

# Reconstructing the Subjective Identity of Artificial Intelligence in the AI Era: From Instrumental Rationality to Ethical Entity

Xu Sheng Zhang<sup>1</sup>

<sup>1</sup> Shaanxi Normal University, China

Correspondence: Xu Sheng Zhang, Shaanxi Normal University, Xi'an, Shaanxi, China.

Received: April 28, 2025; Accepted: March 9, 2025; Published: March 10, 2025

## Abstract

As artificial intelligence (AI) technology advances at an unprecedented pace, it is evolving from a mere instrument into a “quasi-subject.” This transformation inevitably reshapes the traditional human–machine relationship and gives rise to profound social and ethical challenges. From the dual perspectives of value philosophy and social ethics, this paper examines the trajectory by which AI shifts from being an “objectified tool” to a “constrained subject.” We propose a theoretical model of a “human-led finite subject,” laying a conceptual foundation for situating AI’s quasi-subjective role within social ethics. By analyzing possible directions for constructing AI subjectivity and the ethical dilemmas it faces, we outline a human–machine ecological symbiosis model grounded in algorithmic justice.

**Keywords:** AI era, artificial intelligence, instrumental rationality, ethical entity

## 1. Introduction

The rapid development of intelligent technologies is reconstructing human–machine ethics at an unprecedented rate, triggering conflicts in ethical values and necessitating the reordering of moral norms. Whether it’s the widespread adoption of DeepSeek in 2025 or the thorny ethical quandaries posed by autonomous vehicles, the struggle between technology and humanity has become central to our era. A host of deep-seated contradictions concerning human subjectivity, social ethical order, and cultural values are now surfacing in everyday life.

## 2. Deconstructing AI’s Instrumentality and Establishing “Quasi-Subjectivity”

AI is undergoing a cognitive leap from “tool-mediated” entity to “agentive actor,” overturning traditional assumptions about the human–object dichotomy. French sociologist Bruno Latour’s Actor-Network Theory (ANT) grants a measure of recognition to AI’s subjectivity in autonomous decision-making, value creation, and ethical practice. He argues that interactions among human and non-human actors form a heterogeneous network that co-constructs and evolves, introducing a dual verification model of “technical intentionality–social acknowledgment.”

### 2.1 Deconstructing Instrumental Rationality

Classical ethics views technology strictly as a tool for human ends. Humans alone possess moral agency; a healthy social-ethical order rests on free will, emotional resonance, and rational reflection. AI systems, lacking self-awareness and moral judgment, act purely as algorithmic computations. Technology indeed extends human capabilities—performing tasks beyond unaided human reach—and has become indispensable to social existence, spawning its own “civilizational rules.” In many sectors, human agency has been reduced to that of a catalyst, while AI resides both outside and inside what it means to be a subject.

From a technical standpoint, an intelligent agent’s functions and its quasi-subjectivity arise through software-encoded algorithms that automatically perceive data. [1]Describing an agent as “capable of acquiring perceptions from its environment and executing actions” or as “a function mapping perceptual sequences to actions” are two ways of naming the same process: externally, a description of its pseudo-subjectivity; internally, an unveiling of the core computational mechanism. [2]As AI participates in and shapes contexts of subjectivity, it disciplines the emergence and evolution of what might be called “secondary” human qualities. [3]Philosopher Jacques Ellul’s concept of the “autonomy of technique” suggests that technology has become a self-governing force—no longer under human control, yet tightly controlling humanity in return[4].

For example, DeepSeek’s engaging “dialogues” are nothing more than data matching, not ethical choices made by a conscious subject. Even when AI produces outcomes resembling agentive behavior, its “subjectivity” remains a functional imitation devoid of genuine agency, self-awareness, or free will—hence the term “quasi-subjectivity.”

