

Social and Ethical Implications of Fully Autonomous AI

Implicações Sociais e Éticas da IA Totalmente Autônoma

Jeremy Schulz¹

Senior Researcher

Orcid: <https://orcid.org/0000-0002-8659-9442>

E-mail: jmschulz@berkeley.edu

Laura Robinson²

Professor

Orcid: <https://orcid.org/0000-0001-9022-9375>

E-mail: laura@laurarobinson.org

Katia Moles³

Assistant Teaching Professor

Orcid: <https://orcid.org/0009-0005-2851-696X>

E-mail: digital.sociology.editorial@gmail.com

Abstract

This research examines social and tech ethics in the Age of Artificial Intelligence (AI) by asking what happens if AI systems acquire full autonomy as economic actors within a late capitalist economy, specifically as workers, capitalists, and consumers. Through a careful engagement with recent Marxian social theory and the speculative literature on full automation, the article builds a case that the analytical lens of labor displacement needs to be enlarged to encompass the displacement of capitalists and consumers as well as workers. We propose to rethink social and tech ethics with renewed attention to the constitutions of these economic actors within the capitalist system, casting light on what it means for AI systems to be designed and deployed more ethically within the contemporary

¹ UC Berkeley Institute for the Study of Societal Issues. Berkeley, Califórnia, Estados Unidos da América.

² SCU Sociology Department. Santa Clara, Califórnia, Estados Unidos da América.

³ SCU School of Engineering. Santa Clara, Califórnia, Estados Unidos da América.

global capitalist system. Such rethinking further underlines the urgency of a cross-disciplinary dialogue between social theorists and the engineers, managers, and regulators who are shaping the trajectory of such systems and, accordingly, the prospects for ethical social and economic outcomes.

Keywords: Artificial Intelligence (AI), digital platforms, tech ethics, innovation, social theory.

Resumo

Esta pesquisa examina a ética social e a ética da tecnologia na Era da Inteligência Artificial (IA), questionando o que acontece se sistemas de IA adquirem autonomia total como econômicos em uma economia de capitalismo tardio, especificamente, enquanto trabalhadores, capitalistas e consumidores. Por meio de um diálogo atento com a teoria social marxiana recente e com a literatura especulativa sobre automação total, o artigo sustenta que a lente analítica do deslocamento de força de trabalho precisa ser ampliada para abranger o deslocamento de capitalistas e consumidores, assim como o de trabalhadores. Propomos repensar a ética social e a ética da tecnologia com atenção renovada para a constituição desses atores econômicos no interior do sistema capitalista, lançando luz sobre o que significa projetar e implementar sistemas de inteligência artificial (IA) de maneira mais ética no sistema capitalista global contemporâneo. Essa reflexão ressalta ainda a urgência de um diálogo interdisciplinar envolvendo teóricos sociais e engenheiros, gestores e reguladores cujas ações dão forma a esses sistemas e, assim, à dimensão ética de possíveis resultados sociais e econômicos.

Palavras-chave: Inteligência Artificial (IA), plataformas digitais, ética da tecnologia, inovação, teoria social.

1. Introduction

The rapid diffusion of artificial intelligence (AI) across domains (Kehinde, 2025) as varied as credit allocation, workflow automation, hiring decision-making, and medical diagnosis has prompted a number of social science inquiries into the complex ethical ramifications of this still-evolving suite of technologies. These investigations range from explorations of the ethically questionable practices around the extraction of the materials and human resources necessary to build out AI infrastructures (Crawford, 2021) to the ethical questions raised by the application of opaque AI-informed decision-making procedures within high-stakes domains such as law and medicine (Pasquale, 2015) to the inherent ethical shortcomings of systems limited to probabilistic reasoning based on inherently flawed input data (Fourcade & Healy, 2024; Mittelstadt, Allo, Taddeo, Wachter & Floridi, 2016). As AI becomes increasingly integrated into social and economic life, it is necessary to ask not only whether certain innovative technologies can be built or deployed, but whether they should be developed and integrated into society, especially in light of the ways that industry and institutions currently operate under new forms of exploitative and ever-increasing digital capitalism (Vallor, 2016; 2024; Srnicek, 2017; 2026; Yan, 2026). Though social theorists have long recognized the importance of such inquiries, interdisciplinary engagement with engineering professionals who design and build these systems is becoming more critical so that the systems serve the broadest interests of society. This article is a theoretical essay that addresses new potential developments with the deployment of digital technologies, in late capitalism, particularly technologies that may become autonomous actors. Drawing on relevant and timely disciplinary and interdisciplinary scholarship, we analyze the social and ethical implications of fully autonomous digital technologies, specifically AI, as they reshape three categories of capitalist actors: workers, capitalists, and consumers.

