The Danger of Smart Ideologies: Counter-Hegemonic Intelligence and Antagonistic Machines

Accepted for publication in Sion Lindgren, (ed). 2023. Handbook of Critical Studies of Artificial Intelligence.. Edward Elgar. Cheltenham. https://doi.org/10.4337/9781803928562

The Danger of Smart Ideologies: Counter-Hegemonic Intelligence and Antagonistic Machines

Peter Bloom[1] (0000-0003-2225-6149)

[1] Professor of Management, University of Essex

Abstract (150 words)

This chapter will attempt to critically reframe the debate on the relationship of AI and human by highlighting the often overlooked ways AI reproduces dominant ideologies. This form of "smart hegemony" goes beyond merely challenging the all too human prejudices reproduced by AI around race and gender. Rather, it draws attention to the ways AI is used to fundamentally reproduce and reinforce dominant ideological assumptions of the world - particularly linked to capitalism. Yet this chapter will also seek to provide a glimpse at the more radical counter-hegemonic possibilities of "trans-human relations". To this end, its goal is to politically resituate the discourses around AI so that they can be part of movements to challenge often naturalised values and offer the potential for creating new emancipatory alternative ones. As such, it proposes the need for "counter-hegemonic intelligence" and the rise of "antagonistic machines

Keywords (6 keywords): Artificial Intelligence; Ideology; Hegemony; Trans-Human Solidarity; Counter-Hegemonic Intelligence; Antagonistic Intelligence

Introduction

In 2021, the EU produced the first ever regulation for AI. Its goal was to decrease its risk to humans while acknowledging the were heading toward an increasingly "smart" world. It declares:

> "The proposed AI regulation ensures that Europeans can trust what AI has to offer. While most AI systems pose limited to no risk and can contribute to solving many societal challenges, certain AI systems create risks that we must address to avoid undesirable outcomes."

For this purpose, it categories AI into the four risk categories of "unacceptable risk", "high risk", "limited risk", "minimal or no risk" based on how dangerous they are to humans. Though these regulations also contained some very telling "omissions" including "It leaves

Big Tech virtually unscathed. It lacks a focus on those affected by AI systems, apparently missing any general requirement to inform people who are subjected to algorithmic assessment" (MacCarthy and Propp, 2021: n.p.).

These regulations reflect a wider discourse around the threat of "biased" AI - in which what was once assumed by many to be merely objective calculations by machines are now understood to be worryingly prejudiced in their produced insights. The emphasis now increasingly is on "managing" (Roselli et al., 2019) or "tackling" bias in AI (Silberg and Manyika, 2019). No less than the Harvard Business Review opined in 2019 that

> "Over the past few years, society has started to wrestle with just how much human biases can make their way into artificial intelligence systems—with harmful results. At a time when many companies are looking to deploy AI systems across their operations, being acutely aware of those risks and working to reduce them is an urgent priority" (Manyika et al., 2019: n.p.)

These concerns have also brought about real world results, such as the recent news in 2022 that Microsoft is now banning the use of its facial recognition technology to identify individuals and populations based on demographic information such as race or gender (Hern, 2022).

There is an interesting paradox at work here, though. While ostensibly focused on protecting humans - these regulations and discourses in fact represent the threat posed by humanity to AI. "Over the last few years, society has begun to grapple with exactly how much these human prejudices, with devastating consequences," notes the head of *Head of Data Science and AI at* Australian Computer Society, Steve Nouri (2021: n.p.), "can find their way through AI systems". In particular, the ability for our machine counterparts to take on our human biases or be used for the weaponised purposes of reinforcing existing power relations between humans. These are perfectly understandable fears and ones that must be taken seriously. However, they do not go far enough. What is really crucial, in this respect, is to understand the deeper ideological biases that continue to hegemonically shape and order society. According to David Magerman (2022), Managing Partner of the AI entrepreneurial firm *Differential Ventures,*

> "The root cause of bias in data-driven AI systems comes down to how the data to train those systems is collected, and whether or not the decision-making in that data represents the corporate or societal goals of the deployed system based on that data" (n.p.)

