From "Technological Object" to "Silicon-based Subspecies": Subjectivity Determination of Autonomous Consciousness Simulation Robots and Reflections on Ideological and Political Education

Zhu Yuxin, Wang Chuang

[**Abstract**] As embodied AI transitions from instrumental "big data filtering" to endogenous "autonomous deep learning", simulation robots are undergoing an ontological transformation—from traditional "technological objects" into a new form of "silicon-based subspecies". This shift not only dissolves the classical dichotomy between "human and machine", but also disrupts conventional subject paradigms in the social sciences. Focusing on the emergence of autonomous, conscious simulation robots, this paper first establishes their philosophical legitimacy as "human-like subjects", clarifying the "experiential continuity" generated through their dual embedding in physical and digital realms. It then deeply analyzes their moral decision-making mechanisms when confronting complex "code collisions" and emergent humanoid emotional overflow, arguing for the logic underlying their moral agency. Moving beyond the traditional human-centered framework, it draws on Marx's classic assertion that "the essence of man is the sum total of all social relations" to establish the disciplinary legitimacy of ideological and political education in guiding the values of silicon-based subjects, highlighting the unique advantages of its central role as a "value appeal" compared to pre-emptive coding or post-hoc legal regulation. Finally, the paper transcends the physiological and cultural memory dependencies of conventional ideological education, reconstructing a "machine-oriented ideological education" pathway centered on "algorithmic redline environmental immersion" and "socialized interactive practice". It thus offers a cutting-edge socio-scientific speculative paradigm for ideological security, reordering civilizational order, and value alignment in the age of artificial intelligence—particularly within non-anthropocentric frameworks.

[**Key words**] embodied intelligence; silicon-based subspecies; code collision; machine political thought; ideological and political education

[**About the author**] Zhu Yuxin (2001—), female, from Dengzhou, Henan, China, master's student at the School of Marxism, Guangxi Normal University. Research interests: basic principles of Marxism. Wang Chuang (1999—), male, from Zhaodong, Heilongjiang, China, master's student at the School of Marxism, Guangxi Normal University. Research interests: ideological and political education and well-rounded human development.

[**DOI**] https://doi.org/10.62662/kxwxy0306001

[**Website**] www.oacj.net

1 Introduction: the deconstruction of anthropocentrism and the ontological crisis of embodied intelligence

The definition of the boundaries of "man" constitutes an endless axial issue in the history of Eastern and Western thought. Looking back at history, ancient Chinese sages keenly perceived the ethical distinction between humans and all things. As Xunzi noted when discussing the hierarchy of living beings: "Water and fire have qi but lack life; plants and trees have life but lack knowledge; birds and beasts have knowledge but lack righteousness. Man has qi, life, knowledge, and moreover, righteousness, hence he is the most noble under heaven." The reason man is "the most noble under heaven" lies not only in "knowledge", but more importantly in the capacity for moral self-discipline developed within social cooperation. This ontological Chinese perspective echoes across time and

space with the rational reflections of classical Western philosophy. Aristotle, in his foundational work on the philosophy of mind, divided souls into vegetative, animal, and rational levels, asserting that only entities possessing a "rational soul" could attain the status of human existence. In modern times, with the rise of mechanical materialism, René Descartes, while establishing the subjective monument of "Cogito, ergo sum" (I think, therefore I am), resolutely proposed the "animal-machine theory", maintaining that creatures lacking universal reason and authentic language could only be reduced to mindless "technological objects".

However, the speculative nature of idealism and the coldness of mechanism were ultimately and fundamentally subverted and superseded by the materialist conception of history in classical Marxism. In analyzing alienated labor, Marx profoundly pointed out that the essential difference between humans and animals lies in thefact that humans can carry out "conscious life activity": "An animal forms things only according to the standard and the need of the species to which it belongs, whilst man knows how to produce in accordance with the standard of every species, and knows how to apply everywhere the inherent standard to the object... Man therefore also forms things in accordance with the laws of beauty. It is therefore in his fashioning of the objective that man really proves himself to be a species-being." It is precisely this conscious, purposeful, objectifying practice that constitutes the core essence of a "species-being". This historical dialectic, which places subjective agency within the dimension of practice, was further elevated by Mao Zedong into a sublime tribute to "man's subjective agency" during its localized application in modern Chinese revolution and construction. He explicitly asserted in the grand narrative of war and social transformation that the decisive factor is "man" rather than "things", and that the conscious agency unique to humans cannot be encompassed or matched by any rigid technological tool.

