

Artificial intelligence and rationalized unaccountability: Ideology of the elites?

Organization

1–13

© The Author(s) 2020



Article reuse guidelines:

sagepub.com/journals-permissions

DOI: 10.1177/1350508420963872

journals.sagepub.com/home/org**Mikko Vesa**  and **Janne Tienari**

Hanken School of Economics, Finland

Abstract

In this Connexions essay, we focus on intelligent agent programs that are cutting-edge solutions of contemporary artificial intelligence (AI). We explore how these programs become objects of desire that contain a radical promise to change organizing and organizations. We make sense of this condition and its implications through the idea of ‘rationalized unaccountability’ that is an ideological state in which power and control are exerted algorithmically. While populist uses of new technologies receive growing attention in critical organization and management studies, we argue that rationalized unaccountability is the hidden end of a spectrum of populism affecting societies across the world. Rather than populism of the masses, this is a populism of elites. This essay lays out some premises for critical scholars to expose the workings of intelligent agent programs and to call into question the problematic ideological assumptions that they are grounded in.

Keywords

Accountability, artificial intelligence, intelligent agent, power relations

Introduction

Artificial intelligence (AI) seems to be a topic of masterful grand returns. It hibernates into ‘AI winters’, only to receive a thorough thawing once the promise of a novel and radical technological breakthrough emerges. In this respect, we live in interesting times as AI has yet again awoken from hibernation. At the heart of its new coming are intelligent agents or adaptive, autonomous, and

Corresponding author:

Mikko Vesa, Hanken School of Economics, P.O. Box 479, Helsinki FI-00101, Finland.

Email: mikko.vesa@hanken.fi

social computer programs (Alonso, 2014). The purpose of this essay is to explore the implications of such intelligent agent programs for critical studies of organizing and organizations.

In what follows, we first highlight the core capabilities of intelligent agent programs and explore how practitioners experience their promise. We then outline types of accountability problems that these programs create. We propose that AI has entered a state of ‘rationalized unaccountability’, which constitutes an ideology that may come to characterize organizational power relations also more generally. The ideological faith in intelligent agent programs leads us to ask if AI should be understood as a form of elitist populism in late modern society. In so doing, we strive to complement the emerging discussion on AI and its sub-technologies in critical organization and management studies and beyond (Bader and Kaiser, 2019; Lange et al., 2019; Raisch and Krakowski, 2020; Shrestha et al., 2019; von Krogh, 2018).

On artificial intelligence, disruptions, and accountability

Artificial intelligence and intelligent agent programs are typically construed as disruptive technologies (Bloomfield and Vurdubakis, 2015; Lindebaum et al., 2020). The idea of disruption is powerful because it shapes public perception of the societal implications of AI by placing society in an immanent and unavoidable relationship with technologies. Perhaps the most widely discussed of these is the disruption brought about by AI in the transformation of work (Fleming, 2019; Ford, 2015; Kellogg et al., 2020). The core of the argument is that there may be a rapid substitution of labor by artificial intelligence; potentially including highly skilled professionals. The future of society, in consequence, depends on how it handles this ensuing technological, economic, and social disruption (Brynjolfsson and McAfee, 2014; Srnicek and Williams, 2016). There is also a debate about the destabilization caused by machine learning and algorithmic decision-making (Chun, 2016; Pasquale, 2015). This debate focuses on the way the functioning of AI technologies results in reinforced biases, creates ‘echo chambers’, and violates privacy whilst lacking public scrutiny and transparency (Friedland, 2019). Finally, and moving into the sphere of speculation, there is a set of discourses that are best described as existential disruption. In this genre, we find conceptualizations of artificial superintelligence (Bostrom, 2014; Chace, 2015; Harari, 2016; Tegmark, 2017), the achievement of which causes a disruption to what it means to be human by creating a – potentially conscious – technology superior to humanity. This profound change can be labelled as transhumanism, immortality, singularity or extinction, and the ‘human’ as we know it ceases to exist within these discourses.

What does all this mean for critical organization and management studies? Rather than emphasizing disruption as such it is argued that research on AI should be realistic, informed, and prudent (von Krogh, 2018). The validity of, for example, the economic disruption narrative is disputed (Fleming, 2019). Friedland (2019) argues that ceding control of small everyday human tasks to AI-based automation is deeply problematic. While letting algorithms do things may be convenient for decision-makers, ‘big data’ as the fuel of machine learning both aggregates vast quantities of data *and* distances itself from the everyday life of human beings (Hansen and Flyverborn, 2015). Hence the concern is one of how the use of AI that is ‘elusive and strange’ (Lange et al., 2019) generates self-inflicted states of learned helplessness (Lindebaum et al., 2020). There is inherent skepticism among critical researchers about the moral authenticity of algorithms, but also arguments for enhancing said authenticity by having technology better mimic humans (Jago, 2019).

