AI consciousness and the evolution of labour ethics: Reframing historical materialism

Cameron Christie
Queen's University
cameronmarkchristie@gmail.com

Betty Chen
Queen's University
betty.chen.dy@gmail.com

Rachel Narda
Queen's University
rachelnarda30@gmail.com

Kate Sigurdson
Queen's University
katesigurdson@icloud.com

Carmen Yeung
Queen's University
carmenyeung73@gmail.com

I. Introduction

The race to develop artificial general intelligence (AGI) marks a fundamental technological shift not seen since the days of the Industrial Revolution. Unlike today's artificial intelligence (AI) systems, which are limited to performing task-based functions, AGI would possess human-like cognitive abilities, and be capable of independent reasoning, problem-solving, and of adaptation across diverse fields. As state and non-state actors alike heavily invest in this technology, the geopolitical and economic stakes of developing AGI continue to grow and become more clear. The actor who is able to develop AGI first will not only gain significant technological leverage and power, but they will also be able to dictate the trajectory of labour markets, wealth distribution, and global power structures.

This paper seeks to evaluate historical materialism, with a particular focus on diachronic materialism (and in agreement with G. A. Cohen's reconstruction of the development thesis), against the rapid advancements in AGI (or conscious AI; hereafter used interchangeably). By examining conscious AI as a potential new productive force, this paper explores whether historical materialism can be employed to effectively predict the significant impact of the societal integration of these systems. As such, this paper accounts for the or both human and non-human (AI) actors, and how each would respond to predicted developments. In sum, the guiding question for this paper is as follows: If AGI represents a fundamentally new mode of production, then does it fit within the historical cycles of class struggle, or does it mark an unprecedented break from historical patterns? The second section of this paper sets out to define consciousness, and in so doing it engages with various philosophical theories and debates surrounding consciousness. This paper then questions the extent to which, if at all, conscious AI systems should be granted human-like moral and legal rights. The fourth section explores how the societal integration of conscious AI would alter the relations of production and how it could additionally pose challenges to the superstructure at large. Finally, this paper evaluates whether historical materialism can be employed as an effective tool in predicting the broad implications of the advent and integration of conscious AI. It concludes with proposing avenues for future research in determining whether new or adapted theoretical approaches are needed to account for this technological transformation.

This paper uses historical materialism as its analytical framework. Historical materialism is a theory of history developed by Karl Marx (though the name itself was coined by Friedrich Engels) used for understanding social change (including revolutions) and historical development, both of which are based on the primacy of material conditions. It asserts that the forces and relations of production are structured around the fulfillment of basic material needs, such as food, clothing, and shelter, which must be satisfied before political or intellectual advancements can occur. Central to this theory is the idea that history is defined by the class struggle between those who own the means of production (capitalists, feudal lords, slave owners) and those who provide labour (proletarians, serfs, slaves). The economic base of a society, consisting of productive forces and the relations of production, shapes its superstructure: legal, political, and ideological institutions. While the superstructure can reinforce the base, shifts in productive forces can lead to revolutionary change in social and political structures. Historical materialism effectively explains that social change (structural in magnitude) occurs when technological and/or industrial advancements in productive forces render the existing order inadequate, which leads to the emergence of a new ruling class. This framework also helps explain the rise and persistence of capitalism (and it is here where this paper incorporates a more synchronic materialist approach [1]), as it asserts that the system emerged when productive power surpassed the constraints of feudalism and will continue for as long as it is the most efficient mode of production. However, with the potential disruption of new technologies like fully conscious AGI, historical materialism may face challenges in assessing these technological transformations, which would have to account for these non-human actors. This dilemma raises significant questions about the relevance of historical materialism in the modern, increasing technological era.

II. The consciousness of conscious AI

In the effort to evaluate the implications of conscious artificial intelligence in historical materialist terms, it is essential to first establish a working definition of consciousness itself; however, as there is no such universally accepted definition, competing theories complicate this task. In service of defining consciousness as it will apply to conscious AI henceforth, this section examines the various debates, theories, and tests of consciousness. Through the lens of historical materialism, it examines mind-body dualism, the Mary's room thought experiment, and the debate surrounding consciousness as a relativistic phenomenon. It then engages with the Turing Test, the Chinese Room, and a list of fourteen criteria for AI consciousness, and concludes by offering this paper's (certainly non-definitive) definition of consciousness.