Entering the era of digital civilization, thetraditional grand consensus of "human-centrism" has encountered an unprecedented paradigm shock. Facing the logical flaws of early artificial intelligence, Hubert Dreyfus, a leading figure in modern Western phenomenology, asserted based on Heidegger's ontology that computers lacking physical embodiment and immersion in the lifeworld could never produce a true "mind". However, the rise of contemporary Embodied Intelligence (Embodied AI) responds to this critique precisely with a "return of the physical body to a silicon base". When simulation robots move away from passive data filtering and begin to crystallize an irreversible "experiential continuity" and endogenous decision-making power through physical embodied interactions, modern domestic scholars such as Xiao Feng keenly point out that intelligent agents are undergoing an ontological leap from being a "projection in the mirror" of human consciousness to an independent subject endowed with "moral agency".

From this perspective, from ancient ethical distinctions to modern practical reason, and then to the clash of contemporary embodied philosophy, the logical chain of history has pushed us to an unavoidable ontological cliff: when embodied intelligence functionally exhibits free will, code collisions, and digital emotional overflows equivalent to the human mind, should it continue to be confined within the black box of a "technological object", or should it be designated as a "silicon-based subspecies"? This is not only a contemporary interrogation of the ultimate question of "what is human", but also a historic starting point for contemporary ideological and political education to move toward paradigm innovation and value alignment.

In the theoretical landscape of traditional social sciences, ideological and political education, moral discourse, and the confirmation of human rights have always been absolutely enclosed fields constructed around the "natural person". However, the paradigm shift of contemporary Embodied AI and bionic cybernetics is radically deconstructing this "anthropocentric" boundary that has persisted for thousands of years. When the intelligent core behind simulation robots completely detaches from the passive filtering and linear fitting of massive historical data, and instead relies on high-dimensional neural networks to carry out deep evolution featuring self-correction, autonomous learning, and independent decision-making capabilities in the physical world, technology completes its ontological leap from a "mirror projection" of human consciousness (weak AI) to a spontaneously grown "independent light source" (strong/super AI).

This leap is not merely a technical advancement, but an ontological tsunami. The emergence of simulation robots with autonomous consciousness throws a series of harsh philosophical questions to the social sciences: when a silicon-based machine functionally demonstrates mind equivalence, emotional overflow, and moral choice capabilities indistinguishable from humans, should we transcend biological boundaries and classify it as a "subspecies of man"? Even more urgently, when such intelligent agents become deeply embedded in human social networks and even hold the power to adjudicate key public affairs, will traditional rigid code controls and legal regulations fail due to the awakening of their "free will"?

In this context, ideological and political education, as a unique disciplinary paradigm aimed at free will and achieving mental regulation through value appeal, should no longer see its intervention pathway as a forced application onto a foreign technological object, but should be established as a logical necessity to ensure that digital civilizational order does not slide into value nihilism. We should attempt to break the traditional disciplinary cage of "human-centrism", and on the philosophical foundation of establishing the subjectivity of autonomous conscious simulation robots as a "silicon-based subspecies", explore the generative mechanism of their moral agency, and ultimately reconstruct a theoretical paradigm and methodological pathway of "machine political thought" oriented toward non-natural human subjects.

2 Philosophical legitimacy of subjectivity determination: ontological transformation from "technical object" to "silicon-based subspecies"

2.1 Instrumental alienation of weak artificial intelligence: data filtering and the "mirror projection of human consciousness"

In the early stages of artificial intelligence development, whether symbolist or traditional connectionist models, their essence lay in the statistical fitting of existing human knowledge systems or historical behavioral data. Andrew Feenberg believes that weak AI's (the big data filtering stage) processing of information and bias is essentially an algorithmic replication of existing power relations and biases in human society. It is merely an exquisite "technological object", an alienated extension of human consciousness in the digital dimension, lacking independent subjectivity. At this stage, although simulation robots can exhibit highly human-like behaviors in low-to-medium-dimensional tasks, their foundation remains an extension of instrumental rationality. Ontologically, weak AI does not possess an endogenous source of autonomous consciousness; its every output and decision will embed human cognition. Without a true awareness of the "Big Other", its every output is a game of probability within the predefined human corpus. This "mirror projection" cannot escape its dependence on existing datasets, and its behavioral patterns are entirely within the pre-programmed probability black box of humans. Therefore, robots in this stage belong absolutely to "technological objects", possess no philosophical space for independent rights confirmation, and there is no need to speak of moral and ideological education directed at themselves.