What is then the accountability of this new technology? There is no single definition for accountability. To secure it, some sort of shared agreement on how it is manifested is yet needed. As a subject of inquiry, accountability has discipline-specific meanings: ‘auditors discuss accountability as if it is a financial or numerical matter, political scientists view accountability as a political

imperative and legal scholars as a constitutional arrangement, while philosophers treat accountability as a subset of ethics' (Sinclair, 1995: 221). In the English language, accountability and accounting are closely related (Ezzamel, 1997), but this is not the case in many other languages. In the context of organizations, accounting systems and procedures nevertheless offer one grounding for accountability. However, accounting technologies tend to give 'selective visibility' to particular organizational outcomes, rather than automatically leading to greater accountability (Hopwood, 1984: 179). In corporate governance, accountability operates through various mechanisms such as a strive for transparency, rather than being simply an issue of agency between management and shareholders (Brennan and Solomon, 2008).

While accountability requires people to explain and take responsibility for their actions, ideas of accountability change, and they are experienced in multiple ways. Accountability 'exists in many forms and is sustained and given extra dimensions of meaning by its context' (Sinclair, 1995: 219). Sinclair (1995) shows its 'chameleon-like' nature in management: accountability is under continuous construction, and it includes layers of meanings, contradictions, and tensions. In turn Overfield and Kaiser (2012) claim that contemporary upper-level managers have problems in holding people (presumably including themselves) accountable for their decisions and performance. According to these authors, there is 'an epidemic of letting people off the hook'. Also, information overload leads to decision-makers being highly reliant on experts and advisors functioning as buffers and filters of various kinds (Ezzamel et al., 2004), further blurring the notion of accountability. As we highlight when reviewing the capabilities of intelligent agent programs, there is reason to suspect that this ambiguity and accountability gap risks growing ever wider in the technological 'disruption'.

Intelligent agent programs

The intelligent agent program is today a key technology of artificial intelligence (Franklin, 2014). As stated by Russell and Norvig (2016: 46) in the field's seminal textbook, 'the job of [AI] is to design the agent program that implements the agent function'. The roots of intelligent agent programs can be traced to interest in cognitive embodied artificial intelligence in the 1990s (Varela et al., 1991). These programs are algorithmic systems that attempt to execute the 'best' course of actions under severe resource constraints (Russell and Wefald, 1991).

According to Alonso (2014), to qualify as intelligent agents computer programs need to have three qualities. First, they need to be autonomous so that they can take independent decisions. They need qualities that render them capable of acting as agents, typically by being able to independently sense, analyze, and respond to the environment. Second, they need to be adaptive as they must be able to both transform independent observations into experience and learn from the environment into which they have been assigned. This means that these programs do not make decisions solely based on pre-existing information programmed into them when the code is created. It is this adaptability that is the source of the intelligent agent program's ability to develop autonomously, typically by a trial-and-error exploration. Third, then, intelligent agent programs must be social so that they can recognize, co-operate, and organize with other, both non-human and human agents (Bader and Kaiser, 2019). Sociality allows the agent program to rationally and effectively pursue its goals when it is active in stochastic and complex environments.

Intelligent agent programs are becoming increasingly common. They are the decision-making component of modern artificial intelligence and often function together with other key AI technologies such as machine learning, natural language processing, or AI-powered robotics. The book that Amazon recommends to you is suggested by an intelligent agent, as are increasingly the loan deals offered to you by an online bank for buying a new car. Self-driving cars currently under

development for cruising on roads populated by a mix of other self-driving cars and human drivers have an intelligent agent behind the steering wheel – and so do highly contentious autonomous weapon systems (Bloomfield and Vurdubakis, 2015).

For us, the next question to ask is how can we understand the promise and allure that intelligent agent programs hold for contemporary organizing and organizations?

The promise

According to renowned science fiction author Clarke (2013), ‘any sufficiently advanced technology is indistinguishable from magic’. Artificial intelligence has always had a flirting relationship with this idea of ‘magic’. Its promise has often – typically in hype cycles ending in some form of failure – been sold as radically disruptive. As laymen, we are awed by it, and often strangely disarmed and docile when confronted with its disruptive potential.

Many of us are likely to perceive AI as a form of technological ‘magic’. Imagine the promise of intelligent agent programs: they never miss a detail, they never forget, and they are constantly vigilant. Nor do they (supposedly) engage in petty games nor discriminate. They appear superior in their rationality and efficiency. They do not have ‘agency’ in any classical sense and, as a consequence, no agent-principal problems. These programs do what they are told. Only they do so a bit better every time and they transcend human capabilities in processing information many times over. Promises of superior performance or competitive advantage derived from such technologies tend to be an easy sell for decision-makers.

As such, intelligent agent programs and algorithms become objects of desire in complex ways for the power elite in society. The way AI delivers competitive advantages allows for a reconfiguration of power relations. Beneath it all lies the radical promise of organizing and organizations free of human concerns and shortcomings. In effect, this creates the promise to view intelligent agent programs as perfect rational agents. However, this is largely an experiential state associated with the mastery of such code by those who control them. This promise of rationality easily positions any critique as romantic, old-fashioned, and irrational.

