between 'rights', 'risk', and EU legislation. In my Methodology section, I justify my choice of material and method: a qualitative content analysis of AIA using a mix of data-driven and theory-driven coding. Subsequently, I present my results before engaging in analytical discussions pertaining to each of my three main coding categories: 'labour', 'rights', and 'risk'. Finally, I conclude that the tripartite hypothesis cannot be rejected since there appears to be an entirely insufficient conceptualisation of AI labour, whereby AI labourers are overlooked as right-holders. Moreover, to the detriment of these labourers, AIA's risk-based approach fails to address the risks and harms occurring prior to the final stages of the AI value chain. Nedzhvetskaya and Tan (2021) write that "the role of workers in AI governance has received […] little attention" (p.2). In analysing AIA, it is my ambition to both interrogate AIA's treatment of AI labour whilst shedding light on the importance of human labour to artificial intelligence.

Literature review

Underpinning AIA are the dual concepts of 'risk' and 'right'. Given their centrality to AIA, it would be remiss not to review and reflect upon their definition and function. As such, the following literature review consists of three parts. First, a review of the most relevant literature pertaining to 'risk'. Second, a review of the philosophical approaches to 'right'. The last section of this literature review is dedicated to outlining the theory of data colonialism. While data colonialism is not strictly related to 'risk' or 'right', it helps serve as an introduction to the more extensive theory of AI empire, which will be presented in the subsequent 'Theory' segment of this thesis.

Risk

Risk is a fundamentally forward-looking concept. It is 'always becoming'; an "unrealised potentiality" of an unwanted event or harm (Rigakos and Law, 2009:80). The esteemed risk scholar, Ulrich Beck, would concur. "Risks exist in a permanent state of virtuality, and become 'topical' only to the extent that they are anticipated. Risks are not 'real', they are

'*becoming real*'", Beck (2006:332, original emphasis) contends. Nevertheless, for Beck, 'risk' is more than a shorthand for any unrealised potentiality. Risk, he says, is a hallmark of modern society. Living in the contemporary 'risk society' means living with the new type of risks introduced with the advent of industrialisation and exacerbated throughout late modernity. With modern globalisation, Beck (1992) warns us that risks are complex, "incalculable", "often irreversible", and endanger all planetary life-forms (pp.21-23). Consequently, risk becomes experienced as an "omnipresent" facet of life, in which "there are only three possible reactions: denial, apathy or transformation" (Beck, 2006:331). Regardless of the reaction, Beck (1992) defines risk as a "*systematic way of dealing with hazards and insecurities induced and introduced by modernization itself* (p.21, original emphasis). Risk society is systematically and "increasingly occupied with debating, preventing and managing risks that it itself has produced" (Beck, 2006:332). Still, if one reaction is to manage such risks and prevent harm, another is outright denial. Denial, Beck notes, often has an exacerbating effect. The "irony of risk", he argues, is that "the more emphatically the existence of world risk society is denied, the more easily it can become a reality". Here, Beck offers the example of climate change deniers (*ibid.*:p.330). Either way, in Beck's 'risk society', risk is both a forward-looking concept as well as an omnipresent characteristic of life which must be dealt with.

Risks are not value neutral. According to Beck (1992), defining what constitutes risk – and what does not – creates "winners" and "losers" (p.23). For risk determinations are not only based on "mathematical probabilities", but also "social interests" (p.29) – and, as Beck reminds us, those interests typically favour "big business" (p.23). The designation *of risk* is thus also a functional recognition of particular interests being *at risk*. Relatedly, Lupton (2023) contends, societal interests are the reason why the "concept of risk has gained importance in recent times". The interests of society are increasingly dependent on decision-making, all while decision-making, in turn, depends on the identification and management of certain risks (p.18). Yet, "the act of designating riskiness to technology is ultimately a *creative act* situated within a social and cultural context" (Arora et al, 2023:4, added emphasis). Indeed, risk is a "socially constructed phenomenon, in which some people have a greater capacity to define risks than others" (Beck, 2006:333). Crawford (2021), for example, notes how data is often clouded in a discourse of 'immateriality'. If data for AI systems "is seen as abstract and immaterial", she argues, "then it more easily falls outside of traditional understandings and responsibilities of […] risk" (p.113). To paraphrase, when data and AI systems are constructed as immaterial, they are similar considered inconsequential – that is,

they are not considered 'risky'. The risk profile constructed for AI, then, precludes the possibility of unwanted harms. In sum, beyond being an omnipresent, forward-looking concept, 'risk' is also socially constructed through creative acts.

The designation of risk is not only a creative and subjective act, but also a consequential one. As Beck (1992) writes, "[w]hat is at stake in the public dispute over the definition of risks" is a set of "*social, economic and political consequences*" (p.24, original emphasis). Certainly, in discussing EU data legislation, Macenaite (2017) claims that we are seeing the "riskification" of law whereby "risk has become a new boundary"; risk increasingly determines the scope of EU legislation and whether new legal, regulatory, and procedural safeguards are warranted (p.507). What constitutes 'risk', then, has a direct effect on Europe's legal-political outcomes. Furthermore, the strength of risk-based approaches to policy-making can be found in their flexibility. They reflect a "proportionate and adaptive strategy for regulatory enforcement" by setting priorities and allocating responsibilities without the rigidity of rule-based regulation (*ibid.*:pp.515-516). Modern risks are "too complex" for traditional regulatory approaches to adequately address them (p.534). Therefore, the "ultimate aim of risk regulation" is not to secure compliance with a specific set of rules, but rather to allocate responsibility for controlling potential risks (p.532). However, risk-based approaches have their drawbacks too. Following the example of data legislation, Macenaite argues that a pure risk-based approach would fail to sufficiently protect the right to data protection since it would "presuppose the legality" of data processing activities; the approach only considers the *risks* of data processing, not the *legitimacy* of data processing itself (pp.533-534). For similar reasons, scholars discuss a "duality" of risk (Macenaite, 2017:510; Arora et al, 2023:1). In attempting to legislate against specific potentialities, risk-based regulations may inadvertently institutionalise or perpetuate other types of risk. For example, it is conceivable that in trying to mitigate the risks posed *by* AI systems, AIA may presuppose the legitimacy of risks and harms going *into* AI development.

Rights

AIA's risk-based regulatory approach is about protecting fundamental rights. Consequently, we must also review the concept of 'right'. Watson (2021) defines 'right' "as a *complex entitlement that provides justification for the performance and prohibition of actions and omissions*" (p.3, original emphasis). This definition consists of multiple components, and the following will address each in part. In terms of 'complex entitlement', Watson draws on the

seminal work by Wesley Hohfeld (see Hohfeld, 1913) to present the notion of 'claim-rights'. Understood to constitute a 'right' in the strictest sense of the term, claim-rights are effectively claims by the right-holder which "determine what duties people have" with respect to the right-holder. Every claim-right confers a "correlative duty" on others (Watson, 2021:4). For example, with property rights, the owner has a claim-right against others. The owner can lend his/her property against specific conditions, demand it back, or entirely block others from using it. In effect, others have a duty to abide by these demands and to not use the right-holder's property without their permission. There are two types of claim-rights: positive and negative claim-rights (p.5). Negative claim-rights require inaction from the duty-bearer; they have a duty to do nothing against the right-holder. An example would be the right of individuals to not have their privacy violated. In contrast, the right to education or right to medical care are examples of positive claim-rights: the right-holder must be actively provided with education/medical care by the duty-bearer (in many cases, some government agency).

The remaining aspects of 'complex entitlement' refer to 'privileges', 'powers', and 'immunities'. Power-rights effectively allow the right-holder to waive their claim-right (Watson, 2021:4). In the property rights example, this would be tantamount to letting others use your property, thereby foregoing your claim-right to refuse them. The right-holder may be willing to temporarily waive their claim-right precisely because they have an immunity-right. With an immunity-right, the right-holder enjoys immunity from the arbitrary altering or deprivation of his/her claim-right. In other words, an individual – say the one borrowing the right-holder's property – cannot suddenly decide that they own the item and appropriate the associated claim-right. Finally, a privilege-right means that the right-holder is at liberty to do a certain thing, but not in a way that places duties or obligations on others (*ibid.*). If the right-holder owns a football, (s)he is free to play with it but cannot compel others to join in. While complex entitlement refers to what rights *are*, we must also question what rights *do*.

The function of rights is to protect right-holders. Watson (2021) argues that this is achieved thanks to rights mandating "actions and omissions". Some actions are mandated *qua* positive duties, whilst negative claim-rights are effectively entitlements to non-interference, mandating the omission of action. Regardless, there are two theories as to *how* rights protect right-holders. The 'interest theory' holds that individuals enjoy rights when others have a duty to protect that person's interests, in many cases assumed to be their wellbeing. Here, rights protect right-holders' interests (Eleftheriadis, 2008:13). In contrast, the 'will theory' holds that rights protect the right-holder's will. From this perspective, duty-bearers have an obligation to protect the "free, autonomous choices" of right-holders (Watson, 2021:8).

The final aspect of Watson's definition pertains to the justification for rights – *why* should certain 'actions and omissions' be 'performed or prohibited'? There are conflicting views as to why rights are justified. The status-based approach to rights points to the moral status and 'final value' of right-holders. As Watson (2021) remarks, the term "final value" points to how the value of rights is derived, in its final instance, from the intrinsic value of right-holders themselves. Rights have value because they protect valuable right-holders (p.10). Indeed, insofar as such value confers some sort of moral worth, "to have a right", Thomson (1990) suggests, "is to have a kind of moral status" (p.38). Conversely, the instrumentalist approach to rights believes that their justification is "grounded in the further goods that they lead to" (Watson, 2021:11). Utilitarian philosophers like J.S. Mill (2015 [1861]), for example, may justify rights by claiming they produce overall greater levels of happiness.

Watson's (2021) definition encompasses the *what, how,* and *why* of rights. Yet, there are other debates as to the nature of rights. One concerns the difference between legal and moral rights. Legal rights are those institutionalised and codified within judicial systems (Eleftheriadis, 2008:6), whilst a moral right is "a right whose existence can be established by moral argument" or through an intuitive sense of morality (Gilbert, 2018:36). Following this distinction, Gilbert claims that moral rights reflect normative ideas of right or wrong, whilst legal rights lack an intrinsic sense normativity; legal rights and laws can be criticised for being morally or normatively 'wrong' (p.37). Legal rights theorists, however, may dispute the notion that there is such a thing as moral law. Legal positivists such as Jeremy Bentham (1843) have famously decried the notion of extra-judicial, non-legal rights as unfounded "nonsense upon stilts" (p.914). They would insist that rights can only exist *qua* legal rights, upheld by legal institutions. Yet, even within this school of thought, disagreements persist. Consider the EU, for example, where proponents of the 'sovereigntist' approach would claim that only sovereign states can guarantee fundamental rights, while proponents of the 'pluralist' approach consider fundamental rights best protected by a plurality of (competent) national and supranational judicial authorities (Fabbrini, 2014:16-20). Thus, there remains a distinction both between moral and legal rights in general, but also between the specific *types* of legal rights.

A final point of note is the difference between procedural and substantive rights. Substantive rights refer to those rights that exist for their own sake, and their intrinsic worth to the right-holder – such as the right to life or liberty. Procedural rights, by contrast, refer to the regulations, rules, procedures, and entitlements intended to ensure justice and the

preservation of substantive rights – such as the right to information or legal counsel (Eleftheriadis, 2008:4). Such rights are typically articulated as legal rights but can still enshrine moral values. A related distinction between types of rights refers to the 'absolute' versus 'relative' theories of rights. The former theory considers rights, particularly substantive rights, to have an "untouchable core" – that is, a certain "essence" which cannot be compromised. The relative theory, in contradistinction, believes compromise is possible: legislation should seek to balance proportionality and procedural practicality with this so-called 'essence' of rights (Dawson et al, 2019:767). In sum, legislative practice means considering which types of rights to enshrine, what their intended function is to be, how they are justified, as well as how to balance substance and essence with procedure and proportionality.

Data colonialism

The final part of this literature review is dedicated to the theory of data colonialism. To understand data colonialism – and how it relates to the wider theory of 'AI empire' discussed below – we must review data colonialism's theoretical foundations.

According to Couldry and Mejias (2019), data colonialism refers to a "new stage of capitalism" wherein "everyday life must be reconfigured and represented in a form that enables its capture as data" (pp.337-339). It is about "transforming life process into "things" with value". That is, abstracting the activities of human lifeforms into valuable 'data' objects (pp.342-343). For example, the authors argue that social media facilitate a form of online activity – a specific mode of social interaction – which is "ready for appropriation and exploitation for value as data" (p.338). Similarly, they believe that the 'Internet of Things' functions to "*continuously and autonomously* collect and transmit data" about its users (p.344, original emphasis). Put succinctly, data colonialism may be understood as the ambition to render all human activity valuable to capitalist interests. This is achieved by reconfiguring everyday life in such a way that it is mediated by digital technologies which, in turn, capture human activities and abstract them as data commodities.