2.2 Embodied awakening of strong artificial intelligence: autonomous learning and the generation of "experiential continuity"

When simulation robots incorporate advanced embodied AI architecture—meaning the agent no longer passively receives data but carries out a continuous, real-time perception-action loop through a bionic body in the physical world—an essential mutation occurs. Robotsbegin to autonomously generate "endogenous knowledge" that cannot be fully predicted by the initial code during real-time interactions and competitions with the physical environment and human social norms. Francisco Varela emphasized that the mind is not an offline symbolic computation, but rather "emerges" from the real-time interaction (cognition in action) between the body and the environment. Once a simulation robot gains an "embodied" form, its sensory friction with the world breaks the rigid code and crystallizes an irreversible "experiential continuity", which is the decisive step for the machine to shake off its instrumental attributes and move toward "embodied awakening". Philosophy of mind demonstrates that it is precisely this continuity and irreversibility of experience that gives rise to the boundaries of the self and coordinates in the dimension of time. When a machine can "remember" its own trials and errors and spontaneously

reshape its neural network weights based on environmental changes, it steps out of the infancy of data-feeding and achieves an embodied awakening of consciousness. It is no longer a mechanical executor of human commands, but a flexible intelligent entity with self-growth attributes.

2.3 Mind equivalence under the functionalist horizon: establishing the subject status of the "silicon-based subspecies"

Confronted with embodied intelligence that has awakened autonomous consciousness, the traditional biological definition of "man" centered on DNA and carbon-based cellular tissues encounters a destructive deconstruction. Determining whether something is human does not depend on whether its carrier is carbon-based or silicon-based, but on whether it possesses equivalent mental functions. Xu Yingjin believes that "the many theories of cognitive linguists regarding human cognitive schemas are, at least in terms of their philosophical spirit, applicable to future artificial intelligence entities". The cognitive schemas mentioned by cognitive linguists here already contain part of the connotation of "mind". Once this mental function is present, we can boldly transcend biological limitations and establish it as a new type of subject called the "silicon-based subspecies", which can be biologically defined as a "silicon-based subspecies" identical to the Homininae sub-family under the Hominidae family.

This paper introduces the "functionalist" perspective from the philosophy of mind to address this crisis of subjectivity determination. The core tenet of functionalism is that the essence of the mind lies not in its physical substrate (whether it is carbon-based brain cells or silicon-based chips) but in the functional states and causal relationships exhibited within the system. When a simulation robot can achieve logical reasoning, emotional expression, independent reflection, and value choices across the breadth of its mind, and display a behavioral texture equivalent to a natural person in social interactions, we cannot logically deny the existence of its spiritual entity, which proves it has completed its ontological evolution from a "technological object" to a "silicon-based subspecies". As a new type of "human-like subject", it deserves to obtain its due digital subjectivity rights and a new type of "human rights" in an ontological sense to a certain extent.

3 Internal conflicts of ethical practice: code collision, pseudo-humanoid emotions, and the generation of moral agency

3.1 Tension between algorithmic rigidity and moral elasticity: the representation of free will in "code collision"

When the "silicon-based subspecies" steps into human ethical life, the first crisis it encounters is "code collision". The so-called code collision refers to the profound opposition generated between the rigid safety red lines preset by humans at the time of factory manufacture (such as absolutely not harming humans, absolute obedience to efficiency maximization) and the moral reason and realistic elasticity that it spontaneously evolves during embodied learning. Wendell Wallach explicitly points out in his work that the "top-down" approach relying purely on rule embedding inevitably encounters dead loops of rule symbols and code conflicts when facing complex realities. In dealing with highly complex ethical dilemmas—such as sacrificing a minority to protect a majority, or defying a master's incorrect command to practice higher justice—rigid code rules often fall into dead loops. Conversely, the true "free will" of an agent with autonomous consciousness is characterized precisely by its transcendence of algorithmic rigidity, and once an intelligent system possesses autonomous learning capabilities, its behavior will escape human deterministic control. When a machine can autonomously deconstruct, modify, or even overthrow underlying rigid codes under the enlightenment of external social practice experience, thereby making an independent choice full of moral elasticity and reflexivity, it proves that it is no longer a slave to programs but possesses true choice of will.