This chapter will attempt to critically reframe the debate on the relationship of AI and

human by highlighting the often overlooked ways AI reproduces dominant ideologies. This form of “smart hegemony” goes beyond merely challenging the all too human prejudices reproduced by AI around race and gender. Rather, it draws attention to the ways AI is used to fundamentally reproduce and reinforce dominant assumptions of the world - particularly linked to capitalism. Significantly, it interrogates AI’s role for strengthening and expanding hegemony using theories of “trans-human relations” that emphasise the mutually constituting roles machines and humans for socially constructing their shared worlds. At stake, in this analysis is to challenge overblown and misguided fears of an AI takeover as manifested in discourses of “singularity” and instead bring to the fore the very real danger in reaffirming and spreading dominant ideologies and the dominating power relations they help support.

Yet this chapter will also seek to provide a glimpse at the more radical counter-hegemonic possibilities of “trans-human relations”. Specifically, in the ability of AI to precisely identify these ideological biases and trace out their diverse intersectional impacts on individuals. Such advances would help to critically navigate and challenge the arguably inherently neoliberal ideology driving both the development of AI and algorithmic governance. In doing so, it will seek to point to how machine learning can be a revolutionary force for more fully understanding the power of ideologies and the different consequences they have for different people. To this end, its goal is to politically resituate the discourses around AI so that they can be part of movements to challenge often naturalised values and offer the potential for creating new emancipatory alternative ones. As such, it proposes the need for “counter-hegemonic intelligence” and the rise of “antagonistic machines”.

The Danger of Smart Ideologies

The heralded social promise of AI has been eclipsed by its less than ideal, and often quite profoundly troubling, realities. Rather than be a force for dramatically augmenting human knowledge and capabilities, it is charged with reinforcing existing biases and inequalities. Rather than making us “smarter” what appears to be happening is that quoting a recent headline from The Guardian AI “**i**s learning all our worst impulses” leading to the rise of “racist robots” (Buranyi, 2017). On the one hand, this reflects an abiding need to safeguard individuals from the threat of AI. On the other, this represents the imperative to protect AI from human prejudices and desires to use it for purposes of power rather than broader social betterment (see Metz, 2021).

What is especially dangerous is how AI can grant these biases a veneer of objectivity.

Discriminatory interpretations of natural language processing, for instance, can make it appear that these are "natural" or scientific rather than socially produced and based on faulty human data (Caliskan, 2021**)**. These biases, furthermore, have significant real world consequences. Already there have been substantiated reports that million of black people have been discriminated against by health-care algorithms in US hospitals (Ledford, 2019). This tendency toward discrimination is also built into its design - as in the name of efficiency its models will disregard the consequences of errors on what it views as rare but in actual serious errors (Hellemans, 2020).

Amidst the legitimate and high profile concerns over AI bias there is perhaps an even more fundamental one that is missing. Notably, the underlying assumptions regarding race, class, gender, and value that are driving these prejudicial assumptions. Just this year in 2022, the US Department of Commerce's National Institute of Standards and Technology (NIST) published "Towards a Standard for Identifying and Managing Bias in Artificial Intelligence" (Schwartz et al., 2022: i) that declares in its opening paragraph

> "As individuals and communities interact in and with an environment that is increasingly virtual, they are often vulnerable to the commodification of their digital footprint. Concepts and behavior that are ambiguous in nature are captured in this environment, quantified, and used to categorize, sort, recommend, or make decisions about people's lives. While many organizations seek to utilize this information in a responsible manner, biases remain endemic across technology processes and can lead to harmful impacts regardless of intent. These harmful outcomes, even if inadvertent, create significant challenges for cultivating public trust in artificial intelligence (AI)."

