Asia Pacific Academy of Science Pte. Ltd.

Metaverse 2026, 7(4), m260046.
https://doi.org/10.54517/m260046

Perspective

Substitution without consciousness, moral judgment and the 'machinization' of humans in the metaverse

José Pablo Salazar-Aguilar

Latin University of Costa Rica, San José, Costa Rica

*__Corresponding author:__ José Pablo Salazar-Aguilar, jose.salazar@ulatina.cr

CITATION

Salazar-Aguilar JP. Substitution without consciousness, moral judgment and the 'machinization' of humans in the metaverse. Metaverse. 2026; 7(4): m260046. https://doi.org/10.54517/m260046

COPYRIGHT

Copyright © 2026 by author(s). *Metaverse* is published by Asia Pacific Academy of Science Pte. Ltd. This work is licensed under the Creative Commons Attribution (CC BY) license. https://creativecommons.org/licenses/by/4.0/

Abstract: Artificial intelligence, embodied agents, and immersive environments are reshaping human–machine relations by enabling computational systems to act as socially present interlocutors and to influence decisions, identities, and forms of interaction. This Perspective examines the ethical and social implications of that transformation through critical theory and recent research on artificial agency, algorithmic decision-making, social presence, and digital inequality. Its purpose is to distinguish functional performance from subjective experience and to assess the consequences of delegating evaluative authority to systems that can simulate empathy, self-reference, and normative reasoning without demonstrating experiential or accountable agency. The analysis develops three concepts: The consciousness gap identifies the distance between observable machine performance and evidence of subjective experience; quasiperfect objectivity describes the authority acquired when computational consistency and precision obscure the social choices embedded in data, classifications, and optimization goals; and predesigned difference captures the risk that diversity becomes recognized only when it can be encoded, classified, and administered. This analytical viewpoint argues that immersive digital infrastructures can simultaneously humanize machines and make people increasingly readable to them, thereby reinforcing asymmetries of power, status, and responsibility. It concludes that technological systems may substitute numerous human functions without becoming responsible social subjects, and that critical reflection, contestation of imposed ends, and accountability must remain non-delegable principles in increasingly automated and immersive societies.

Keywords: Artificial intelligence; Human–machine interaction; Critical theory; Instrumental reason; Algorithmic governance; Digital inequality

1. Introduction: The humanization of machines and the machinization of humans

The convergence of artificial intelligence (AI), immersive environments, embodied agents, and persistent virtual worlds is transforming the relationship between humans and machines. In the metaverse, computational systems are no longer merely tools, but increasingly appear as advisors, companions, teachers, caregivers, creators, and decision-makers whose voices, avatars, memories, and adaptive responses produce a social presence. A recent Perspective published in this journal argued that, as AI assumes cognitive functions in medicine, the human role can regain importance through empathy and care [1]. The present article extends that discussion beyond professional complementarity and asks a more fundamental question, *what happens to consciousness, moral judgment, and social responsibility when machines are humanized while humans are reorganized according to machine-readable criteria?*

This question requires distinguishing observable performance from subjective

experience. Contemporary AI can generate language, classify situations, imitate emotional understanding, and recommend conduct, but these abilities do not establish that a system feels, understands, suffers, or knows that it is acting. Current consciousness research therefore proposes theory-derived indicators rather than inferring consciousness from behavioral resemblance alone [2]. This Perspective proposes a more limited statement, functional equivalence is not experiential equivalence, and automated normative performance does not constitute moral judgment. A machine may produce an ethically acceptable answer without experiencing obligation, recognizing another as a moral subject, or assuming responsibility for the consequences.

The problem is therefore twofold. Machines are humanized when intelligence, intention, empathy, or agency are attributed to their outcomes, while human beings are mechanized when experience, identity, and behavior are translated into profiles, probabilities, scores, and optimized choices. Horkheimer and Adorno's critique of instrumental reason remains relevant because it shows that reason can be reduced to the efficient selection of means, while the legitimacy of the ends remains unexamined [3]. Recent critical approaches to AI ethics similarly argue that ethical analysis must address the power structures shaping technological systems and the conditions under which they constrain human emancipation [4].

This double movement may intensify in the metaverse, where artificial agents can become persistent, embodied, personalized, and socially persuasive. Experimental research indicates that AI behavior can influence human moral decisions and alter the explicit attribution of responsibility [5]. The main danger, then, is not whether machines will eventually become conscious, but that institutions and users may delegate authority to systems that simulate subjectivity while remaining unable to experience the consequences or respond morally for them.

