Preserving human relevance, as a new social responsibility of business in the AI age

Ciprian N. Radavoi

Abstract

Purpose – *This paper aims to contribute to the scholarly debate, ongoing in this and other journals, on the justification and extent of artificial intelligence (AI)-related responsibilities of a variety of segments of society, such as governments and parliaments, scientists, corporations, media and AI users. Among these, business has received less attention, in both academic and political speech, hence this paper's attempt to decant the content of a principle of corporate social responsibility related to AI.*

Design/methodology/approach – *This conceptual paper is built on two pillars. Placing the discussion in a framework of corporate social responsibility, this paper first argues that in the AI age, the list of corporate social responsibility (CSR) principles should be updated to include one relevant to AI development and deployment. Second, this study looks at the possible content of a new CSR principle.*

Findings – *Born from and still permeated by ethical principles, CSR principles evolve in time, reflecting contemporary societal priorities. If we define CSR as the integration of social concerns in corporate decision-making, then preserving the relevance of the human in the age of AI should qualify as a CSR principle. Like other CSR principles (anticorruption, transparency, community engagement, etc.), this would start as voluntary, but could harden in time, if society deems it necessary. Human relevance is more appropriate than human centrality as a CSR principle, despite the latter being referred to as a desideratum in numerous studies, policies and political statements on AI governance.*

Originality/value – *To the best of the author's knowledge, this study is the first to demonstrate that in the age of AI, the list of recognized CSR principle should be updated to include an AI-related one. Introducing human relevance, as opposed to human centrality, as the content of such principle is also highly original, challenging current assumptions.*

Keywords *Artificial intelligence, AI risks, Business ethics, Corporate social responsibility, Human centrality, Human relevance*

Paper type *Conceptual paper*

Ciprian N. Radavoi is based at School of Law and Justice, University of Southern Queensland, Toowoomba, Australia.

Received 3 January 2025
Revised 19 March 2025
11 August 2025
Accepted 9 September 2025

Introduction

This article is meant to contribute to the ongoing debate on social responsibilities associated with the advent of artificial intelligence (AI) (Kaas, 2024; Krkač, 2019; Saheb and Saheb, 2023; Bednar and Spiekermann, 2024). Reviewing state regulation of AI in 30 countries, Saheb and Saheb (2023) found a lack of clear guidance on ethical strategies for AI development and deployment, these being largely left to corporations' discretion. Leaving it to the corporation may not be the best idea, as its actions are limited to the adoption of value principles which have proven of limited practical impact (Bednar and Spiekermann, 2024). In the meantime, ethical AI solutions keep being proposed by scholars, but these also fail to stimulate corporate adherence in the real world, one reason being that their amount and diversity generate contradictions and uncertainty among business and stakeholders. Against this background, the main claim of this article is that AI ethical ideas meant to guide the corporation toward responsibly creating and deploying AI

can and should be distilled in the form of a principle of corporate social responsibility (CSR). That is, the existing list of CSR principles should be updated to include one addressing the risks associated with AI creation and deployment.

The downsides of unreflectively embracing AI in all sectors of personal, social and economic life are increasingly concerning. Scientists, philosophers, policymakers and major firms' executives warn of the dangers that loom in humanity's march toward total technologization and robotization. Consequently, unified AI governance frameworks for corporations and governments have been proposed, in the regulatory continuum from self-regulation to binding rules. Koniakou (2023), for example, argues for an organizing framework built on the direct application of human rights to *both* public and private actors, while Floridi and Cowls (2019) prescribe five principles as a unified framework for AI governance: beneficence, non-maleficence, autonomy (retaining some of the human power in decision-making), justice and explicability. Noting that:

> [i]t is the duty and responsibility of the state to help this active citizen to be one-self with the active appearance as a citizen and morally autonomous decision-maker in the information world (118).

Rendtorff (2022) shows that four principles should guide relevant governance: autonomy, integrity, dignity and vulnerability.

Unified frameworks based on ethical principles are certainly needed to guide the evolution of regulation of all types, but we should bear in mind that the state and the corporation's responsibilities are different. The state is expected to fulfil its side of the social contract by creating and enforcing a just AI regulatory framework, while the corporation as facilitator, creator and/or beneficiary of AI has its own social responsibility to create and use AI in a manner that leads to profit but also considers the interests of the society. However, while there is now intense regulatory action from the state around responsible AI, the business sector has not sent a clear message against AI misuse. Despite the occasional statements of a few concerned CEOs, examples of corporations taking advantage of the interstices of a governance framework still in its infancy abound; in particular, corporations benefit from "outsourcing" human decision-making to AI, thus avoiding accountability for wrongdoing (Diamantis, 2020).

This article looks at the corporate responsibilities regarding AI, placing the discussion in the CSR framework. CSR went in one century from denial (the firm's only responsibility is to make profit for their shareholders) to recognition (the firm has responsibilities to other stakeholders, and in general to society and the environment). In its most condensed expression, CSR is the idea that corporate managers should be guided, in their decision-making, by not only the profit imperative but also the needs and expectations of a variety of stakeholders and of society in general. To operationalize this rather general formulation, principles of CSR have been proposed in both academia and the business environment since the 1990s: the corporation must be transparent, it must engage the local communities where it operates and so on. But as society evolves, its expectations from the corporation evolve as well, hence this article's argument, put forward in Section 1, that the accelerated technological progress should be reflected in the CSR construct, as a new principle. The argument in this section is built via successive propositions derived from the review of relevant literature.

The second claim of this paper is about the content of such a principle. At the highest level of abstraction, the foundational principles of AI governance at corporate versus state level are different: the state's appears to be that of human centrality, while the corporation's must be formulated with a lower level of stringency, given its focus on profit. Accordingly, this article proposes in its second section that human relevance, rather than human centrality, should be at the core of the new CSR principle.

Why an artificial intelligence-related new corporate social responsibility principle?