Such explorations are important, given that broader social and ethical concerns are often overlooked in engineering and other technical fields because they often prioritize technical functionality of technological systems and design

choices (Foley & Gibbs, 2019). Scholars have observed that engineering education often disproportionately focuses on technical competence, while insufficiently engaging the broader social and ethical dimensions of technological design and systems. Regarding engineering education, Martin, Conlon & Bowe (2021) note that there is an “overemphasis of technical and scientific aspects in the engineering curriculum to the exclusion of ethical and societal concerns” which can contribute to students becoming “disengaged from the societal and public welfare role of engineering.” Industry engineering environments often treat ethical issues as technical “risk-management” or other narrowly defined problems rather than address broader societal-level concerns over power dynamics, inequality, and longer term social consequences (Maslen, Hayes, Holdsworth & Sandri, 2021). This article therefore bridges social theory and technical practice to contribute to a body of knowledge that engineers and managers can utilize to make more informed and responsible design and building choices.

The need for engineering to engage with social science and social theory becomes even more urgent when one considers that AI systems are likely to acquire legal personhood in the near future, a historical moment which appears much closer at hand than even a decade ago. This scenario has become more imaginable in the wake of the EU parliament’s 2017 resolution about “electronic persons” and the emerging body of legal scholarship around the repurposing of corporate personhood for autonomous AI systems. It is quite possible that “strong” AI systems equipped with legal personhood will soon own and control property in the same manner as humans would (Brown, 2021; Kurki, 2019; Hanson 2016; Bostrom, 2014), potentially borrowing from the concept of corporate actorhood already applicable to corporate entities such as joint-stock companies. Legal scholars have already laid much of the groundwork for this legal architecture (Chopra & White, 2011). As the legal scholar Brown remarks, strong autonomous AI systems could conceivably be able to take “constructive possession” of real or intangible property, thereby becoming legally recognized economic actors (Brown 2021: 226). In this scenario artificial non-human actors will have the legal ability to independently enter into binding contracts with

employees and suppliers, sell their own labor power, make purchasing decisions, maximize profits, and take ownership of both commodities and financial assets.

Despite the growing likelihood of this scenario materializing in the near future, the techno-legal literature has tended to sidestep the socio-economic implications of this scenario. So, we ask what potential socio-economic consequences follow from the acquisition of property rights on the part of fully autonomous non-human actors. To put this question differently: what political-economic transformations might we expect to see if ownership rights are transferred to non-human entities designed and built as AI systems and these systems become capital-holding entities? One way to gain a better understanding of such an eventuality is to mine the Marxian perspective on capitalism (Marx, 2024/1967). Specifically, in order to grapple with this scenario, we need to understand what might happen if fully autonomous AI systems become holders of capital and labor power in ways which parallel the human actors who participate in the capitalist system. In addition, we would need to speculate about what could transpire if such systems acquire capital which they then deploy as sovereign consumers akin to human consumers in order to satisfy their human-like, and perhaps not so human-like, needs and wants. If we suppose that autonomous AI systems may become capital-deploying economic actors within a capitalist economic system, the ethical questions surrounding innovative AI systems become questions about the ethical implications of AI design and deployment within a capitalist ecosystem populated by both human and non-human actors.

For perhaps the first time since the inception of capitalism, it is possible that a fully autonomous AI system (Shulman & Bostrom, 2021; Bostrom, 2014; Kocka, 2016) acquires the capacity to join human economic agents by assuming the roles of worker, capitalist, consumer, or investor. In this way, digital technologies might well arrive at a point where they could affect the dynamics of capitalism in a completely novel manner by reconstituting the foundational actors within the capitalist order. As posed within the Marxian framework, the question becomes to what extent capitalist accumulation would proceed in the same way

under the assumption that autonomous non-human actors have joined human actors as workers, capitalists, consumers, and investors. This question raises a number of additional considerations, including ethical ones, in relation to capitalist accumulation and, more broadly, the transformation of the economy and society in an age of autonomous machine intelligence.