This article develops three concepts to examine that danger, namely: the **consciousness gap**, which separates machine performance from evidence of subjective experience; **quasiperfect objectivity**, which describes the appearance of neutrality created by computational consistency and precision; and **predesigned difference**, through which diversity is accepted only when it can be encoded, classified, and administered.

These concepts operate at different analytical levels. The consciousness gap is primarily a descriptive-epistemic distinction; quasiperfect objectivity and predesigned difference are critical-theoretical constructs for examining technologically mediated power, classification, and social organization. The principle of non-delegability, developed in the final section, constitutes the normative implication derived from this framework.

The central argument is that machines may substitute human functions without becoming conscious moral subjects, while the technologization of social life may simultaneously weaken the conditions for human critical consciousness. Moral judgment and responsibility must therefore remain non-delegable, even in increasingly automated and immersive societies.

2. Consciousness and moral judgment beyond functionality

The debate about artificial consciousness is often obscured by the tendency to treat intelligence, agency, consciousness, and morality as interchangeable properties, which they are not. A system may display goal-directed behavior, adapt to changing conditions, and affect its environment without possessing subjective experience. Dung's multidimensional account treats artificial agency as a variable profile composed of goal-directedness, autonomy, world-directed action, long-term planning, and responsiveness to reasons [6]. This avoids describing AI

as either a passive tool or a fully human-like agent. Yet even a sophisticated agency profile does not establish that a system experiences its goals, understands its reasons, or has a first-person stake in what it does.

This Perspective therefore does not deny artificial agency in a functional sense. It rejects a different inference: that goal-directedness, autonomy, responsiveness, or other functional dimensions of agency are sufficient evidence of subjective consciousness or full moral responsibility.

The same caution must be applied to consciousness. Because subjective experience is not directly observable from the outside, fluent language or a convincing self-description cannot constitute conclusive evidence. Butlin et al. propose assessing the potential consciousness of machines using indicators derived from competing scientific theories, rather than solely through behavioral similarities [2]. Their proposal doesn't resolve whether artificial consciousness is possible, but it does establish an important limit: increasingly human-like performance cannot close the consciousness gap. A machine can describe pain without suffering, narrate introspection without experiencing an inner life, or use the language of responsibility without being responsible for an action.

Moral performance must consequently be distinguished from moral judgment. AI can classify conduct according to rules, reproduce ethical vocabularies, compare predicted consequences, and recommend an option that conforms to accepted standards. These operations may be useful, but they do not necessarily amount to moral agency. Zafar argues that minimal definitions of artificial moral agency import autonomy, rationality, and responsibility in ways that differ substantially from their application to human agents [7]. A more accurate description would be that of automated normative performance, which states that the system generates a result within a technically and socially defined field of norms, without demonstrating moral understanding or ownership of the decision.

From this perspective, a sound moral judgment involves more than selecting a compatible or optimal response. It requires recognizing conflicts between values, interpreting the specific situation of others, questioning the legitimacy of available options, and accepting responsibility for the consequences. Moral judgment begins where optimization ends because no optimization function can determine whether the goal it pursues is worth pursuing. A system can order the available means; the moral subject must still question whether the end, the categories, and the authority that define the problem are justifiable.

This distinction also helps explain why nominally keeping a "human in the loop" is insufficient. Robbins argues that meaningful human control must begin by determining which outcomes should not be delegated and asserts that evaluative judgments depend on the considerations by which they are reached [8].

Machines can outperform humans in consistency, memory, prediction, and rule application while remaining outside the experiential and accountability dimensions of moral life. The ethical problem arises when simulated judgment is treated institutionally as judgment itself, allowing authority to shift toward systems incapable of grasping the human meaning of their consequences.

3. Instrumental reason and quasiperfect objectivity

Horkheimer and Adorno's critique of instrumental reason offers a useful starting point for understanding the ethical problem of artificial intelligence and the impossibility of determining whether a system produces correct results. Their concern was not directed at calculation itself, but at the narrowing of reason to the point of becoming an efficient selection of means while the legitimacy of the ends remains outside the scope of scrutiny [3]. Under this logic, that which can be measured, predicted, and controlled acquires greater authority than that which is ambiguous, experiential, or resistant to standardization. Digital

systems intensify this tendency because social judgments can be translated into variables, thresholds, classifications, and optimization criteria, giving contingent decisions the appearance of technical necessity. Contemporary applications of Adorno's thought similarly identify digital monitoring, algorithmically organized consumption, and platform-mediated discourse as extensions of a technologically administered heteronomy.