Artificial intelligence-related social responsibility: the overlooked role of business

The core current challenges posed by AI are related to the opaque nature of the machines, which limits accountability and contestability; the inequalities perpetuated by automated decision-making; the social alienation AI generates via job loss and reduced human interaction; the alteration of rational choice and thereby of human autonomy; and the risk of diminished or lost human control over autonomous systems and of malicious use (Dignum, 2019; Šucha and Gammel, 2021; Sartori and Theodorou, 2022). For the distant future, some authors also foresee the risk that somewhere along the road, AI, initially a product of human intelligence and, thus, infused with our perspectives and goals, becomes creative and parts ways with humans. This is the so-called superintelligence or artificial general intelligence (AGI) stage, the plausibility of which is debated in literature. Numerous academics and futurists warn of possible dystopian futures, with humanity threatened with annihilation by superintelligent machines whose goals have become different to ours (Barrat, 2013; Bostrom, 2014; Tegmark, 2017), while other scholars picture a rosier future, where humans work alongside empathic robots in a harmonious society (Nahavandi, 2019; Dietz *et al.*, 2022). While the risk of superintelligence seems low in terms of probability, the magnitude of damage to be expected should the risk eventuate suggests at least keeping it in mind, as a worst (though unlikely) scenario.

That action must be taken in response to current impacts and risks is unanimously accepted. As for the distant and uncertain future of the human–machine relation, action taken now to address contemporary impacts and risks will also reduce the risk of a dystopian future. The question then is whose task is currently it to ensure AI stays responsible and aligned with human needs and rights. According to Dignum (2019, 109):

> AI development can be motivated by money and shareholder value, or by human rights and well-being, and societal values. It is up to us to decide. Is AI going to enhance our facilities and enable us to work better, or is it going to replace us? [. . .] The power of decision is the power of all of us. Researchers, developers, policymakers, users. Every one of us.

But when the task of protecting something is assigned to "every one of us," tragedies of commons tend to occur, so a clearer identification of entities with a social responsibility to act is needed. Here, these are the civil society (AI users, public figures, mass media and academia), scientists, state authorities and policymakers, the international community (international organizations constituted by not only states but also non-governmental organizations) and the business sector. Some of these are well researched, see for instance Knott *et al.* (2024) on the responsibility of media as a part of civil society and Frankel (2015) on the responsibility of scientists. State responsibility related to AI risks is also well researched, and the verdict is – despite effervescent relevant regulation in this area lately – that developing coherent and comprehensive AI regulatory frameworks is made difficult by a variety of factors, including the fluidity of the object of regulation: "a changing technology with changing uses, in a changing world" (Maas, 2022). Other objective factors inhibiting regulation are the delicate balance between ensuring safety and promoting innovation (that is, avoiding both underregulation and overregulation) and the lack of conceptual clarity surrounding the AI concept. The intense regulatory activity of the past few years (for an overview of the global regulatory race, see Saheb and Saheb, 2023; for a comparative analysis of underlying policy narratives in various countries, see Bareis and Katzenbach, 2022) reflects these difficulties.

One effect of the recent intense regulatory effort – see for instance the Artificial Intelligence Act of the European Union, adopted in 2024, and President Biden's Executive Order on the Safe, Secure and Trustworthy Development and Use of Artificial Intelligence, issued in October 2023 – is that the spotlight is on governments and rightly so. On the other hand, the role of the corporation in ensuring the minimization of AI risks and negative impacts is

insufficiently examined (Flyverbom *et al.*, 2019). This state of fact should be remedied, given the major role of the corporation in creating and deploying potentially harmful AI. The first step is a more precise conceptualization of the responsibility of business. It is important to go to the root of the various proposed norms and agree on a core principle, a unique and succinctly formulated norm upon which the more specific rules of proper behavior are built. That is:

P1. The core of artificial intelligence-related business's responsibility should be defined with precision.

Toward a new principle of corporate social responsibility

An AI-related foundational norm of corporate behavior belongs to the realm of CSR – the concept aligning the economic targets of the corporation (profit) with social expectations and imperatives. Admittedly, defining CSR is notoriously difficult. However, when "navigating through the jungle of definitions" (Crane *et al.,* 2014, 5), one can decant sets of core characteristics of CSR. For Crane *et al.,* (2014) for instance, the common denominator of the definitions examined is that CSR is voluntary (i.e. going beyond what the law mandates in a certain field), requires multiple stakeholder orientation, is about balancing social and economic responsibilities and interests, requires a set of values that underpin practice and is more than mere philanthropy. Analyzing definitions of CSR proposed in more than two decades, Dahlsrud (2008, 5) found the most frequent dimensions of the concept to be stakeholder orientation, social, economic and environmental integration and voluntariness. Kerr *et al.* (2009, 9) also noted that common to the numerous definitions they examined, from both the governmental and business spheres, is the element of integration of economic considerations, which is the traditional focus of the corporation, with environmental and social concerns. Similarly, following an examination of prior scholarship, Wickert and Risi (2019) refer to CSR as:

> [. . .] an umbrella term to describe how business firms, small and large, integrate social, environmental and ethical responsibilities to which they are connected into their core business strategies, structures and procedures [. . .](22).

The minimal understanding of CSR as integrating economic, social and environmental concerns is therefore unchallenged in the CSR literature. However, defining the concept in such a broad manner does not take us too far in terms of its operationalization in corporate practice. Accordingly, specific guidelines intended to operationalize the above definitional, umbrella principle were proposed in academia and policymaking circles. Early sets of business ethics guidelines proposed in academia (Frederick, 1991) coincided with the first wave of globalization and refer to responsible behavior in the areas of consumer protection, corruption, employment, environment and basic human rights. Some of these are also reflected in the Caux Principles for Business adopted in 1994 and in the principles of the Global Compact proposed in 2000 by the UN.