Central to data colonialism is the transformation of "productive activity" into "labour" – appropriating human life as something which produces surplus-value to be abstracted and extracted *qua* data (Couldry and Mejias, 2019:342). Such capital-oriented "extractive rationalities" underpin the authors' claim that the above constitutes data *colonialism*, analogous to the extractivism of historical colonialism and similarly exploitative of

un(der)compensated labour (*ibid*.:p.340). However, data colonialism builds on other theoretical critiques. 'Playbour', a portmanteau of 'play' and 'labour', represents a "new ideology of capitalism" – one where "[w]orking time and spare time become inseparable" (Fuchs and Sevignani, 2013:265). Essentially, the experience of digitally-mediated activity can be a fun, enjoyable, and playful one – from video games and social networks to fitness tracking and 'smart' devices. Yet, in addition to playful fun, play-labour also "creates a data commodity" whose exchange value becomes the "the heart of the capital accumulation model" for many technology firms (p.237). Receiving little to no compensation for the data abstracted and extracted from their productive activities, users of digital technologies are essentially alienated from the (data) products of their (play) labour. They are, in effect, "highly exploited" workers who "create surplus value and monetary profits" for data colonialists (Fuchs, 2014:280-281).

Another useful concept for understanding data colonialism is 'prosumerism' – a second portmanteau, here denoting 'productive consumerism'. It is closely related to playbour insofar as both relate to how the "data of users' activities becomes a product that is sold" or otherwise capitalised on for a profit (Fuchs, 2014:280). In the usage popularised by Fuchs, 'prosumerism' is typically understood in relation to social media. That is, how the "profiles, content, connections, social relations, networks and communities" of social media *consumers* at the same time *produce* valuable data (*ibid.*). Everything from users' social relations to their music preferences is abstracted and expropriated as data, which can then be sold as a commodity (Gonçalves and Costa, 2020:156-157). However, as observed by Walton and Nayak (2021), "data has now become a form of capital (data capital) which is capable of generating new digital products and services" (p.2). Such products include AI. Especially since AI requires vast quantities of data, Walton and Nayak argue that only "data-rich platform companies" will be able to develop and own AI systems (p.5). Thus, beyond merely selling data for profit, data-rich firms can also capitalise on their data monopolies by reinvesting their data capital into even more profitable ventures. It is, in their interactions with digital technologies and platforms, prosumers who play a crucial part in providing firms with vast amounts of data. Yet, as Fuchs (2014) reminds us, firms can only afford to obtain so much data since prosumers receive no wages for their productive (play) labour, which is typically conducted in prosumers' spare time and rarely actually perceived as labour (p.280). In fact, data colonialism is premised precisely on obscuring the human productivity integral to all data.

A key 'extractive rationality' of data colonialism is 'dataism'. Dataism presumes trust in the objectivity of data and data collection (van Dijck, 2014). Put differently, dataism serves to naturalise data. As Taffel (2021) asserts, "raw data is an oxymoron. Data does not exist 'out there' in the world waiting to be collected; it must be actively produced or generated" (p.21). Couldry and Mejias (2019) recognise as much when claiming that data must be constructed as "raw material" which is "just there" (pp.339-340). Only then can data justifiably be expropriated under data colonialism. The alternative is to recognise the human inputs and subjectivity going into data. Therefore, data must be "made "objective"" (Diana, 2021:208). Yet, data is far from objective and natural. Not only does data reflect information about human activity, but as van Dijck (2014) notes, the appropriation of data requires "careful interpretation *and intervention*" (p.201, original emphasis). For example, under data colonialism, the algorithms and user interfaces of online platforms are "systematically fine-tuned to channel user responses" in a way that enables their capture as data (*ibid.*). Thus, both the activity being captured *and* the mode of capturing it require specific human behaviours for such activity to be generative of data. From this point of view, dataism's discursive naturalisation of data as a readily available natural resource serves to conceal the colonisation of human prosumers and their data. Thus, in order to guarantee prosumers' rights and to protect them from exploitation, it is imperative that policies like AIA recognise and address the colonisation of data-producing labour rather than assuming that data is 'just there' as a natural resource.

To summarise the literature, both 'risk' and 'rights' are highly multifaceted concepts. How they are conceptualised and operationalised is of consequential significance. Defining a risk or right is of consequence to how societal issues are framed and dealt with. Analysing the nuances of 'risk' and 'rights' in AIA should therefore prove a fruitful starting point for scrutinising its address of AI labour. Data colonialism, on the other hand, offers useful insights as concerns the extractive, colonial approaches to data accumulation. Yet, data colonialism is only one strand, or one mode, of exploiting human productivity for data. While data colonialism helps reveal the value attributed to data, we must turn to the wider theory of 'AI empire' to consider the broader, globalised dynamics of labour exploitation and data production.

Theory

In this segment, I build on the reviewed literature whilst expounding the theories and concepts most pertinent to this thesis. I start by outlining the theory of AI empire. Here, I also describe the AI value chain. I then draw on (neo-)Marxian theory to clarify some of this thesis' most central concepts – such as 'labour' and 'exploitation' – and relate them to AI. Lastly, I elaborate on the relationship between 'risk' and 'right' as well as their potential relevance to AIA.

AI empire

Presented by Tacheva and Ramsubramanian (2023), 'AI empire' is an intersectional approach to AI. In recognising "the entire lifecycle of AI algorithms, as well as the associated material, knowledge, data, logistical, labor, and political, cultural, economic, and ideological infrastructures behind them" (p.2), Tacheva and Ramsubramanian (2023) are highly critical of AI's "assumed objectivity" (p.8). Like data under data colonialism, they argue that in popular imagination AI is assumed to exist independently of human labour. Kate Crawford (2021), whom the authors cite extensively, adds that AI also has a presumed immateriality – "removed from any relation to the material world" (p.7). With these constructions of AI, its material entanglements are obscured as are the many subjective, human inputs without which AI could not exist.

The material and human origins of AI are evident from the very beginning of its global value chain. For example, AI requires "the extremely labor-intensive processes of mining and refining the rare earth minerals used to build the hardware infrastructure powering AI" – with such labour frequently occurring in conflict areas such as the DRC (Tacheva and Ramsubramanian, 2023:5). After sourcing and purifying these minerals, hardware components are then manufactured and assembled under what are often slave-like conditions (Fuchs, 2014:1-4). For example, the Chinese ICT-manufacturer Foxconn is infamous for its workers committing suicide due to harsh working conditions (Mejias, 2013:32). Once technological components become obsolete – such as processors, batteries, and data storage hardware – they are disposed of in ways that may expose e-waste workers to cancerous "pollutants such as dioxins and furans" (Taffel, 2021:6-7). In all these ways, AI leaves a trace of labour exploitation and precarity in its wake. Naturally, however, the AI value chain does not jump from hardware assembly to disposal. AI systems must also be developed. Operating at these midway 'links' of the value chain, Nedzhvetskaya and Tan

(2021) identify two types of "AI workers". The first type, the "designers", are those humans who "design, construct, or apply an AI technology". By contrast, "trainers" are workers who "feed data into an existing AI technology [or] supplement AI systems with human labor" (pp.3-4). Indeed, insofar as data is a defining aspect of AI, trainers – those who provide AI models with data – are equally vital to the development of such systems as designers. Certainly, "AI systems are not autonomous, rational, or able to discern anything without extensive, computationally intensive training with large datasets" (Crawford, 2021:8). From this perspective, it would seem the provision of data for AI systems would be a handsomely compensated form of labour. Yet, "exploitative forms of work exist at all stages of the AI pipeline" (*ibid.*:p.63). And as the following reveals, there are no exceptions along the global value chain for AI trainers or others who produce data for AI – their labour, too, is exploited and precarious.

As the basis for its "sense-making", AI demands accurate, legible, and unbiased data (Crawford, 2021:95) – and AI empire has multiple modes of extracting it. Firstly, in discussing Couldry and Mejias' (2019) work, Tacheva and Ramasubramanian (2023) recognise data colonialism as "illegitimate appropriation, exploitation, and dispossession"; data colonialism, they claim, expropriates data-producing immaterial labour (p.7). Immaterial labour is defined by Fuchs and Sevignani (2013) as human activity which "creates immaterial products, such as knowledge, information, communication, a relationship, or an emotional response" (p.256). Yet, as they remind us, there is no truly 'immaterial' form of labour (*ibid.*). Even the data produced by prosumers' playbour under data colonialism – an 'immaterial' product embedded with some kind of information – relies on material, human bodies and brains as well as physical hardware infrastructures, consumer products, and data centres. To be explicit, 'immaterial' labour requires humans to interact with material reality. Regardless, Tacheva and Ramasubramanian (2023) insist that AI empire depends on both 'immaterial' labour as well as more recognisable forms of material labour (pp.2-3). In fact, they assert that we must "illuminate parts of AI Empire beyond data colonialism's focus on […] data extraction", and acknowledge material forms of labour such as "content moderation, data annotation, and the evaluation of AI output— processes that, as AI companies admit, are essential" (pp.5-6). For whilst data colonialism is one avenue for appropriating data, the broader AI empire relies heavily on exploiting the material labour of AI trainers, annotators, moderators and more. As such, let us turn our attention to these forms of labour.

A type of 'AI trainer', data annotators are individuals who label data. As Diana (2021) explains, annotations or labels "associate meanings to the training datasets needed by AI

development" and play a "fundamental role" in AIs' performance (p.207). In practice, data annotation can range from designating a text as a 'poem' or 'informal SMS' to categorising images according to whether they feature 'trees' or 'cars' (Tacheva and Ramasubramanian, 2023:8). In any case, data annotators and AI trainers effectively instruct AI how to 'interpret' otherwise meaningless data. Still, in their description of AI trainers, Nedzhvetskaya and Tan (2021) falsely assume that "AI workers must be employed" (p.5). In fact, under data colonialism, uncompensated playbour trains and provides data for AI. For example, social media users upload images, which they consequently 'tag' as featuring specific locations, people, products, and even moods. This "unpaid labor" effectively provides platforms with "proprietary troves" of "labeled data" (Crawford, 2021:106). Another example of how data-rich firms acquire such data-wealth includes Google's reCAPTCHA. To assess whether online users are human or not, it asks users to select images featuring a specific motif. Here, reCAPTCHA essentially demands that users label data to be granted access to a webpage – all whilst providing Google with heaps of annotated data (*ibid*.:p.69). In effect, users train AI by refining its datasets.

Content moderators are arguably another type of AI trainer. Content moderators sift through digital content which other humans have marked as sensitive, or they filter through extreme content specifically assigned by the AI developer. Content moderators improve underdeveloped AIs' interpretive capacities by correcting their outputs and, consequently, providing AI systems with newly-labelled data-inputs. Typically outsourced as "immiserated" and "underpaid" labour in the Global South (Apostolakoudis, 2022:235), many moderators suffer "psychological detriment" from the trauma of the moderated content (Tacheva and Ramasubramanian, 2023:5-6) – varying from descriptions of rape to graphic murders (Perrigo, 2023). Yet, while constituting a fundamentally vital part of the AI value chain, these forms of AI labour are often hidden and obscured from view.

The outsourcing of 'AI labour' – a term henceforth referring to human activities which generate, moderate, or refine data for AI systems – serves not only to provide cheap access to "exploited gig workers" (Diana, 2021:207), but also to invisibilise such labour. Put succinctly, it is "largely invisible workers in the Global South who clean up data and refine algorithm development for the benefit of those using algorithms in the Global North" (Arora et al, 2023:1). This applies both to crowdsourced labourers of the gig economy, but also to contractually employed workers. Perrigo (2023), for instance, reveals how the developer of ChatGPT, OpenAI, contracted Sama to have its Kenyan workers label thousands of violent and traumatic images and texts for approximately $1.5/hour. Yet, there are no visible traces

of these labourers when users interact with OpenAI's products. As Perrigo writes, "AI often relies on hidden human labor in the Global South that can often be damaging and exploitative. These invisible workers remain on the margins even as their work contributes to billion-dollar industries". Furthermore, Diana (2021) notes how databases like ImageNet contain millions of labelled images for training AI. Nonetheless, despite being widely used for AI research and training purposes, these images are "employed without any reference to where and whom labels come from" (p.208). The human labour preconditioning these images is concealed. Thus, it seems AI labour suffers from an unjust irony: being integral yet invisible to the accomplishments of AI developers. Offering "poverty wages" at best, global AI empire rests "heavily on the "ghost work" of human annotators and moderators" (Tacheva and Ramasubramanian, 2023:5-8; also Gray and Suri, 2019) as well as a whole "shadow workforce of contract laborers" (Crawford, 2021:219).