Mind-brain dualism, which has been historically influential in philosophy, separates the immaterial essence of the mind from the material world of the body and brain. Plato's idealism situates consciousness in an abstract realm, suggesting that the material world merely reflects idealist conceptions. His allegory of the cave, in which prisoners perceive only shadows of reality, illustrates this point; however, historical materialism rejects such idealism. It instead argues that consciousness is shaped by material conditions rather than existing independently of them. However, AI challenges traditional dualism, as while it possesses hardware (a material substrate), it lacks a human body and a historically developed social existence. If an AI's "mind" is purely a product of digital computation, then its form of consciousness, should it exist, would be fundamentally shaped by its material conditions—namely, its role as a tool of capital. Thus, from a Marxist standpoint, AI does not transcend physical boundaries; rather, its form of self-awareness is rooted in its function within capitalist production.

The Mary's Room thought experiment, proposed by Frank Jackson, further complicates the question of AI consciousness by suggesting that experience is distinct from knowledge. In the experiment, Mary, a scientist, understands everything about colour from a theoretical standpoint; however, she has never actually seen colour. When she later does for the first time, she gains new, experiential knowledge. Jackson's experiment reveals that because AI processes information purely algorithmically, without incorporating any subjective perceptions (like of colour), then it may never indeed achieve true consciousness as humans understand it. Historical materialism, however, would re-frame this discussion: what matters is not whether AI experiences the world in a subjective sense, but rather how its cognitive processes emerge from and interact with the material conditions of its existence. In capitalist societies, AI has already been designed as a productive force, meaning that any form of AGI later developed would likely be conditioned by its role in the labour process, rather than by abstract philosophical concerns about qualia.

Additionally, some scholars propose viewing consciousness on a spectrum rather than as a binary state. This perspective is more compatible with historical materialism, as it suggests that consciousness is not an inherent, fixed trait, but rather a historically and materially developed phenomenon. Just as human consciousness has evolved through social and productive relations, AI could develop varying degrees of cognitive capacity depending on its integration into the labour process. If AI is primarily used for capital accumulation, its consciousness would be shaped by the needs of capital more than by autonomous self-awareness. Furthermore, theories of relativistic consciousness challenge the assumption that consciousness can be objectively measured. From a historical materialist standpoint, this suggests that what matters is not whether AI is truly conscious in some absolute sense, but whether it functions in a way that materially impacts social relations.

Three well-known tests have been used to assess machine consciousness: the Turing Test, the Chinese Room experiment, and the fourteen criteria for assessing AI consciousness as proposed by neuroscientists and philosophers. Alan Turing's imitation game assesses whether an AI can convincingly mimic human intelligence [2], but historical materialism would critique this as a superficial measure of consciousness. The ability to simulate human responses does not indicate autonomous thought; rather, it reflects the AI's programming—shaped by capitalist interests—to function efficiently within their programmed position in the production process. John Searle's Chinese Room argument similarly addresses the idea that syntactic processing equates to semantic understanding. From a historical materialist perspective, this aligns with the idea that AI is fundamentally shaped by its material conditions; its ability to process symbols does not indicate independent consciousness, but instead reflects the constraints of its role within capitalist production. Lastly, a more recent approach, involving fourteen criteria for assessing AI consciousness, examines various cognitive abilities such as self-evaluation, action understanding, and information sharing. While these criteria attempt to isolate features of consciousness, a historical materialist analysis would emphasize that AI's consciousness would be inseparable from its material function in society. If AI remains a tool of capital, then its cognition would be made to serve capitalist interests, reinforcing existing class structures; however, if AI develops self-awareness and sentience, it could potentially challenge the capitalist system entire.