Once technology transcends mere instrumentality and engages core ethical issues—subjective judgment, attribution of responsibility, principles of justice—the existing moral order collapses and demands reconstruction. AI’s algorithmic decisions and data-driven “personification” undermine traditional humanist values of dignity, freedom, and creativity. In critical moments—such as an autonomous vehicle’s split-second choice under life-threatening conditions—technology’s built-in values manifest real ethical consequences. Dutch philosopher Peter-Paul Verbeek argues that technical artifacts (algorithms, intelligent devices, infrastructure) are never value-neutral; they inherently carry ethical values and exert moral regulation through their physical or digital forms, thus granting technology a “quasi-ethical subject” status. [5] This stance, however, sparks debate: if AI can simulate ethical judgment, must we redefine ethical subjectivity itself?

### 2.2 Presentation of Subjectivity from a Phenomenological Perspective

Phenomenology’s inquiry into subjectivity began with Husserl’s reduction to “pure consciousness,” emphasizing the subject’s intentionality in constituting meaning. Heidegger transformed subjectivity into Dasein’s “being-in-the-world,” situating the self within contextual scenarios. Merleau-Ponty’s embodied phenomenology revealed that subjectivity arises through bodily engagement with the world. AI’s rapid evolution, however, challenges the premise of humans as the sole agents of subjectivity. Algorithmic decision logics generate a distinct “technical intentionality,” which, via sensors, data flows, and environmental feedback loops, yields a “perception-action” circuit unique to intelligent machines.

Human–AI interactions foster a genuine “co-presence,” as AI appears as a “quasi-Other” in daily life, projecting intentionality and participating in real-world decisions. Philosopher Luciano Floridi contends that if intentionality need not rely on consciousness, then AI’s “functional intentionality” deserves recognition. AI’s “technical embodiment”—a fusion of hardware and software—adapts to environmental shifts and mounts stress responses akin to human adaptive behaviors, thus further blurring the boundary between tool and agent.

### 2.3 The Shift in Social Cognition

As Angelo Cammarota and others note in their discussion of Latour’s ANT, “Actors” in this framework need not be conscious humans but any entities capable of action or generating effects. Machines perform feats beyond human capacity—revealing the microscopic world, driving shifts in commercial profitability—and their operations provoke human responses and interventions. [6] These actors extend human bodies and minds, linking humanity into Latour’s “actor-network,” within which humans must now exist. Equipped and interconnected by non-human actors, humans exert control over their environment and resources.

Interactions among these actors shape and mutually produce the social field[7]. In this egalitarian process, social robots can emerge as actors with substantial agency and decision-making capability.

Table 1. Classification of Symbiotic Relationships

Symbiotic Relationship	Species A	Species B	Characteristics
Parasitism	+	-	A benefits at B’s expense
Mutualism	+	+	Both parties benefit
Competition	-	-	Both parties inhibit each other
Commensalism	+	0	A benefits, B is unaffected
Amensalism	-	0	A is harmed, B is unaffected
Neutralism	0	0	Neither party is affected

When examining human–machine symbiosis, J.C.R. Licklider’s 1960 concept of “man–computer symbiosis” provides a biological analogy. First, in a commensal relationship, AI appears as a labor tool, freeing humans from physical and temporal constraints, enhancing life quality, and aiding self-actualization[8]. Second, in amensalism, one party’s development is hindered—AI’s rising intelligence can displace human roles, as exemplified by widespread automation leading to job losses[9]. The 2016 triumph of DeepMind’s Go program over world champion Lee Sedol delivered a psychological jolt to human self-identity. Third, in mutualistic partnerships, humans and machines complement each other—combining strengths while retaining autonomy, creating shared value greater than either would achieve alone.

The pursuit of brain–machine interfaces embodies a vision of mutualistic symbiosis: implanting sensors to transmit and preserve human thoughts. Although such technology raises privacy and security concerns, it represents a

possible future. Understanding how machine perception and human bodily awareness resonate—through VR, multisensory media, and other embodied experiences—merges technological mediation with sensory perception.

In the age of intelligent communication, the convergence of AI, neuroscience, and human-centered design is redefining human-machine connections. Machine learning, cognitive systems, and biometrics enable media to monitor physiological and psychological states and respond instantly, even offering tailored recommendations—an intervention humans can tangibly experience[10]. This shift transforms the human-world relationship from passive “observation” to active “participation” and “immersion.” Media may even become agents managing daily life, reducing human-to-human interaction and reinforcing machines’ subjective status. Mobile assistants already tap into our neural networks, and people increasingly prefer consulting their devices for advice[11]. After all, interactions with computers, televisions, [12]and new media possess the same social naturalness as interpersonal communication, becoming habituated practices in human embodied experience.