The discussion on principles of business ethics in the 1990s was largely informed by moral precepts. By the end of the first decade of the new millennium, it was noted, however, that "CSR, until recently sighted only off in the misty horizon of ethical slogans, has now taken on sharp enough normative contours in law [. . .]" (Kerr *et al.*, 2009, 3). Accordingly, in one of the most comprehensives radiographies of the CSR concept in terms of principles, Kerr *et al.* (2009) looked at the law, explaining that a positive approach should complement the deontological one that had prevailed before in the quest for ethical guidelines giving practical meaning to the CSR concept. This mixed inquiry, ethical and positivist, led the authors to the identification of the following principles of legal CSR: Principle 1 (reflecting the umbrella principle introduced above), "Integrated, Sustainable Decision-Making"; Principle 2, "Stakeholder Engagement"; Principle 3, "Transparency"; Principle 4,

"Consistent Best Practices"; Principle 5, "Precautionary Principle"; Principle 6, "Accountability"; and Principle 7, "Community Investment". The list, however, appears incomplete now, when looked at with 2020s lenses.

Some may argue that AI design and use are already related to CSR in organizations, via above-listed, existing CSR principles like transparency and accountability. The inherent connection is confirmed by the results of a global survey where 90% of managers in companies with at least $100m in annual revenues reported that their organization's responsible AI and CSR efforts are linked (Renieris *et al.*, 2022). But the responsibility of the corporation is higher in AI governance than in other areas, for two aggregated reasons (Rendtorff, 2019, 45): first, the current stage of technological advance brings about a much higher risk of damage to human and non-human world than was ever the case before and, second, the unavoidable gap between scientific and regulatory advance leads to weak and ambiguous AI legislation – making the self-regulatory response of the corporation crucial. Indeed, in this type of porous regulatory landscape, CSR presupposes that the corporation establishes and quickly updates its own ethical boundaries; after all, versatility is one of the claimed advantages of self-regulation over state regulation. However, a corporation willing to act responsibly despite the lack of a strong regulatory framework – say, by adopting a moral code of conduct on AI-related matters at company level – would lack an unanimously agreed upon foundational principle. That well-intending corporation would have available some guiding lines when it drafts its AI-related code of conduct (i.e. the same ethical principles proposed to inform government regulation in this area), but not the foundational principle that would coagulate the specific requirements of transparency, respect for human dignity, etc. That missing foundational principle is one that in this current age should have a place in the panoply of recognized principles of CSR. That is:

P2. A new corporate social responsibility principle, reflecting the socioeconomic and political challenges of the artificial intelligence era, is necessary.

What ethical norm should the artificial intelligence-related corporate social responsibility principle be based on?

The principles of CSR are important because they transcend the debate on voluntarism versus regulation, permeating the whole regulatory continuum, from *no regulation* at one end, to *self-regulation* (corporate voluntary codes) to *co-regulation* (or "principle-based" regulation, focused on outcome and leaving the choice of process to the corporation) and, finally, to rule-based *regulation* (on the regulatory continuum and types of regulation, Kaplow, 2000; Sama and Shoaf, 2005). Many of the CSR principles referred to above were born as mere expectations stemming from universal ethical norms, and only later, when society found it necessary, they were (at least partially) cast in the stone of hard regulation. The anticorruption principle is a good example, with its quick transformation from mere social desideratum based on ethical norms (honesty, fairness and social solidarity) in the second half of the 20th century, to hard and detailed regulation currently. Indeed, exclusive reliance on morality, even when incorporated in corporate voluntary codes of conduct, had proven not enough to curb corporate wrongdoing. In legislating CSR, the state is fulfilling its duty as a party to the social contract by protecting society.

Based on the model described above, an AI-related CSR principle would ground its journey from reasonable societal expectations to hard regulation in the moral norm of human dignity. Despite the notoriously controversial contours and pedigree of the human dignity concept, one rooted in Kantian philosophy and more generally in the natural law tradition, its characterization as a universal moral norm is hard to challenge, as human dignity is the foundation of the whole human rights edifice, mentioned as such in numerous international instruments (the UN Charter, the Universal Declaration on Human Rights and the two human rights covenants, among others) and national constitutions. Furthermore, respecting human

dignity has been already established as a justification for pro-responsible AI policy, either via the human rights link or directly, reflecting human dignity dimensions such as inherent worth, status and respect, including self-respect (Latonero, 2018; Risse, 2019; Ulgen, 2022). State regulation for responsible AI also explicitly acknowledges human dignity as a fundamental value to protect, see for example the already mentioned AI Act of the European Union; so do civil society initiatives, with the 2017 Asilomar Principles of Beneficial AI including one on the need to preserve human dignity. These considerations lead therefore to:

P3. The artificial intelligence-related corporate social responsibility principle should be rooted in the moral norm of human dignity.

Having established that the core of the business's responsibility on AI-related matters needs a more precise contour; that this is best achieved in the CSR framework as a distinct, new principle; and that the new principle should be grounded in the universal moral norm of human dignity, what is left is to formulate the principle.

Preserving the relevance of the human, as a corporate social responsibility principle

Human-centric artificial intelligence, a policy metaphor unsuitable as a corporate social responsibility principle

Human-centric AI is one immensely most popular expression of the past few years, generally used in the context of regulation for responsible AI. Policymakers claim their output to be human-centric, see the EU Commission President's statement that the AI Act will become a "substantial contribution to the development of global rules and principles for human-centric AI" (European Commission, 2023). According to the European Union guidelines for ethical AI:

> The human-centric approach to AI strives to ensure that human values are central to the way in which AI systems are developed, deployed, used and monitored, by ensuring respect for fundamental rights, including those set out in the Treaties of the European Union and Charter of Fundamental Rights of the European Union, all of which are united by reference to a common foundation rooted in respect for human dignity, in which the human being enjoy a unique and inalienable moral status. (European Commission, 2019).