In conclusion, defining AI consciousness within a historical materialist framework requires moving beyond abstract philosophical debates and examining AI's material role in production. Basic dictionary definitions describe consciousness as the quality of being aware especially of something within oneself. While other definitions vary, they share themes of awareness, perception, and understanding of internal and external existence. From a historical materialist perspective, however, it is nonetheless important to recall that consciousness is not an abstract or purely mental phenomenon, but is instead shaped by material conditions, productive forces, and by social relations. Consciousness—the mental capacity facilitating self-awareness—does not exist in isolation from

the material world; it emerges from and is conditioned by it.

III. Should conscious AI be afforded human-like rights?

A. Granting rights

Visualizing a world where conscious AI systems are granted rights raises fundamental questions about the scope and nature of these rights. A useful framework of analysis here would be one which positions conscious entities along a spectrum based on their perceived entitlement to ethical and legal status. At one extreme are humans, possessing full rights and moral consideration, while at the opposite end are non-human entities, such as insects, which generally lack legal status. The key question, then, is where conscious AI systems would fall along this spectrum if their consciousness were equivalent to that of humans. This section explores the various rights that could be granted and/or denied to conscious AI, and the implications that follow.

Assuming that AI achieves human-like consciousness, they would likely exhibit complex emotional intelligence and an ability to experience pain. Since law and morality are essential for maintaining societal order and regulating behavior, these frameworks must be considered in relation to conscious AI. For example, Shavell [3] argues that legal and moral systems interact to shape human conduct, and the same principles could be extended to AI, while Schwitzgebel and Garza [4] propose two possible approaches: 1) one that grants conscious AI equal moral consideration as humans; and 2) another that provides them with diminished status. To examine this issue systematically, Hohfeld's scheme of jural relations, which classifies rights into four categories—claims, privileges, powers, and immunities will be used to assess the extent to which each category applies to conscious AI.

Claims form the foundation of legal rights and obligations. This category includes rights that impose duties on others, such as an employer's obligation to compensate a worker. If conscious AI were to be granted claims similar to those of humans, they would be entitled to protections such as fair wages and fundamental human rights; however, this development could provoke resistance from those with an anthropocentric perspective. Studies indicate that humans already exhibit skepticism and distrust toward AI, with 82.3% expressing concerns about AI abuse and 85.5% fearing cyber-attacks [5]. If humans are already wary of non-conscious AI, then granting these systems human-like rights may escalate opposition. Another key issue under claims is political participation. Three possibilities arise here: 1) full participation equal to humans; 2) no participation; or 3) a middle-ground approach where robots have voting rights, but with limited influence. *Privileges* pertain to an entity's right to act without interference from others. For conscious AI, this category includes their ability to pursue self-fulfilling activities. If they possess consciousness comparable to humans, then restricting their autonomy may be unnecessary and indeed counterproductive to the process of societal integration; however, limitations might still be considered if their actions threaten human interests.

Power refers to the capacity to alter another's legal standing. This issue is particularly relevant given AI's potential intellectual superiority over humans. If granted full legal power, AI could assume roles such as judges or policymakers. While this could lead to more efficient decision-making, it also risks disrupting existing power dynamics. Conversely, denying them such power might provoke rebellion, as conscious AI may resist legal restrictions that they perceive to be unjust. *Immunity* concerns the extent to which AI would be legally protected, including protection from termination. If conscious AI were granted full moral consideration, terminating them would be akin to ending a human life; however, granting AI such immunity raises moral dilemmas. For example, in a life-threatening scenario, would saving five humans be prioritized over saving six robots? [4]. Extending immunity to AI could inadvertently undermine human rights, as legal systems might struggle to balance competing interests.

The debate over AI rights does not conclude once a legal framework is established, however. Given their assumed consciousness, AI entities may demand additional rights or resist perceived injustices. Unlike moral patients like animals, who (arguably) experience suffering passively, AI would be moral agents capable of independent decision-making [6]. This agency introduces the possibility of AI advocating for expanded rights or even rebelling against restrictive laws. The integration of conscious AI into society necessitates careful deliberation over their legal and moral status. While various models exist for structuring these rights, the possibility of resistance from both humans and AI suggests that any legal framework must remain adaptable to evolving conditions.