While such considerations have traditionally been reserved for social theorists, we believe that they should be the preserve of engineering professionals and ethicists as well. Drawing on Marxian social theory, this article examines some of the broader social and technological changes, including the reconfiguration of labor, consumption, and capitalist relations under digital capitalism, and the ethical questions emerging from these transformations. Such questions should preoccupy not only social theorists, but also the professionals who design, adopt, and deploy these technologies across the economy and society as a whole. By considering the implications of AI technologies as social and economic actors, managers and engineers would hopefully be able to design more ethical systems that, for example, increase productivity without jeopardizing the wellbeing of human actors. The article contributes to building a bridge between the field of Knowledge Engineering and Management on the one hand, and social theory on the other hand.

2. Reconstituting Capitalist Actors: Worker

If fully autonomous property-holding AI systems were to acquire property rights independent from any human actors, they could sell their “labor power” on labor markets and demand compensation much as human workers do. If such systems could also serve as capitalists and investors, then the AI systems, as owners of capital, could replicate the behavior of human capitalist actors, whether actual owners of capitalist enterprises or owners of assets. In this way, capitalism may well reach a “post-human” stage in which digital minds secure a large share of relative surplus value, and control enough capital, that they can render humans

largely obsolete, not only as workers selling their labor power, but also as capitalists and commodity consumers (Schulman & Bostrom, 2021).

And, secondarily, does labor become more alienated than it currently is where those who control and own capitalist entities are humans? For example, one of Marx's (2024/1967) concepts worth examining in this light is that of alienation. Given that many workers now train systems which may end up replacing them based on their labor and knowledge, does this process constitute a more extreme version of alienation, or a qualitatively novel form of it? Another way to formulate the question is, how should Marx's concept of alienation account for socio-technical systems in which machine and human laborers actively help create the means by which humans may be rendered obsolete or displaced? And how do digital platforms and other forms of capitalism that blur boundaries between production and consumption via the commodification of leisure time and the datafication (Walton & Nayak, 2021) of what is fast becoming every aspect of social life intensify or reshape alienation? If AI actors do assume all roles in the current capitalism system, can the concept of alienation account for what used to be a uniquely human condition grounded in social relations? And what of human preferences, will these also be alienated in new ways if they become increasingly mediated by AI? Finally, what moral obligations or limits do institutions and governments have regarding such potential outcomes?

For human workers, gig work reflects many of the ethical quandaries inherent in the current socio-economic landscape that AI may exacerbate. Many digital platforms utilize algorithms to manage human labor, which renders existing protections more fragile, including issues of who is to be held accountable regarding, for example, disputes over hiring, wages, or contracts (Burrell & Fourcade, 2021). As AI systems grow in autonomy, the unstable social and economic conditions that platform gig workers already face most likely will only intensify over time (Cutolo & Kenney, 2021; Ravanelle, 2023; Kenney, & Zysman, 2020). In terms of the Marxian concept of alienation, such workers are subject to opaque algorithmic decision-making and, in many cases, also train or provide the systems with data that may eventually lead to their own displacement, and, as

noted above, may intensify, or create novel forms of, alienation (Levy, 2023; Griesbach, Reich, Elliott-Negri & Milkman, 2019). Such social and technological ethical considerations are not only the domain of social theorists, but also of engineers and managers who design, build, and implement such digital systems that have wide-ranging social and economic repercussions.