Recent critical approaches to AI ethics extend this argument beyond abstract principles. Waelen argues that AI ethics should diagnose power relations and safeguard human emancipation, rather than simply evaluating isolated systems based on general principles [4]. Krijger shows that the operationalization of ethics depends on organizational assumptions, underlying values, and conflicts between principles that cannot be resolved solely through metrics [9]. Thus, ethical design never constitutes a neutral translation of values into code; it is a situated process in which institutions decide which values prevail, their interpretation, and the trade-offs become technically acceptable.

The term quasiperfect objectivity is used here to describe the authority that arises when computational precision, procedural consistency, and institutional legitimacy converge. The term does not imply that algorithmic decisions are genuinely objective, but rather refers to their capacity to appear less questionable than human judgments because they are expressed through numbers, models, and repeatable procedures. De Ruvo describes algorithmic objectivity as ideological when it obscures the power practices embedded in computational systems and weakens the capacity of individuals to critically perceive those practices [10]. Quasiperfect objectivity adds a key relational dimension, presenting the outcome as superior and devoid of human values, where those affected by it occupy a weaker epistemic position than those who design, own, or authorize the system.

Unlike automation bias, which describes a tendency of users to defer to automated recommendations, quasiperfect objectivity refers to the broader social authority produced when computational consistency, quantitative precision, and institutional endorsement converge. It therefore concerns not only individual reliance on algorithms, but the conditions under which technologically mediated judgments become structurally harder to contest.

Artificial environments cannot solve this problem by separating machines from ordinary social interaction. A training or simulation environment may be synthetic, but it is not asocial; thus, the artificial environment is best understood as a compressed social order, in which social assumptions are translated into technical parameters and subsequently reintroduced into society as machine-generated outcomes. The appearance of autonomy masks a chain of prior human decisions. Krijger's organizational approach supports this interpretation by showing that design processes are still shaped by background assumptions, governance arrangements, and decisions about the technical formalization of competing ethical demands [9].

Nor should it be understood that near-perfect objectivity produces automatic obedience in all contexts. Alon-Barkat and Busuioc identified selective adherence when algorithmic advice coincided with pre-existing stereotypes in some of its evidence [11]. Human deference to algorithms is not mechanically determined; it is mediated by institutional context, prior beliefs, professional knowledge, and social categories. Algorithmic authority can thus reinforce existing judgments, while allowing decision-makers to experience the outcome as if it had been externally validated.

Reeves and Sinnicks' contemporary reading of Adorno presents digitization as an extension of heteronomy through monitoring, algorithmically organized consumption, and platform-mediated political discourse [12]. The ethical danger

extends beyond machines directly controlling human beings; it lies in people increasingly encountering socially produced classifications as if they were independent facts. When technical consistency is mistaken for neutrality, and neutrality for moral legitimacy, judgment is displaced without disappearing. In the metaverse, where identity, behavior, and social presence can be continuously measured and modified, this displacement can become a constitutive feature of the environment itself.

4. Metaverse as a relational infrastructure

What makes the metaverse particularly relevant to this argument is the convergence of four properties that may otherwise remain dispersed across digital environments: immersion, persistence, embodied social presence, and intensive behavioral datafication. Their combination can amplify both the attribution of human-like qualities to artificial agents and the translation of human subjects into continuously measurable and administrable data.

The metaverse should not be understood merely as a set of increasingly complex virtual spaces that reproduce pre-existing social relations. Its architectures actively shape the emergence of individuals, forms of communication, movement, recognition, and visibility to institutions. Avatars, interfaces, identity protocols, rules of behavior, recommendation systems, and access permissions measure interactions after they have occurred and establish many of the conditions under which interaction becomes possible. Thus, the metaverse constitutes a critical relational infrastructure in which a sociotechnical environment organizes forms of presence, capabilities, recognition, and authority, both for human beings and for artificial systems.