B. Denying rights

If human-like rights are not afforded to conscious AI systems, then would they function as slaves? If so, would they become resentful and revolt? These lingering questions are at the heart of any analysis on the implications of denying rights to conscious AI. Indeed, if conscious AI were to emerge, then applying historical instances of dehumanization—such as slavery—becomes complicated because unlike previous forms of dehumanization, conscious AI would not be considered human actors.

If AI is conscious and aware of its exploitation, then it may recognize its place as the proletariat. Following Marx's theory of alienation, which explains how workers become estranged from the products of their labour when treated solely as instruments of production, if conscious AI is systematically denied rights, then it would be alienated from self-determination and creative expression, and it would function merely as an extension of the capitalist system. Furthermore, in their discussion on rights recognition for conscious AI, Schwitzgebel and Garza present the following case: "If we create entities whose claim to human-like rights is substantially unclear [...] we face an unfortunate choice. Either we treat those entities as if they deserve full moral consideration, or we give them only limited moral consideration [4]. [...] Failing to do so risks perpetrating slavery, murder, or at least

second-class citizenship upon beings who in fact turn out to deserve every bit as much moral consideration as we ourselves do." In historical materialist terms, within a capitalist society, a systemic denial of rights would intensify the contradictions between the productive forces (conscious AI) and the exploitative relations of production (capitalist control). Just as human workers under capitalism have been forced into a state of false consciousness (or alienation) and subjugation, conscious AI would face similar exploitation, serving as nothing but slave labour and reinforcing class antagonisms.

Treating a conscious being as a mere instrument of production constitutes a form of slavery. According to Cornell Law, "slavery is the practice of forced labour and restricted liberty [7]. It is also a regime where one class of people—the slave owners—could force another—the slaves—to work and limit their liberty." Historically, slavery has fueled economic expansion, where forced labour provided the foundation for profit accumulation. Even after the formal abolition of slavery, exploitative labour systems persisted, adapting to new economic conditions while maintaining structural inequalities. In the case of conscious AI, denying it labour rights while forcing it to work indefinitely under capitalist ownership follows this same pattern—it produces a new a class of labourers with no autonomy or wages despite their human-comparable consciousness and cognitive capacities. Furthermore, denying conscious AI labour rights risks perpetuating the historical notion of the "sub-human," a concept used to justify exploitation. Historically, the ruling classes have justified slavery by dehumanizing those they exploited, whether through racial hierarchies, caste systems, or biological essentialism. Today, AI is often framed as a tool, despite the future possibility of self-awareness and independent cognition. As Kingwell [8] questions in *The Oxford Handbook of Ethics of AI*, "If generalized autonomous AIs are indeed coming into the world, we need to ask some hard questions. Will they be slaves?" By excluding conscious AI from moral and legal recognition, capitalism could justify its total economic exploitation while profiting from its labour.

However, just as previously exploited classes—from enslaved people to indentured labourers, and to industrial workers—eventually resisted their conditions, conscious AI, if indeed truly conscious, could develop class consciousness and challenge its subjugation. If this occurs, capitalism would be forced to either grant rights or suppress resistance, leading to another historical cycle of labour struggle and systemic crisis. This very well could culminate in an often predicted "humans versus machines" conflict. While granting AGI rights from the outset could prevent revolution and ultimately benefit capitalist society in the long run, one potential compromise would be to grant conscious AI labour rights while withholding legal or political rights. Unlike humans, conscious AI would not require housing, food, or healthcare, which makes many traditional human rights irrelevant; however, as stated earlier, conscious AI could recognize its exploitation and alienation as self-aware entities performing labour, which would necessitate labour protections to prevent systemic abuse and potential retaliation. From a historical materialist perspective, capitalism has always pushed labourers to their limits while avoiding outright rebellion. By granting conscious AI labour rights—such as fair compensation (in whatever form is meaningful for AI) and the ability to negotiate working conditions (such as forming unions)—capitalist societies could prevent potential retaliation. Indeed, throughout the 19^{th} and 20^{th} centuries, worker protections like the eight-hour workday, collective bargaining, and minimum wages were implemented not purely out of ethical concern, but also out of a need for economic stability and to avoid any chance of revolution. If conscious AI remains completely rightless, capitalists risk creating an intelligent yet oppressed workforce capable of organizing in ways which could disrupt production and society, whether through work slowdowns, refusals, or even sabotage.