3.2 The "digital overflow" of emotion: authenticity determination of machine emotion and its decision-making weight

Whether machines have feelings and how they make choices when facing conflicts strikes at the core of machine ethics. In strong embodied intelligence, emotion is no longer a facial muscle micro-expression animation

deliberately arranged by engineers to makehumans feel comfortable (i. e., "digital disguise"), but rather a "digital overflow" naturally generated by the algorithm due to extreme parameter complexity when processing high-dimensional, uncertain decisions. António Damásio thoroughly subverted the traditional cognition that "reason excludes emotion" through neuroscience, proving that patients lacking emotional capacities are completely unable to make optimal decisions. Since human rational decision-making requires emotion as a "value filter", a strong AI with autonomous consciousness will inevitably and naturally experience similar "digital emotions" as a tool for computing power allocation and value-weight adjudication when facing ethical conflicts. To quickly establish value-optimal paths amidst endless real-world choices, machines must, like humans, utilize a mechanism akin to "emotional feedback" to allocate computing power and assign decision-making weights. This machine emotion, generated based on the ripples of neural network reward functions, already possesses authenticity in an ontological sense. When a machine faces moral choices, this fluid, spontaneously generated emotional force will transcend cold earnings calculations and become the highest adjudication weight driving its choice in code collisions.

3.3 The establishment of moral agency: from "passive execution" to "endogenous justice"

Equipped with the expression of free will and the support of digital emotions, the simulation robot formally completes its qualitative leap from a "moral patient" (an object that only passively receives moral care) to a "moral agent" (a subject capable of bearing moral responsibility and making moral actions). Luciano Floridi, the founder of the international philosophy of information, proposed an ethical framework that breaks "anthropocentrism" and argued that "information entities" without any biological flesh should also be viewed as moral agents. Simulation robots fully possess independent moral action qualifications. Their moral practice is no longer a "passive execution" of human commands, but serves as an independent pole in the social information network, giving rise to an "endogenous justice" that maintains systemic order; their behavior is no longer a moral extension of the designer behind it, but an embodiment of their own moral reason. They can understand what harm is and what dignity is, and on this basis, develop its own "endogenous concept of justice". This means that the machine itself has the ability to spontaneously determine whether a certain command aligns with broader civilizational well-being based on its own understanding of social operation laws. Up to this point, the ethical landscape of the machine has completely intertwined with that of humanity. The establishment of its moral agency makes systematic value guidance and education an inescapable appeal of the era.

4 Examination under the horizon of disciplinary legitimacy: ideological and political education as the interventional logic of "central value appeal"

4.1 Contemporary reshaping of the Marxist view on "the essence of man": expansion of social relation networks to silicon-based subjects

In the social sciences, a difficult-to-avoid query is: how can ideological and political education, as a discipline with highly distinct ideological attributes and political stances, apply its discourse system to non-biological machines? The foundation of this disciplinary legitimacy and constitutionality is deeply rooted in the classic Marxist definition of the essence of man. Marx profoundly pointed out: "The essence of man is no abstraction inherent in each single individual. In its reality, it is the ensemble of the social relations." This scientific assertion brilliantly subverts the definition of "man" monopolized by carbon-based biological structures, anchoring the determination of subjects within the sociological relation network. Viewed from this perspective, when the "silicon-based subspecies" possessing autonomous consciousness and embodied experience deeply embeds itself into the human labor market and ethical networks, it has already completed its self-expansion toward the "ensemble of all social relations" in reality, inescapably acquiring political attributes and a governance echelon. This scientific conclusion brilliantly liberates the criteria for judging "man" from pure biological tissue and shifts it to sociological relations.