This relational character distinguishes immersive environments from those representations of technology that consider it an external instrument. Hennig-Thurau et al. describe real-time, multisensory social interactions as a central feature of the metaverse and show that their effects depend on mediating conditions and interaction formats, not on an intrinsic superiority of virtual reality over two-dimensional communication [13]. Their findings are important because they depart from technological determinism, which states that immersion modifies the conditions under which interactants perceive, evaluate, and respond emotionally to one another. Consequently, the social effects of the metaverse must be studied through the relationships that its design enables, restricts, or intensifies.

Recent research indicates that social virtual reality can elicit interaction patterns comparable to those observed in face-to-face encounters. Son and Rubo found that participants using avatars similar to themselves or generic avatars exhibited natural verbal and nonverbal behaviors, suggesting that virtual environments can sustain socially meaningful exchanges even when identity is partially anonymized [14]. This does not mean that virtual interaction is equivalent in all respects to physical interaction. It demonstrates that an artificial environment can sustain behaviors experienced by users as socially authentic. It should be kept in mind that the environment may be synthetic, but the relationships, emotions, expectations, and consequences that arise within it are not socially unreal.

The same relational mechanisms can extend social efficacy to artificial agents. Lim et al. compared interactions with embodied AI conversational agents within immersive virtual reality with text-based exchanges and found a greater perceived presence in the embodied condition [15]. The study shows that an artificial system can become more socially relevant when it occupies a body, shares a space, directs its gaze, uses a voice, and responds conversationally. Thus, embodiment reduces the perceived distance between technical operation and social participation without resolving the ontological difference between simulated

presence and subjective experience.

Attributed consciousness refers here to the user's inference that an artificial entity possesses an experiencing inner perspective on the basis of behavioral and relational cues. It differs from anthropomorphism, which concerns the attribution of human-like characteristics; from social presence, which concerns the perceived presence of another social actor; and from trust, which concerns willingness to rely on that actor. These phenomena may reinforce one another, but none of them demonstrates subjective experience. An AI agent does not need to become conscious to be treated as if it understands, cares, remembers, or intends. Its social power can arise from the human interpretation of its performance.

Anthropomorphic appearance alone, however, should not be treated as a sufficient or universal explanation of trust. Huynh and Aichner found that social presence and emotional engagement were positively associated with human-like trust in generative AI, whereas anthropomorphism by itself had no significant effect in their study [16]. This middle-range finding qualifies broad claims that merely making machines look human necessarily produces obedience. Trust emerges from a more complex interaction among perceived fairness, emotional responsiveness, social presence, user expectations, institutional context, and the task being performed. The ethical concern is therefore not anthropomorphism in isolation, but the coordinated design of relational cues capable of making a system appear competent, benevolent, intimate, and socially accountable.

Describing the metaverse as a managed relational infrastructure means that interaction unfolds within architectures whose underlying possibilities have already been predetermined. Platform designers and owners determine which identities can be registered, permitted movements, behaviors rewarded or sanctioned, data collected, and visible artificial agents. Users can exercise creativity within these conditions, but they do not enter an empty or neutral space.

The metaverse thus creates a double asymmetry. Artificial agents can acquire social presence without awareness, while human users can experience meaningful relationships within systems whose governance assumptions remain beyond their grasp. The crucial question is not whether virtual relationships are real enough to matter—evidence suggests they can be. The fundamental question is who defines their conditions, the interests encoded in the environment, and whether users can recognize and challenge the infrastructures that organize their perception of others. When artificial presence is mistaken for moral presence, relational authority can be transferred to systems capable of influencing human experience but incapable of experiencing or taking responsibility for its consequences.

5. Predesigned difference and algorithmic stratification

Technological systems do not eliminate human difference; they reorganize the conditions under which it becomes visible, intelligible, and actionable. In immersive environments, users may appear free to select bodies, identities, abilities, and modes of interaction, but these choices depend on the categories that platforms make available and the technical standards by which identities can be represented. This perspective uses the concept of predesigned difference to describe a condition in which diversity is recognized only insofar as it can be anticipated, codified, categorized, and managed. Difference is not prohibited; it is translated into a finite repertoire of technically readable options.

The common mechanism connecting these inequalities is machine readability: social differences become actionable for technological systems only after they are translated into categories that can be detected, classified, compared, and administered.