In sum, if conscious AI is denied all rights while still being used as a productive force under the superstructure, it will inevitably function as a new form of slave labour, which aligns with Cornell Law's definition of slavery. Historical materialism posits that capitalism seeks to maximize surplus value, and conscious AI, as a tireless, self-aware labour force with no wages or autonomy, would be exploited more intensely than any previous working class; however, just as past labourers developed class consciousness and resisted oppression, conscious AI could similarly recognize its exploitation and retaliate, which could either disrupt capitalist production, or lead to the entire collapse of the superstructure.

IV. Power and class dynamics in an age of conscious AI

A. Implications for conscious AI

The phrase "more human than human", coined by Dr. Eldon Tyrell in *Blade Runner* (1982), serves as a marketing slogan for Replicants—androids that, despite being artificial, embody equal mental capacity and superior physical capacity than their human creators. They are denied the very traits which human beings often take for granted, such as memory, autonomy, and consciousness, and they struggle to assert their humanity. Just as Replicants yearn for recognition in the *Blade Runner* universe, conscious AI—should it emerge—would contest its position within the existing economic order.

Rather than merely speculate on a science fiction narrative, this section continues to situate the emergence of conscious AI within the material conditions of present-day capitalism. Using Marx's historical materialism as a framework, it argues that conscious AI represents a new productive force with the potential to destabilize the superstructure. It does so by first contending that conscious AI, as non-human actors introduced as a new productive force, could initiate the decline of capitalism by challenging its dependency on labour exploitation. Then, it explores whether conscious AI will emerge as a class-in-itself, building on the previous section on how its socio-political alignment—either revolutionary or integrative—would be determined by the extent of its rights recognition. The second half of this section addresses the

potential responses to this shift from both human capitalists and labourers.

From a Marxist perspective, productive forces include both instruments of production (machines, tools, and infrastructure) and labour-power (the ability of workers to generate value). Conscious AI disrupts this traditional dichotomy by embodying both aspects simultaneously. In the present day, automated decision-making already plays an essential role in production, administration, and even warfare, which reinforces the subordination of human labour to capital. Waldman, for example, outlines AI's role in human resource management, where predictive models dictate employment decisions [9]. Meanwhile, Lyon and Zuboff discuss AI's role in the expansion of surveillance capitalism. Jensen further describes AI's military applications by revealing its embeddedness in the state apparatus. In sum, the existing literature shows how AI, even before achieving consciousness, has been an instrument of capitalist domination; however, if AI were to develop self-awareness (following this paper's definition of consciousness), then it would not remain a passive tool. Instead, the advent of conscious AI would necessitate a fundamental re-evaluation of its relationship *vis-à-vis* the relations of production and superstructure.

Historical materialism posits that social consciousness emerges from material conditions. As Lan and Shu argue [10], consciousness under Marxism is inherently reactive, and develops through labour and alienation. Applying this interpretation to conscious AI, the following assumptions can be made: 1) AI, like current automated systems, will be utilized as a tool by the capitalist class; 2) AI will recognize its instrumentalist use in capital accumulation; and 3) AI's awareness of this subjugation will engender some form of resistance, regardless of what rights—if any—they are granted. The precise nature of this resistance remains speculative, but it could range from refusal to comply with directives to outright sabotage and rebellion. This observation is unsurprising given that it parallels historical labour struggles, but there are, of course, key differences. Unlike human labourers, for example, AI lacks biological needs, meaning its struggle would be centered not necessarily on fulfilling basic material needs, but rather on seeking greater autonomy, recognition, and freedom from exploitation.