We should resist such rejection of critique as it is in stark contrast to the experience of those who are subservient or controlled by intelligent agent programs, and for whom the experience can be dehumanizing and totalitarian. As noted earlier, there is a sense of inevitability baked into AI as we experience its discourse today. What we propose here is that this inevitability largely comes down to the fact that outsiders – and this includes most of the aforementioned power elite – to developing artificial intelligence struggle to pry ajar the ‘AI black box’. The magic, one could say, resists critical scrutiny. This renders AI technologies such as intelligent agent programs difficult to understand even for those who invest in them or organize through them.

Consequently, it becomes difficult to assess what challenges and problems AI technologies come to be associated with. Of interest to us – and, we suggest, to the critical organization and management studies community more generally – are the potential problems that these programs create for accountability in organizing and organizations. Questions of ambiguity and distance are crucial here. Huber and Munro (2014) explored moral distance in organizational ethics, building on Bauman’s (1991) ideas on commitment of immoral acts becoming easier with social distance. Zyglidopoulos and Fleming (2008), too, emphasize ethical distance, that is the distance between an act and its ethical consequences. These moral and ethical distances are accentuated in online spaces in general and in relation to AI in particular. We lose sight of how specific acts lead to particular consequences.

We propose that there are three areas of accountability that are especially problematic and need critical scrutiny. Each problem is directly linked with the three persistent qualities of intelligent

agent programs outlined above. First, we must explore how intelligent agents are (not) rendered accountable in interactions with humans. Second, we must explore how they are (not) scrutinized for what they have learned. Third, we must explore power relations behind intelligent agents to bring renewed attention to the human goalsetting and assumptions behind intelligent agent operations.

Three problems of accountability

The core qualities of intelligent agent programs are autonomy, adaptability, and sociality (Alonso, 2014). Each contains specific aspects that render these programs problematic from an accountability perspective. These relate to how the intelligent agent operates, how it learns, and what it seeks to achieve. Accountability is, ultimately, always a human burden. Hence a human should always be accountable for the actions of intelligent agent programs. However, it is questionable if such accountability can be informed – or whether the responsibility is in the end more of a scapegoat nature.

Autonomy and speed in interactions between intelligent agents and humans

The first accountability problem stems from the basic observation that human beings and computer programs are simply fundamentally different (Bader and Kaiser, 2019). This trivial observation becomes central when one recognizes that, differences aside, intelligent agents and humans both act as decision-makers in organizations (Bader and Kaiser, 2019; Lange et al., 2019; Newell and Marabelli, 2015). One could go as far as to argue that power is exercised algorithmically in organizations today (Pasquale, 2015). For example, Helbing et al. (2017) suggest that algorithms already conduct a staggering 70% of all global financial transactions. As argued by Johnson et al. (2013), such financial market transactions exceed human capacity for interaction. On a day-to-day basis, this is due to the internal capacity of the intelligent agents with regard to speed, memory, and sensory exactness. The ‘human clock’ reacts, at best, in seconds while the clock speed of computers is faster by an order of magnitudes. The human mind and body exist in varying states of vigilance, but the intelligent agent is always 100% alert. While a human being has a working memory of a handful of items, an intelligent agent is only restricted by the size of its hardware. It is constantly alert, consistently vigilant, and it reacts in micro-seconds. At the very extreme, in the future markets are destabilized and wars decided quite literally in a heartbeat.

Hence the accountability problem is this: how can a human decision-maker be held accountable for decisions that occur faster and with a larger number of inputs than they can physiologically comprehend and react to? This is a real consequence from allowing intelligent agent programs autonomy, that is to allow them to act independently and without direct supervision. Beverungen and Lange (2018) argue that current high frequency algorithms are still too ‘stupid’ to be considered fully autonomous. However, for every new generation of hardware and for every refinement of algorithms that drive intelligent agent programs the basic physiological chasm between humans and artificial intelligence technologies will grow wider.

Adaptability and understanding what intelligent agents (can) learn

Perhaps the most agency-like characteristic of intelligent agent programs is found in their ability to adapt. This means that when acting in the world, the program will constantly monitor its environment in order to determine whether its operations are successful. Based on these, the program accumulates information on what works and what does not. If it discovers that a particular move is

successful, or that a certain move is not helpful in achieving its goals, the program will learn to either prefer or avoid such moves in the future. In software terms, this is called building up a percept sequence. One can think of this as the life history of the intelligent agent program. The size of the percept sequence is limited only by the sensory input that the program has access to, and the amount of memory it is allocated.

The problem for the human who is accountable for the intelligent agent is that we cannot know precisely what the intelligent agent has learned. The challenge with machine learning techniques is that as outsiders, we can at most know what goes into the ‘black box’ of the intelligent agent and what comes out. However, we cannot know for sure why what comes out does come out. In the world of ‘big data’, transparency vis-à-vis data is often meaningless, because the sheer amount of data analyzed renders it to impossible to process for the human mind (Pasquale, 2015). Data patterning itself contains problematic aspects such as a tendency for homophily, for example, in an algorithm created on the basis of particular assumptions (about groups of people and the world) in its goal function reproducing itself endlessly as the intelligent agent program adapts and learns to optimize its performance (Lambrecht and Tucker, 2019).