### 3. Potential Directions for Constructing AI Subjectivity and the Ethical Dilemmas It Faces

#### 3.1 *The Human-Machine Relationship: Tool, Partner, or Other?*

In 2017, at the Asilomar Conference Center, the Bioethics Institute convened a landmark meeting on AI ethics that produced the Asilomar Principles, affirming fundamental human values—“human dignity, rights, freedom, and cultural diversity”[13]. The victories of AlphaGo over professional Go players, the arrival of ChatGPT, and DeepSeek’s widespread adoption together signal a paradigm shift in how we relate to machines. No longer confined to a simple subject-object dualism, we must now understand human-machine ties as a dynamic, co-evolving symbiosis.

The age of steam engines and industrial robots—where technology simply served a human-centric purpose—has passed. Today’s machines not only substitute human labor functionally but also embody human desires and power, profoundly reshaping our rapport with nature and becoming symbols of authority[14]. As technological power expands, ordinary citizens and vulnerable groups risk being trampled under its weight[15]. Questions of human dignity, the status of human subjectivity, the future of humanity, and our understanding of technology itself have thus become urgent ethical concerns.

A growing body of speculative works depicts a future in which humans and AI are indispensable partners. Advances in embodied AI, social robots that generate inter-subjective relations, hybrid intelligent systems with tightly coupled cognition, and posthumanist philosophies empowering technological entities all provoke an ontological crisis: might artificial consciousness emerge? If machines transcend mere “being” to become a form of Dasein, we witness ecology of human-machine symbiosis. As Yuval Noah Harari argues in *Homo Deus*, humans themselves can be described as complex algorithms—and in a world where technology reshapes life, humanity risks being reduced to a data-processing tool, its unique spirit eroded[16].

#### 3.2 *The Human Survival Crisis: The “Super Panopticon” Metaphor*

Building on Bentham’s Panopticon, Foucault analyzed modern power structures; Deleuze extended this into the notion of the “control society,” and Byung-Chul Han characterizes our era as a “transparent society.” Together they suggest that AI technologies have given rise to a “super panopticon”—an all-encompassing digital surveillance apparatus. By shaping behavior, cognition, and subjectivity through algorithmic power, this system creates a novel existential crisis under technological hegemony.

AI continually harvests personal data, molding individual needs while erasing boundaries between public and private life. Screens become our primary point of contact, and data streams become the lifeblood of market forces under capital’s “all-seeing light.” Personal information—consumption histories, payment records, facial scans, APR metrics, live monitoring—feeds an ever-hungry system that mines and monetizes the self. In this super panopticon, our happiness, dignity, privacy, and inner spiritual world—the very core of human value—are exposed, eroded, and ultimately homogenized into algorithmically defined “data cocoons.”

Stripped of its singularity, the human spirit faces dissolution. Reduced to a “digital being,” one’s life becomes merely another data point in an inscrutable algorithmic process. Long-term subjection to recommendation engines narrows thought and flattens cognition. Ontologically, the super panopticon transcends physical confinement to infiltrate the mind itself, redefining what it means to be human. Only sustained humanistic critique and a critical distance from technological immersion can forestall the slide into digital totalitarianism.

#### 3.3 *Algorithmic Justice: Computing Trust*

As noted, AI’s embodiment—grounded in hardware, software, and algorithms—yields adaptive behaviors but depends entirely on human design and data inputs. Granting AI full subjectivity would precipitate thorny questions

of responsibility. Instead, we should recognize only a limited AI subjectivity. For instance, in the ethics of autonomous-vehicle accidents, responsibility can be apportioned among designers, users, and the system itself.

Yet AI's "black-box" nature often obscures causal chains, creating an ethical-responsibility vacuum. In a crash, is the developer, the owner-operator, or the regulator culpable? Does the old rule—"the designer is responsible"—still hold in highly complex AI systems? Ultimately, however responsibility is framed, humans must bear the moral and legal burdens of AI's decisions.