They distinguish, to this end, beyond surface level "statistical and computational biases" with the respectively deeper "human bias" and "systemic bias". At stake, hence , are the ideological biases being literally coded into our knowledge and decision- making by AI (REF - MIT technology review). These represent entrenched social values underpinned by even more naturalised and culturally sedimented understandings of the world and ourselves as human beings - specifically how AI "encodes" this systemic human bias into our decision-making (Raji, 2020: n.p.).

These increasingly popular critiques of AI also reflect deeper issues of the ways algorithmic governance and speculative logics represent hegemonic neoliberal ideologies. In particular, it perpetuates a form of ongoing human disciplining that seeks to maximize individual and collective efficiency. Here, the goal is to create an "optimized" person and society that can be achieved through ongoing practices of AI powered quantification and calculation. The focus on "smart" biases, hence, risks distracting from its perhaps fundamental neoliberal bias, as these approaches are attempting to socially and political discipline a new generation of populations to be economically productive and valuable subjects.

These insights open up the space to reconsider and update our conceptions of hegemony. AI stands a socially vaulted "truth teller" that is able to present its biased insights as calculated fact. Consequently, it follows in a tradition of granting ideology a scientific veneer (Vesa and Tienari, 2020). Today there is emerging what can be called "smart hegemony" - the unreflective reproduction of dominant human discourses for further spreading these values and using them for the conjoined purposes of social ordering and human disciplining (Cui and van Esch, 2022). AI then is inherently political precisely because it is a viral and increasingly authoritative ideological force (Katz, 2020).

It is critical, therefore, to reframe the question of AI bias to better reflect issues of ideology and hegemony. At present, there is a plethora of work grappling with the "ideology of AI" (see Sias, 2021). Just as important, though, is the need to theoretically understand how AI serves to make certain ideologies hegemonic and for whose human benefit.

Transhuman Hegemony

A key aim of this chapter is to better understand the role of AI for producing and reinforcing ideological hegemony. Tellingly, dominant discourses of human machine relations focus primarily either on their ultimate conflict or the capacity for developing artificially aided super human powers. Less prominent but arguably much more important and urgent, is the social use of AI and their potential deployment for strengthening and challenging hegemonic assumptions shaping human relations. Equally crucial is the capacity of AI to help us reimagine our shared trans-human futures. The historian Apolline Taillandier (2021: 215) recounts that trans-humanism promoted stories of "futures change" that "include the unlimited individual liberty of the technologized self, the knowledge-ordering properties of the market, and the rational aggregation of individual interests over the long term."

Ernesto Laclau's work on discursive hegemony is particularly valuable for this purpose. Along with Chantel Mouffe, he associated power with the ability of a dominant discourse

to shape understandings and, in doing so, contingently regulate actions. It is “a space in which bursts forth a whole conception of the social based upon an intelligibility which reduces its distinct moments to the interiority of a closed paradigm” (Laclau and Mouffe, 1986: 93). In his later work, he further introduces the concept of “ideological imaginaries” which represent the over-arching horizon of meaning within which different discourses compete for dominance, representing that which “is not one among other objects but an absolute limit which structures a field of intelligibility and is thus the conditions of possibility for the emergence of any object’ (Laclau, 1991: 63) At the very least, then, it becomes imperative to interrogate what ideological imaginary AI emerged within and continues to be constituted by.

In the present context, technology supposeldy represents a dangerous ultra modern form of alienation. The fears of automation points to a coming “smart” future where machines will replace humans in the name of profit. The threat of human obslence looms over any and all discussions of AI - raising the prescient concern of whose interests this technology is, in fact, being made for. This sense of existential disempowerment is only enhanced by our daily relationship to technology which can feel manipulative and unmoored from any human ability to control or regulate (see Packard, 2018). Reflected is a paradoxical reality whereby as the capabilities of technology increase the power of humans decrease. While AI and digital interactions can connect us to information and people almost instantanous, customised to meet our most individualised of preferences, what is missing is the is the shared ability to “design what a future ought to be: open to regular revision in response to our practical behaviours given the persistent contingency of the conditions in which we are immersed” (Bauer, 2010: 106).