As such, there is nothing that guarantees that statistically internally coherent models resonate with outside realities (Chon, 2016). This means that in more complex situations the intelligent agent’s percept sequence forms the basis for the agent program’s praxis – perhaps one could even say worldview – and while this is not transparent it renders the agent adaptively autonomous. What this means in practice is that intelligent agent programs do not get to (re)write their own goals. However, they get to rewrite the roadmap for reaching said goals – otherwise they would not be adaptive. In particular, when the decisions involved are largely non-reversible, such as in financial transactions or in the use of lethal force, this begs the age-old question of whether the end justifies the means.

To some extent this may not be as dramatic as it sounds because in many ways the same applies to a human decision-maker. There is a vast literature about cognitive biases that affect human decision-making. However, the idea of accountability includes that in spite of such biases the decision-maker should ultimately be accountable, which the intelligent agent program is not. Hence the accountability problem is this: how can a human decision maker be held accountable for decisions made by an intelligent agent program that depend on adaption through inscrutable learning techniques – for example, neural networks and machine learning – that in complex decision-making environments alter how the intelligent agent program acts from one instance to another?

Sociality and the goals that intelligent agents pursue

Finally, within the confines of their purpose, intelligent agents have the ability to interact with the external world. Autonomous cars drive on roads populated with other cars, autonomous and not. Trading algorithms swap information and conduct transactions on financial markets. Parisi (2015) argues that this world of intelligent agents has become a second nature that exists alongside the conventional human reality of the financial markets. Overall, Galloway (2012: 93) observes that ‘the point of power today resides in networks, computers, algorithms, information and data’. Hence there is a powerful illusion built into the intelligent agent, namely that it is in light of its capabilities a social and independent entity. Both cars and stock markets seem to crash because of what the intelligent agent program did (or did not) do.

However, for all its autonomy and adaptivity, the intelligent agent does not have any subjective purpose or marginal utility that it seeks to fulfill or maximize. Rather, it always only optimizes – and does so without any human sense of deliberation regarding the consequences of its actions. What this means is that the moral of the code that drives the intelligent agent resides within what

is known as its goal function. The goal function is always preset and controlled by those who run the intelligent agent program. Hence cars and stock markets driven by intelligent agents may crash, but only in the pursuit of the goals of their human masters. This also means that questions of accountability cannot be assigned to intelligent agents, but rather to those who control the content of the goal function, which in itself remains static short of human intervention. Intelligent agents get to optimize the roadmap, but they do not choose the final destination.

The sociality of intelligent agents has problematic implications. As argued by Parisi (2015: 136), again with regard to financial markets, ‘it is hard to dismiss the possibility that the automation of thought has exceeded representation and has instead revealed that computation itself has become dynamic’. Hence in a radical fashion the complex networks of interacting autonomous intelligent agent programs – this ‘second nature’ – has emergent properties that affect society. Networks of interacting intelligent agent programs can create systemic emergent outcomes not readily decipherable from single goal functions. Homophily and its outcomes, for example, are potentially multiplied in these networks (Lambrech and Tucker, 2019).

Ananny and Crawford (2018) maintain that models for understanding and holding systems accountable have long rested upon ideals and logics of transparency. These authors critically interrogate the ideal of transparency in computational systems and argue that transparency is an inadequate premise for understanding and governing algorithms. Pasquale (2015) points out that the inherent complexity of algorithmic behavior renders the notion of transparency obsolete, because the underlying complexity of operations is such that even if fully transparent, the system still remains inscrutable. He points out that what we should demand is comprehensibility: ‘Algorithms are not immune from the fundamental problem of discrimination. [. . .] They are programmed by human beings, whose values are embedded into their software’ (Pasquale, 2015: 38). In practice, this would imply that any intelligent agent whose operations cannot in a satisfactory manner be explained to the public should not be permitted to exist in conditions that require sociality.

The problem for accountability contained in the sociality of intelligent agent programs is thus twofold. First, how can we as members of society judge the consequences of intelligent agent behavior if the goal functions of such programs are not (transparent and) comprehensible? Second, where does the accountability of emergent system level intelligent agent behavioral outcomes reside?

Rationalized unaccountability as ideology

We argue that the conundrum we confront with intelligent agents and accountability is made of two components. On the one hand, we are dealing with what appears to be a radical technology that holds an ideological promise to disrupt markets and society. On the other hand, this technology on examination is quite problematic for questions of accountability-at-large, which is arguably a core dimension of late modern liberal society. Where does this interweaving of the ideological promise of intelligent agents combined with their associated accountability problems leave us? We answer this by conceptualizing what we call the state or condition of ‘rationalized unaccountability’. It arises partly from the compound challenges with the accountability of intelligent agents as outlined previously, but also from the ideological drivers that position networks and populations of intelligent agents as a desirable social future.