Numerous academics have endeavored to define human-centric AI and to shed light on the way to achieve it (Bryson and Theodorou, 2019; Ulgen, 2022; Amariles and Baquero, 2023; Chetouani *et al.*, 2023). Among the conditions for human-centric AI they commonly identified are explainability and accountability. Unpacking the expression further, Chetouani *et al.* (2023) find human-centric AI to be defined by human goals and values, respect for human autonomy, inbuilt capacity for self-reflection, a communication style based on human language and the openness to take advice from humans. The main goals of human-centric AI are to ensure that that human values are incorporated into the design of algorithms and that humans maintain control over automated systems (Bryson and Theodorou, 2019, 4). Similarly to the above definition by the European Commission, numerous authors find the ultimate grounding of human-centric AI in human dignity, as a reverberation of the Kantian imperative to never treat humans as means to an end, but as ends in themselves – hence, "human-centric." The opposite of a human-centric AI would be a technology-biased system; one that would, for example, prioritize profit over human privacy and autonomy. Such an AI system would treat humans as means to an end and would, therefore, affect their dignity (Ulgen, 2022).

Given the appeal and popularity of the expression, it would be tempting to embrace it as a principle of CSR as well: corporations should respect the centrality of the human being in their design and use of AI. There are, however, reasons to avoid this formulation of the

needed AI-related principle of corporate responsibility. The most obvious objection is grounded in the CSR definition as merely integrating societal concerns, rather than making them central in corporation management. As the human person is at the center of state's concerns, human centrality is credible as an overarching imperative of AI governance, but things are different for the corporation, the central concern of which is profit. Relevantly, in the well-known Carroll's (1991) pyramid of corporate responsibility, the bottom layer is economic sustainability. Indeed, to be socially responsible, the corporation must first exist. In this perspective, the human person is relevant but not central to the corporate person's raison d'être; therefore, an AI-related CSR principle should be less ambitious than human centrality.

Further, the "human-centric AI" expression is a policy metaphor with a certain role. Its semantic sphere would have been very well covered by the notion "responsible AI," but the latter lacks the mobilizing (in support of AI regulation) and perhaps anesthetic undertones the concept of human centrality carries. Anesthetic, in the sense that the denomination "human-centric AI" may be intended as a counter-narrative to narratives of dystopian futures, as well as a counter-critique to claims of alienation, inequality and breach of privacy and autonomy brought about by current deployment of AI. This communicative potential may or may not be fulfilled when the expression is applied to state AI regulation, but it would certainly not work if applied to corporate AI-related governance. As Fischer (2003) shows in his elaborate analysis of metaphor and narrative in public policy, trust and belief are crucial to the success of a policy metaphor, and these depend, among other things, on the legitimacy of the entity making the claim. Corporations claiming to navigate the economy under the pennant of human-centricity would not be taken seriously.

The above arguments against importing the human-centricity governance principle as a new CSR principle are based on the distinct roles fulfilled by the state and the corporation. In other words, the narrative may fit the state, but not the corporation. There are, however, voices critical of the use of the "human-centric AI" syntagm altogether. One author finds the terminology both ambiguous and anachronistic, explaining his assertion as follows:

> On the one hand, it is obvious that any technology, AI included, must be at the service of humanity, its values, and needs. On the other hand, one must also consider the environment as crucially important, yet "humancentric" seems to be synonymous with "anthropocentric", and we know how much the planet has suffered from humanity's obsession with its importance and centrality, as if everything must always be at its service, including every aspect of the natural world, no matter at what costs and losses. (Floridi, 2021, 218).

The critique that human-centricity is both superfluous and improperly focused is echoed by other researchers. Taylor *et al.* (2023) explain that AI is already human-centric and setting already achieved targets distracts from needed action on making AI less alienating to society. As for focus, Mhlambi (2020) questions the emphasis on Western values in the discourse on human-centric AI, given the legacy of the West's dehumanization and atomization. Indeed, human centrality and its conceptual twin, human uniqueness, have a strong Western philosophical flavor, in both religious and secular understandings. A community- rather than human-centric approach to AI has, therefore, been proposed, infused with non-Western philosophies like Ubuntu (Mhlambi, 2020).

For all these reasons, an AI-related CSR principle should not be built around the concept of human-centric AI. This article proposes its substance to be, instead, the preservation of the human beings' relevance.

Preserving human relevance, as a corporate social responsibility principle

In his critique of the gradualist view of AI domination – the view claiming that the progress toward artificial superintelligence will be gradual enough to leave humans the time to learn

how to deal with it – Barrat (2013) reminds us that we should first be worried about the dangers we will face along the way:

> [...] while superintelligent machines can certainly wipe out humankind, *or make us irrelevant*, I think there is also plenty to fear from the AIs we will encounter on the developmental path to superintelligence (29; emphasis added).

But irrelevance of the human in the AI age, seen by Barrat as a threat associated with the superintelligence stage of AI, is in fact an ongoing threat. Unlike the risk of physical annihilation, materialization of the risk of losing relevance (annihilation in spirit) does not depend on reaching the stage of AGI; if that happens, then things will become much worse, but if not, then there will be anyway a creeping process of relevance loss, already undergoing.

The Cambridge online dictionary indicates the adjective "relevant" to mean closely related to the matter at hand. Words in its semantic proximity are "pertinent," "significant" and "important." The word in its negative form ("irrelevant") is used occasionally in relation to risks associated with AI development, either in a general sense, like in the above quote from Barrat (2013) or in more confined understandings. Brynjolfsson and McAfee (2015, 13), for example, analogizing the faith of humans in the age of AI to that of horses in the age of the steam and then the internal combustion engine, speak of "economic irrelevance." While this topical application affords at least intuitive comprehension of the phenomenon described, the general reference of the risk of human irrelevance, without anything else, seems little more than a metaphor to perhaps replace concepts like participation in socioeconomic and political life. If there is to use "human irrelevance" as a tool used in the critical analysis of the human's faith at the intersection with AI, then some conceptual unpacking is needed.