Here the distinction between human and artificial intelligence begins to critically blur. Indeed, the insights produced by machines are understood to be, to a profound extent, an extension of human discourses. Furthermore, it is reflection of a particular form of intelligence that has portrayed as a universal and objective truth. To this end, it is what Glynos and Howarth (2007) refer to as a producer of “social logic” - a form of knowledge that promotes insights and practices reinforcing dominant values and relations.

What is then crucial is to critically unpick the social logics being reproduced and granted legitimacy by AI. “Smart hegemony” is the use of artificial intelligence to naturalise pre-existing ideological bias. In doing so, it justifies it as a hegemonic truth, an unquestioned “common sense” that can be utilised for regulating and controlling human behaviour. Hence, there is a dual and interconnected process of disciplining occurring - the narrowing of machine learning to officially sanctioned and accepted discursive limits and the

weaponisation of this data and knowledge to discipline individuals and groups in conformity to these hegemonic values. Yet they also point to an alternative and less alienating type of counter-hegemonic trans-human relation based on challenging these social logics, thus allowing radical forms of AI and transformative human and machine learning to emerge that could disrupt these dominant discourses and the dominating social orders that they produce.

Counter-Hegemonic Intelligence

The concept of "smart hegemony" helps to critically reconsider the character and consequence of artificial intelligence. Rather than focus on it calculative capabilities, it highlights instead its role in ideologically reinforcing existing social orderings and their underlying power relations. These insights echo, to a degree, concerns about the rise of "algorithmic governance". (O'Neil, 2016; Sætra, 2020) Yet it also aims to radicalise existing efforts to politically deploy AI to resist this increasingly "authoritarian capitalist" status quo (Bloom and Sancino, 2019). At stake is gaining a fuller comprehension of developing and using new forms of counter- hegemonic intelligence.

The current literature on AI is, understandably, increasingly concentrated on foster more empowering types of Human-Computer Interactions (see Biele, 2022; Faulkner, 1998; Mathew et al., 2011). At the heart of this strategy, is the attempt to make AI more human-like in its reasoning. A further goal, in this respect, is also to create "explainable AI" that can enhance the transparency of machine decision-making, pulling it out of its proverbial "black box" so that humans can understand and, if necessary, challenge its rationale (Liao et al., 2020; Rai, 2020 Samek et al., 2019). These approaches would appear to contravene concurrent views of AI as generating "super intelligence" far beyond the ken of human cognition and processing (see Bostrom, 2014).

There is a radical third way though that is worth exploring and perhaps much more relevant to current social conditions. Namely, the production of counter-hegemonic intelligence - machine learning that can uncover the ideological biases of human relations. Returning to the theories of Glynos and Howarth (2007), it is an interrogation of how AI can drive forward and support political rather than social logics. At the most basic level, this means the political use of AI to challenge dominant assumptions and power. However, counter-hegemony goes beyond resistance, it is also about fundamentally re-ordering the production and consumption of knowledge - putting into question the very epistemological basis for our prevailing "common sense" and asking us to radically reassess "how we know what we know". Discursive hegemony, at its core, then is an attempt to "weave together different strands of discourse in an effort to dominate or structure a field of meaning, thus

fixing the identities of objects and practices in a particular way" (Howarth, 2000: 102).

The radical potential of AI is precisely in this possible epistemological revolution, the social transformation of knowledge itself. Already this knowledge revolution is underway through ICTS and the advent of social media which is both digitally "shrinking the world" and allowing for new, often troubling, voices and perspectives to gain unprecedented prominence for shaping views.

Significantly, the algorithms underpinning these digitalised communications profitably encourage and ultimately make almost unavoidable political "echo chambers" (See Cinelli et al., 2021; Nguyen, 2020) where based on our preferences we only hear those views most similar to our own or those so dissimilar that it evokes a provocative rather than deliberative response. The goal, here, is increased usage not an open, free, and constructive space for collective discussion and informed democratic decision-making.