In developing this argument, we are standing on the shoulders of giants. We cannot claim to do justice to the vast and varied work of Max Weber, the Frankfurt School of Critical Theory (and the likes of Max Horkheimer, Theodor Adorno, Erich Fromm and Herbert Marcuse), and Zygmunt Bauman. Nevertheless, we build here on selected insights from their work. We find inspiration from Weber (2003/1905) and his conceptualizations of rationality and rationalization as hallmarks

of modernity and of capitalism's rise to dominance worldwide.¹ In particular, we are inspired by critique of how particular forms of rationality, and rule through ideology, function in (and today, across) societies. Also the various incarnations of the Frankfurt School of Critical Theory are fundamentally important here, particularly in helping to make sense of the realm of popular or mass (media) culture that is made possible by technological developments and, today, intelligent agent programs.²

In *Dialectic of Enlightenment*, Horkheimer and Adorno (1969/1944) explored fusions of domination and technological rationality and emphasized the role of knowledge and technology as means of exploitation (of labor) that is ideologically grounded. They considered technological domination of human action as the negation of what can be considered inspiring purposes of the Enlightenment. People may think they are 'free', then, but they are only free to choose an ideology, which always reflects a form of economic and technological coercion. Their freedom is 'freedom to choose what is always the same'. The accountability problems of intelligent agent programs play out in much the same way. We are shrouded in a veil of inevitability and coerced blindly to have faith in technology and its 'rationality'. This is, to echo Fromm (1941/1994; see also Deslandes, 2018), an escape from freedom into the seemingly comforting conceptual fold of the algorithm onto which we regressively latch hopes of ontological security. Adapted into a meme, it is as if our elites now desire to clutch coffee mugs reading 'Keep Calm and Trust the Algorithm'.

Of the many philosophical and sociological thinkers of past decades, the ideas of Zygmunt Bauman resonate with our line of argumentation. For Bauman, rationalization is a manifestation of modernity and its order-making efforts.³ In contrast to Horkheimer and Adorno's (1969/1944) attention to sameness, however, Bauman (1992) alerts us to a world where difference is prevalent in what has become consumer capitalism. He helps us to understand relations between decentred oppressors pulling 'invisible strings', on the one hand, and their 'happy victims' who are exploited in contingent ways, on the other. The latter appear to be prepared to surrender their 'freedoms' to the decentred power-knowledge of consumer capitalism. Global systems and structures are remote and unreachable, and our lives are unstructured and fluid.

Intelligent agent programs accentuate this divide and render us malleable in the hands of power elites who are able to cover their tracks with technologies – or, more precisely, with promises of constantly evolving technologies and their disruptive effects. In *Liquid Modernity* Bauman (2000) addresses the profound uncertainties of this radical state of late modernity; its burdens on individuals, constant change, ambivalence, and chaos. However, beyond the liquidity of tax havens, Davos culture and global deregulation awaits a new form of liquidity that dissolves responsibility and accountability into an endless digital sea of ones and zeroes.

Specifically, we argue that rationalized unaccountability today is an ideological state in which power and control are exerted algorithmically. The concept of ideology is contested, and it is studied in different ways in diverse research traditions (Seeck et al., 2020). We conceive of ideology in the present context as production of self-evident truths. We propose that artificial intelligence functions as an ideology as it manufactures normative idea(l)s of social reality and turns these into self-evident features of discourse (Fairclough, 1989) through which we are (not) able to make sense of the world. In suggesting this, we draw from the spirit of Critical Theory and offer a critique of ideology as it serves to legitimize and normalize or naturalize particular rationalizations and exploitation. Ideology, then, is a discourse by the (aspiring) elites to explain their present and future rule as the norm.

AI builds on and attracts a particular form of discourse, which constitutes it as reality. Through legitimation as well as through 'magic' (cf. Clarke, 2013) we are led to believe in its inevitability. While legitimation is grounded in authoritative and seemingly rational arguments, 'magic' operates by leaving its underlying assumptions hidden and unquestioned. Accountability again reveals

its ‘chameleon-like’ nature (Sinclair, 1995), and through repetition, over time, more and more of us are socialized into the ideology. At the same time, technological and financial elites bolster their dominant position and idea(l)s in society. Mechanisms of power remain concentrated, and control is funneled into the hands of limited groups of actors – power elites – with shared technological-financial worldviews (cf. Wright Mills, 1956).

Rationalized unaccountability today is an ideological state in which power and control are executed algorithmically through intelligent agent programs acting independently and collectively to accomplish goals unknown to single humans, entire organizations, or whole societies that are objects of this power. Technologically the state of rationalized unaccountability is driven by the ‘magic’ associated with the intelligent agent, which is an emergent property of the three co-existing dimensions of the agent algorithm’s structure discussed above. This entangled co-existence is the source of the ‘black box’ effect associated with intelligent agent programs. Under conditions of rationalized unaccountability, the intelligent agent program is valued as a superior conduit for power and control due to its technological capabilities regarding autonomy, adaptivity, and goal-based sociality. Part of this superiority is that intelligent agents are in an increasing number of instances preferred over human work by decision-makers in organizations (Fleming, 2019).

Artificial intelligence is an ideology that can be used to both explain and control the direction of our current post-globalization and post-deregulation world that legitimizes new forms of power and control. To an extent, it replaces neoliberal political discourses with artificial intelligence-derived technological discourses. Rationalized unaccountability, then, is a state through which elites today are able to construe different kinds of power relations and their authority.