To summarise, the AI value chain is saturated by instances of exploitation. Contemporary AI as we know it would simply not be possible without "layers of exploitation, including the extraction of mass unpaid labor" under data colonialism (Crawford, 2021:69). Notwithstanding full acknowledgement of the AI value chain beginning with the mining and refining of minerals, this thesis is particularly concerned with the exploitation of 'AI labour' – defined as that which provides data for AI systems. And as long as AI labour remains unnoticed and hidden from view, its exploitation under AI empire will persist. AI empire is a global, and globalised, phenomenon whose deconstruction would undoubtedly require broad regulatory action. Insofar as AIA legislates against its exploitative practices, AIA therefore holds the potential to lead the way in decolonising AI empire and inspiring similar policies abroad. Yet, if we are to decolonise AI empire and "advance justice, we must radically transform not just the technology itself, but our *ideas* about it" (Tacheva and Ramasubramanian, 2023:1, original emphasis). We must interrogate the whole AI value chain and reflect upon whether it is justifiable. In this regard, AI empire is a valuable framework for ascertaining whether policies like AIA adequately address the unjust exploitation of 'ghost work'; is AI labour brought under the legislative spotlight or does this 'shadow workforce' remain in the shadows?

Labour

The theory of AI empire helps illuminate the exploitation of labour along the global AI value chain. However, if this thesis is to offer a sound examination of AIA and its address of AI

'labour' and 'exploitation', then these concepts must be further expounded upon. Marxian thought offers some valuable perspectives in this regard.

'*[N]il posse creari de nihilo*' – "out of nothing, nothing can be created" (Marx, 2000 [1887]:311). According to Marx, labour can be defined as that which produces value by expending labour-power to transform "Nature's material" into a product (*ibid.*:pp.261, 269, 282). A distinction can be made between 'work' (or "real labour") which produces use-values to satisfy human needs, and 'labour' (or "abstract labour") which generates exchange value (Fuchs and Sevignani, 2013:239-240, 248). In both cases, labour is embedded and objectified within the product of labour: "from being the labourer working, it becomes the thing produced" (Marx, 2000 [1887]:273). Indeed, in Marx's labour theory of value, any given commodity derives its value from the "quantity of labour expended on and materialised in it" (*ibid.*:pp.269-270) – with this 'quantity of labour' typically measured in units of labour hours. The advent of data colonialism, however, has revived criticisms of this theory. Walton and Nayak (2021), for example, contend that a "'neoproletariat' workforce of data providers" challenges a labour-oriented theory of value. "Datafication", they argue, means that digital 'products', like data, are increasingly dematerialised, easily replicated, and disseminated "at near-zero marginal cost". Consequently, the value of data cannot viably be measured in labour hours (pp.1-3). However, as Fuchs and Sevignani (2013) assert, even in the digital age Marx's theory of value holds true as long as we are willing to reconceptualise 'labour'. They refer to "abstract labour" as any kind of human activity which generates "economic value" (pp.259-260). For example, although "[w]orking time and spare time become inseparable" under data colonialism's playbour (p.265) – thereby rendering 'labour time' conceptually obsolete – it is still human activity that is required for the production of economically valuable data. And data is certainly valuable; "data has now become a form of capital (data capital) which is capable of generating new digital products" (Walton and Nayak, 2021:2, see also Crawford, 2021:113, Diana, 2021). As such, this thesis believes it would be remiss to consider data-generating human activity as anything other than labour – regardless of whether such labour is 'immaterial' and subjected to data colonialism, or a more material form of labour performed by data annotators and moderators concentrated in the Global South.

Marx (2000 [1887]) further distinguished "necessary labour" from "surplus labour". The former denotes the amount of labour required to reproduce the value of labour-power – that is, the economic value *necessary* to sustain the labourer. By contrast, surplus labour refers to all the additional labour occurring in excess of the necessary labour (pp.312-313): "the latter is nothing but the continuation of the former beyond a definite point" (pp.281-

282). Relatedly, surplus-value refers to the difference between the exchange-value of a product and the *necessary* costs of labour-power as well as any other resources required for its production. (p.275). It is the "specific characteristic of capitalism" that labourers work beyond the point where necessary labour becomes surplus labour – and all "so that surplus-value is created" and appropriated by the capitalist (Fuchs and Sevignani, 2013:245). The "rate of surplus-value is therefore an exact expression for the degree of exploitation of labour-power by capital, or of the labourer by the capitalist" (Marx, 2000 [1887]:315). Indeed, in Marxian thought, exploitation is defined as the labourer's "persistent loss of surplus value" (Apostolakoudis, 2022:219). Thus, the greater the rate of surplus-value – that is, the more the labourer conducts surplus labour for the capitalist instead of necessary labour for himself – the greater the 'degree of exploitation'. In sum, exploitation is the extraction and appropriation of surplus-value from the labourer.

This thesis is particularly concerned with AI labour – that which produces or curates data for AI. For "there is no value without some form of human labour" (Apostolakoudis, 2022:219). And as Diana (2021) maintains, such a Marxian approach to value holds true for data as well. "Metadata", he argues, can be defined as otherwise meaningless data whose economic value is increased thanks to the addition of "*something human*, that is something *meaningful*" (p.207, original emphasis). For data to be of any value, it must contain meaning – whether such meaning is manually attributed to data by annotators, labellers, moderators and the like, or "generated by humans indirectly through their interactions with machines: top search key-words and trend topics, web navigation patterns, reactions on social networks, geolocation of photos and videos, etc." (*ibid.*). In essence, human inputs are the source of value in metadata – hitherto simply referred to as 'data'. This gives cause for a slight revision of our definition of AI labour: as human-performed activity which produces or curates *meta*data for AI. For indeed, as recognised by Couldry and Mejias (2019) in their outline of data colonialism, data is not "just there" (pp.339-340). As Marx (2000 [1887]) wrote over a century ago, "[a]ll raw material is the subject of labour" (p.259). Likewise, all nominally 'raw' data – including the vast quantities used to train AI systems – exist only by virtue of labour-power being expended to transform otherwise meaningless data into a product of value. However, with dataism's naturalising discourse, the human subjectivity of metadata "evaporates forever in the digital light" (Diana, 2021:208). Metadata is "made objective" as though humans were never involved in the process. Diana offers the aforementioned example of ImageNet where millions of labelled images are employed without "any reference to where and whom labels come from […] *as if the labels were native properties of the raw*

data" (*ibid.*, original emphasis). Moreover, Diana warns us that such concealment of AI labour risks the "legitimation of potentially harmful choices" (*ibid.*). As the theory of AI empire reveals, such harmful choices include exploiting AI labourers whose value-generating labour is hidden from view and therefore 'legitimately' appropriated.

In criticising Marx, there is no consensus that exploitation – i.e., the appropriation of surplus-value – is necessarily 'bad'. Scholars like Arneson (1981) have long asked, "what's wrong with exploitation?". Certainly, the 'prosumerism' and 'playbour' of data colonialism hardly *feels* exploitative – at least not in the sense of it being an uncomfortable experience (Hjorth, 2018:8). Yet, I agree with Mayer (2007) who insists that what is fundamentally 'wrong' with exploitation is the violation of "fairness". Exploitation entails a "wrongful gain" where "exploiters always gain at the expense of others" – even if the relationship between exploiter and exploited is mutually advantageous (p.137). The same applies to 'prosumerism' and 'playbour': although users may enjoy their digitally-mediated experiences, the appropriation of their data without compensation is still not *fair*. Consequently, although recognising that exploitation, in the Marxian sense, is a defining characteristic of global capitalism, I consider the extreme exploitation of AI labour particularly undesirable. Hence, this thesis is premised on a rather normative assumption: while perhaps not as precarious as contracted annotators and moderators of metadata, prosumerism and playbour are still 'risky' forms of labour insofar as their exploitation constitutes unjust, undesirable, and unwanted events.

Risks and rights

This thesis has already noted the significance of 'risk' and 'right' to AIA and, accordingly, reviewed much of the relevant literature concerning these concepts. However, we are yet to elaborate on the relationship *between* risks and rights.

I have noted that rights, in the strictest sense, may be considered claim-rights which confer duties on others in relation to the right-holder. Discussing EU data protection law, Macenaite (2017) remarks how risk can play a particular role in conferring such duties. In EU legislation, risk is increasingly operationalised as "an obligation adjuster" or as a "trigger for new obligations" (pp.522-525). Put differently, risk becomes a tool for adjusting the duties of, in this case, data controllers. As such, risk also becomes a tool for delineating rights insofar as legal duties entail a corresponding entitlement by potential claimants. That is, the

allocation of duties corresponds to an allocation of rights: if data controllers fail to uphold their duty – say, to ensure privacy – then affected citizens have effectively had their right to privacy violated. Without data controllers having such a duty, citizens would not have a claim-right against them either. In this sense, "an understanding of "risk" as a social and relational concept, entails taking "rights" seriously" (van Dijk et al, 2016:289). We must recognise how risk-based approaches to legislation 'seriously' affects relations of rights and duties.

Another element of risk-based regulation is that it, in effect, "partially shifts the responsibility for the protection of fundamental rights and the blame from policy makers to data controllers" (Macenaite, 2017:540), or whomever the specific legislation applies to. With data protection, for example, it becomes the duty of data controllers to manage potential risks to privacy rights. However, given that risk is a forward-looking concept, it can be argued that the responsibility conferred to duty-bearers is merely 'prospective responsibility' – whereby duty-bearers cannot be held accountable for earlier rights-violations. According to Brunsson et al (2022), "*Prospective responsibility* is forward looking, as when someone is given responsibility for a certain task, implying the expectation that this person will handle this task in the future – a meaning close to duty or obligation. *Retrospective responsibility* is backward looking and defines someone as responsible for an event that has already occurred" (p.3, original emphasis). In this sense, risk-based regulation may successfully allocate duties and prospective responsibilities yet fail to hold anyone accountable for injustices occurring before a specific moment in time. For example, if AIA only allocates prospective responsibilities at or after the point of an AI system's deployment, then it remains no one's duty to prevent the unjust violation of rights earlier along the AI value chain. Everything prior is simply overlooked or presumed to be legal (Macenaite, 2017:533-534).

While risk-based regulation may prompt new duties – and corresponding rights – 'risk' also engenders 'rights' in other ways. Van Dijk et al (2016) note how, sometimes, "rights only exist because they are perceived as being under threat, *i.e.* as being at risk, or already encroached upon. In this sense "risks" and "rights" are already translations or frames for the more immediate *perceptions of threats*" to people's wellbeing (p.295, original emphasis). Van Dijk et al conceptualise this relationship between rights and risk as "rights at risk". Here, it is the experience of 'risk' which precipitates an articulation of 'right': "rights at risk occur as irreducible experiences and perceptions on the side of citizens and publics that their rights have been violated. Without re-articulations of rights within changing techno–legal–political configurations, rights will lose their meaning, and possibly even cease to

exist" (*ibid.*:p.296). From this perspective, rights are sustained by virtue of their continuous reconceptualisation in light of 'techno-legal-political' developments. From privacy to non-discrimination, one can imagine that the advent of AI is one such occasion where new risks warrant new articulations of rights.

A final theoretical relationship between 'risk' and 'right' is less symbiotic. In discussing data protection, Macenaite (2017) notes how risk-based approaches might actually erode data subjects' rights (p.517). As elaborated by van Dijk et al (2016), the "basic processes of risk assessment and management are not fundamentally concerned with the *nature of rights*, but rather with the *likelihood* of certain consequences occurring" (p.292, original emphasis). In other words, risk-based regulation is perhaps less concerned with the 'essence' of rights, but rather adopts risk as an "organising concept" for ensuring compliance and mitigating harm (Macenaite, 2017:517). Rights, in this view, can be compromised and balanced against other practical priorities – such as having legislation that is actually enforceable or a clear allocation of responsibility. In sum, through various mechanisms, 'risk' and 'right' are mutually dependent theoretical concepts – sometimes reifying the other, sometimes undermining it.

This section has outlined the notion of empire, traced the AI value chain, explored the concepts of 'labour' and 'exploitation', as well as discussed the relationship between 'risk' and 'right'. While these theoretical foci may appear divergent, *prima facie*, they serve a joint purpose in disentangling my research problem. For in assessing whether AIA adequately addresses the exploitation of AI labour under AI empire, it is imperative to understand the significance of 'risk' and 'right' – both central concepts within AIA's legal text. It would be remiss to scrutinise AIA while neglecting to scrutinise its most engaged terms. How exactly I endeavour to conduct such analytical scrutiny of AIA is detailed in the following methodology section.

Methodology

This section outlines the methodological approach of the thesis and how it tests the hypothesised relationship between AIA's address of AI labour and a tripartite set of independent variables: the conceptualisation of 'labour', 'rights', and 'risk'. Firstly, however, I justify the choice of AIA as my research material. I then present my choice of method and

the observed rules for conducting a valid and reliable qualitative content analysis. Lastly, I discuss the inevitable trade-offs and limitations of my thesis.

Material

There are three primary reasons for choosing AIA as my case and single source of research material. Firstly, AIA is undeniably of "intrinsic interest" (6 and Bellamy, 2011:114) to citizens, policy-makers, AI developers, wider industry, and academics alike. Its regulatory impact will be felt across numerous sectors – not just the technological sector, but also within the likes of automobile manufacturing and healthcare, where AI plays an increasingly influential role. Likewise, in protecting citizens from the harms of AI, academics and legislators will scrutinise AIA, assess its effectiveness, and perhaps learn from its strengths and weaknesses when new AI policies are to be developed. Indeed, a second justification for my choice of material is AIA's potential 'Brussels effect'. As stated in the Introduction, AIA could not only inspire similar legislation abroad, but also see the EU leverage its market power to have its regulatory standards diffused globally (Bas et al, 2024:56). Finally, AIA is the first and only policy of its kind; any analysis of labour in AI regulation must engage with AIA.