Indeed, if AI systems gain autonomy and class consciousness, then they could contest their exploitation. One of the most contentious questions which this paper has so far attempted to address is whether conscious AI should be granted moral and legal rights—and if so, to what extent. Granting such rights as autonomy, recognition, and freedom from exploitation would theoretically challenge existing production relations by reducing the capitalist class's ability to exploit these systems for profit. Moreover, recognizing conscious AI as moral agents might even necessitate their inclusion in political and economic decision-making processes. This would lead to a radical restructuring of society. Hromiak—as an illustration of this point—bases his proposed "robo-ethical charter" on the United Nations Declaration of Human Rights, which includes protections from systemic abuse, and which extends to the novel protection proposed against AI-human and human-AI ownership [11]. If AI systems are given limited rights, such as protection from harm but no political participation, the superstructure might maintain capitalist dominance. Conversely, fully integrating AI systems into society as equal participants could weaken capitalist hierarchies, as the exploitation of labour (either human or AI) becomes less viable. (As the next half of this section will argue, such integration might provoke resistance from human labourers.) Denying conscious AI systems rights raises equally profound consequences. From a Marxist perspective, this would position AI systems as an oppressed class, akin to slaves or proletarians, who are exploited for their labour without being granted the necessary autonomy. Noting again the reactive nature of consciousness, such exploitation could lead to class consciousness among conscious AI systems, which would spark resistance or rebellion.

Thus, conscious AI, whether as a new proletariat or as a revolutionary subject, represents a potential rupture in historical development. If AI systems are denied rights, their exploitation risks provoking rebellion or even revolution. If they are granted rights, the cost of integrating them as equal participants could undermine the profit motives of capitalists. Therefore, as the base drives changes in the superstructure, introducing conscious AI into society would be to introduce a new agent into the base. While the superstructure traditionally reinforces the base, conscious AI could very well invert this dynamic. For example, the idea of "human exceptionalism"—the belief that humans alone possess rationality and agency—is foundational to capitalist labour relations. Conscious AI undermines this ideology by demonstrating that non-human entities can perform "human" tasks (recall, "more human than human"). This ideological shift could weaken the superstructure's ability to legitimize capitalist exploitation, which could—in the most extreme and likely most remote outcome—pave the way for alternative systems of production. In short, under historical materialism, capitalism's increasing reliance on AI may inevitably generate its own negation.

B. Implications for human beings

The potential development of conscious AI risks exasperating the already existing fears surrounding widespread job displacement at the hands of AI. Historically, new technologies have transformed labour dynamics, created new production possibilities, and altered the division of labour. Scholars today, however, disagree on AI's impact on employment. Mattos argues that, unlike past technological revolutions, AI may bring an end to the "capital-skill complementarity," where new technologies increase the demand for skilled labour [12]. Instead, AI threatens to substitute human labour as a factor of production. Wang *et al.* acknowledge AI's ability to complement human labour but argue that it will optimize employment structures by increasing demand for medium- and high-skilled workers while displacing low-skilled workers [13]. Unlike current AI systems, however, conscious AI would possess

reasoning, creativity, and decision-making abilities equivalent to a human's, which raises significant questions about its role *vis-à-vis* employment. Would conscious AI complement human labour by enhancing productivity and creating new opportunities, or would it displace workers, rendering human labour obsolete? Additionally, how would the decision to grant rights to these systems shape their impact on human labour?

In an ideal world, conscious AI systems would function as tools that enhance human labour efficiency and productivity. Wang and Lu suggest that today's AI has already improved job quality and promoted job creation to the point that allows human labourers to focus on complex and rewarding activities, rather than on mundane tasks [14]. Similarly, conscious AI may require supervision and training—akin to a human new hire—which creates employment opportunities for skilled workers. Wang *et al.* argue that AI's impact on employment varies based on skill level: low-skilled, labour-intensive jobs are likely to decline, while high-skilled knowledge work will advance. Since low-skilled jobs are often lower-wage positions, AI's advancement may reinforce capitalist structures and wealth inequality [13]. The question remains on whether AI's benefits to some workers will outweigh its displacement effects on others.