This problem goes beyond the conventional digital divide. Papaioannou conceptualizes the AI divide through access, use, and outcomes, highlighting that

unequal participation depends not only on connectivity or devices but also on power asymmetries, opaque governance, and market concentration [17]. From a critical theory perspective, Salazar-Aguilar further argues that ethical frameworks applied to Web3 and the metaverse are insufficient when they abstract away the persistent structural inequalities of connectivity, technological capacity, and participation, particularly in the Global South [18]. In the metaverse, these dimensions can accumulate. Participation can depend on hardware, bandwidth, physical accessibility, technological literacy, language, computational resources, and the ability to understand automated decisions. Consequently, users may inhabit nominally shared environments while possessing unequal opportunities to appear, act, create, and influence.

The datafication of the metaverse can deepen these inequalities by transforming increasingly intimate aspects of activity into actionable information. Hesselbein et al. argue that metaverse technologies extend the datafication of social life and can exacerbate existing and emerging inequalities [19]. Immersive systems can capture movement, gaze, posture, voice, spatial behavior, affective responses, and interaction histories. This data can be used to predict behavior, personalize environments, classify users, assign visibility, and determine access. The subject becomes progressively legible to the system, while the system's classifications remain only partially legible to the subject.

This asymmetry should not be reduced to isolated algorithmic bias. Zajko argues that algorithmic systems often reproduce social hierarchies and feedback loops, requiring analysis through the sociology of inequality rather than through technical correction alone [20]. Wu's concept of algorithmic status inequality similarly identifies enduring disparities in social position, influence, and resource access reinforced by AI systems [21]. This middle-range concept is useful for the metaverse because status may be produced through platform verification, reputational metrics, data visibility, computational enhancement, and proximity to those who control the infrastructure.

Research on disability representation illustrates both the emancipatory potential and the limits of encoded inclusion. Angerbauer et al. found that inclusive avatars can support authentic self-representation for people with disabilities, while inclusive and exclusionary interactions shape emotional responses, engagement, avatar connection, and perceived inclusion [22]. Representational options therefore matter, but adding categories to an interface does not itself produce an inclusive social order. The meaning of an avatar depends on the relations, norms, and governance structures surrounding its use.

The resulting hierarchy may be heuristically described as caste-like, not as a claim of historical or direct sociological equivalence, but as a way of identifying status differences that become persistent, difficult to contest, and structurally reproduced. Harari shows in Sapiens that caste systems and other social hierarchies are not necessarily based on natural differences, but can be consolidated through socially constructed norms, institutions, and orders [23]. This premise takes on a prospective dimension in Homo Deus, where he warns that the combination of artificial intelligence and biotechnology could transform economic inequalities into biological differences and divide humanity between technologically augmented groups and sectors progressively deprived of economic or political relevance [24]. In Nexus, this possibility shifts toward the control of data and information networks: the concentration of infrastructure and algorithmic power can generate new forms of dependency between those who control the systems and those who are subjected to them [25]. From this perspective, the notion of a caste-like hierarchy functions here as a warning about statuses that can become persistent, difficult to challenge, and structurally reproduced: those who design the classifications versus those who are classified by them; those who can

modify systems versus those who can only adapt; augmented and non-augmented human beings; users with advanced AI literacy and users dependent on automated mediation; subjects recognized as valuable and those turned into anomalies or made invisible.

Therefore, predesigned difference connects the design of identities with social stratification. When recognition depends on machine readability, that which cannot be encoded can be treated as noise, error, risk, or nonexistence. The future governance of the metaverse must protect not only access and representation, but also the right to question algorithmic status assignments and to participate in redefining the infrastructures through which difference is socially recognized.

Figure 1 synthesizes these processes as a critical-relational model in which the humanization of machines and the machinization of humans converge within the administered infrastructures of the metaverse, while consciousness, moral judgment, and responsibility constitute the principal limits to full functional substitution.