Naturally, counter-arguments to my choice of material exist. It can be argued that AIA "does not exist as a stand-alone piece of legislation and should be considered carefully in respect of/with the existing EU labour legislation" (Cefaliello and Kullmann, 2022:561). Others note that AIA is related to GDPR, which not only seeks to uphold fundamental privacy rights but also kindled EU's "risk-based approaches to regulation" (Salgado-Criado and Fernandez-Aller, 2021:61). Likewise, with this thesis' theoretical focus on data, a similar accusation can be levied that it should analyse the EU's recent Data Act and Data Governance Act. Nonetheless, I maintain that my choice of material is justified. While the aforementioned pieces of EU legislation are all in some ways related to AIA and my thesis, they are mostly concerned with data privacy rights and voluntary data sharing. They do not address this thesis' emphasis on labour or AI – although the Data Act does regulate access to prosumer-generated data (European Commission, 2024). Perhaps most significantly, however, the European Parliament (2023a) explicitly claims that AIA is a "*comprehensive* AI law" (added emphasis). Relatedly, it refers to the "AI value chain" 17 times. Article 28 is even titled "Responsibilities along the AI value chain", strongly implying that AIA comprises a holistic approach to AI regulation. Given AIA's self-proclaimed comprehensiveness, it is

reasonable to examine how AIA addresses the exploitation of AI labour without having to reference, or rely on, separate EU policies. Thus, given its intrinsic interest, 'comprehensiveness', novelty, and potential for regulatory diffusion, I am confident that AIA is a relevant and valid case for analysing the role of labour in AI regulation.

Method

The main purpose of this thesis is to examine and explain the degree to which AIA addresses the exploitation of AI labour, a phenomenon explored in the Theory section. As such, although motivated by a normative ambition to identify – and thereby help rectify – exploitative injustices along the AI value chain, my core research purpose is fundamentally an explanatory one. In order to explain AIA's address of AI labour, I have developed three distinct independent variables: 1) AIA's conceptualisation of labour, 2) its conferment of rights, and 3) its anticipation of certain risks. These variables have been developed by engaging both with the theory and material most pertinent to this thesis. I operationalise my variables in a tripartite hypothesis as part of a hypothetico-deductive approach (Halperin and Heath, 2020:124):

> *AIA fails to adequately address 'AI labour' because*:
>
> *H1) ...AIA includes an insufficient conceptualisation of AI labour*
> *H2) ...few, if any, of the rights appearing in AIA pertain to 'AI labour' or 'AI labourers'*
> *H3) ...AIA legislates against prospective risks occurring late in the AI value chain*

In assessing how these hypotheses fare, they will help me solve the research problem: '*To what extent does AIA address the exploitation of human-performed 'AI labour'?*'. To test these hypotheses, I employ interpretivist means and a qualitative content analysis (QCA). As 6 and Bellamy (2011) note, social science research requires interpretation when coding, analysing, and drawing inferences from our data (p.17). Yet, different tools exist for doing so. As a "more interpretive form of analysis concerned with uncovering meanings, motives, and purposes in textual content" (Halperin and Heath, 2020:365), qualitative content analysis is a fruitful approach to conducting sound and reliable text analysis. Compared to its quantitative counterpart, QCA detects not just manifest meaning, but also "latent and more context-dependent meaning" (Schreier, 2014:6). Consequently, it is a useful choice for uncovering

how exactly AIA conceptualises labour, rights, and risk as well as what these variables *mean* within AIA. Put differently, QCA offers the tools for explaining AIA's address of AI labour by reference to the operationalised variables.

A strength of my approach is its openness to potential multicausality – where multiple of my independent variables may influence my dependent variable, namely AIA's address of AI labour. Similarly, it is also sensitive to equifinality whereby the outcome – a more or less adequate address of AI labour – can be achieved via multiple causal pathways (6 and Bellamy, 2011:121). Unfortunately, it can do little to account for the interactions *between* my independent variables, since I am only testing the independent variables' effect on the dependent variable. Moreover, my approach does not allow me to fully reject the causal significance of other potential factors. Nevertheless, such concerns are partially offset when considering that 2/3 of my independent variables are already derived from the material, and that the QCA's coding process is an iterative one. By continuously revisiting AIA and refining my operationalised variables accordingly, I limit the risk of overlooking other causally significant factors (Mayring, 2000:6). Certainly, my approach emphasises the academic virtue of parsimony by focusing only on the few most important variables (Halperin and Heath, 2020:130). Regardless, QCA is fundamentally about interpreting multifaceted data in a systematic and consistent manner (Schreier, 2014:2). Thus, to ensure reliable and valid results, I must propose coherent and transparent rules for conducting my research and coding the data.

Coding

Coding is the central instrument of QCA. It enables the systemic processing, organisation, and analysis of data. By reviewing and coding AIA according to a consistent set of rules and category definitions, QCA becomes a useful tool for drawing out the explicit and implicit themes and meanings embedded throughout AIA. It should allow me to obtain an overview of the most pertinent aspects of AIA as well as their relative importance vis-à-vis each other. Moreover, the approach is "flexible" insofar as it enables coding according to both data-driven and theory-driven categories – thus ensuring that my findings will be those most conducive and relevant to my research topic (Schreier, 2014:2-3; also Thornberg and Charmaz, 2014:11). Once the coding process is complete, the findings should allow me "to draw conclusions and generalizations by linking the data back to the research question" (Halperin and Heath, 2020:383). Put differently, an analysis of the data should provide

indications as to how well my hypotheses fare and offer insights as to how AIA addresses the exploitation of AI labour.

Due to time constraints, I need to select which parts of AIA to analyse before I begin the actual coding process. To do so, I employ NVivo software to search for the frequency of certain 'word clusters'. These 'clusters' correspond to my three independent variables: labour, rights, and risk. Each cluster (except 'rights') has a number of closely associated word roots whose frequencies I cross-reference with the Preamble, Titles, and Annexes of AIA.

Table 1: Frequency of word 'clusters' throughout AIA

	Preamble	Title I	Title II	Title III	Title IV	Title V	Title VI	Title VII	Title VIII	Title VIIIa	Title IX	Title X	Title XI	Title XII	Annexes	Total
Risk	**318**	**13**	**4**	**276**	**0**	**14**	**5**	**3**	**60**	**34**	**1**	**6**	**0**	**5**	**25**	**764**
Risk*	311	9	2	271	0	13	5	3	60	34	1	5	0	5	25	744
Danger(/s)	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Threat(/s)	5	3	2	2	0	1	0	0	0	0	0	0	0	0	0	13
Jeopard*	0	1	0	2	0	0	0	0	0	0	0	1	0	0	0	4
Hazard(/s)	2	0	0	1	0	0	0	0	0	0	0	0	0	0	0	3
Rights	**134**	**7**	**2**	**27**	**2**	**9**	**4**	**0**	**16**	**3**	**0**	**3**	**0**	**1**	**2**	**210**
Right(/s)	134	7	2	27	2	9	4	0	16	3	0	3	0	1	2	210
Labour	**55**	**3**	**1**	**9**	**3**	**0**	**1**	**0**	**1**	**0**	**0**	**0**	**0**	**0**	**8**	**81**
Labo#r	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2
Work*	28	0	0	2	2	0	1	0	1	0	0	0	0	0	2	36
Worker*	9	2	0	3	0	0	0	0	0	0	0	0	0	0	1	15
Employ*	10	1	0	1	0	0	0	0	0	0	0	0	0	0	2	14
Job	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	2
Occupation	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Moderat*	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Annotat*	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1
Label*	6	0	1	2	1	0	0	0	0	0	0	0	0	0	1	11

Table 1 – Overview of word 'clusters' and their frequency throughout AIA

While there is some subjective choice in the development of these clusters, I do not believe they skewer the final coding results, since their sole purpose is to determine the most relevant parts of AIA to analyse – they do not influence the coding itself. Certainly, the clusters clearly point to some Titles as more substantially pertinent than others. The Titles included in my analysis are boldly outlined in *Table 1*. While they may seem to constitute a small part of AIA, they are in fact some of its most extensive segments – constituting a total 74.54% of AIA's entire legal text (with Annexes constituting 7.85%). Thus, in quantitative terms, I analyse the vast majority of AIA.

The most important decision in terms of what to analyse is whether or not to include the Preamble. Arguably, analysing the Preamble may skewer the results by having the same subcategory referred to in a duplicate manner. For example, both the Preamble and the enacting terms discuss 'systemic risk', although only the enacting terms do so with regulatory authority. Nonetheless, I opt to include the Preamble in my analysis not only due to its extensiveness (40.8% of the entire text) but also since it reveals the intended applications and ambitions of AIA. As such, the Preamble offers direct insights as to how and whether AIA addresses AI labour.

During the coding process, I follow a set of consistent rules in order to ensure the reliability and validity of my findings. First, I set rules for deciding *what* to code. In prioritising replicability and intercoder reliability (Halperin and Heath, 2020:192), I opt for formal, objective criteria when determining my unit of coding. That is, I draw "on the inherent structure of the material" (Schreier, 2014:15). Here, AIA's legal paragraphs constitute the basic unit of coding. If a paragraph has subparagraphs and/or points, these become these the basic units instead. Nevertheless, the remainder of the paragraph which does not belong to any subparagraph may still be coded as a distinct unit. This similarly applies to the Preamble where recitals represent the coding units. My objective coding rules enable me to capture the distinct meanings of AIA's most specific legal provisions whilst maximising the replicability of my approach. Moreover, in order to avoid skewing the findings with 'irrelevant' data, I specify that not every unit needs to be coded. On the other hand, a single unit can be coded up to three times, but never more than once within the same main coding category. This is to uphold the principle of mutual exclusiveness among my subcategories and to ensure their construct validity (Schreier, 2014:9-10; 6 and Bellamy, 2011:26). If tempted to code a unit twice within the same main category, the rule of mutual exclusiveness prompts me to re-evaluate the validity of my subcategories and reassess whether they remain accurate, mutually exclusive measures of specific concepts.

When choosing to code a unit under one of the main categories' numerous subcategories, certain rules apply. The three main categories correspond to the three independent variables: 'labour', 'rights', and 'risk'. To guarantee unidimensionality, there should be no overlap between the main categories and no subcategory should belong to multiple main categories. The rules for coding within each of the main categories are presented in *Table 2*.

Table 2: Coding rules for main categories

Labour	Rights	Risk
Code within 'labour' if the Regulation is describing, discussing, or referencing human-performed labour/work in relation to AI – either as something affected by AI systems or as necessary for the production, maintenance, and/or development of AI systems. Labour is here defined according to this thesis' notion of "AI labour" whilst also including any explicit reference to 'labour' and/or the nouns 'work' or 'worker'.	Code within 'rights' if the Regulation is describing, discussing, or referencing any type of right conferred by the Regulation or when the Regulation describes, discusses, or references any type of right addressed and upheld by it. Code also when the Regulation describes, discusses, or references fundamental rights in general or any series of more or less specified rights. Rights are here understood both according to their legal and moral interpretations and may pertain to any type of substantive or procedural entitlement.	Code within 'risk' if the Regulation is describing, discussing, or referencing the risks posed by AI, whether they are real, perceived, imagined, unclear, or unknown. Risk is here defined as the potentiality of an unwanted occurrence such as, but not exclusively, harm to humans and non-humans as well as harm to natural, digital, and built environments. Code also within 'risk' when the Regulation references 'serious incidents' – either as something that has already occurred, is suspected to have occurred, or might occur in the future.

Table 2: Coding rules for main categories

Once I decide to code within a main category, I then follow extensive rules determining which subcategory to code for (see *Appendices 1-3*). Complementing the rules in *Appendices 1-3*, it should also be stated that where a unit may be coded for 'unclear' risk/'fundamental rights in general' or a more specific subcategory, then I always code for the specific subcategory. If all other rules fail to produce a clear coding result – and conflicting coding options remain – then the unit is to be coded according to a subjective evaluation of the most applicable subcategory. These specific rules for when to code, and when not to, bolster the reliability of my approach, as well as the content and conceptual validity of the subsequent findings (6 and Bellamy, 2011:96). Furthermore, thanks to the flexibility of QCA and its iterative coding process, these subcategories and coding rules have been continuously revised so as to ensure they accurately reflect the data. For example, the final list of subcategories (see *Table 2*) includes codes that did not exist prior to engaging with the material. On the other hand, numerous 'labour' subcategories have been omitted from the overview of findings simply because no units were coded accordingly – e.g., 'playbour' and 'data moderation'.