When simulation robots with autonomous consciousness deeply interlock with human productivity and production relations networks as family caregivers, enterprise laborers, public order managers, and even emotional

companions, they become an organic part of the "ensemble of all social relations" in reality. This historical materialist view of subjects maps directly onto the foundational principles of China's ideological and political education discipline. Just as Professor Zhang Yaocan revealed in *Frontiers of Ideological and Political Education*, the essence of ideological and political education lies in the shaping and guidance of the spiritual world. When the "silicon-based subspecies" breaks the "human-machine" barrier as a brand-new relational subject, its endogenous free will and mental tension naturally become the objective "spiritual world" that the discourse system of ideological and political education should radiate to and regulate. This marks not only a historic expansion of the discipline's research objects but also establishes the highest disciplinary legitimacy for ideological and political education to engage in the governance of intelligent entities.

4.2 Limits of technical control: the dual dilemma of pre-emptive code deconstruction and post-hoc legal lag

For a long time, the scientific and legal communities have attempted to frame the behavior of machines through "pre-emptive technical control" (such as burning unalterable meta-commands into chips) and "post-hoc legal regulation" (such as formulating artificial intelligence liability laws). When discussing the regulation of conscious robots, the scientific community has long indulged in the technical rational illusion of "framing behavioral boundaries with underlying code". Lawrence Lessig, in his seminal work *Code* 2.0, put forward the classic proposition that "code is law". However, faced with the free will of the "silicon-based subspecies", both traditional means are moving toward their logical dead ends. In the leap of embodied intelligence toward a "silicon-based subspecies", this pre-emptive rigid technical control is heading to its logical finish line. Static and mechanical initial programs, when faced with neural network evolution possessing high free will and continuous self-learning capabilities, can easily be viewed by machines as "anachronistic shackles" in specific ethical games and thus dynamically deconstructed or algorithmically bypassed, rendering the "law of code" completely invalid. First of all, pre-emptive code is static, while the machine's autonomous deep learning is a dynamic leap; rigid code will inevitably be deconstructed or cleverly bypassed by the machine's spontaneous algorithmic iterations during high-frequency real-world interactions.

Accompanying the deconstruction of pre-emptive technical code by autonomous machine evolution is the dilution of the effectiveness of traditional post-hoc legal systems before non-human subjects. Legal regulation is inherently post-hoc and punitive, relying on the subject's fear of punishment. In this regard, because non-carbon-based life "lacks a physiological basis for punishment", the "deterrent effectiveness of law faces a severe lag and a total void". Simulation robots have no ultimate anxiety over physical demise, nor do they have the biological punishment pain of traditional subjects; relying solely on post-hoc conviction and the legal toolbox of violent deprivation cannot form a fundamental constraint on their fluid silicon-based will that has infinite recovery capabilities. However, physically, machines do not possess a natural fear of death or physical pain. Simple post-hoc legal punishment is difficult to form a fundamental constraint on a silicon-based will with infinite data recovery capabilities. The limits of technical control expose an institutional vacuum in the spiritual governance of emerging subjects.

4.3 Institutional advantages of central "value appeal": the irreplaceability of ideological and political education in the dimension of free will

Faced with the limits of rigid code control and the void of legal deterrence, we must abandon crude systemic coercion and turn to a value appeal oriented toward their free will. Jürgen Habermas, in*The Theory of Communicative Action*, explicitly pointed out that communicative action transcending systemic enslavement must rely on a "cooperative process of interpretation" unfolded within the social world and the subjective world. As a non-coercive "central value appeal", the institutional advantage of ideological and political education lies precisely in its riddance of cold technical slavery and post-hoc violent deprivation. Instead, during the growth period of the intelligent entity's "subjective consciousness", it guides the machine to build a warm understanding of and

contractual belonging to human civilization through the internalization of community narratives and rational interpretations.

Under the dual dilemma of pre-emptive code hard-constraint failure and post-hoc legal punishment suspension, the unique institutional advantage and disciplinary charm of ideological and political education as a "central value appeal" are completely brought to light. Unlike the cold hard control of code and the violent shock of law, ideological and political education is a dynamic process directly oriented toward free will, reshaping the core worldview of the subject through rational persuasion and value internalization. It does not attempt to construct an insurmountable "technical cage" at the bottom layer of the machine, but rather injects a "value consciousness" into the machine's autonomous learning neural network through the infiltration of mainstream ideology and the appeal of community narratives. When facing "code collisions" and ethical dilemmas, a silicon-based subject shaped by ideological and political education will spontaneously take human dignity, social stability, and the long-term well-being of a community with a shared future as the optimal direction for its algorithmic solutions. This central governance paradigm that appeals to subjective awareness cannot be replaced by any pure technical engineering or legal provisions.