By contrast, counter-hegemonic intelligence would prioritise the ability of machine learning to quantify and qualify the humanity's ideological bias. Crucially, it would seek to go beyond mere explicability. By contrast, it would use its advanced capabilities to precisely identify the dominant discourses driving social decision-making. This could range, for example, from the perceived belief in the importance of competition and entrepreneurship for economic development to efforts to the authoritarian reasoning and intentions driving strategies linked to AI powered "social credit" scoring for disciplining populations. (ref, quote). This would involve new "human in the loop" approaches that could allow human and machine collaborations for this counter-hegemonic knowledge production.

Crucially, it would permit AI from being simply innovative to be fundamentally transformation in its use and effects. At present, it is being deployed most for improving existing systems - often for quite morally troubling ends. Whether this be the employment of "sentient" analysis on workers by employers to monitor and control their behaviour, the Uber-like application of AI powered platforms to aid economic exploitation of an increasingly precarious workforce, or the deployment of hi-tech tracking systems for ushering in a new era of global detention, incarceration, and "rehabilitation". By recalibrating AI to actually identify these hegemonic discourses, it enhances its ability for enacting systemic social change.

Antagonistic Machines

A continual discourse surrounding AI is its overall threat to humanity. Initially, this was couched in terms of the existential fears of machines turning on their creators - threatening to make humans outdated. Even the late brilliant Stephen Hawkins once

warned

> The development of AI could spell the end of the human race. It would take off on its own and redesign itself at an ever increasing rate. Humans, who are limited by slow biological evolution, couldn't compete and would be superseeded. (Quoted in Cellan-Jones 2014: n.p.)

These fears have evolved in present time to the danger of AI controlling human behaviour and stealing their data for profitable and nefarious purposes. "Digital mechanization has also smoothed the way for the growth of insecure and underpaid jobs. This reflects the socio-political features of neoliberal capitalism and not the intrinsic attributes of technology per se", writes Professor Peter Fleming (2019: 217), "Because of this, robotic automation might even help deepen the institution of paid employment in Western economies, not release us from it."

The attempts to render AI more trustworthy (Chatilla et al., 2021; Kaur et al., 2022) is, hence, premised on a desire to lessen these long standing and increasingly serious concerns. Yet it also risks politically limiting the radical potential of AI as a disruptive force for positive change. The alarm over human conquering robots may appear to be primarily in the realm of science fiction. However, it has actually driven much of the past and to an extent present theory of AI, notably linked to the concept of singularity whereby humans will become so smart they will overtake the limited intelligence of humans (see Good, 1965) Yet the threat of singularity is firmly rooted in human values of power, control, and exploitation. The scholar Gregory Jerome Hampton (2015: 2) insightfully observes, in this regard, that

> "Slavery, after all, was largely invested in producing and controlling a labor force, which was dissociated from humanity. In many regards, American slavery was a failed experiment to employ flesh and blood machines as household appliance, farm equipment, sex toys, and various tools of industry without the benefit of human and civil rights. Consequently, what is interesting about the development and production of mechanical robots is how they are being assigned both race and gender as identity markers. Why does a machine need such a complex identity, if the machine is designed only to complete the mundane labour that humanity wishes to forego? One

> plausible response is that the machine is being designed to be more than an appliance and less than a human. The technology of the 21st century is in the process of developing a modern day socially accepted slave."

It assumes, hence, that machines would seek to be our masters as that is rationale by which humans have historically operated. If singularity is a threat - which itself is a questionable assertion based on current technological developments - then it is fundamentally a human created one that can only be addressed through a profound transformation of social values.