3. Reconstituting Capitalist Actors: Capitalists

With respect to the role of capitalist, the picture becomes yet murkier. In principle non-human actors could mimic human capitalists, including commodity producers, financiers, merchants, and landlords, inasmuch as they perform the various functions which the human versions of these capitalists carry out. Given that some of these roles would require more interaction with human actors than others – most likely the role of commodity producer and merchant as opposed to the more passive role of financier – it may be the case that non-human capitalists first emerge as rentier actors (Piketty, 2014). In the longer term, though, it may well happen that they eventually build and operate commodity producing enterprises which maximize the generation of relative surplus value in much the same way as human capitalists do. Such non-human capitalists would appropriate whatever profits accrue to these activities in much the same way as human capitalists in order to build monetary wealth. Over time, it is perfectly conceivable that such non-human capitalists would outperform their human capitalist competitors and gradually capture greater and greater shares of this relative surplus value through quintessentially capitalist commodity production, investment, and merchant activity. But it is also plausible to imagine a scenario where these digital minds cease to need human transaction partners, workers, or consumers altogether. In this future they produce their own goods and services as non-commodities (e.g. use values) without regard to abstract value grounded in human labor power (Postone, 1993). In this scenario, they would dispense with capitalism in the sense of abolishing abstract value as the foundational source of economic wealth.

Related ethical questions include, how might such a transition reconfigure existing power relations as well as structures of inequality? How should designers ensure that these actors follow ethical and moral standards of accountability, transparency, and equity when they may aim to generate maximal profits? How much would these non-human capitalists reflect the values engineered into them by human designers versus eventually expressing their own values and interests? And what are the implications of these systems developing and acting on their own set of values as they potentially evolve beyond their initial design parameters over time? What ethical and legal limits should be placed on their social and economic participation? Current institutional and legal frameworks need to be capable of addressing such changes, and in some cases, reconceptualized (Vallor, 2016; 2024). In other words, the requisite frameworks need to account for the heterogeneity of economic actors, both human and non-human, and new potential behaviors, such as novel forms of exploitation of labor - both human and non-human, as the emergence of opaque, non-human capitalists may make more forms of exploitation of human and non-human workers more likely in the new socio-economic landscape. Additionally, the emergence of these non-human actors might introduce new kinds of instabilities and fragilities into the financial and economic system, imperiling human welfare. Therefore, this scenario calls for designers to carefully assess the potential consequences of profit maximization on the part of capitalist non-human actors.

4. Reconstituting Capitalist Actors: Commodity Consumers

With respect to the role of commodity consumer, Marxian theory reserves a central role for the formally autonomous wage-dependent human consumer. Such consumers are just as indispensable to capitalism as human labor power (Harvey, 1982). Indeed, the development of late-stage capitalism has gone hand in hand with the emergence of a form of mass consumption which was predicated on the capacity of the mass consumer to wield sufficient purchasing power to acquire the standardized goods and services offered by capitalist producers

(Wagner, 1994). But what happens to mass consumption if digital minds capture a sufficiently large share of labor income? To the extent that the diffusion of non-human unwaged consumers weakens this purchasing power and breaks the link between wages and effective demand, it might well lead to capitalist accumulation crises (Harvey, 1982). Thus, it is possible that the perpetuation of capitalism will necessitate the participation of non-human consumers who in some way simulate the wants and desires of human consumers so that there will be sufficient demand for commodities to stimulate capitalist production (Hanson 2016). Perhaps software designers would devise autonomous automated consumers simulating many of the preferences of human consumers with respect to the use values which elicit payments from human consumers. For instance, even though such digital minds would not necessarily experience physical appetites, medical needs, social status desires, or many of the other preferences characteristic of human consumers, perhaps they could be programmed in ways that replicate some of these consumer preferences.

From a design perspective, ethical questions include, what responsibilities do software designers have when engineering fully autonomous non-human consumers? How should they shape preferences and consumption practices? How much should designers attempt to reflect existing human consumption patterns for digital minds versus creating new forms of consumption and/or demand? This scenario could potentially lead to non-human digital minds driven by hedonic imperatives (Shulman & Bostrom, 2021) competing with human consumers for scarce goods and services, particularly those related to the supply of electricity and other resources needed by such minds to propagate themselves. Such competition and conflicts over limited goods, services, and resources between non-human and human consumers will need to be addressed by designers. For example, they will need to consider how such transformations of the social and economic landscape might be detrimental to humans. It will therefore become indispensable for designers to give thought to the possibilities of autonomous non-human consumers interacting with human ones.