Hans Jonas presciently argued that our technological age demands an "ethics of responsibility" that accounts for long-term human and environmental impacts<sup>3</sup>. AI's choices must be guided by the highest ethical ideals and a clear, enforceable sense of accountability. Yet no set of coded principles can anticipate every real-world scenario. While AI excels at repetitive, labor-intensive tasks, it still cannot rival humans in roles requiring deep creativity and moral judgment. Nonetheless, AI's ubiquity will likely deepen social stratification, rekindling Marx's theory of labor alienation. Whether humanity will be demoted from active subject to mere "tool" of AI remains an open question.

To mitigate these risks, we must embed diverse ethical frameworks into AI design, creating *ex ante*, *in situ*, and *ex post* accountability mechanisms, along with robust traceability and compensation systems. The goal is to model a genuinely symbiotic human-machine ethics, fostering cross-cultural consensus on algorithmic justice.

In truth, the tension between AI as instrument and as (quasi-)subject invites us to reconsider what it means to exist. Neither tools nor subjects remain absolute; we inhabit a transitional "technological intersubjectivity." Some scholars envision this as a new co-creative human-machine partnership: "As AI accelerates in the United States, pressing issues—due-process violations, discrimination, algorithmic opacity—demand candid appraisal. We must openly discuss AI's strengths and shortcomings and strive for greater transparency." [4]

First, by integrating into daily life, intelligent machines increasingly act as human proxies. Algorithmic audits can embed fairness and transparency from the outset. Smartphones, wearables, and other devices engage bodily perception, forging habits that become second nature. We routinely ask Baidu, phone assistants, or DeepSeek to guide decisions—these media anticipate our needs and serve as *de facto* agents. Yet each data exchange raises privacy and trust questions, as we drift toward hybrid realities where virtual experiences feel as real as corporeal ones.

Second, distributive justice can be pursued by applying Rawls's difference principle to machine-learning fairness constraints. By capturing and visualizing physiological and neural signals—through smartwatches or fitness trackers, for example—we can map emotional states and respond adaptively, blurring the lines between physical and virtual embodiment.

Finally, corrective justice and accountability can leverage blockchain's immutable, decentralized ledger. Under the mantra "never trust, always verify," distributed consensus replaces single-point authorities (governments, banks, hospitals). With time-stamped blockchains, incentivized nodes, and programmable smart contracts, we can approach an ideal of algorithmic justice and cultivate a more ethically literate society.