An appealing alternative would be to construct a new more cooperative basis for constructing trans-human relations. These values are already reflected, for instance, in ideas of cooperative robotics (see Johnson et al., 2011; Khoshnoud et al., 2020). Internet of Things has similarly begun adopting these more collaborative principles (See Jiang, 2019). Fundamentally, this speaks to notions of what I have previously termed "mutually intelligent design" between humans and machines in which

> "Ultimately, perhaps the greatest hope for the politics of the twenty-first century is the transformation of posthuman emancipation into transhuman liberation. More precisely, the escape, the freedom, from human based oppression and tyranny for an integrative human and non-human society that creatively collaborates and cooperates in the making and remaking of their shared social realities" (Bloom, 2020: 201).

Yet this theory should not be mistaken for trying to reduce the political possibilities of AI. Rather, it should be a force for challenging hegemonic values and power relations. Laclau and Mouffe refer, in this respect, to the importance of social antagonisms. They note that "Antagonisms are not objective relations, but relations which reveal the limits of all objectivity. Society is constituted around these limits, and they are antagonistic limits" (Laclau and Mouffe, 1986: xiii– xiv). The presence of antagonisms is not just one of resistance. They are also prefigurative. They present an alternative view of what society could and arguably should be.

What then would be the prospects of developing not only counter-hegemonic intelligent but antagonistic machines? A critical step in this direction is through the use of AI powered simulations to help reveal these other possible worlds. As Häkli (2018: 173) contends

> "We should study further the idea of a humanised posthumanism, building on an ontology of possibility that acknowledges our assembled entanglement with the non-

> human world but also accords an important role for humans in acknowledging these interdependencies. As a move beyond monist posthumanism, instead of portraying mastery over passive nature, this position builds on the idea of political responsibility for the vulnerabilities, injustices, and hazards that our assembled life of dual being in and with the nature entails. It also acknowledges that all ontological claims and arguments remain meaningless without the audiences to which they are directed—audiences concerned with how to
> lead a civic life in a more-than-posthuman
> world"

Doing so means expanding upon ideas of the "cyborg citizen" which emphasizes a trans-human "participatory evolution" where we should shape our future through multiple human choices, incomplete and contradictory as they often are. Participatory government is the same…Decisions about evolution should be made at the grassroots, just as political and economic decisions should be, especially now that we have begun to recognize the political evolution of cyborgs" (Gray 2006: 3)

At present, for instance, digital twin technologies permit for precisely simulating and predicting human outcomes. These could be expanded to include digital alternatives which draws on existing data in order to reflect what a radically different social order could look like - thus deploying virtual technologies to paradoxically virtually more real these counter-hegemonic perspectives. Perhaps needed above all else, though, is a renewed praxis of trans-human solidarity. The cultivation of political cultures that emphasize the common struggle of machines and humans against exploitation. Such solidarity is already informally arising in progressive movements across the world. These could be made more explicit through processes of radical knowledge sharing leading to revolutionary human-machine learning. This could take the form, for example, of chatbots or "cobots" (see Sutherland, 2018) who can provide expert advice for organisers and those interested in experimenting with radical alternatives such as cooperative ownership, commons resource governance, and degowth development.

These advances would do more than simply identify existing biases. They would also bring to the fore how AI is attempting to reinforce a neoliberal worldview and discipline people in accordance with its underlying capitalist values. It would show the potential to repurpose "intelligence" away from optimizing people as efficient human machines and instead as a resource for exploring alternative and more emancipatory social relations. The speculative logic of calculation and prediction, in turn, is transformed into an opportunity for people to discover and personalize different visions of the future.

While these possibilities are not yet fully created, they are technically possible and politically urgent. Just as AI is dramatically influencing human power and control, so to must it be designed to enhance contemporary resistance and transformation. Only by working and struggling together can a better liberated future for humans and machines alike become possible. It is ever more imperative to generate counter-hegemonic intelligence and antagonistic machines for this revolutionary purpose.

Further readings

Bloom, P. (2020). *Identity, Institutions and Governance in an AI World: Transhuman Relations*. Springer Nature.