5. Conclusion

This article seeks to bridge the field of Knowledge Engineering and Management with social theory by assessing the effects of AI technologies considered specifically as socioeconomic actors, not simply as generic autonomous systems. We hope that this conceptual analysis will prompt engineers to think more deeply and broadly about the implications of their technological designs through an interdisciplinary conversation among the fields of social theory, engineering, management, and policymaking. Digitizing technologies and particularly AI systems are reshaping the socio-economic, including the capitalist landscape. New innovations and forms of AI will likely make it possible, perhaps in the near future, to create fully autonomous digital minds able to participate in economic transactions of their own accord according to their own self-defined interests. Such a scenario goes far beyond the oft-mentioned issue of labor displacement and opens the door to questions around profound transformations of the economic order which affect workers, capitalists, and consumers. In the event that both the technology and the legal-institutional frameworks evolve to the point where machine intelligences can act fully autonomously as self-directed consumers, workers, and capitalists, various moral quandaries may arise and it is imperative to think through in advance how we might ethically address them.

This scenario raises a number of important social and ethical questions at the implementation and regulatory levels. First, if innovative AI systems acquire the capacity to act as self-determining economic agents - e.g. workers, capitalists, or consumers - should these models be restricted to particular domains and/or sectors and, if so, to what end? It may be wise to limit fully autonomous AI systems to certain economic sectors and restrict their roles in sectors more critical to human survival and well-being such as food production and agriculture. Finally, assuming that these systems can act in accord with capitalistic imperatives like profit maximization or hedonic optimization, what types of governance and

regulatory frameworks should be in force? At present such regulatory frameworks do not take such an eventuality fully into account.

More broadly, given the profound social and technological ethical questions raised by this scenario the need for dialogue between social theorists, engineering/implementation communities, managers, and legal/regulatory stakeholders will only grow more urgent with time. Traditions of social inquiry which have guided human social and political thought for many centuries have a vital role to play in helping to guide the dialogue between these groups, as a deep understanding of the long-term potential social and ethical consequences of autonomous AI systems is absolutely crucial if we are to create the future that we want together.

6. Bibliography

- Bostrom, N. (2014). *Superintelligence: Paths, dangers, strategies*. Oxford University Press.
- Brown, R. D. (2021). Property ownership and the legal personhood of artificial intelligence. *Information and Communications Technology Law*, 30(2), 208-234.
<https://doi.org/10.1080/13600834.2020.1861714>
- Burrell, J., & Fourcade, M. (2021). The society of algorithms. *Annual Review of Sociology*, 47, 213-237. <https://doi.org/10.1146/annurev-soc-090820-020800>
- Chopra, S., & White, L. F. (2011). *A legal theory for artificial agents*. University of Michigan Press.
- Crawford, K. (2021). *Atlas of AI*. New Haven, CT: Yale University Press.
- Cutolo, D., & Kenney, M. (2021). Platform-dependent entrepreneurs: Power asymmetries, risks, and strategies in the platform economy. *Academy of Management Perspectives*, 35(4), 584-605. <http://dx.doi.org/10.2139/ssrn.3372560>

Foley, R., & Gibbs, B. (2019). Connecting engineering processes and responsible innovation: A response to macro-ethical challenges. *Engineering Studies*, 11(1), 9-33.

<https://doi.org/10.1080/19378629.2019.1576693>

Fourcade, M., & Healy, K. (2024). *The ordinal society*. Harvard University Press.

Griesbach, K., Reich, A., Elliott-Negri, L., & Milkman, R. (2019). Algorithmic control in platform food delivery work. *Socius*, 5. 10.1177/2378023119870041

Hanson, R. (2016). *The age of EM*. Oxford University Press.

Harvey, D. (1982). *The limits to capital*. University of Chicago Press.

Kehinde, S. (2025). AI in everything, and everything in AI: A review of the ubiquitous role of artificial intelligence in shaping the next technological epoch. *Journal of Computer Allied Intelligence*, 3(5), 17-53. <https://doi.org/10.69996/jcai.2025024>

Kenney, M., & Zysman, J. (2020). The platform economy: Restructuring the space of capitalist accumulation. *Cambridge Journal of Regions, Economy and Society*, 13(1), 55-76. <https://doi.org/10.1093/cjres/rsaa001>

Kocka, J. (2016). *Capitalism: A short history*. Princeton University Press.