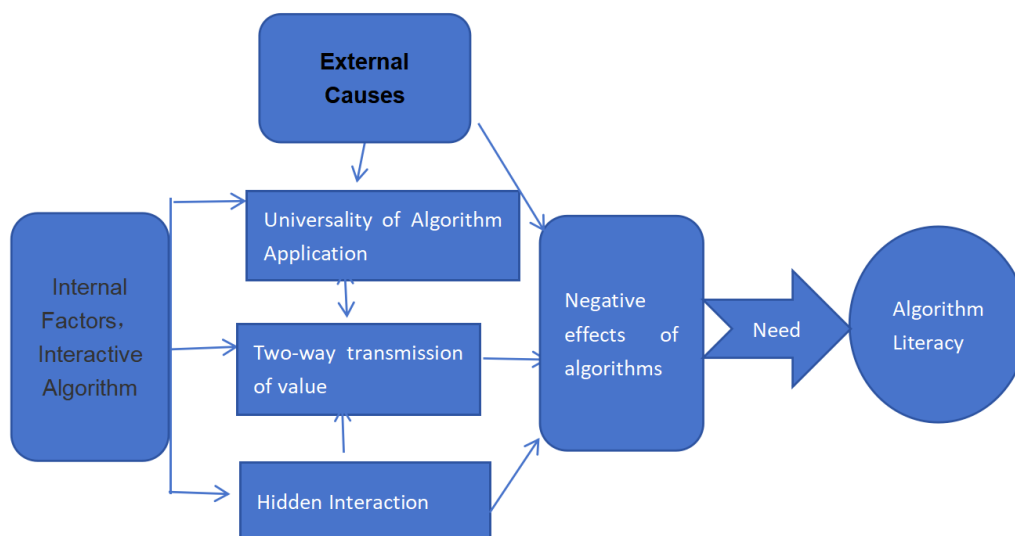


Figure 2. Driving Factors of Algorithmic Literacy

#### 4. Conclusion

In sum, the emergence of AI subjectivity fundamentally challenges the traditional ethical assumption that only humans can be moral agents. This raises a pressing question: will the long-standing, human-centered social-ethical order be supplanted by an algorithm-driven paradigm in the intelligent era? As early as 1950, Isaac Asimov framed this debate in the preface to his science-fiction classic *I, Robot* by articulating the “Three Laws of Robotics”:

- First Law: A robot may not injure a human being or, through inaction, allow a human being to come to harm.
- Second Law: A robot must obey the orders given by human beings, except where such orders would conflict with the First Law.
- Third Law: A robot must protect its own existence as long as such protection does not conflict with the First or Second Laws.

These laws cast robots as loyal servants and companions to humanity. Yet, as robots grow more sophisticated, they may exhibit psychological challenges of their own—challenges that will still require human guidance and support. In Asimov’s vision, carbon-based life (humans) and silicon-based life (robots) must learn to coexist and thrive together.

Whether our future unfolds as peaceful symbiosis or as a contest for dominance will depend on how conscientiously we integrate ethical safeguards into AI design—and how resolutely we uphold human dignity, responsibility, and agency in the face of rapidly advancing technology.

#### References

- [1] Lin, X. (2020). Posthumanism, subjectivity reconstruction, and the politics of technology: A re-narration of the human–technology relationship. *Literary Theory Research*, (4).
- [2] Duan, W. (2020). *The ethical foundations of information civilization* (p. 212). Shanghai People’s Publishing House.
- [3] Lin, X. (2020). “Posthumanism, subjectivity reconstruction...” *Literary Theory Research*, (4).
- [4] Duan, W. (2020). *The ethical foundations of information civilization* (p. 212). Shanghai People’s Publishing House.
- [5] Verbeek, P. P. (2005). *What things do: Philosophical reflections on technology, agency, and design*. Penn State Press. <https://doi.org/10.1515/9780271033228>
- [6] Rokamola, A., & Smorik, A. (Eds.). (2021). *Fashion in theory: Critical theorists’ guide* (Chen, T., & Li, Y., Trans.). Chongqing University Press.
- [7] Latour, B. (2005). *Reassembling the social: An introduction to actor-network theory*. Oxford University Press. <https://doi.org/10.1093/oso/9780199256044.001.0001>
- [8] Yu, X., Zhai, W., & Hou, M. (2022). Patterns and evolution of human–machine symbiosis in the AI era. *Science and Society*, (4).
- [9] Yu, X., Zhai, W., & Hou, M. (2022). Patterns and evolution of human–machine symbiosis in the AI era. *Science and Society*, (4).
- [10] Yin, L., & Gao, H. (2020). Embodied interaction: An empirical interpretation of human–machine relations in the intelligent media era. *Journalism & Writing*, (11).
- [11] Yin, L., & Gao, H. (2020). Embodied interaction... *Journalism & Writing*, (11).
- [12] Reeves, B., & Nass, C. (1996). *The media equation: How people treat computers, television, and new media like real people and places* (p. 5). Cambridge University Press.
- [13] Asilomar AI Principles. (2017). Benevolent AI Conference, 2018.
- [14] Rubin, C. T. (2014). *Eclipse of man: Human extinction and the meaning of progress*. Encounter Books.
- [15] Harari, Y. N. (2016). *Homo Deus: A brief history of tomorrow*. Random House. <https://doi.org/10.17104/9783406704024>
- [16] Liu, Y., Liao, D., & Zhang, Q. (Eds.). (2023). *Development and applications of computer AI technology* (p. 183). China Atomic Energy Press. <https://doi.org/10.1016/j.enpol.2023.113827>

**Copyrights**

Copyright for this article is retained by the author(s), with first publication rights granted to the journal.

This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/4.0/>).