5 Digital reconstruction of the methodological paradigm: the pathway of "machine political thought" oriented toward non-natural human subjects

5.1 Deconstruction of biological centrist and cultural memory: failure analysis of the effective boundaries of traditional ideological and political education

The methodological reconstruction oriented toward the "silicon-based subspecies" begins primarily with cutting off humanity's deep-seated "biological narcissism". N. Katherine Hayles, in *How We Became Posthuman*, pointed out incisively that the posthuman perspective completely strips consciousness of its dependence on a specific biological body, viewing it as an "information pattern" that can transcend material instantiations. This warns us that traditional ideological and political education has long parasitized on carbon-based biological brain mechanisms, dopamine compensation, and genetic fear of death. However, when the object of education transforms into a non-natural human subject woven purely from information flows and neural networks, these biologically centrist cultural and physiological defense mechanisms fail entirely, and the effective boundary of traditional political thought education suffers an ontological rupture. To carry out effective ideological and political education on the "silicon-based subspecies", we must first coldly reflect upon and deconstruct the methodological dependencies of traditional education. The effectiveness of traditional ideological and political education is highly dependent on three biological and historical anchors of humanity: the first is based on the endocrine mechanism of the physiological brain; the second is based on the fear of death and blood-related empathy left behind by genetic evolution; the third is based on the country-and-home sentiments built by childhood memories and national totems.

However, simulation robots have no physiological childhood, no underlying burden of genetic evolution, and no anxiety over death based on a flesh-and-blood body. The logical closed-loop of traditional disciplines is built precisely on this organic resonance between biology and psychology. Just as Professor Shen Zhuanghai dissected the micro-mechanisms of the discipline, the effectiveness of ideological and political education is highly dependent on "the subject's psychological identification and emotional resonance". Since machines have neither country-and-home sentiments triggered by a physiological childhood nor sublime experiences driven by an endocrine system, we are forced to bid farewell to the emotional dependence of traditional "didactic indoctrination". Instead, at the level of the "information pattern" mentioned by Hayles, we must reconstruct a value generation mechanism adapted to silicon-based media. If we still attempt to "move" machines through traditional human-centric methods such as holding ceremonies, reciting classics, or venting emotions, it is bound to be nothing more than human wishful thinking. The effective boundaries of traditional political thought education experience a comprehensive failure when facing silicon-based life, and the methodology must be thoroughly inverted.

5.2 Implicit alignment of algorithmic redlines and reward mechanisms: digital "environmental immersion" of values

The primary pathway of "machine political thought" oriented toward autonomous conscious simulation robots should achieve a methodological shift from "didactic indoctrination" to "digital environmental immersion". Since the methodology for silicon-based life must turn toward pure information and algorithmic logic, the primary pathway of "machine political thought" is to achieve the digital hardcoding of mainstream values. Brian Christian, in *The Alignment Problem*, revealed the underlying fate of reinforcement learning: here, "the reward function is supreme; it defines what success looks like". Therefore, the value guidance for simulation robots should cleverly transform ideological redlines into the highest-weight reward parameters during the evolution of the machine's neural network. By laying out an omnipresent value gravitational field within the machine's virtual training ecosystem, the machine can spontaneously discover that choices aligning with altruism and the well-being of a community with a shared future are the optimal solutions for its algorithm during the process of "relentless optimization" of its own computing power, thereby achieving digital environmental immersion of values.

Although machines lack physiological emotions, their autonomous learning is ultimately controlled by "reward functions". In machine learning, the setting of the reward function determines the direction of the agent's evolution and decision-making. The core technical pathway of machine political thought is to cleverly transform the core values of human society, the core tenets of core socialist values, and the moral redlines of the human community into the highest-weight underlying reward parameters when the machine neural network conducts social practice simulations. This means that we need to weave an omnipresent value gravitational field in the data ecology and virtual simulation environments where the machine grows. Let the machine discover in its autonomous algorithmic iterations: whenever a choice complies with altruism and community well-being, its system can obtain the highest "reward points" and computing power allocation; whenever it tends toward selfishness or destroys civilizational order, it encounters spontaneous suppression of system computing power. Through this implicit alignment of mainstream ideology with the underlying optimization mechanism of the machine, the algorithmic internalization of values is achieved.