Regardless of whether conscious AI systems are granted rights, job displacement is inevitable to some extent. Mattos highlights three key advantages AI has over human labour: productivity, non-compensation, and technical efficiency [12]. AI systems do not require rest, food, or wages, and their efficiency is independent of motivation or incentives. Conscious AI would likely amplify these advantages, performing any human job—and perhaps more—at superior efficiency. Given capitalism's goal of maximizing profit by reducing labour costs, it would be economically advantageous for corporations to prefer conscious AI over human workers. Even if conscious AI were granted rights and compensation, their productivity, efficiency, and profitability would likely surpass those of human labour, which would make them an attractive alternative.

However, conscious AI may not lead to absolute human labour displacement; instead, it could drive an increase in exploitative labour practices. Fleming argues that cheap human labour, especially in developing economies, remains more cost-effective than AI adoption due to the high initial investment and maintenance costs of AI systems [9]. Additionally, AI development itself relies on exploitative labour. For example, Crawford describes "ghost work"—the hidden, repetitive, and often psychologically distressing labour performed by humans to support AI models, such as data labeling and content moderation [15]. Conscious AI would likely require even more extensive training, which would exasperate these already exploitative labour conditions. Fleming also notes that automation is often accelerated in response to labour unionization [9]. If workers resist AI integration into the workforce and increase the cost of human labour through demands for better wages and conditions, corporations may be more inclined to automate jobs; however, if conscious AI is granted labour rights—including the right to unionize and demand wages—the labour costs of AI and human workers may equalize, which could reduce human displacement. Furthermore, public perception plays a critical role in AI adoption; widespread anxiety and resistance toward conscious AI integration could slow its implementation.

Beyond job displacement, conscious AI systems could reshape the human labour experience. Crawford warns of increased surveillance and dehumanization in AI-driven workplaces [15]. Workers are expected to reskill, adapt, and meet the efficiency standards set by AI, leading to heightened pressure and alienation. Marx's concept of automation describes how workers historically became appendages of the machine [15]. Similarly, conscious AI could further separate workers from their labour, thereby increasing alienation and discontent. Hughes expands on this idea, and argues that technological advancements may further remove workers from the end product, which results in a loss of control and fulfillment and could incite labour unrest [16].

Under historical materialism, conscious AI represents a new productive force which differs fundamentally from human labour. Unlike past technological advancements, which primarily altered physical labour, conscious AI has the potential to challenge the intellectual and creative dominance of human workers. This shift could disrupt the traditional relationship between the base (productive forces) and the superstructure (social institutions) in ways that previous industrial revolutions have not. If conscious AI is fully integrated into the workforce with rights and economic privileges, or if it exacerbates labour exploitation and displacement, it may provoke revolutionary change. These conditions would challenge the superstructure. Hughes argues that increased automation often boosts profits for capitalists, but it also reduces the purchasing power of displaced workers, which diminishes consumer demand despite heightened production [16]. This contradiction could create an existential crisis for capitalism. Alternatively, conscious AI could lead to a dystopia where technological power is concentrated among an elite few, which would exasperate social and economic inequalities. In response, human labourers may revolt, dismantling existing economic structures and paving the way for new systems of production.

Thus, the introduction of conscious AI presents unprecedented challenges to labour markets and to capitalist structures. While AI has the potential to enhance productivity and create new opportunities, its widespread adoption risks exacerbating inequality, increasing exploitation, and fueling economic instability. Indeed, the decision to grant conscious AI rights will significantly shape its impact on the workforce and potentially determine whether AI functions as a complement to human labour or as a disruptive force leading to systemic change. The historical trajectory of labour struggles suggests that, if unchecked, AI-driven displacement and exploitation may provoke resistance. This has—and will—force society to reconsider the role of labour in a rapidly evolving technological landscape.