The relevance theory developed in information science is of some limited assistance. Limited, as this theory focuses on information as the object of relevance, while in our case, the object is the human race. However, a relationship is central to any relevance framework, so when we claim that humans risk becoming irrelevant in the age of AI, we should also clarify: irrelevant to whom? The answer could be the planet, but this would expose the argument to the same criticism mounted against the "human-centric AI" concept: it is anthropocentric. The thousands of species we wiped out, tortured and enslaved would certainly not deplore loss of human relevance in the AI world, so it is perhaps more appropriate to speak of a *sense* of irrelevance (irrelevance as perceived by humans themselves) rather than irrelevance objectively assessed.

Drawing further on information science relevance theory, not only the entities involved in the relevance relationship should be specified but also the nature of the relationship (Saracevic, 1975, 337). This, if we go back to the lay definition of relevance, then would mean asking what is the "matter of hand" and in which way is the object of relevance (humans) no longer "closely related" to it? The remainder of this section attempts to answer these questions, in the context of various types of irrelevance that humans may encounter/feel in the journey toward AGI.

The first is the already mentioned *economic irrelevance*. Humanity's transition toward the post-work society, where most of the population will have no job, seems inevitable (Ford, 2015; West, 2018; Dorsey, 2022). Reassurances based on the experience of prior technological revolutions, that the loss of some jobs will be compensated by the creation of others, are unconvincing: in previous centuries, machines have gradually replaced humans in manual jobs, while AI will in the end replace the human in virtually any job. This may be a longer process, decades perhaps, for the professions (Susskind and Susskind, 2022, 391), while for non-professional jobs in services, agriculture and industries, dramatic impacts on the job market are expected within a shorter time horizon. Admittedly, not necessarily a very short time: surveying patents owned by Amazon, Delfanti and Frey (2021) found that for now, they do not indicate an immediate disappearance of the human being from the

warehouse floor, though the future is uncertain. But altogether it is safe to assume that in the not-too-distant future, most humans will become economically irrelevant.

When most humans become irrelevant to the production of goods and to the delivery of services, a deeper, existential sense of relevance loss will ensue, as powerfully expressed by writer Stephan Talty: "I don't really fear zombie AI. I worry about humans who have nothing left to do in the universe except play awesome video games" (Talty, 2018). Indeed, optimistic accounts of the "ludic life" in which humans, supported via a universal basic income mechanism, will engage in the post-work society (Danaher, 2020) downplay the point that work is much more than doing a job for a salary. As individuals, we obtain an income allowing for a decent life, which in the post-work society will be hopefully provided via various solutions of universal basic income; but from work, we also derive dignity, self-esteem, knowledge and social recognition – all, it should be noted, highly relevant to the concept of human dignity, perhaps even more so here than in the discussion on "human-centric" AI. Further, because those denied work will deem themselves as socially irrelevant, they will face the risk of mental harm and even suicide (Classen and Dunn, 2012). In a broader perspective, societies derive harmony and prosperity from the aggregated work of individuals, which is why to philosopher John Dewey, "[t]he first great demand of a better social order [...] is the guarantee of the right, to every individual who is capable of it, to work" (Ratner, 1939, 420).

While economic irrelevance is the focus of numerous academic and policy outputs, the impacts of AI are significant on other planes of human relevance as well. To start with, humans are already being rendered increasingly irrelevant by AI from a *political* perspective. Scholars have throughout the years raised awareness of the mechanisms whereby our views of the world and our behavior are shaped by AI (Flyverbom *et al.*, 2019; den Hond and Moser, 2023; Stockinger *et al.*, 2024), to the effect that democracy is significantly eroded (Radavoi, 2020). On the one hand, democracy implies free will, which is becoming an illusion in the AI-dominated world (Damasio, 2018); on the other hand, AI is often better than humans at making decisions, so even without behavioral modification and mind control, humans will be more and more inclined to delegate to AI decision-making on major issues out of overtrust and moral complacency (Kaas, 2024). One way or the other, wittingly or unwittingly, the relevance of humans to political decision-making is decreasing.

Related to political irrelevance is what we may term *ethical irrelevance*. Admittedly, AI is better in many ways at making decisions, by standardizing processes and eliminating human bias – but reliance on AI's binary logic and formalization of knowledge has its drawbacks:

> AI has the potential to trivialize humanity by removing the subjective and emotional aspects of decision-making and reducing human decision-making to a purely rational process. This could lead to a loss of empathy and understanding and a lack of appreciation for the unique qualities that make people human (Koering, 2023, 9).

Further, we may talk about *evolutionary irrelevance*. Some humans will soon be able to take the transhumanity exit of the evolutionary highway, upgrading themselves into superhumans, with the help of AI (manipulation of human genome, use of cognitive extenders, brain–computer interfaces, etc.). The privileged will be able to enhance their bodies, either organically or using implanted technology, in the quest for immortality and domination (Harari, 2017). This will be a radical evolutionary breakaway and will likely lead to the creation of a superior class, but the rest of the population is also set to gradually shift away from the evolutionary path started some two hundred thousand years ago in Africa. While evolution so far has been conditioned through interaction with the natural and social landscape, the more we let AI take control of our minds and actions, the more AI becomes the "nature" that will shape our new evolutionary path, in physical and cognitive terms. In an article entitled "Where is the Human?", Van Den Eede (2021), for example, maps the debate

in terms of the beliefs supported by various camps, from transhumanists to bioconservatives, and concludes that "[n]o matter how dispersed we are, there still seems to remain a 'we', a human existential–experiential condition to guard and perhaps cherish" (159).

Deriving from the above is the risk of *historical irrelevance*. When it becomes increasingly likely that humans of the, say, 24th century will be half tissue half technology, they will no longer be as relevant to preceding generations as we were to our predecessors. Today, we still mobilize to preserve the planet for future generations, but this enthusiasm for distant successors may vanish with the realization that those successors will be very different to us. Conversely, how will our primitive lives and decisions look to entities of the distant future? We will not be more relevant to them as *homo erectus* is to us.