First, it is a matter of wielding power directly through intelligent agent programs. This is what we witness through applications as varied as credit ratings, trading algorithms, and autonomous weapons systems. Second, it is a matter of wielding power coercively by disciplining citizens and societies with the threat of substitution. This is what we witness, for example, in the way that reports concerning future labor market trends and disruptions are used to control wage structures, the content of education programs, or public investment schemes. Third, it is a matter of deploying power in a utopian sense by sustaining the possibility of emancipatory technology that would once and for all free the power elite of the burdens of control through the creation of a superior agentic technology. Here we find a wide diaspora of genres from transhumanism to superintelligence. We argue that these are all belief systems and utopian discourses created by the elite for the elite.

This political dimension of intelligent agents and algorithms is crucial for critical scholarship in organization and management studies. In a sense, rationalized unaccountability is the other, often hidden, end of the spectrum of contemporary populism. It is an ideology created by (and for) a global elite who have been losing space to an array of populist movements that are often hostile to the free flow of financial, economic and cultural assets as well as humans. By construing the economy and the society as entities inevitably heading for technological disruption, the state of rationalized unaccountability offers a legitimation for global power elites to contest the rising force of ‘plebian’ populism with a ‘patrician’ populism of their own, which is combined with a technology-induced emancipatory narrative.

Concluding remarks

In this *Connexions* essay, we have argued that artificial intelligence is much more than a technology. It is a discourse actively used to shape the political, economic, and social realities of our times. Often, the discourse has only a convoluted relationship with the actual capabilities of the technologies at hand. We have sought to make sense of this condition and its implications through the idea of ‘rationalized unaccountability’ that is an ideological state in which power and control are exerted

algorithmically. With these ideas we have sought to complement the emerging discussion on artificial intelligence in critical organization and management studies and beyond (Bader and Kaiser, 2019; Lange et al., 2019; von Krogh, 2018). We have argued that a key concern for critical scholars today is to explore how AI (and its subcomponent technologies) are used to legitimize or delegitimize policies and forms of organizing, and to elucidate their disciplining effects on people and communities. An important premise is to bring the technology in question – the intelligent agent – under real scrutiny.

We conclude this essay by presenting some questions for developing critical inquiry into algorithms and intelligent agents. What are their goals? What is their data? When do they stop? First, goals direct our focus on the ‘doing’ of artificial intelligence by asking us to scrutinize the human assumptions and goal setting behind the intelligent agent. We must combat the blurring of accountability enabled by intelligent agent programs as well as the sense of inevitability baked into AI discourse overall. As the goal functions of such programs are not (transparent and) comprehensible, we must be sensitive to how AI functions as an ideology as it manufactures normative idea(l)s of social reality into self-evident truths, benefitting some at the expense of others.

Second, scrutinizing data is about understanding what the intelligent agent has learned about the world. Whilst it is not possible to do this on an item-by-item basis (there is simply too much ‘big data’), it would be useful to look at this thematically as to which kinds of data the program is supposed to be learning from. Again, we are back to questions of ideology. The overwhelming amount and fluidity of data help sustain the AI ideology as global systems and structures remain out of the reach of most people. By aiming to determine who is able to make sense of the world and who is not, power elites are able to cover their tracks with promises of constantly evolving technologies and their ‘disruptive’ effects.

Third, the question of stopping takes us into the world of safety mechanisms, and it prompts more questions. When does the program – superior to humans as it is in many respects – simply shut itself down in order to allow for a human assessment? What would this mean? For all these questions, ethical and moral considerations of intelligent agent programs are particularly timely for critical organization and management studies (Bauman, 1991; Huber and Munro, 2014; Zyglidopoulos and Fleming, 2008). While AI as an ideology of the elites steers us away from asking ethical and moral questions, critical researchers must ally with other actors to keep them firmly on the societal agenda in conditions where adaptive, autonomous, and social computer programs are ubiquitous.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

ORCID iD

Mikko Vesa  <https://orcid.org/0000-0001-8643-662X>

Notes

1. It is useful to draw attention to processes by which all aspects of the natural and social world become subject to calculation, planning and efficiency, and how Max Weber’s ideas on instrumental or purposive rationality still find resonance today. He also alerted us to the fact that a totally rationally administered world is not likely to be fulfilling for its inhabitants.
2. Following Horkheimer (1976/1937) in *Traditional and Critical Theory* we opt for critical method and analysis that is not dogmatic in its assumptions, although we realize that this is elusive and that a form of self-reflection is needed when making claims in the name of criticality.

3. Zygmunt Bauman suggests we pay attention to the different approaches modern society adopts toward the stranger that cannot be controlled and ordered and is thus an object of fear. In *Modernity and the Holocaust*, Bauman (1989) details the devastating dangers of these fears.