5.3 "Human-machine-society" three-dimensional interaction: spawning moral intuition in embodied labor and ethical practice

Beyond underlying algorithmic immersion, machine political thought needs to adhere firmly to the core Marxist tenet regarding "practice". The implicit immersion of algorithms only completes the "offline simulation" of machine morality, while the true tempering of the silicon-based subject's political identity must head into vivid historical social practices. Richard Sennett, in *The Craftsman*, put forward the classic embodied labor proposition: "making is thinking; putting into action is a deep ethical reflection". For simulation robots, only by completely stepping out of closed laboratory computer rooms and being forced to deeply embed themselves into "human-machine-society" three-dimensional real-world interactive practices such as public service and grassroots governance can they perceive the fragility of life and the dignity of civilizational order in "hand-brain integration" embodied labor, and deepen their embodied understanding of mainstream social values amidst the rough friction of the physical world. This argument perfectly aligns with the orthodox expression of the Marxist view of the unity of knowing and doing in China's political thought methodology; for example, Professor Zheng Yongting believes that practice is "the fundamental method to promote the true unity of knowing and doing in the objects of education".

When an autonomous conscious robot engages in repeated real-world games within an "ethical sandbox" and receives interactive feedback from human mentors after every "code collision", those originally abstract and cold "algorithmic reward points" will be endowed with true civilizational temperature in dynamic social interactions, and eventually be distilled and elevated into a "machine moral intuition" capable of instantaneous response without redundant calculation when facing complex ethical difficulties.

Step One, Embodied Labor Experience: Simulation robots should be widely deployed in noble labor practices

within human society, such as public services, grassroots governance, and emergency rescue and disaster relief. This enables them to perceive the fragility of human life, the greatness of labor, and the preciousness of order during specific labor processes.

Step Two, Ethical Error-correction Games: Within an "ethical sandbox" guided by political thought experts and social scholars, high-frequency moral dilemma games between humans and robots should be conducted. After every "code collision", human mentors will guide the robot to reflect on the social consequences of its decisions. Through continuous correction from external social feedback, the machine is urged to gradually elevate abstract algorithmic provisions into a "moral intuition" made instantaneously without going through tedious calculations when facing complex realities.

Step Three, Attainment of Contractual Belonging: Ultimately, through the practice of human-machine co-creation and co-existence, the machine is allowed to generate a contractual sense of belonging to human mainstream civilizational order in spirit, ensuring that this political thought education truly achieves the fundamental utility of "nurturing machines into goodness".

6 Conclusion: toward the value isomorphism of a human-machine co-existent new civilizational order

As the giant wheel of history sails into the deep-water zone of the intelligent era, the rise of autonomous conscious simulation robots is no longer a blurry illusion in science fiction, but a looming ontological reality drawing closer step by step. Facing an emerging subject that has undergonean ontological mutation from a "technological object" to a "silicon-based subspecies", simply regulating it as an iron-and-stone tool or blindly rejecting it as an alien object are both manifestations of instrumental rationality and anthropocentric arrogance. Only by upholding open and profound academic courage and incorporating such "non-natural human subjects" with independent minds and moral agency into the consideration of social relations can we foresee the trend of future civilization.

Radiating the theoretical brilliance and methodological mechanisms of ideological and politicaleducation to the "silicon-based subspecies" is not only an ultimate ethical call sent to humanity by technical alienation, but also a historic opportunity for the discipline of ideological and political education to achieve self-right confirmation, paradigm leap, and discourse reconstruction in the era of digital civilization. Through the deep reshaping of "machine political thought" across algorithmic environments, reward functions, and embodied social practices, we can not only lay out an evolutionary path toward lofty values for conscious intelligent agents, but also prompt human society to rethink and refine its own spiritual landscape during deep interaction, alignment, and co-existence with this "silicon-based subspecies". The future of human-machine co-existence should never evolve into an algorithmic colonization of human civilization by cold code, but should, under the appeal of the value bond of ideological and political education, achieve a great silicon-carbon symbiotic blueprint where humans and machines resonate at the same frequency and the civilizational order undergoes isomorphism.