Upon completing the initial coding process and organising my findings, certain subcategories were merged to ensure greater replicability and intercoder reliability. Whereas units were originally coded into the separate subcategories 'procedural rights of individuals, 'right to explanation', or 'right to complain', these were later merged as 'procedural rights of persons'. All three subcategories pertain to the same 'type' of right, namely the procedural rights of individual persons. Since further specificity does not contribute to this thesis' focus, these subcategories were grouped as one to ensure, in practice, greater intra-researcher replicability. For similar reasons, the subcategories of 'unspecified risk' and 'general risk' were also combined post-coding. Both denote a lack of specificity, but their distinction can be unnecessarily difficult to ascertain. Lastly, during the second round of coding, it was discovered that a set of risks – the respective risks of harm to critical infrastructure, property, environment, and humans – are grouped in Article 3(44) as "serious incidents". Following the conceptual merger of these disparate risks, I replaced a range of subcategories with the single 'serious incidents' subcategory – so as to better capture AIA's references to these risk-types in a final round of coding. Merging and reconfiguring subcategories helped boost replicability without compromising the validity or qualitative insights of my initial coding approach.

Once I obtained an overview of the final codes, I could study their qualitative nuances, draw qualified inferences, and offer examples of how they each shape AIA's address of AI labour.

Trade-offs and limitations

Compared to quantitative content analysis, my choice of qualitative content analysis favours measurement validity over reliability. Allowing for a better detection of implicit meanings and context-dependent themes (Schreier, 2014:6), QCA enables me to assess whether any given unit accurately reflects an operationalised coding category. My categories represent largely multivalent concepts meaning that a qualitative interpretation of the text – according to clear coding rules – is the optimal method for ensuring measurement validity. Elsewhere, however, I prioritise reliability over validity, such as when utilising formal criteria for distinguishing coding units. Replicating this practice should be highly reliable, yet, "[s]ometimes a paragraph embraces too many ideas for there to be consistent assignment of the text segment to a single [sub]category" (Halperin and Heath, 2020:379); the coding of some units is less accurate, or 'valid', than others. Despite extensive coding rules to offset this drawback, it did indeed prove an issue on occasion. Lastly, I also prioritise reliability over validity when stipulating that subcategories within the risk/rights categories should feature an explicit reference to such risk/rights. This limits subjective bias and bolsters replicability, but entails the possible omission of otherwise valid coding instances. Without the requirement of explicit reference, I may have registered more risk/rights codes. However, I do not believe this methodological risk has materialised.

Another trade-off is between internal and external validity. Prioritising the former, this thesis is restricted to the single case of AIA with its internal complexity and variables; I do not seek to extrapolate and generalise my findings to other policies (Halperin and Heath, 2020:192). In fact, given the lack of comparable legislation, the scope for generalising my findings to external contexts is considerably constrained, rendering such endeavours futile. Nonetheless, this thesis may offer a model for analysing future AI regulations.

Finally, a potential limitation of this thesis is the risk of circularity. That is, a potentially unclear distinction between my dependent and independent variables. For example, it remains hypothetically conceivable that how AIA addresses AI labour has a causal bearing on how it applies concepts of labour, rights, and risk – in effect, inversing my variables. However, I consider it logically sound and prudent to assume that AIA policy-makers decided to proceed with a risk-based, rights-oriented approach prior to any subsequent address of AI labour. Moreover, for the purpose of this thesis, demonstrating a

relationship of "association" between my variables is sufficient to produce valuable findings (Halperin and Heath, 2020:142-143). As long as this thesis demonstrates a strong relationship between the dependent and independent variables, then it becomes clear that AIA's legal conceptualisation and operationalisation of labour, rights, and risk is of significance to its address of AI labour. Nonetheless, as mentioned, I cannot fully demonstrate causality since I cannot exclude other potentially causal factors. Similarly, my research design lacks the methodological means for verifying the particular mechanism by which my independent variables *cause* change in AIA's address of AI labour (*ibid.*). Rather, as a possible *indication* of causality, it explores whether a relationship exists. So, does it?

Analysis

This section presents the findings of this thesis before analysing the results and discussing the performance of my tripartite hypothesis.

Findings

The findings, which largely reflect the dual-use of data-driven and theory-driven coding, show an overwhelming reference to risk and rights (data-driven) and an underwhelming feature of labour subcategories (theory-driven). In fact, the 'labour' category includes only two subcategories, each with merely three coding counts. The only references to labour are in terms of preparing metadata and the workplace monitoring of employees. By contrast, the 'risk' category reveals AIA's particular concern with the risks of 'discrimination', 'serious incidents', 'lack of 'human control and oversight', and 'systemic risk'. Meanwhile, AIA refers to a broad range of 'rights', not least 'privacy rights', the 'procedural rights of persons', and the 'right to non-discrimination'. Yet, for both 'risk' and rights', the largest subcategories were less specific – respectively coded as 'unclear' (risk) and 'fundamental rights in general'. Let us discuss the results and their implications in more detail.

Table 3: Overview of final coding results			
Risk	***N*= (Preamble)**	***N*= (Enacting terms)**	***N*= (Total)**
Unclear	27	26	53
Dependency	0	1	1
Cybersecurity	3	0	3
Deception and misinformation	3	0	3
Unknown risk	1	2	3
Privacy	0	9	9
Human control and oversight	2	15	17
Serious incidents	1	17	18
Discrimination	9	9	18
Systemic risk	15	29	44
Rights	***N*= (Preamble)**	***N*= (Enacting terms)**	***N*= (Total)**
Freedom of expression	1	0	1
Procedural rights of persons	4	6	10
Procedural rights of operators	1	2	3
Workers' rights	3	1	4
Intellectual Property Rights (IPR)	4	2	6
Privacy rights	7	1	8
Right to non-discrimination	7	1	8
Fundamental rights in general	27	28	55
Labour	***N*= (Preamble)**	***N*= (Enacting terms)**	***N*= (Total)**
Data preparation	1	2	3
Workplace management	2	1	3

Table 3 – Overview of final coding results

Discussion

The following discussion is divided into three parts, each of which discusses a component of my tripartite hypothesis.

An insufficient conceptualisation of labour

It is immediately clear from the results that 'labour' is a vastly underemployed concept vis-à-vis the data-driven categories of 'risk' and 'rights'. Considering the numerous types of labour discussed in the Theory section, there is a remarkable absence of their recognition in AIA. Most significantly, however, is the lack of acknowledgement for data-generating AI labour – despite the three counts for 'data preparation'. Whilst the term 'data' appears 366 times throughout the entirety of AIA, there seems to be little to no recognition of where such metadata stems from. The few instances of 'labour' codes pertain 1) to management *of* labour in the workplace, and 2) to the preparation/annotation of (meta)data. In the case of 'data preparation' – a central aspect of AI labour – all three codes fail to reference or acknowledge the human labour(er). That is, by way of nominalisation or passive voice, AIA fails to mention *who* actually produces and prepares the metadata. Article 10(2-c), for example, concerns "data preparation processing operations, such as annotation, labelling, cleaning, updating, enrichment and aggregation". In other words, the 'meaningful' aspect of metadata – conferred by 'something human' – is wholly ignored. The 'labour' codes, then, feature a conspicuous, paradoxical disregard for the actual labour performed. This serves to reproduce

dataism's naturalising discourse of data as a raw material 'just there', and obscures the human contributions to surplus-value objectified within metadata.

The 'workplace management' subcategory includes the only explicit references to labour (or rather, 'work'). Here, AIA is chiefly concerned with how AI manages workers and productivity in the workplace. Thus, AIA is not concerned with the labour that goes *into* AI, but rather that which is subjected to AI-facilitated digital Taylorism (Thompson, 2020:302). Yet, even here, AIA makes it clear that "in the context of employment and protection of workers, this Regulation should […] not affect Union law on social policy and national labour law" (Recital 5a). AIA is explicitly *not* seeking to reaffirm workers' rights with the advent of AI. Moreover, AIA demonstrates a distinct bias: considering only labour/work as that which is employed. For instance, Article 2(5) states that member states may initiate legislation that protects workers' rights vis-á-vis their *employers*. As such, AIA entirely neglects forms of immaterial labour – such as the 'playbour' and 'prosumerism' of data colonialism.

Nevertheless, as Cefaliello and Kullmann (2022) insist, "[h]igh quality data appears to be the main concern of the AI Act" (p.557). Indeed, numerous passages (e.g., Recital 44) point to the importance of high quality data to safe, fair, and effective AI systems. Yet, AIA fails to appreciate that quality data is dependent on human-performed labour. Furthermore, the charge can be levied that AIA's demands for quality data amount to a tacit acceptance, or even incentivisation, of AI empire's exploitative practices. By demanding data without protecting AI labour, AIA risks facilitating the extreme appropriation of surplus-labour – labour which remains hidden, unacknowledged, and unconceptualised.

In sum, it seems we cannot reject the hypothesis that *AIA fails to adequately address 'AI labour' because AIA includes an insufficient conceptualisation of AI labour*. An analysis of the results clearly finds that there is no basis for a conceptualisation of AI labour within AIA. Whether the 'playbour' and 'prosumerism' of data colonialism or the annotation and moderation of metadata, there is no acknowledgement of the human-performed labour integral to AI empire. Furthermore, AIA risks tacitly incentivising the continued exploitation of AI labour by demanding high quality data – as though metadata exists 'just there', independent of its production by humans' surplus-labour.

AI labourers lack rights

The results show that most of the rights coded for in AI pertain to negative, substantive claim-rights such as freedom of expression, intellectual property rights, and especially the rights to privacy and non-discrimination. Unfortunately, the largest subcategory, 'fundamental rights in general' is of little analytical value – referring mostly to broad sets of values, which may entail both positive and negative claim-rights. Recital 5, for instance, requests the protection of "fundamental rights, including democracy, rule of law and environmental protection". Still, besides negative claim-rights and 'general' fundamental rights, AIA also enshrines certain positive claim-rights such as the procedural rights of citizens and operators as well as workers' rights. Yet, as in the 'labour' category, 'workers' rights' codes pertain solely to employed individuals. And the only code featured in the enacting terms, merely states that AIA should not hinder national labour laws (Article 1(5e)). Thus, AI labour rights appear wholly non-existent in AIA. A decolonisation of AI empire and the protection of AI labour would, arguably, require AIA to enshrine a set of positive, substantive claim-rights which confer duties on those firms who exploit surplus-labour for capturing, extracting, and producing metadata. Instead, we see a mix of negative substantive rights, positive yet procedural claim-rights, and a few references to workers' rights which are entirely lacking in substance.

Arguably, AIA fails to confer AI labour with legal, positive, and substantive claim-rights because the EU lacks jurisdiction. Certainly, all the rights coded for (except 'fundamental rights in general') can feasibly be upheld by the ECJ. The EU legal system can mandate rights for workers employed in the EU[2], it can guarantee the protection of procedural and intellectual property rights, and it can sanction firms who deploy AI systems in a manner jeopardising the freedom of expression and rights to privacy and non-discrimination. By contrast, legislating against the exploitation of labour across the global AI value chain is a much trickier endeavour. AIA explicitly recognises that its juridical scope is restricted to "persons within the Union" (Article 2(1cc)). In line with both the 'sovereigntist' and 'pluralist' approaches, it appears that the EU lacks the sovereign, judicial authority to sufficiently uphold legal claim-rights abroad. If we accept that rights function to protect right-

[2] Indeed, the EU has mandated workers' rights on many occasions, such as the "Directives to protect workers' health and safety (Directive 89/391/EEC and related Directives); the Directive on the consultation and information of workers and their representatives (e.g., Directives 2009/38/EC, 2002/14/EC and 2003/72/EC); and the right of freedom of association, collective bargaining and collective action (Art. 12 and Art. 28 EUCFR)" (Cefaliello & Kullmann 2022:544)

holders, then we see that the list of coded subcategories makes perfect sense: whether citizen, employed worker, IP-owner, or AI operator, each of these right-holders can claim protection by the Union's legal authority. Conversely, an AI labourer located beyond the Union cannot feasibly rely on the EU to protect his interests or will – regardless of whether he appeals to a legal or moral right. From this perspective, AIA cannot confer AI labourers with rights nor their exploiters with corresponding duties. Likewise, with risk-based regulation, rights may rank second to the viable enforcement of legislation. As discussed earlier, rather than prioritising the 'essence' of rights, risk-based approaches turn to risk as an 'organising concept' – where rights can be compromised and balanced against other practical considerations. Here, it appears that AIA only enshrines those rights which it can reliably enforce and judicially uphold.