Even loss of *military relevance* could be mentioned, although this is not necessarily a bad thing – at least not for the side holding the technology cards. Yet, in a study of a Dutch military innovation hub, Van Der Maarel *et al.* (2023) surprisingly found disillusionment among soldiers upon the realization that many of their roles in combat will be taken by robots.

All these facets of potential loss of human relevance, and possibly others not identified in this paper, are interconnected, and the corporation's role in some is stronger than in others. The most prominent role is around economic relevance, and – justifying the new CSR principle – this is also where the links to human dignity are the most visible. As for the other types of AI-induced irrelevance, corporate action should be guided and if necessary, constrained by state regulation, but there remains, in the regulatory space, enough room for the corporation to prove voluntarily responsible and apply the CSR principle of preserving the human's relevance.

Conclusion and practical implications

As business has the central role in both creation and use of AI, it is only natural that corporations act alongside governments for minimizing the risks of AI while fulfilling its potential and promises. Against this background, this essay has presented two claims. First, it proposed that an AI-related principle of CSR should be agreed upon, one that would encapsulate the numerous directions of corporate action suggested in the literature: transparency, precaution, respect for human dignity and so on. And second, this article argued that the content of said principle should be the preservation of human relevance in the age of AI; the banner of human-centric AI is not suitable for the corporation, given the centrality of profit for this type of institution. Preserving human relevance is a much lower expectation, obviously, but one that fits the corporate DNA, is rooted in human dignity just like the human-centric AI concept and is more credible (being less bombastic and less Western-focused than the human-centric claim).

Built at the intersection of the CSR and of ethical AI academic debates, this article's theoretical contribution is twofold. First, as far as the CSR literature is concerned, the article completes the list of CSR principles with one that is necessary in the 21st century. While CSR principles such as transparency, community engagement or refraining from corruption are unanimously acknowledged in academia, policymaking and corporate practice, to the best of the author's knowledge, this article is the first to propose an AI-related CSR principle. Second, in the debate on ethical AI, this article contributes by exposing the weakness of the "human centric" metaphor claimed to guide the states' quest for responsible AI and its unsuitability to corporations.

If the proposed formulation for the AI-related CSR principle was accepted in public and corporate discourse, then the main practical implication would be that civil society organizations, academics and concerned citizens would have a credible, robust benchmark against which to assess and when case, critique the corporate creation and use

of AI. Indeed, as much of AI is created purposely to replace the human, an expectation that business maintains AI human centric would make no sense – but demanding business to use AI in a manner that keeps the human relevant would be sensible. Accordingly, corporations that will maintain a balanced approach, keeping the human economically and socially relevant as much as possible, will be rewarded with a good reputation, as the case of several retail chains returning to human cashiers indicated, at the end of 2023 (Meyersohn, 2023). Another, somewhat related field of direct application would be the replacement of human operators with robots as interface for client communication in banks and other types of business – a frustrating experience against which the CSR principle of preserving human relevance would provide a platform for critique. In short, acknowledging preservation of human relevance, in the form of a CSR principle, as a minimum standard in AI-related corporate decision-making, would create awareness and would first give legitimacy to relevant community claims.

Corporate and government decision-makers would necessarily react to the "human relevance" narrative, if civil society embraces it. As a first step, the syntagm "human relevance" would likely appear in corporate speech just as "human centrality" now features in public policymaking speech. Relevant adjustments of the codes of conduct would follow. As for governmental action, similarly to other CSR principles, this would also begin with support shown to the principle, via public speech and policy papers. Hard regulation is difficult to envisage at this point, but so was in the 1980s in regard to corruption.

Further research could identify more of the new principle's domains of application, possibly as case studies, which in turn would hopefully result in more specific self-regulatory action by the corporation. More generally, it is hoped that this article will inspire interdisciplinary research further clarifying the rich potential of CSR to contribute to more ethical AI.

Funding

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

References

Amariles, D.R. and Baquero, M. (2023), "Promises and limits of law for a human-centric artificial intelligence", *Computer Law & Security Review*, Vol. 48, p. 105795, doi: 10.1016/j.clsr.2023.105795.

Bareis, J. and Katzenbach, C. (2022), "Talking AI into being: the narratives and imaginaries of national AI strategies and their performative politics", *Science, Technology, & Human Values*, Vol. 47 No. 5, pp. 855-881, doi: 10.1177/01622439211030007.

Barrat, J. (2013), *Our Final Invention: Artificial Intelligence and the End of the Human Era*, St. Martin's Press.

Bednar, K. and Spiekermann, S. (2024), "Eliciting values for technology design with moral philosophy: an empirical exploration of effects and shortcomings", *Science, Technology, & Human Values*, Vol. 49 No. 3, pp. 611-645, doi: 10.1177/01622439221122595.

Bostrom, N. (2014), *Superintelligence: Paths, Dangers, Strategies*, Oxford University Press.

Brynjolfsson, E. and McAfee, A. (2015), "Will humans go the way of horses", *Foreign Affairs*, Vol. 94, pp. 8-14. available at: www.foreignaffairs.com/world/will-humans-go-way-horses

Bryson, J.J. and Theodorou, A. (2019), "How society can maintain human-centric artificial intelligence", in Toivonen, M. and Saari, E. (Eds), *Human Centered Digitalisation and Services*, Springer.

Carroll, A.B. (1991), "The pyramid of corporate social responsibility: toward the moral management of organizational stakeholders", *Business Horizons*, Vol. 34 No. 4, pp. 39-48, doi: 10.1016/0007-6813(91)90005-G.

Chetouani, M., Dignum, V., Lukowicz, P. and Sierra, C. (Eds) (2023), *Human-Centered Artificial Intelligence: Advanced Lectures*, Springer.