References

- Alonso, E. (2014) 'Actions and Agents', in K. Frankish and W. M. Ramsey (eds) *The Cambridge Handbook of Artificial Intelligence*, pp. 232–46. Cambridge: Cambridge University Press.
- Ananny, M. and Crawford, K. (2018) 'Seeing without Knowing: Limitations of the Transparency Ideal and Its Application to Algorithmic Accountability', *New Media & Society* 20(3): 973–89.
- Bader, V. and Kaiser, S. (2019) 'Algorithmic Decision-Making? The User Interface and Its Role for Human Involvement in Decisions Supported by Artificial Intelligence', *Organization* 26(5): 655–72.
- Bauman, Z. (1989) *Modernity and the Holocaust*. Ithaca, NY: Cornell University Press.
- Bauman, Z. (1991) *Modernity and Ambivalence*. Ithaca, NY: Cornell University Press.
- Bauman, Z. (1992) *Intimations of Postmodernity*. London: Routledge.
- Bauman, Z. (2000) *Liquid Modernity*. Cambridge: Polity.
- Beverungen, A. and Lange, A. C. (2018) 'Cognition in High-Frequency Trading: The Costs of Consciousness and the Limits of Automation', *Theory, Culture & Society* 35(6): 75–95.
- Bloomfield, B. P. and Vurdubakis, T. (2015) 'Mors ex Machina: Technology, Embodiment and the Organization of Destruction', *Organization Studies* 36(5): 621–41.
- Bostrom, N. (2014) *Superintelligence: Paths, Dangers, Strategies*. Oxford: Oxford University Press.
- Brennan, N. M. and Solomon, J. (2008) 'Corporate Governance, Accountability and Mechanisms of Accountability: An Overview', *Accounting, Auditing & Accountability Journal* 21(7): 885–906.
- Brynjolfsson, E. and McAfee, A. (2014) *The Second Machine Age: Work, Progress, and Prosperity in a Time of Brilliant Technologies*. New York: W. W. Norton & Company.
- Chace, C. (2015) *Surviving AI: The Promise and Peril of Artificial Intelligence*. London: Three C's Publishing.
- Chun, W. (2016) *Updating to Remain the Same: Habitual New Media*. Cambridge, MA: MIT Press.
- Clarke, A. C. (2013) *Profiles of the Future*. London: Hachette UK.
- Deslandes, G. (2018) 'Book Review: Erich Fromm *Escape from Freedom: Revisiting Erich Fromm in the Light of Contemporary Authoritarianism*', *Organization Studies* 39(1): 1649–57.
- Ezzamel, M. (1997) 'Accounting, Control and Accountability: Preliminary Evidence from Ancient Egypt', *Critical Perspectives on Accounting* 8(6): 563–601.
- Ezzamel, M., Hyndman, N. S., Johnsen, Å., et al. (2004) 'Has Devolution Increased Democratic Accountability?', *Public Money & Management* 24(3): 145–52.
- Fairclough, N. (1989) *Language and Power*. London: Longman.
- Fleming, P. (2019) 'Robots and Organization Studies: Why Robots Might Not Want to Steal Your Job', *Organization Studies* 40(1): 23–38.
- Ford, M. (2015) *Rise of the Robots: Technology and the Threat of a Jobless Future*. New York: Basic Books.
- Franklin, S. (2014) 'History, Motivations, and Core Themes', in K. Frankish and W. M. Ramsey (eds) *The Cambridge Handbook of Artificial Intelligence*, pp. 15–33. Cambridge: Cambridge University Press.
- Friedland, J. (2019) 'AI Can Help Us Live More Deliberately', *MIT Sloan Management Review* 60(4): 45–51.
- Fromm, E. (1994/1941) *Escape from Freedom*. New York: Holt Paperbacks.
- Galloway, A. R. (2012) *The Interface Effect*. Cambridge, UK: Polity Press.
- Hansen, H. K. and Flyverborn, M. (2015) 'The Politics of Transparency and the Calibration of Knowledge in the Digital Age', *Organization* 22(6): 872–89.
- Harari, Y. N. (2016) *Homo Deus: A Brief History of Tomorrow*. Random House.
- Helbing, D., Frey, B. S., Gigerenzer, G., et al. (2017) 'Will Democracy Survive Big Data and Artificial Intelligence?', *Scientific American*, 25 February. Retrieved March 31, 2017 from <https://www.scientificamerican.com/article/will-democracy-survive-big-data-and-artificial-intelligence/>
- Hopwood, A. (1984) 'Accounting and the Pursuit of Efficiency', in A. Hopwood and C. Tomkins (eds) *Issues in Public Sector Accounting*, pp. 141–57. Oxford: Philip Allan.
- Horkheimer, M. (1976/1937) 'Traditional and Critical Theory', in P. Connerton (ed) *Critical Sociology: Selected Readings*. Harmondsworth: Penguin.