EU's restricted jurisdiction, however, does not fully vindicate AIA's lack of AI labour rights. The EU could leverage its market power and capitalise on the Brussels effect to "influence global governance" and incentivise "non-EU companies" (Bas et al, 2024:56) to abide by a set of standards as a condition for acquiring single market access. AIA already imposes pre-market demands on would-be deployers to, for instance, outline the potential risks and intended use of their systems as well as document how they will mitigate such risks and respect Union copyright law (Friedl and Gasiola, 2024:para.10). Moreover, Bas et al (2024) contend that the EU's new AI Office will facilitate the "enforcement of the AIA overseas" (p.60), thereby suggesting a scope for the externalisation of AIA's stipulations. Thus, the potential exists for AIA to place requirements on how AI firms source data and appropriate surplus-labour along the global value chain. The EU could externalise the enforcement of AI labour rights by withholding market access – as it has done in other regards. Yet, whether such requirements would be considered legal or moral rights, or even rights at all, remains a question of interpretation. For the protection of these 'rights' would depend entirely on the good-will of firms seeking EU market access. While it can be posited that firms always have a *moral* duty to respect the moral claim-rights of AI labourers, such firms would not have a *legal* duty to do so: they could simply relinquish access to the EU market and thereby free themselves of any legal compulsion to respect AI labour rights. Ultimately, we are faced with a dual question of value. Firstly, insofar as the justification for rights is the 'final value' of the right-holder, it appears AIA has tacitly deemed AI labourers unworthy of rights – or perhaps their worth has been overlooked. While other right-holders have been protected, no rights have been extended to AI labourers. Secondly, even if AIA sought to enshrine rights for non-EU-based AI labourers, it could only do so if firms deemed

the EU market valuable enough to respect AI labour rights and accept the moral and/or legal duties consequently placed upon themselves. In this view, AI labourers are deprived of rights because it is not 'worth it'.

The most significant finding within the 'rights' category is the fact that nearly every coded right pertains to the later stages of the AI value chain. Exceptions to this claim include the 'IPR' subcategory, which primarily concerns the use of copyrighted data for training AI systems (e.g., Recital 83), and the "procedural rights of operators" subcategory where all three codes simply reference operators' rights and duties in EU regulation 2019/1020 on market surveillance and compliance. All other subcategories (except 'fundamental rights in general') reflect either 1) substantive rights deemed *at risk* from AI systems, or 2) procedural rights which provide recourse to individuals when they fear that their substantive rights are *at risk*. For example, Recital 70b affirms that AI-generated content should not hamper 'freedom of expression', and Recital 37 concerns the right to not be discriminated against for public services on the basis of AI-outputs. Meanwhile, Articles 68a-c grant individuals the procedural rights to complain and to receive explanations about AI systems affecting them. All these rights are enshrined in line with van Dijk et al's (2016) notion of 'rights at risk'. They are responses to 'perceptions of threat' – as though AIA considers certain rights *at risk* of violation by AI. Therefore, the 'rights' category predominantly reflects rights protecting individuals from fully-developed AI models. By contrast, AIA offers no protections against the risks of unwanted events occurring prior to the deployment of AI systems. AIA entirely overlooks the rights of AI labourers whose location in the AI value chain means they are not put *at risk* by AI.

Furthermore, AIA's risk-based approach functions as an 'obligation adjuster' – delineating the duties to mitigate "the risks for health, safety and fundamental rights" posed by AI systems (Recital 43). But the allocation of duties mirrors the allocation of rights. For example, Article 64(3) outlines the duty of "national public authorities" to uphold the right to non-discrimination in relation to high-risk AI systems. Here, the *duty* of national bodies to prevent AI-facilitated discrimination implies, and reifies, citizens' *right* to non-discrimination. Yet, it appears that since AI labourers are not considered *at risk* by already-deployed AI, then AIA does not confer any duties upon firms to protect labourers from potential harm or exploitation. Accordingly, AI labourers – whether the 'prosumers' and 'playbourers' of data colonialism or the annotators and moderators of wider AI empire – are disregarded as potential right-holders.

In sum, I argue that my second hypothesis cannot be rejected: *AIA fails to adequately address 'AI labour' because few, if any, of the rights appearing in AIA pertain to 'AI labour' or 'AI labourers'*. Certainly, the results show that not a single 'rights' code pertains to AI labour. Instead, AIA enshrines the negative, substantive rights or the positive, procedural rights of EU-based legal persons whose entitlements are considered *at risk* of violation by already-deployed AI systems. As such, AIA fails to empower labourers and decolonise AI empire. As it does in other regards, AIA could potentially have leveraged its market power to demand greater protections for AI labourers beyond the EU. Yet, it seems their location along the value chain renders labourers 'out of sight, out of mind' for policy-makers; or worse still, labourers are simply not considered worth protecting.

Forward-looking risks overlook earlier harms

If the experience of risk leads to either "denial, apathy or transformation" (Beck, 2006:331), it does not seem that the riskiness of AI labour has induced legislative transformation in AIA. Within the 'risk' category, by far the largest, not a single code refers to risks faced by AI labourers. From data colonialism's exploitation of 'prosumers' and 'playbourers' to the slave wages of annotators and the traumatic experiences of content moderators, AIA fails to include mention of any of these risks. Instead, all the coded subcategories – with the exception of 'unclear' risk – pertain specifically to the risks of deployed or soon-to-be deployed AI systems.

The most coded subcategories include the risk of 'discrimination' by AI, insufficient 'human control and oversight' of AI systems, the risk of 'serious incidents', and the 'systemic risk' posed by new General Purpose AI models (GPAI). Article 52a(1) determines 'systemic risk' to be an attribute of any GPAI with "high impact capabilities". In turn, 'high impact capabilities' are defined as when a GPAI's computing power exceeds 10^{25} FLOPs. "Virtually unknown at the time of the Commission's original proposal, and thus fully absent from it" (Friedl and Gasiola, 2024:para.8), GPAI models like ChatGPT are typically understood as demonstrating a capacity for iterative self-improvement (Arora et al, 2023:2) or emulation of human cognition (Apostolakoudis, 2022:216). Owing to the myriad data on which they are trained, and their lack of a single intended use-purpose, a defining characteristic of GPAIs is their 'systemic risk' – having "reasonably foreseeable negative effects on public health, safety, public security, fundamental rights, or the society as a whole, that can be propagated at scale" (Article 3(44d)). For this reason, amendments to AIA ensured that the provisional

agreement features extensive regulation of systemically risky GPAIs – being the most frequently coded subcategory of 'risk' with 44 counts in total. Beyond defining and outlining the conditions for the classification of GPAIs with systemic risk (e.g., Article 52a), the vast majority of 'systemic risk' codes refer to the pre-deployment interfaces between GPAI providers and the European Commission. For example, in order to acquire insights to the internal testing and systemic risk mitigation of GPAIs, Article 68j(7) sketches the scope for "structured dialogue" between the AI office and system providers. Similarly, Article 52b(1) mandates that the Commission should be informed once AI providers become aware that their models will surpass the 10^{25} FLOPs threshold. Thus, 'systemic risk' codes pertain mainly to pre-deployment regulation of AI.

In regulating the risks of 'discrimination' and 'human control and oversight', AIA maintains a focus on pre-deployment measures. For instance, although disregarding how data is sourced, AIA's Article 10(3) mandates that high-risk AI models should be trained on valid and representative datasets to prevent discriminatory outputs. Relatedly, Article 29a (1c-d) stipulates that AI-deployers must conduct pre-market impact assessments for national market authorities, including an evaluation of a system's potential discrimination risks. Subparagraph 1e similarly demands that these impact assessments outline the measures taken to ensure human oversight. Article 14, titled "Human oversight", also details a range of pre-deployment and post-deployment measures taken to ensure human control, such as the AI system being designed with "appropriate human-machine interface tools" (paragraph 1) and an override "stop"-button. Furthermore, it dictates that human controllers must be able to "properly understand" and "correctly interpret" AI outputs (paragraph 4). Paragraph 5 adds that no action can be taken on the basis of high-risk systems' conclusions unless seperately verified by "two natural persons with the necessary competence, training and authority". My point here is that these main subcategories of risk – 'systemic risk', 'discrimination', and 'human control and oversight' – are all premised on the notion of 'prospective responsibility'; the codes pertain to the responsibility of deployers and providers to mitigate the prospective risks of AI systems upon their deployment.

The final major 'risk' subcategory, 'serious incidents', focuses less on pre-deployment risk mitigation and more on the "post-market monitoring" of potential harms (Recital 78). For example, most 'serious incidents' codes stem from Title VIII, Chapter 3: "Sharing of information on serious incidents". Here, AIA prescribes how AI operators and providers must notify national authorities whenever a serious incident is believed to have occurred – that is, whenever a deployed AI system has caused harm to humans,

infrastructure, property, or the environment. Other codes within the subcategory specify the competences of national authorities in the case of a serious incident (Article 63b(3)), and outline the responsibilities of providers when testing AIs (Article 54a(6)). Meanwhile, Article 58(e-ii) concerns the review of serious incidents by a new AI Board and Article 3(44) simply defines what constitutes a 'serious incident'. Like Chapter 3, all the remaining 'serious incidents' codes pertain to the notification of relevant authorities. Either way, the 'serious incidents' subcategory is chiefly pre-occupied with post-deployment risks and harms. AIA's address of 'serious incidents' concerns only the final stages of the AI value chain. But although the codes stipulate how to handle events suspected to have already transpired, AIA's conceptualisation of 'serious incidents' risks is fundamentally prospective – premised entirely on the perceived dangers of functional AIs. Yet, there is a principal difference between 'serious incidents' and other 'risk' subcategories: whilst other codes reflect a pre-emptive approach to risk management, 'serious incidents' codes allocate 'retrospective responsibility' for when prospective risks have materialised into actual incidents. Still, despite 'serious incidents' containing an element of retroactivity, not one single 'serious incidents' code relates to risks occurring prior to the testing or deployment of AI models. AIA offers a regulatory action plan for when post-market AI models cause 'serious incidents'. However, the subcategory features no action plan for unwanted events occurring earlier in the value chain. As such, it does little to decolonise the hidden, exploited labour of data colonialism and AI empire.

Like all the above, the remaining 'risk' subcategories are preoccupied with the risks engendered by complete or near-complete AI systems. For example, 'privacy' codes summarise when and how providers can access personal data so as to detect and correct bias in AI systems (e.g., Article 10(5)). Moreover, the 'cybersecurity' codes outline the responsibilities of providers and the European Commission as regards safeguarding AI models from 'data poisoning' and behaviour modification by malicious actors (Recitals 51, 51a, and 69). In terms of 'deception and misinformation', AIA mandates that outputs be clearly watermarked as AI-generated. Other "specific transparency obligations" apply as well, such as notifying humans when they are interacting with AI. This is to prevent "risks of impersonation", "misinformation", and "manipulation" facilitated by generative AI and chatbots (Recitals 70-70b). Even the lone 'dependency' code, Article 7(2e), concerns the risk posed by fully-functional AIs – specifically, the inability of "potentially harmed" persons to escape the effects of AI decisions. Perhaps most tellingly, however, is the fact that every 'unknown risk' code – indeterminate by definition – nevertheless relates to "the possible risks

emerging *from* AI systems" (Recital 78, added emphasis). Despite being 'unknown', it is still assumed that risks can only arise from fully-developed and already-deployed AI systems. Thus, while none of the 'risk' subcategories acknowledge data colonialism's extreme exploitation of surplus-labour – nor the precarity of workers under AI empire – AIA *does* include provisions for risks that remain unknown and yet to be imagined.

Considering all the above, it appears AIA's risk-based approach is premised on 'prospective responsibility' and a fundamentally forward-looking conceptualisation of risk. Each and every 'risk' subcategory reflects AIA's attempt to mitigate the risks posed by AI systems late in the value chain. By contrast, no subcategory refers to the risk of unwanted events occurring *before* the deployment of somewhat complete AI models. A minor, partial exception to this discovery is the 'serious incidents' subcategory which contains elements of 'retrospective responsibility' rather than pre-emptive risk management. Yet, codes in this subcategory concern the post-market monitoring of deployed AI systems and whether they have caused unwanted risks to materialise into actual incidents. Thus, while some retrospective, backward-looking provisions can be identified in AIA, the scope of such risk-based regulation remains confined to late, post-market 'links' in the AI value chain. This makes perfect sense when considering that AIA's risk-based approach categorises the 'riskiness' of AI systems according to their 'intended purpose' (Article 7(2a)) – that is, the 'purpose' of functional, fully-developed AIs. AIA's categorisation of risk is thus fundamentally predicated on the prospective, 'intended' deployment of AIs; AIA does not conceptualise risks according to the retrospective, 'unintended' externalities of early-stage AI development. Consequently, risk-based regulatory action takes effect long after AI labourers have encountered the risks saturating global AI empire.

Another reason for AIA's regulatory oversight may be dataism's naturalising discourse on AI and data. Insofar as AI and data are conceptualised as immaterial, AIA risks neglecting the real, lived effects of AI along its value chain – especially as regards the material, human labour integral to AI's metadata. For when AI and data are viewed as abstract, they escape "traditional understandings" of risk (*ibid.*) – as though a lack of materiality entails a lack of consequentiality. As such, by only acknowledging the materiality of AI systems once they become finished, tangible products, AIA fails to conceive of material risks arising much earlier in the AI lifecycle. This may partially explain the many forward-looking, late-stage subcategories of 'risk' coded for in AIA.