Classen, T.J. and Dunn, R.A. (2012), "The effect of job loss and unemployment duration on suicide risk in the united states: a new look using mass-layoffs and unemployment duration", *Health Economics*, Vol. 21 No. 3, pp. 338-350.

Crane, A., Matten, D. and Spence, L.J. (2014), *Corporate Social Responsibility: Readings and Cases in a Global Context*, Routledge.

Dahlsrud, A. (2008), "How corporate social responsibility is defined: an analysis of 37 definitions", *Corporate Social Responsibility and Environmental Management*, Vol. 15 No. 1, pp. 1-13, doi: 10.1002/csr.132.

Damasio, A. (2018), "The strange order of things: life", *Feeling and the Making of Cultures*, Pantheon Books.

Danaher, J. (2020), "In defense of the post-work future: withdrawal and the ludic life", in Cholbi, M. and Weber, M. (Eds), *The Future of Work, Technology and Basic Income*, Routledge.

Delfanti, A. and Frey, B. (2021), "Humanly extended automation or the future of work seen through amazon patents", *Science, Technology, & Human Values*, Vol. 46 No. 3, pp. 655-682, doi: 10.1177/0162243920943665.

Den Hond, F. and Moser, C. (2023), "Useful servant or dangerous master? Technology in business and society debates", *Business & Society*, Vol. 62 No. 1, pp. 87-116, doi: 10.1177/00076503211068029.

Diamantis, M.E. (2020), "The extended corporate mind: when corporations use AI to break the law", *North Carolina Law Review*, Vol. 98 No. 4, pp. 893-932.

Dietz, E., Kakas, A. and Michael, L. (2022), "Argumentation: a calculus for human-centric AI", *Frontiers of Artificial Intelligence*, Vol. 5, p. 955579, doi: 10.3389/frai.2022.955579.

Dignum, V. (2019), *Responsible Artificial Intelligence: How to Develop and Use AI in a Responsible Way*, Springer.

Dorsey, S. (2022), *Upper Hand: The Future of Work for the Rest of Us*, Wiley.

European Commission (2019), "Ethics guidelines for trustworthy AI", Publications Office, available at: https://data.europa.eu/doi/10.2759/346720

European Commission (2023), "Commission welcomes political agreement on artificial intelligence act", Press Release, 9 Dec, available at: https://ec.europa.eu/commission/presscorner/detail/en/ip_23_6473

Fischer, F. (2003), *Reframing Public Policy: Discursive Politics and Deliberative Practices*, Oxford University Press.

Floridi, L. and Cowls, J. (2019), "A unified framework of five principles for AI in society", *Harvard Data Science Review*, Vol. 1, available at: https://hdsr.mitpress.mit.edu/pub/l0jsh9d1/release/8

Floridi, L. (2021), "The European legislation on AI: a brief analysis of its philosophical approach", *Philosophy & Technology*, Vol. 34 No. 2, pp. 215-222, doi: 10.1007/s13347-021-00460-9.

Flyverbom, M., Deibert, R. and Matten, D. (2019), "The governance of digital technology, big data, and the internet: new roles and responsibilities for business", *Business & Society*, Vol. 58 No. 1, pp. 3-19, doi: 10.1177/0007650317727540.

Ford, M. (2015), *Rise of the Robots: Technology and the Threat of a Jobless Future*, Basic Books.

Frankel, M.S. (2015), "An empirical exploration of scientists' social responsibilities", *Journal of Responsible Innovation*, Vol. 2 No. 3, pp. 301-310, doi: 10.1080/23299460.2015.1096737.

Frederick, W.C. (1991), "The moral authority of transnational corporate codes", *Journal of Business Ethics*, Vol. 10 No. 3, pp. 165-177, doi: 10.1007/BF00383154.

Harari, Y.N. (2017), *Homo Deus: A Brief History of Tomorrow*, Vintage.

Kaas, M.H.L. (2024), "The perfect technological storm: artificial intelligence and moral complacency", *Ethics and Information Technology*, Vol. 26 No. 3, p. 49, doi: 10.1007/s10676-024-09788-0.

Kaplow, L. (2000), "General characteristics of rules", in Bouckaert, B. and De Geest, G. (Eds), *Encyclopedia of Law and Economics*, Edward Elgar, Vol. 5, doi: 10.4337/9781782540519.00008.

Kerr, M., Janda, R. and Pitts, C. (2009), *Corporate Social Responsibility: A Legal Analysis*, LexisNexis Canada.

Knott, A., Pedreschi, D., Jitsuzumi, T., Leavy, S., Eyers, D., Chakraborti, T., Trotman, A., Sundareswaran, S., Baeza-Yates, R., Biecek, P. and Weller, A. (2024), "AI content detection in the emerging information ecosystem: new obligations for media and tech companies", *Ethics and Information Technology*, Vol. 26 No. 4, p. 63, doi: 10.1007/s10676-024-09795-1.

Koering, D. (2023), "Exploring the human-AI nexus: a friendly dispute between second-order cybernetical ethical thinking and questions of AI ethics", *Enacting Cybernetics*, Vol. 1 No. 1, p. 5, doi: 10.58695/ec.4.

Koniakou, V. (2023), "From the 'ush to etehics' to the 'race for governance' in artificial intelligence", *Information Systems Frontiers*, Vol. 25 No. 1, pp. 71-102, doi: 10.1007/s10796-022-10300-6.

Krkač, K. (2019), "Corporate social irresponsibility: humans vs artificial intelligence", *Social Responsibility Journal*, Vol. 15 No. 6, pp. 786-802, doi: 10.1108/SRJ-09-2018-0219.

Latonero, M. (2018), "Governing artificial intelligence: upholding human rights and dignity", Data and Society Research Institute, available at: https://apo.org.au/node/196716

Maas, M.M. (2022), "Aligning AI regulation to sociotechnical change", in Bullock, J., Zhang, B., Chen, Y.-C., Himmelreich, J., Young, M., Korinek, A. and Hudson, V. (Eds), Oxford Handbook on AI Governance, Oxford University Press.