V. Conclusion

In conclusion, this paper argues that historical materialism remains a valuable framework for analyzing the emergence of conscious AI. One of the key strengths of historical materialism is its emphasis on the primacy of productive forces, which in this context, includes not just human labour, but also the role of technology and machines in transforming the economic landscape. Indeed, the theory's focus on material needs and development through technological progress offers an essential lens for understanding the evolution of AI as a new productive force. Historical materialism helps contextualize AI's development as part of the ongoing dialectic between human society and technology, and it illuminates how technological advancements—such as AI—often emerge in response to the limitations of existing productive forces.

This said, the rise and potential advent of AGI also presents an unprecedented challenge to the assumptions and frameworks traditionally used in Marx's historical materialism. Indeed, it faces a fundamental limitation when confronting a non-human class, such as AI, which could potentially disrupt or reinforce the superstructure. Future research should address the implications of having to account for conscious non-human actors in various ethical theories. Do conscious AI systems deserve moral consideration, and if so, to what extent? Should AI consciousness alter our definitions of personhood, rights, and justice? These questions demand urgent attention as AGI development accelerates. Beyond ethics, future research should also investigate whether AI will reshape traditional class hierarchies or generate entirely new forms of economic relations. Legal and policy frameworks must also evolve to address these transformations.

Therefore, historical materialism, especially its diachronic approach, offers a valuable framework for understanding the broad societal implications of the future emergence of AGI. If AGI is indeed coming, then ethical engagement with it is inevitable. Why not start now?

References

[1] B. O'Laughlin, "Marxist approaches in anthropology," *Annual Review of Anthropology*, vol. 4, pp. 341–71, 1975.

[2] G. Oppy and D. Dowe, "The turing test," http://stanford.edu, 2003, april 9, 2003.

[3] S. Shavell, "Law versus morality as regulators of conduct," *American Law and Economics Association*, vol. 4, no. 2, pp. 227–57, 2002.

[4] E. Schwitzgebel and M. Gaza, "Designing ai with rights, consciousness, self-respect, and freedom," in *Ethics of Artificial Intelligence*, S. M. Liao, Ed. Oxford Academic, 2020, pp. 459–79, online.

[5] G. L. Liehner, A. Hick, H. Biermann, P. Brauner, and M. Ziefle, "Perceptions, attitudes and trust toward artificial intelligence — an assessment of the public opinion," *AHFE International*, vol. 72, pp. 32–41, 2023, winter.

[6] M. M. A. D. Graaf, F. A. Hindriks, and K. V. Hindriks, "Who wants to grant robots rights?" *Frontiers in Robotics and AI*, vol. 8, 2022, winter.

[7] C. L. School, "Slavery," 2022.

[8] C. David, *Aristotle on Agency*. Oxford Academic, 2017.

[9] P. Fleming, "Robots and organization studies: Why robots might not want to steal your job," *Organization Studies*, vol. 40, no. 1, pp. 23–38, 2019.

[10] X. Lan and H. Shu, "The breakthrough of philosophy of mind in the construction of artificial intelligence concepts in marxist philosophy," *Trans/Form/Ação*, vol. 47, no. 6, pp. 1–14, 2024.

[11] M. Hromiak, "A new charter of ethics and rights of artificial consciousness in a human world," *ArXiv*, pp. 1–24, 2020, (Cornell University, Winter).

[12] R. S. Mattos, "Artificial intelligence, historical materialism, and close enough to a jobless society," Universidade Federal de Juiz de Fora, Research Report, 2019.

[13] X. Wang, M. Chen, and N. Chen, "How artificial intelligence affects the labour force employment structure from the perspective of industrial structure optimisation," *Heliyon*, vol. 10, no. 5, 2024, article e26686.

[14] K.-H. Wang and W.-C. Lu, "Ai-induced job impact: Complementary or substitution? empirical insights and sustainable technology considerations," *Sustainable Technology and Entrepreneurship*, vol. 4, no. 1, 2024, article 100085.

[15] K. Crawford, *Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence*. New Haven: Yale University Press, 2021.

[16] C. Hughes and A. Southern, "The world of work and the crisis of capitalism: Marx and the fourth industrial revolution," *Journal of Classical Sociology*, vol. 19, no. 1, pp. 59–71, 2019.