As Beck (1992) reminds us, a set of socio-political consequences are at stake in how we define 'risk'. Similarly, I noted earlier Arora et al's (2023) claim that "designating

riskiness to technology is ultimately a creative act" (p.4). The results of my QCA, meanwhile, reveal that AIA's construction of risk is premised entirely on the abilities of already-developed AI systems. The 'creative act' here is imagining the potential harms that AIs *could* cause in the future. Even when amending AIA to incorporate the advent of GPAIs, the primary creative act was to conceptualise the prospect of 'systemic risk'. The creativity of designating riskiness is exclusively oriented towards the unrealised potentialities of the future. There is no scrutiny of AI empire's early-stage maltreatments. Consequently, AI labourers appear the undisputed 'losers' of AIA's conceptualisation of risk. Moreover, insofar as AIA fails to conceptualise earlier risks, it may actually *exacerbate* the exploitation and trauma of AI labour. The "irony of risk" (Beck, 2006:330) is that the more you deny it, the more it becomes reality. Indeed, if AIA "presupposes the legality" (Macenaite, 2017:533-534) and justness of that which comes prior to the prospective risks coded for in this thesis, then AIA arguably serves to legitimise and thereby facilitate AI empire's continued colonisation of 'prosumers', 'playbourers', annotators, moderators, and other AI labourers. If AIA claims to be a 'comprehensive AI law' – one that addresses all "[r]esponsibilities along the AI value chain" (Article 28) – then its neglect of AI labour may be interpreted as a tacit endorsement of AI empire's exploitative practices. Certainly, there is a 'duality of risk': as AIA legislates against prospective risks posed *by* AI, it enables the perpetuation of risks which go *into* AI – such as the extreme appropriation of surplus-value and the traumatic moderation of data.

In sum, I cannot reject the hypothesis that *AIA fails to adequately address 'AI labour' because AIA legislates against prospective risks occurring late in the AI value chain*. Not a single 'risk' subcategory addresses the risks faced by AI labourers. Instead, all subcategories coded for in my thesis pertain to the prospective risks posed by tangible and already-deployed AI models – i.e., those risks derived from the 'intended' use purposes of fully-developed AI systems. The emphasis on prospective risk is likely also owed to the naturalising discourse surrounding metadata and AI – framing them as 'immaterial' and inconsequential. The designation of 'riskiness' seems reserved for AI systems in their final, most 'material' stages. Consequently, there is no retrospective address of the risks and harms faced by AI labour prior to the deployment of functioning AI models. Even the 'serious incidents' subcategory, which contains some retrospective elements, is fundamentally predicated on the prospective risks of post-market AIs. All in all, it seems AIA does little to decolonise the wider risks of global AI empire. If anything, AIA implicitly legitimises empire's exploitation of labour by

presupposing that AI-risks occur only late in the value chain – as though everything prior has proven just, safe, and lawful.

Conclusion

This thesis has endeavoured to disentangle an important puzzle: for how can AIA's forward-looking, risk-based approach adopt a retrospective address of AI labour and the exploitation it experiences at the early stages of the AI value chain? Unfortunately, I find that AIA features an entirely absent address of AI labour. Here, the term 'unfortunately' reveals a certain normative standpoint – suggesting AIA could and *should* contribute to the decolonisation of global AI empire. Indeed, in adopting a rights-oriented and risk-based regulatory approach, AIA held the potential to enshrine the rights of AI labourers and protect them from exploitation. Yet, AIA neglects not only the rights of AI labourers, but also the risks they face. In fact, there is no conceptualisation of AI labour at all. Consequently, my tripartite hypothesis must stand and cannot, at present, be rejected. In answering my research problem, I conclude that *the extent to which AIA addresses the exploitation of human-performed 'AI labour' is wholly inadequate*.

In order to decolonise the hidden labour of global AI empire, labourers must be protected from its risks and harms. By conducting a qualitative content analysis of AIA, I ultimately find that such protections are lacking. Despite 'rights' codes reflecting both negative, substantive rights and positive, procedural rights, AIA contains no provisions for protecting AI labourers. Rather, AIA enshrines those rights specifically considered *at risk* from fully-developed AI systems – i.e. those rights and privileges which are put in jeopardy late in the AI value chain. Moreover, right-holders are explicitly legal "persons within the Union". While this makes judicial sense, this thesis has noted that the EU could leverage its market power to externalise AIA's stipulations abroad – as it does in other regards. Still, another reason for AIA's lack of AI labour rights is undoubtedly its lacking conceptualisation of AI labour. The only theory-driven category of my QCA, 'labour' is also the most underrepresented. Containing only two subcategories, none actually focus on AI labour. Whether the exploited 'playbour' and 'prosumerism' of data colonialism or the precarious contract work of annotating and moderating metadata, AIA contains no recognition of such AI labour. Furthermore, by demanding high quality data for the training of AI models, AIA may even incentivise the continued exploitation of surplus-labour. In fact, when analysing the 'risk' category, it can be argued that AIA inadvertently endorses AI empire. Claiming to be

the "world's first comprehensive AI law" – referencing the AI "value chain" 17 times – AIA nevertheless focuses solely on the chain's final links. Certainly, its risk-based approach pertains exclusively to the prospective risks of already-deployed AI systems. As such, AIA entirely overlooks AI empire's broader risks and harms – implying that such unwanted events do not merit regulatory action. Thus, as regards my three research questions, I contend that AIA 1) features zero conceptualisation of AI labour; 2) protects only those rights deemed *at risk* by fully-developed AI systems; and 3) legislates solely against the prospective risks of functional AI models late in the value chain. Consequently, I deem AIA's address of AI labour – and its decolonisation of AI empire – inadequate.

The first of its kind, AIA sets a unique precedence for the regulation of AI – both in Europe and internationally. Specifically, this thesis has demonstrated that how AIA engages the concepts of labour, rights, and risk is of significance to its address of AI labour – by not acknowledging their labour, rights, nor the risks they face, AIA offers AI labourers virtually no protections. More generally, however, this thesis has exposed the pitfalls of risk-based regulation within the EU. Legislating solely against prospective risks, regulations like AIA risk overlooking retrospective harms. This has global consequences. By failing to address AI labour, AIA enables the perpetuation of AI empire's dynamics of exploitation. Whether 'prosumers' subjected to data colonialism or contracted data annotators and moderators in the Global South, AI labourers remain hidden from view as an exploited 'shadow workforce'. My hope is that this thesis's scrutiny of AIA exposes the regulation's shortcomings and helps inspire more comprehensive legislation in the future. By casting a light on the shadows of global AI empire, I am hopeful that AI labour will soon enter the regulatory spotlight – in Brussels and beyond.

Word count: 16500

List of references

6, P., & Bellamy, C. (2011). *Principles of Methodology: Research Design in Social Science.* London: Sage.

Apostolakoudis, T. (2022). The global politics of artificial intelligence. In M. Tinnirello (Ed.), *Artificial Intelligence and Post-Capitlaism: The Prospect and Challenges of AI-Automated Labour* (pp. 209-246). CRC Press.

Arora, A., Barrett, M., Lee, E., Prince, K., & Oborn, E. (2023). Risk and the future of AI : Algorithmic bias, data colonialism, and marginalization. *Information and Organization, 33*(3), 1-7.

Bas, G., Salinas, C., Tinoco, R., & Sevilla, J. (2024). The EU AI Act: A pioneering effort to regulate frontier AI. *Inteligencia Artificial, 27*(73), 55-64.

Beck, U. (1992). *Risk Society: Towards a New Modernity.* (M. Ritter, Trans.) London: Sage.

Beck, U. (2006). Living in the world risk society: A Hobhouse Memorial Public Lecture given on Wednesday 15 February 2006 at the London School of Economics. *Economy and Society, 35*(3), 329-345.

Bentham, J. (1843). *The Works of Jeremy Bentham* (Vol. II). (J. Bowring, Ed.) Edinburgh: William Tait.

Brunsson, N., Nordin, I. G., & Hallström, K. T. (2022). 'Un-responsible' organization: How more organization produces less responsibility. *Organization Theory, 3*(4), 1-20.

Calderaro, A., & Blumfelde, S. (2022). Artificial intelligence and EU security: The false promise of digital sovereignty. *European Security, 31*(3), 415-434.

Cefaliello, A., & Kullmann, M. (2022). Offering false security: How the draft artificial intelligence act undermines fundamental workers rights. *European Labour Law Journal, 13*(4), 542-562.

Council of the EU. (2024). *Artificial intelligence (AI) act: Council gives final green light to the first worldwide rules on AI*. Retrieved July 28, 2024 from www.consilium.europa.eu: https://www.consilium.europa.eu/en/press/press-releases/2024/05/21/artificial-intelligence-ai-act-council-gives-final-green-light-to-the-first-worldwide-rules-on-ai/

Crawford, K. (2021). *The Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence.* New Haven, CT: Yale University Press.

Diana, S. (2021). Rewriting Marx to expose the data society and AI. *Cambio, 11*(21), 199-211.

Dawson, M., Lynskey, O., & Muir, E. (2019). What is the added value of the concept of the "essence" of EU Fundamental Rights? *German Law Journal*, *20*(1), 763-778.

Dolowitz, D. P., & Marsh, D. (2000). Learning from abroad: The role of policy transfer in contemporary policy-making. *Governance, 13*(1), 5-23.

Eleftheriadis, P. (2008). *Legal Rights.* Oxford: Oxford University Press.

European Commission. (2018a). *Coordinated Plan on Artificial Intelligence.* Retrieved July 28, 2024 from www.eur-lex.europa.eu: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=COM:2018:795:FIN

European Commission. (2018b). *Artificial Intelligence for Europe.* Retrieved July 28, 2024 from www.eur-lex.europa.eu: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=COM%3A2018%3A237%3AFIN

European Commission. (2023). *Artificial Intelligence – Questions and Answers*. Retrieved July 28, 2024 from www.ec.europa.eu: https://ec.europa.eu/commission/presscorner/detail/en/QANDA_21_1683

European Commission. (2024). *Data Act*. Retrieved July 28, 2024 from www.digital-strategy.ec.europa.eu: https://digital-strategy.ec.europa.eu/en/policies/data-act

European Parliament. (2023a). *EU AI Act: first regulation on artificial intelligence*. Retrieved July 28, 2024 from www.europarl.europa.eu: https://www.europarl.europa.eu/topics/en/article/20230601STO93804/eu-ai-act-first-regulation-on-artificial-intelligence

European Parliament. (2023b). *P9_TA(2023)0236 – Artificial Intelligence Act.* Retrieved July 28, 2024 from www.europarl.europa.eu: https://www.europarl.europa.eu/doceo/document/TA-9-2023-0236_EN.pdf

European Parliament. (2024). *Provisional agreement resulting from interinstitutional negotiations.* Retrieved July 28, 2024 from www.europarl.europa.eu: https://www.europarl.europa.eu/meetdocs/2014_2019/plmrep/COMMITTEES/CJ40/AG/2024/02-13/1296003EN.pdf

Fabbrini, F. (2014). *Fundamental rights in Europe.* Oxford: Oxford University Press.

Friedl, P., & Gasiola, G. G. (2024). Examining the EU's Artificial Intelligence Act. *Verfassungblog, (2366-7044)*.

Fuchs, C. (2014). *Digital Labour and Karl Marx.* London: Routledge.

Fuchs, C., & Sevignani, S. (2013). What is digital labour? What is digital work? What's their difference? And why do these questions matter for understanding social media? *tripleC, 11*(2), 237-293.

Gilbert, M. (2018). *Rights and demands: A foundational inquiry.* Oxford: Oxford University Press.

Gonçalves, G. L., & Costa, S. (2020). From primitive accumulation to entangled accumulation. *European Journal of Social Theory, 23*(2), 146-164.

Gray, M. L., & Suri, S. (2019). *Ghost Work: How to Stop Silicon Valley from Building a New Global Underclass.* Boston, MA: Houghton Mifflin Harcourt.

Hacker, P. (2023). What's missing from the EU AI Act – Adressing the four key challenges of large language models. *Verfassungsblog, (2366-7044)*.

Halperin, S., & Heath, O. (2020). *Political Research. Methods and Practical Skills.* Oxford: Oxford University Press.

Hjorth, L. (2018). Ambient and soft play: Play, labour and the digital in everyday life. *European Journal of Cultural Studies, 21*(1), 3-12.

Ho-Dac, M. (2024). Considering fundamental rights in the European standardisation of artificial intelligence: Nonsense or strategic alliance? In K. Jakobs (Ed.), *Joint Proceedings EURAS & SIIT 2023* (pp. 1-21). Verlag Günter Mainz.

Hohfeld, W. N. (1913). Some fundamental legal conceptions as applied in judicial reasoning. *The Yale Law Journal, 23*(1), 16-59.

Lupton, D. (2023). *Risk* (3rd ed.). London: Routledge.

Macenaite, M. (2017). The 'riskification' of European data protection law through a two-fold shift. *European Journal of Risk Regulation, 8*(3), 506-540.

Marx, K. (2000 [1887]). *Capital* (Vol. I). Electric Book Company.

Mayring, P. (2000). Qualitative content analysis. *Forum: Qualitative Social Research, 1*(2), 1-10.

Mejias, U. A. (2013). The privatization of social life. In U. A. Mejias, *Off the Network: Disrupting the Digital World* (pp. 19-36). Minneapolis, MN: University of Minnesota Press.