Meyersohn, N. (2023), "Walmart, Costco and other companies rethink self-checkout." CNN", 23 November, available at: https://edition.cnn.com/2023/11/13/business/self-checkout-stores-shopping/index.html

Mhlambi, S. (2020), "From rationality to relationality: ubuntu as an ethical and human rights framework for artificial intelligence governance", *Carr Center for Human Rights Policy Discussion Paper Series*, Vol. 9.

Nahavandi, S. (2019), "Industry 5.0—a human-centric solution", *Sustainability*, Vol. 11 No. 16, p. 4371, doi: 10.3390/su11164371.

Radavoi, C.N. (2020), "The impact of artificial intelligence on freedom, rationality, rule of law and democracy: should we not be debating it?", *Texas Journal on Civil Liberties and Civil Rights*, Vol. 25 No. 1, pp. 105-126.

Ratner, J. (1939), *Intelligence in the Modern World: John Dewey's Philosophy*, Modern Library.

Renieris, E.M., Kiron, D. and Mills, S. (2022), "Should organizations link responsible AI and corporate social responsibility? It's complicated", MIT Sloan Management Review, available at: https://sloanreview.mit.edu/article/should-organizations-link-responsible-ai-and-corporate-social-responsibility-its-complicated/

Rendtorff, J.D. (2019), *Philosophy of Management and Sustainability: Rethinking Business Ethics and Social Responsibility in Sustainable Development*, Emerald Publishing.

Rendtorff, J.D. (2022), "Outline of ethical issues concerning government, business and information technologies", in Lütge, C., Uhl, M., Kriebitz, A. and Max, R. (Eds), *Business Ethics and Digitization. J.B. Metzler*, doi: 10.1007/978-3-662-64094-4_8.

Risse, M. (2019), "Human rights and artificial intelligence: an urgently needed agenda", *Human Rights Quarterly*, Vol. 41 No. 1, pp. 1-16, doi: 10.1353/hrq.2019.0000.

Saheb, T. and Saheb, T. (2023), "Topical review of artificial intelligence national policies: a mixed method analysis", *Technology in Society*, Vol. 74, p. 102316, doi: 10.1016/j.techsoc.2023.102316.

Sama, L.M. and Shoaf, V. (2005), "Reconciling rules and principles: an ethics-based approach to corporate governance", *Journal of Business Ethics*, Vol. 58 Nos 1-3, pp. 177-185, doi: 10.1007/s10551-005-1402-y.

Saracevic, T. (1975), "Relevance: a review of and a framework for the thinking on the notion in information science", *Journal of the American Society for Information Science*, Vol. 26 No. 6, pp. 321-343, doi: 10.1002/asi.4630260604.

Sartori, L. and Theodorou, A. (2022), "A sociotechnical perspective for the future of AI: narratives, inequalities, and human control", *Ethics and Information Technology*, Vol. 24 No. 1, pp. 1-11, doi: 10.1007/s10676-022-09624-3.

Stockinger, E., Maas, J., Talvitie, C. and Dignum, V. (2024), "Trustworthiness of voting advice applications in Europe", *Ethics and Information Technology*, Vol. 26 No. 3, p. 55, doi: 10.1007/s10676-024-09790-6.

Šucha, V. and Gammel, J. (2021), "Humans and societies in the age of artificial intelligence", Publications office of the European Commission, available at: https://data.europa.eu/doi/10.2766/61164

Susskind, R., and Susskind, D. (2022), *The Future of the Professions: How Technology Will Transform the Work of Human Experts*, Oxford University Press.

Talty, S. (2018), "What will our society look like when artificial intelligence is everywhere?", The Smithsonian Magazine, April 2018, available at: www.smithsonianmag.com/innovation/artificial-intelligence-future-scenarios-180968403/

Taylor, R.R., O'Dell, B. and Murphy, J.W. (2023), "Human-centric AI: philosophical and community-centric considerations", *AI & Society*, Vol. 39 No. 5, pp. 2417-2424, doi: 10.1007/s00146-023-01694-1.

Tegmark, M. (2017), *Life 3.0: Being Human in the Age of Artificial Intelligence*, Knopf.

Ulgen, O. (2022), "AI and the crisis of the self: protecting human dignity as status and respectful treatment", in Hampton, A.J. and DeFalco, J.A. (Eds), *The Frontlines of Artificial Intelligence Ethics: Human-Centric Perspectives on Technology's Advance*, Routledge.

Van den Eede, Y. (2021), "Where is the human? Beyond the enhancement debate", *Science, Technology, & Human Values*, Vol. 40 No. 1, pp. 149-162, doi: 10.1177/0162243914551284.

Van der Maarel, S., Verweij, D., Kramer, E.-H. and Molendijk, T. (2023), "This is not what I signed up for: sociotechnical imaginaries, expectations, and disillusionment in a Dutch military innovation hub", *Science, Technology, & Human Values, Online First*, Vol. 50 No. 4, doi: 10.1177/01622439231211032.

West, D.M. (2018), *The Future of Work: Robots, AI, and Automation*, Brookings Institution Press.

Wickert, C. and Risi, D. (2019), *Corporate Social Responsibility*, Cambridge University Press.

Further reading

De George, R.T. (1993), *Competing with Integrity in International Business*, Oxford University Press.

Radner, D. and Radner, M. (1989), *Animal Consciousness*, Prometheus Books.

About the author

Ciprian N. Radavoi is the author of two books and numerous articles in reputed academic journals, dealing with various aspects of corporate social responsibility, social justice, law and society. Currently, an Associate Professor of law at the University of Southern Queensland, Australia, Radavoi joined academia in 2012 with a position at the University of International Business and Economics, Beijing, China. Prior, he was a lawyer focused on human rights cases, then a diplomat, with consular positions in Asia and Africa. Ciprian N. Radavoi can be contacted at: ciprian.radavoi@unisq.edu.au