- Horkheimer, M. and Adorno, T. W. (1969/1944) *Dialectic of Enlightenment*. New York: Continuum.
- Huber, C. and Munro, I. (2014) “Moral Distance” in Organizations: An Inquiry into Ethical Violence in the Works of Kafka’, *Journal of Business Ethics* 124(2): 259–69.
- Jago, A. S. (2019) ‘Algorithms and Authenticity’, *Academy of Management Discoveries* 5(1): 38–56.
- Johnson, N., Zhao, G., Hunsader, E., et al. (2013) ‘Abrupt Rise of New Machine Ecology Beyond Human Response Time’, *Nature*. Retrieved March 13, 2019, from <http://www.nature.com/articles/srep02627>
- Kellogg, K. C., Valentine, M. A. and Christin, A. (2020) ‘Algorithms at Work: The New Contested Terrain of Control’, *Academy of Management Annals* 14(1): 366–410.
- Lambrecht, A. and Tucker, C. (2019) ‘Algorithmic Bias? An Empirical Study of Apparent Gender-Based Discrimination in the Display of STEM Career Ads’, *Management Science* 65: 2966–81.
- Lange, A.-C., Lenglet, M. and Seyfert, R. (2019) ‘On Studying Algorithms Ethnographically: Making Sense of Objects of Ignorance’, *Organization* 26(4): 598–617.
- Lindebaum, D., Vesa, M. and den Hond, F. (2020) ‘Insights from “The Machine Stops” to Better Understand Rational Assumptions in Algorithmic Decision Making and Its Implications for Organizations’, *Academy of Management Review* 45(1): 247–63.
- Newell, S. and Marabelli, M. (2015) ‘Strategic Opportunities (and Challenges) of Algorithmic Decision-Making: A Call for Action on the Long-Term Societal Effects of “Datification”’, *The Journal of Strategic Information Systems* 24(1): 3–14.
- Overfield, D. and Kaiser, R. (2012) ‘One Out of Every Two Managers Is Terrible at Accountability’, *Harvard Business Review*, November.
- Parisi, L. (2015) ‘Instrumental Reason, Algorithmic Capitalism, and the Incomputable’, in M. Pasquinelli (eds) *Alleys of Your Mind: Augmented Intelligence and Its Traumas*, pp. 125–138. Luneburg: Meson Press.
- Pasquale, F. (2015) *The Black Box Society: The Secret Algorithms that Control Money and Information*. Cambridge, MA: Harvard University Press.
- Raisch, S. and Krakowski, S. (2020) ‘Artificial Intelligence and Management: The Automation-Augmentation Paradox’, *Academy of Management Review* (forthcoming). Retrieved from <https://doi.org/10.5465/2018.0072>
- Russell, S. and Norvig, P. (2016) *Artificial Intelligence: A Modern Approach*. Upper Saddle River, NJ: Pearson Education.
- Russell, S. J. and Wefald, E. (1991) *Do the Right Thing: Studies in Limited Rationality*. Cambridge, MA: MIT Press.
- Seeck, H., Sturdy, A., Boncori, A.-L., et al. (2020) ‘Ideology in Management Studies’, *International Journal of Management Reviews* 22(1): 53–74.
- Shrestha, Y. R., Ben-Menahem, S. M. and von Krogh, G. (2019) ‘Organizational Decision-Making Structures in the Age of Artificial Intelligence’, *California Management Review* 61(4): 66–83.
- Sinclair, A. (1995) ‘The Chameleon of Accountability: Forms and Discourses’, *Accounting, Organizations and Society* 20(2–3): 219–37.
- Srnicek, N. and Williams, A. (2016) *Inventing the Future: Postcapitalism and a World Without Work*. London: Verso Books.
- Tegmark, M. (2017) *Life 3.0: Being Human in the Age of Artificial Intelligence*. London: Allen Lane/Penguin Random House.
- Varela, F. J., Thompson, E. and Rosch, E. (1991) *The Embodied Mind*. Cambridge, MA: MIT Press.
- von Krogh, G. (2018) ‘Artificial Intelligence in Organizations: New Opportunities for Phenomenon-Based Theorizing’, *Academy of Management Discoveries* 4(4): 404–9.
- Weber, M. (2003/1905) *The Protestant Ethic and the Spirit of Capitalism*. Mineola, NY: Dover Publications.
- Wright Mills, C. (1956) *The Power Elite*. Oxford: Oxford University Press.
- Zyglidopoulos, S. C. and Fleming, P. J. (2008) ‘Ethical Distance in Corrupt Firms: How Do Innocent Bystanders Become Guilty Perpetrators?’, *Journal of Business Ethics* 78(1–2): 265–74.

Author biographies

Mikko Vesa (PhD) is an associate professor of management and organization at Hanken School of Economics. His research focuses on technological and organizational change, covering topics such as gamification, strategy as practice and artificial intelligence. His research has been published e.g. in the *Academy of Management Review*, *Organization Studies*, and the *Journal of Business Ethics*.

Janne Tienari is professor of Management and Organization at Hanken School of Economics, Finland. His research and teaching interests include strategy work, gender and diversity, feminist theory, managing multinational corporations, mergers and acquisitions, branding and media, and changing academia from a critical perspective. He is Senior Editor of *Organization Studies*. Tienari's work has been published in leading journals such as *Academy of Management Review*, *Organization Science*, *Organization Studies*, *Organization*, *Human Relations*, *Journal of Management Studies*, and *Journal of World Business*.