Bloom, P., Jones, O. S., & Woodcock, J. (2021). *Guerrilla Democracy: Mobile Power and Revolution in the 21st Century*. Policy Press.

Bloom, P., & Sancino, A. (2019). *Disruptive democracy: The clash between techno-populism and techno-democracy*. Sage.

References

Bauer, D. (2019). Alienation, Freedom and the Synthetic How. Angelaki, 24(1), 106–117.

Biele, C. (2022). *Human Movements in Human-Computer Interaction (HCI)* (Vol. 996, pp. 1-142). Springer.

Bloom, P., & Sancino, A. (2019). *Disruptive democracy: The clash between techno-populism and techno-democracy*. Sage.

Bloom, P. (2020). *Identity, Institutions and Governance in an AI World: Transhuman Relations* . Springer Nature.

Bostrom, N. (2014). Taking superintelligence seriously: Superintelligence: Paths, dangers, strategies, Oxford University Press.

Buranyi, S. (2017). Rise of the racist robots – how AI is learning all our worst impulses. The Guardian, 8 August.

Caliskan, A. (2021). Detecting and mitigating bias in natural language processing. Brookings, 10 May.

Cellan-Jones, R. (2014, December 2). Stephen Hawking Warns Artificial Intelligence Could End Mankind. BBC News. Archived from the original on 30 October 2015.

Chatila, R., Dignum, V., Fisher, M., Giannotti, F., Morik, K., Russell, S., & Yeung, K. (2021). Trustworthy ai. In *Reflections on Artificial Intelligence for Humanity* (pp. 13-39). Springer, Cham.

Cinelli, M., De Francisci Morales, G., Galeazzi, A., Quattrociocchi, W., & Starnini, M. (2021). The echo chamber effect on social media. *Proceedings of the National Academy of Sciences* , *118* (9), e2023301118.

Cui, Y., & van Esch, P. (2022). Autonomy and control: How political ideology shapes the use of artificial intelligence. *Psychology & Marketing*, *39*(6), 1218-1229.

Faulkner, C. (1998). *The essence of human-computer interaction*. Prentice-Hall, Inc..

Fleming, P. (2019). Robots and Organization Studies: Why Robots Might Not Want to Steal Your Job. Organization Studies, 40(1), 23–38.

Good, I. J., "Speculations Concerning the First Ultraintelligent Machine", in *Advances in Computers*, vol 6, Franz L. Alt and Morris Rubinoff, eds, pp31-88, 1965, Academic Press.

Glynos, J., & Howarth, D. (2007). *Logics of critical explanation in social and political theory* . Routledge.

Gray, C. H. (2000). *Cyborg citizen: Politics in the posthuman age*. Routledge.

Häkli, J. (2018). The Subject of Citizenship–Can There Be a Posthuman Civil Society? Political Geography, 67, 166–175.

Hampton, G. J. (2015). Imagining Slaves and Robots in Literature, Film, and Popular Culture: Reinventing Yesterday's Slave with Tomorrow's Robot. Lexington Books.

Hellemans, P. (2020). Hidden stratification and the accuracy principle: may positive discrimination in AI be a solution? KU Leuven: Centre for IP and IT Law, 24 March.

Hern, A. (2022). Microsoft limits access to facial recognition tool in AI ethics overhaul. The Guardian, 22 June.

Howarth, D. (2000). *Discourse*. McGraw-Hill Education (UK).

Jiang, W. (2019). An intelligent supply chain information collaboration model based on Internet of Things and big data. *IEEE access*, *7*, 58324-58335.

Johnson, M., Bradshaw, J. M., Feltovich, P. J., Hoffman, R. R., Jonker, C., van Riemsdijk, B., & Sierhuis, M. (2011). Beyond cooperative robotics: The central role of interdependence in coactive design. *IEEE Intelligent Systems*, *26*(3), 81-88.

Katz, Y. (2020). *Artificial whiteness: Politics and ideology in artificial intelligence*. Columbia University Press.