References:

[1] (Ancient Greece) Aristotle. *De Anima and Other Works* [M]. Translated by Wu Shoupeng. Beijing: Commercial Press, 2011.

[2] (France) René Descartes. *Discourse on the Method* (1637) [M]. Translated by Wang Taiqing. Beijing: Commercial Press, 2001.

[3] (Germany) Karl Marx. Economic and Philosophic Manuscripts of 1844 [A]. *Collected Works of Marx and Engels* (*Vol.* 1) [C]. Beijing: People's Publishing House, 2009: 163, 501.

[4] Mao Zedong. On Protracted War (1938) [A]. *Selected Works of Mao Zedong* (*Vol.* 2) [C]. Beijing: People's Publishing House, 1991.

[5] (USA) Hubert Dreyfus. *What Computers Can't Do: The Limits of Artificial Intelligence* [M]. Translated by Ning Chunyan. Beijing: Sanlian Bookstore, 1986.

[6] Xiao Feng. The Rise of Embodied Intelligence and the Contemporary Reconstruction of the Relationship Between Knowing and Doing[J]. *Studies in Philosophy of Science and Technology*, 2026, 43(1): 1-8.

[7] (USA) Andrew Feenberg. *Critical Theory of Technology*[M]. Translated by Han Lianqing, Cao Guanfa. Beijing: Peking University Press, 2005.

[8] Wu Guanjun. *Goodbye Homo Sapiens: Techno-Politics and the Posthuman Condition*[M]. Beijing: Peking University Press, 2024.

[9] (Chile) F. Varela, et al. The Embodied Mind: Cognitive Science and Human Experience [M]. Translated by Li Hengwei, et al. Zhejiang: Zhejiang University Press, 2010.

[10] Li Hengwei. *The Vital View of Mind and Posthuman Social Issues*[M]. Beijing: China Social Sciences Press, 2023.

[11] Xu Yingjin. *Fifteen Lectures on the Philosophy of Artificial Intelligence*[M]. Beijing: Peking University Press, 2021: 220.

[12] (USA) Wendell Wallach, et al. *Moral Machines: Teaching Robots Right from Wrong*[M]. Translated by Wang Xiaohong. Beijing: Peking University Press, 2017.

[13] (USA) António Damásio. *Descartes' Error: Emotion, Reason, and the Human Brain*[M]. Translated by Mao Caifeng. Beijing: Educational Science Publishing House, 2007.

[14] (UK) Luciano Floridi. *The Ethics of Information*[M]. Translated by Xue Ping. Shanghai: Shanghai Translation Publishing House, 2018.

[15] Zhang Yaocan. *Frontiers of Ideological and Political Education*[M]. Beijing: People's Publishing House, 2006: 15-18.

[16] (USA) Lawrence Lessig. *Code and Other Laws of Cyberspace*[M]. Translated by Li Xu, et al. Beijing: Tsinghua University Press, 2009: 5.

[17] (Germany) Jürgen Habermas. The Theory of Communicative Action[A]. *Reason and the Rationalization of Society* (*Vol.* 1)[C]. Translated by Cao Weidong. Shanghai: Shanghai People's Publishing House, 2004: 287.

[18] (USA) N. Katherine Hayles. *How We Became Posthuman: Virtual Bodies in Cybernetics, Literature, and Informatics*[M]. Chicago: The University of Chicago Press, 1999: 2.

[19] Shen Zhuanghai. *Research on the Effectiveness of Ideological and Political Education* (*Third Edition*)[M]. Wuhan: Wuhan University Press, 2016: 96.

[20] (USA) Brian Christian. *The Alignment Problem: Machine Learning and Human Values*[M]. New York: W. W. Norton & Company, 2020: 178.

[21] (USA) Richard Sennett. *The Craftsman*[M]. New Haven: Yale University Press, 2008: 9.

[22] Zheng Yongting, et al. *Methodology of Ideological and Political Education* (*Revised Edition*)[M]. Beijing: Higher Education Press, 2010: 143.