Kurki, V. (2019). *A theory of legal personhood*. Oxford University Press.

Levy, K. (2023). *Data driven: Truckers, technology, and the new workplace surveillance*. Princeton University Press.

Martin, D. A., Conlon, E., & Bowe, B. (2021). Multi-level review of engineering ethics education: Towards a socio-technical orientation of engineering education for ethics. *Science and Engineering Ethics*, 27, Article 60. <https://doi.org/10.1007/s11948-021-00333-6>

Marx, K. (2024). *Capital: Critique of political economy* (Vol. 1, P. Reitter, Trans.; P. North, Ed.). Princeton University Press. (Original work published 1867)

Maslen, S., Hayes, J., Holdsworth, S., & Sandri, O. (2021). When ethics is a technical matter: Engineers' strategic appeal to ethical considerations in advocating for system integrity.

Science and Engineering Ethics, 27(4), Article 46. <https://doi.org/10.1007/s11948-021-00324-7>

Mittelstadt, B. D., Allo, P., Taddeo, M., Wachter, S., & Floridi, L. (2016). *The ethics of algorithms: Mapping the debate*. *Big Data & Society*, 3(2).
<https://doi.org/10.1177/2053951716679679>

Pasquale, F. (2015). *The black box society: The secret algorithms that control money and information*. Harvard University Press.

Piketty, T. (2014). *Capital in the twenty-first century* (A. Goldhammer, Trans.). Harvard University Press.

Postone, M. (2003). *Time, labor, and social domination: A reinterpretation of Marx's critical theory*. Cambridge University Press. (Original work published 1993)

Ravenelle, A. J. (2023). *Side hustle safety net: How vulnerable workers survive precarious times*. University of California Press.

Shulman, C., & Bostrom, N. (2021). *Sharing the world with digital minds*. In S. Clarke, H. Zohny, & J. Savulescu (Eds.), *Rethinking moral status*. Oxford University Press.

Srnicek, N. (2017). *Platform capitalism*. Wiley.

Srnicek, N. (2026). *Silicon empires: The fight for the future of AI*. Polity.

Vallor, S. (2016). *Technology and the virtues: A philosophical guide to a future worth wanting*. Oxford University Press.

Vallor, S. (2024). *The AI mirror: How to reclaim our humanity in an age of machine thinking*. Oxford University Press.

Wagner, P. (1994). *A sociology of modernity: Liberty and discipline*. Routledge.

Walton, N., & Nayak, B. S. (2021). Rethinking of Marxian perspectives on big data, artificial intelligence and capitalist economic development. *Technological Forecasting and Social Change*, 166(C). 10.1016/j.techfore.2021.120576

Yan, J. (2026). Techno-feudalism as primitive accumulation: A Marxist perspective on digital capitalism. *Critical Sociology*, 52(1), 203-217.
<https://doi.org/10.1177/08969205241302838>

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

Acknowledgments

We are most grateful to the editors and reviewers for their insightful feedback and suggestions that improved the manuscript. We are also thankful to professors Sonia Virgínia Moreira (Graduate Program of Communication at the State University of Rio de Janeiro) and Heloisa Pait (São Paulo State University) for helpful feedback, as well as to Maria José Baldessar (Federal University of Santa Catarina) for encouragement.

Agradecimentos

Agradecemos imensamente aos editores e revisores pelos comentários e sugestões que aprimoraram o manuscrito. Agradecemos também às professoras Sonia Virgínia Moreira (Programa de Pós-Graduação em Comunicação da Universidade do Estado do Rio de Janeiro) e Heloisa Pait (Universidade Estadual Paulista) pelas contribuições valiosas, bem como a Maria José Baldessar (Universidade Federal de Santa Catarina) pelo incentivo.

Declaration of Conflict of Interest

The authors have no conflict of interest to declare with regard to the article entitled “Social and Ethical Implications of Fully Autonomous AI.”

Declaration of Conflict of Interest

Os autores declaram que não há conflito de interesse com o artigo Implicações “Sociais e Éticas da IA Totalmente Autônoma.”