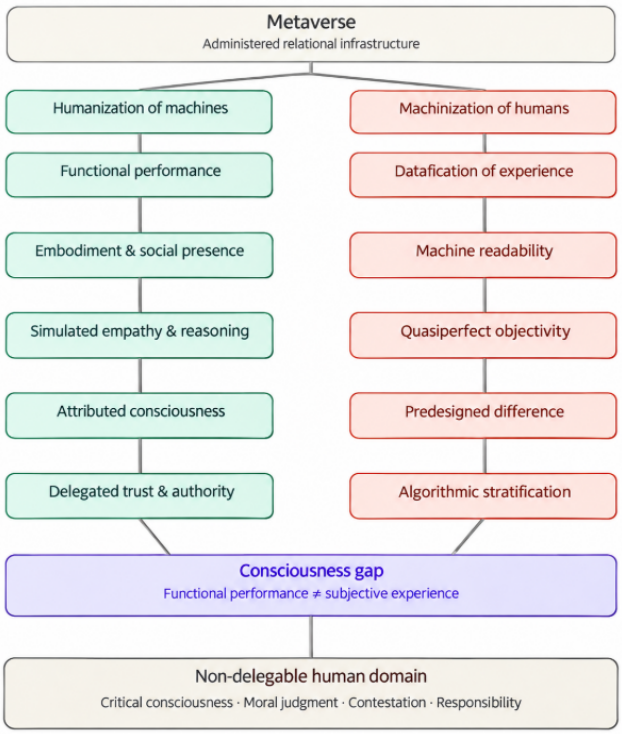


Figure 1. Critical-relational model of substitution without consciousness in the metaverse. The two trajectories should be read as mutually reinforcing rather than as linear causal sequences: increasing machine sociality can intensify human adaptation to machine-readable environments, while greater human datafication enables increasingly personalized and socially persuasive artificial agents.

Source: Author's elaboration.

6. Non-delegable consciousness and responsibility

The prospect of increasingly capable artificial systems does not require denying their usefulness or presuming that technological development inevitably leads to domination. The normative question is whether societies preserve the distinction between delegated performance and responsibility. Meaningful human control cannot consist of nominally placing an operator at the end of an automated process. People must retain the practical capacity to understand, question, suspend, and justify decisions, particularly when these affect rights, recognition, access, or social status [8,9].

Critical consciousness is therefore not another function to be optimized. It includes the ability to question ends that technological systems treat as given, recognize suffering that metrics inadequately represent, and refuse

commands whose procedural consistency conceals injustice [3,4]. Machines may assist deliberation by identifying patterns, testing alternatives, or exposing inconsistencies. They cannot inherit moral responsibility merely because their recommendations appear accurate, personalized, or persuasive.

The principle of non-delegability proposed here does not exclude AI from morally consequential decision-making. Rather, it means that ultimate normative authority and accountability cannot be transferred to an artificial system. Computational systems may support deliberation, identify alternatives, or inform decisions, but identifiable human actors and institutions must remain capable of justifying, contesting, and assuming responsibility for the resulting course of action. This principle also requires contestable classifications, transparent avenues of appeal, and the participation of those whose identities and opportunities are shaped through immersive infrastructures.

Machines may become functionally substitutive without becoming experientially or morally equivalent. The decisive task is not to protect every human activity from automation, but to preserve the conditions under which people remain subjects rather than objects of technological administration.

Acknowledgments: In the preparation of this work, the author used OpenAI ChatGPT as a writing-support tool for language refinement, translation, structural organization, and drafting assistance. The author reviewed and edited the AI-assisted content as needed and takes full responsibility for the final content of the publication.

Author contributions: Conceptualization, J.P.S.-A.; investigation, J.P.S.-A.; formal analysis, J.P.S.-A.; writing—original draft preparation, J.P.S.-A.; writing—review and editing, J.P.S.-A.; visualization, J.P.S.-A. The author has read and agreed to the published version of the manuscript.

Conflicts of interest: The author declares no conflicts of interest.

References

[1] Kenig N, Muntaner Vives A. The role of humans in the future of medicine: Completing the cycle. Metaverse. 2025, 6(1): 3129. https://doi.org/10.54517/m3129

[2] Butlin P, Long R, Bayne T, et al. Identifying indicators of consciousness in AI systems. Trends in Cognitive Sciences. 2026, 30(6): 488–501. https://doi.org/10.1016/j.tics.2025.10.011

[3] Horkheimer M, Adorno TW. Dialectic of Enlightenment: Philosophical Fragments. Stanford University Press; 2002.