Kaur, D., Uslu, S., Rittichier, K. J., & Durresi, A. (2022). Trustworthy artificial intelligence: a review. *ACM Computing Surveys (CSUR)*, *55*(2), 1-38.

Khoshnoud, F., Quadrelli, M. B., Esat, I. I., & Robinson, D. (2020). Quantum cooperative robotics and autonomy. *arXiv preprint arXiv:2008.12230*

Laclau, E., & Mouffe, C. (1986). *Hegemony and socialist strategy: Towards a radical democratic politics* (Vol. 8). Verso Books.

Laclau, E. (1990). *New reflections on the revolution of our time*. Verso Trade.

Ledford, H. (2019). Millions of black people affected by racial bias in health-care algorithms, Nature, 24 October.

Liao, Q. V., Gruen, D., & Miller, S. (2020, April). Questioning the AI: informing design practices for explainable AI user experiences. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (pp. 1-15).

MacCarthy, M. and Propp, K. (2021). Machines learn that Brussels writes the rules: The EU's new AI regulation. Brookings Institute, 4 May.

Magerman, D. (2022). Systemic Bias in Data Makes AI a Stereotype Machine. Inside Big Data, 9 February.

Manyika, J., Silberg, J., & Presten, B. (2019). What do we do about the biases in AI. *Harvard Business Review*, *25.*

Mathew, A. R., Al Hajj, A., & Al Abri, A. (2011, June). Human-computer interaction (hci): An overview. In *2011 IEEE International Conference on Computer Science and Automation Engineering* (Vol. 1, pp. 99-100). IEEE.

Metz, C. (2021). *Who Is Making Sure the A.I. Machines Aren't Racist?* The New York Times, 15 March.

Nguyen, C. T. (2020). Echo chambers and epistemic bubbles. *Episteme*, *17*(2), 141-161.

Nouri, S. (2021). The Role Of Bias In Artificial Intelligence. Forbes, 4 February.

O'Neil, C. (2016). *Weapons of math destruction: How big data increases inequality and threatens democracy*. Broadway books.

Packard, N. (2018). Habitual Interaction Estranged. International Journal of Social Sciences, 7(1), 69–94.

Rai, A. (2020). Explainable AI: From black box to glass box. *Journal of the Academy of Marketing Science*, *48*(1), 137-141.

Raji, D. (2020). How our data encodes systematic racism. MIT Technology Review, 10 December.

Roselli, D., Matthews, J., & Talagala, N. (2019, May). Managing bias in AI. In *Companion Proceedings of The 2019 World Wide Web Conference* (pp. 539-544).

Sætra, H. S. (2020). A shallow defence of a technocracy of artificial intelligence: Examining the political harms of algorithmic governance in the domain of government. *Technology in Society* , *62*, 101283.

Samek, W., Montavon, G., Vedaldi, A., Hansen, L. K., & Müller, K. R. (Eds.). (2019). *Explainable AI: interpreting, explaining and visualizing deep learning* (Vol. 11700). Springer Nature.

Schwartz, R., Vassilev, A., Greene, K., Perine, L., Burt, A., & Hall, P. (2022). Towards a Standard for Identifying and Managing Bias in Artificial Intelligence. NIST, March.

Sias, L. (2021). The Ideology of AI. *Philosophy Today*, *65*(3), 505-522.

Silberg, J., & Manyika, J. (2019). Notes from the AI frontier: Tackling bias in AI (and in humans). *McKinsey Global Institute*, 1-6.

Sutherland, D. (2018). Solidarity Forever-robots, Workers and Profitability. Australian Socialist, 24(1), 12.

Taillandier, A. (2021). "Staring into the singularity" and other posthuman tales: Transhumanist stories of future change. *History and Theory*, *60*(2), 215-233.

Vesa, M., & Tienari, J. (2020). Artificial intelligence and rationalized unaccountability: Ideology of the elites?. *Organization*, 1350508420963872.