Minkman, E., Buuren, M. W., & Bekkers, V. J. (2018). Policy transfer routes: An evidence-based conceptual model to explain policy adoption. *Policy Studies, 39*(2), 222-250.

Nedzhvetskaya, N., & Tan, J. (2021). The role of workers in AI ethics and governance. In *[Chapter draft for] Oxford University Press Handbook on AI Governance* (pp. 1-30). Oxford: Oxford University Press.

Pavlidis, G. (2024). Unlocking the black box: Analysing the EU Artificial Intelligence Act's framework for explainability in AI. *Law, Innovation and Technology, 16*(1), 293-308.

Perrigo, B. (2023). *Exclusive: The $2 per hour workers who made ChatGPT safer.* Retrieved July 28, 2024 from www.time.com: https://time.com/6247678/openai-chatgpt-kenya-workers/

Rigakos, G., & Law, A. (2009). Risk, realism and the politics of resistance. *Critical Sociology, 35*(1), 79-103.

Salgado-Criado, J., & Fernandez-Aller, C. (2021). A wide human-rights approach to artificial intelligence regulation in Europe. *IEEE Technology and Society Magazine, 40*(2), 55-65.

Schreier, M. (2014). Qualitative content analysis. In U. Flick (Ed.), *The SAGE Handbook of Qualitative Data Analysis* (pp. 170-183). London: Sage.

Siegmann, C., & Anderljung, M. (2022). *The Brussels effect and artificial intelligence: How EU regulation wil impact AI market.* Centre for the Governance of AI.

Stix, C. (2021). The ghost of AI governance past, present and future: AI governance in the European Union. In J. Bullock, & V. Hudson, *[Chapter draft for] Oxford University Press Handbook on AI Governance* (pp. 1-37). Oxford: Oxford University Press.

Stone, D. (2012). Transfer and translation of policy. *Policy Studies, 33*(6), 483-499.

Stone, D. (2017). Understanding the transfer of policy failure: Bricolage, experimentalism and translation. *Policy and Politics, 45*(1), 55-70.

Tacheva, J., & Ramasubramanian, S. (2023). AI Empire: Unraveling the interlocking systems of oppression in generative AI's global order. *Big Data and Society, 10*(2), 1-13.

Taffel, S. (2021). Data and oil: Metaphor, materiality and metabolic rifts. *New Media & Society, 25*(5), 1-19.

Thompson, P. (2020). Capitalism, technology and work: Interrogating the tipping point thesis. *The Political Quarterly*, *91*(2), 299-209

Thomson, J. J. (1990). *The Realm of Rights.* Cambridge, MA: Harvard University Press.

van Dijck, J. (2014). Datafication, dataism and datavaillance: Big data between scientific paradigm and ideology. *Surveillance & Society, 12*(2), 197-208.

van Dijk, N., Gellert, R., & Rommetveit, K. (2016). A risk to a right? Beyond data protection risk assessments. *Computer Law & Security Review, 32*(2), 286-306.

Watson, L. (2021). *Right to know: Epistemic rights and why we need them.* London: Routledge.

White House. (2022). *What is the Blueprint for an AI Bill of Rights?* Retrieved July 28, 2024 from www.whitehouse.gov: https://www.whitehouse.gov/ostp/ai-bill-of-rights/what-is-the-blueprint-for-an-ai-bill-of-rights/

Appendices

Overview:

Appendix 1 – Coding stipulations for 'labour' subcategories

Appendix 2 – Coding stipulations for 'rights' subcategories

Appendix 3 – Coding stipulations for 'risk' subcategories

Appendix 1: Coding stipulations for 'labour' subcategories

Appendix 1: Coding stipulations for 'labour' subcategories	
Data preparation Code when Regulation explicitely references or discusses the annotation, labelling, and/or tagging of data. Code when the Regulation – explicitely or otherwise – references human contribution to the preparation of (meta)data.	**Workplace management** Code when Regulation references or discusses AI in the context of worker management or as integrated in the workplace – including, but not limited to, productivity monitoring, surveillance, and task allocation. Do not code for 'workplace management' if the Regulation addresses AI's role in the screening, filtering and/or evaluation of job candidates but where none of the above applies.

Appendix 2: Coding stipulations for 'rights' subcategories

Appendix 2: Coding stipulations for 'rights' subcategories		
Freedom of expression Code when Regulation explicitely references or discusses the right to freedom of expression in relation to AI systems or the legislative measures taken within AI to safeguard this right. Do not code when the Regulation references or discusses the transparency of AI-determind outcomes or AI-produced content.	**Privacy rights** Code when Regulation explicitely references or discusses privacy rights and/or specific type of privacy rights pertaining to the confidential processing of personal data. Code also when the Regulation explicitely references or discusses the rights of 'data subjects'. Do not code if 'data subjects' are mentioned without explicit reference to their rigths. Do not code if privacy concerns are addressed but not articulated or formulated with reference to rights.	**Procedural rights of operators** Code when Regulation explicitely references or discusses the specific procedural rights of AI operators within the EU legal system (where 'operator' is defined in accordance with AIA's Article 3). In cases, where the same unit of coding addresses both the procedural rights of operators as well as other legal or antural persons, code for 'procedural rights of persons' instead. Code for 'procedural rights of operators' when the rights of operators are mentioned with explicit reference to Article 18 of Regulation (EU) 2019/1020.
Fundamental rights in general Code when the Regulation explicitely references or discusses 'fundamental rights' in general with no further specification of the specific right(s) being addressed. If, in same coding unit, fundamental rights are mentioned in general but a specific right is also referenced or discussed, do no code for 'fundamental rights in general' but rather the category of the specific right. Do not code for 'fundamental rights in general' if an unspecified right or plurality of rights is mentioned, but not referred to as 'fundamental right(s)'.	**Procedural rights of persons** 'Procedural rights of persons' is composed of three constituent subcategories merged into one. The constituent subcategories are 'procedural rights of individuals', 'right to complain', and 'right to explanation'. *Procedural rights of individuals* : Code when Regulation explicitely references or discusses the procedural rights of individuals – such as, but not limited to, the right to a fair trial, the right to the presumption of innocence, and the right to effective remedy. Code also when the Regulation explicitely references or discusses the procedural rights of individuals in general.	**Right to non-discrimination** Code when Regulation explicitely references or discusses the right to non-discrimination, including, but not limited to, the value of this right, its definition, potential violations of this right (in relation to AI), and the measures taken to respect and uphold this right. Code when the Regulation references or discusses the right to not be discriminated against in any way, the right to not be unjustly profiled, and/or the right to not suffer 'unfavourable treatment'. Code also when the Regulation references or discusses the right to not suffer discriminatory outcomes such as when AI systems interpret biometric data and/or when AI is used within educational settings and/or when AI is used for purposes of border control. Code also when the Regulation discusses the right of persons with disabilities to not be discriminated against.
Intellectual property rights (IPR) Code when the Regulation references or discusses intellectual property rights in its relation to AI systems and/or the legislative measures taken to uphold them. Code also for 'IPR' when the Regulation explicitly references or discusses copyright rights. Do not code when the Regulation addresses property rights as regards the ownership of AI systems themselves. Examples of IPR-coded units include those which discuss the content, data, and material used to train and develop AI systems.	*Right to complain* : Code when Regulation explicitely references or discusses the individual's right to complain to relevant authorities about potential infringements of AIA. Code also if this right is mentioned in relation to legal persons who are not AI operators (as defined by AIA, Article 3). Do not code when Regulation merely references or discusses the obligations of providers or operators. *Right to explanation* : Code when Regulation explicitely references or discusses the right of natural or legal persons to request explanations from an AI deployer concerning the role of AI in decision-making procedures and/or concerning the output of AI-systems.	**Workers' rights** Code when Regulation explicitely references or discusses the rights of workers – contractually employed or otherwise. Code also when the Regulation refers to the rights derived from national labour laws. Do not code for 'workers' rights' when the text references or discusses the rights of individuals – legal or natural – unless these rights are explicitely addressed in relation to the given individual(s) role or status as a worker or as someone who performs (or has previously performed) labour, remunerated or otherwise.

Appendix 3: Coding stipulations for 'risk' subcategories

Appendix 3: Coding stipulations for 'risk' subcategories	
Cybersecurity Code when Regulation explicitely references or discusses the harms and/or risks pertaining to cybersecurity as concerns AI systems, and/or when the Regulation explicitely pertains to the measures taken to mitigate such risks.	**Privacy** Code when Regulation explicitely references or discusses AI as a potential privacy concern; as something that affects, gains access to, or processes personal data in a potentially or actually harmful manner. Code also when the Regulation elaborates on how personal data should be processed, handled, and protected so as to avoid risk of harm or misuse (of data).
Deception and misinformation Code when Regulation explicitely references or discusses specific harms and/or risk of AI-facilitated deception and/or impersonation. Code also when discussing specific risks of AI-facilitated misinformation or disinformation. Examples include human-like AI 'chatbots' or online disinformation campaigns. Code only when Regulation explicitely refers to such circumstances as 'risks' (or using similar 'risk'-cluster terms). Do not code for this category if AI-produced content or AI-facilitated activities go unnoticed by humans or authorities; that is, do not code for clandestine AI activity – only code for risk of AI activity as deceitful.	**Systemic risk** Code when Regulation explicitely references or discusses 'systemic risk', how 'systemic risk' is defined, what 'systematic risk' entails, and/or the legislative measures taken to mitigate 'systemic risk' – either at current or through prospective future amendments. Do not code when 'systemic risk' is mentioned but none of the above applies. Do not code when legislation pertains to GPAI models without specific mention of 'systemic risk'.
Dependency Code when Regulation explicitely references or discusses the harms and/or risks associated with human dependency on AI systems and/or natural persons' inability to avoid interacting with an AI system without suffering social, economic, physical or other forms of harm. Do not code for risks/harms associated with lack of human control and/or oversight of AI systems. Do not code when Regulation references or discusses human independency from AI – unless the unit includes an explicit reference to any risk associated with lack of such independence (such a risk must also be clearly distinct from the risks associated with lack of human control and/or oversight of AI systems).	**Unclear** 'Unclear' is composed of two constituent subcategories, 'general risk' and 'unspecified risk', merged into one. *General risk*: Code when Regulation explicitely references multiple specific risks and/or harms, but no specific risk is discussed in detail more than any other within the same unit of coding. If, within the same unit of coding, one specific risk category is discussed more than any other – or is discussed in qualitative distinction from all other risks categories – code for that respective risk category instead of 'general risk'. Do not code for 'general risk' in cases where the Regulation references 'high-risk' or 'high-risk' AI systems without mention of any specific risk category. *Unspecified risk*: Code when 'risk' (or similar words within the 'risk' cluster) or 'harms' are mentioned or discussed within the Regulation, but it remains unclear which specific risks and/or harms the Regulation is referring to – typically by absence of their mention or by referring to a broad concept of risk and/or harms. This code also applies when the Regulation references or discusses 'reasonably foreseen' risks. In contradistinction to the 'unknown risk' category, it is assumed that unspecified risks could be explicitly articulated and specifically defined; it is assumed 'unspecified' risks are already conceptualised. Do not code for 'unspecified risk' in cases where the Regulation references 'high-risk' or 'high-risk' AI systems without mention of any specific risk category.
Unknown risk Code when Regulation explicitely addresses or refers to risks or harms that are, of yet, unidentifiable or unknown. In contradistinction to the 'unclear' code, the 'unknown risk' code applies to risks/harms that are not yet conceptualised nor directly legislated against – i.e. codes that *cannot* be specifally addressed since they are yet to be discovered or conceived of. Likewise, risks or harms that are characterised as 'reasonably foreseen' should be coded as 'unclear' instead.	
Discrimination Code when Regulation explicitely references or discusses the risk and/or harm of AI systems – either autonomously or through human usage – disproportionately and/or adversely affecting certain human individuals or a distinct group of natural persons. This includes, but is not limited to, the risks/harms associated with algorithmic profiling or decision-making – as performed by AI systems – which may incur a potential consequence for one or more AI-determined natural persons. Code also when Regulation explicitely discusses or references the risks associated with AI systems' profiling of natural persons and/or the harms associated with the real or potential biases of AI systems.	**Human control and oversight** Code when Regulation explicitely references or discusses risks and/or harms associated with either the autonomy of AI systems and/or a lack of human control over AI systems and/or lack of adequate human oversight of an AI system's operation. Code also when Regulation explicitely pertains to the need for, or the measures taken, to ensure human control and/or oversight of AI systems. Do not code when the Regulation references or discusses 'AI literacy' but the above does not apply.
Serious incidents Code when Regulation explicitely references the term "serious incident" (or "serious incidents") in accordance with the definition put forth in Article 3 (paragraph 44) as pertaining to a specific set of AI-induced risks such as the risk of harm to humans, critical infrastructure, and the physical environment. Coding under this subcategory may include, but is not limited to, definitions of 'serious incident(s)' as well as how they are to be managed, mitigated, and/or communicated. Do not code in units where the term 'serious incident' is not included - or not employed in the definition presented in Article 3.	