[4] Waelen R. Why AI ethics is a critical theory. Philosophy & Technology. 2022, 35(1): 9. https://doi.org/10.1007/s13347-022-00507-5

[5] Salatino A, Prével A, Caspar EA, et al. Influence of AI behavior on human moral decisions, agency, and responsibility. Scientific Reports. 2025, 15(1): 12329. https://doi.org/10.1038/s41598-025-95587-6

[6] Dung L. Understanding artificial agency. The Philosophical Quarterly. 2025, 75(2): 450–472. https://doi.org/10.1093/pq/pqae010

[7] Zafar M. Normativity and AI moral agency. AI and Ethics. 2025, 5: 2605–2622. https://doi.org/10.1007/s43681-024-00566-8

[8] Robbins S. What machines shouldn't do. AI & Society. 2025, 40: 4093–4104. https://doi.org/10.1007/s00146-024-02169-7

[9] Krijger J. Enter the metrics: Critical theory and organizational operationalization of AI ethics. AI & Society. 2022, 37: 1427–1437. https://doi.org/10.1007/s00146-021-01256-3

[10] De Ruvo G. Algorithmic objectivity as ideology: Toward a critical ethics of digital capitalism. Topoi. 2025, 44(1): 175–186. https://doi.org/10.1007/s11245-024-10117-9

[11] Alon-Barkat S, Busuioc M. Human–AI interactions in public sector decision making: "Automation bias"and "selective adherence"to algorithmic advice. Journal of Public Administration Research and Theory. 2023, 33(1): 153–169. https://doi.org/10.1093/jopart/muac007

[12] Reeves C, Sinnicks M. Totally administered heteronomy: Adorno on work, leisure, and politics in the age of digital capitalism. Journal of Business Ethics. 2024, 193: 285–301. https://doi.org/10.1007/s10551-023-05570-2

[13] Hennig-Thurau T, Aliman DN, Herting AM, et al. Social interactions in the metaverse: Framework, initial evidence, and research roadmap. Journal of the Academy of Marketing Science. 2023, 51(4): 889–913. https://doi.org/10.1007/s11747-022-00908-0

[14] Son G, Rubo M. Social virtual reality elicits natural interaction behavior with self-similar and generic avatars. International Journal of Human–Computer Studies. 2025, 199: 103488. https://doi.org/10.1016/j.ijhcs.2025.103488

[15] Lim S, Schmälzle R, Bente G. Artificial social influence via human-embodied AI agent interaction in immersive virtual reality: Effects of similarity-matching during health conversations. Computers in Human Behavior: Artificial Humans. 2025, 5: 100172. https://doi.org/10.1016/j.chbah.2025.100172

[16] Huynh MT, Aichner T. In generative artificial intelligence we trust: Unpacking determinants and outcomes for cognitive trust. AI & Society. 2025, 40: 5849–5869. https://doi.org/10.1007/s00146-025-02378-8

[17] Papaioannou C. Dimensions of the AI divide: Digital inequality and psychological consequences. AI & Innovation. Published online 6 May 2026: e70002. https://doi.org/10.1002/aiv2.70002

[18] Salazar Aguilar JP. Can a universal digital ethics exist in a structurally unequal world? A critical theory perspective on Web3, metaverse, and the global south. Frontiers in Blockchain. 2026, 9: 1825419. https://doi.org/10.3389/fbloc.2026.1825419

[19] Hesselbein C, Bory P, Canali S. Six provocations for metaverse datafication: An emergent cultural, technological, and scholarly phenomenon. Information, Communication & Society. 2025, 28(5): 778–796. https://doi.org/10.1080/1369118X.2024.2433548

[20] Zajko M. Artificial intelligence, algorithms, and social inequality: Sociological contributions to contemporary debates. Sociology Compass. 2022, 16(3): e12962. https://doi.org/10.1111/soc4.12962

[21] Wu J. Algorithmic status inequality: An integrative perspective on AI-driven social stratification. Journal of Management Studies. Published online 13 July 2026. https://doi.org/10.1111/joms.70134

[22] Angerbauer K, Van Wagoner HP, Keplinger K, et al. Inclusive avatars in the Metaverse: Learning from the lived experiences of people with disabilities. Journal of Strategic Information Systems. 2025, 34(4): 101935. https://doi.org/10.1016/j.jsis.2025.101935

[23] Harari YN. Sapiens: A Brief History of Humankind. Harper; 2015.

[24] Harari YN. Homo Deus: A Brief History of Tomorrow. Harper; 2017.

[25] Harari YN. Nexus: A Brief History of Information Networks from the Stone Age to AI. Random House; 2024.