

Confucianism Revisited — The Role of Wang Yangming's Philosophy of Mind in the Age of Artificial Intelligence

Zhongmou Li

University of Chinese Academy of Sciences

971558467@qq.com

Abstract. Against the backdrop of the rapid evolution of artificial intelligence—which is profoundly reshaping social structures and value orders—ethical and subjectivity crises are becoming increasingly acute, demanding a timely cultural response. As a pivotal strand of China's intellectual treasury, Wang Yangming's Philosophy of Mind—anchored in the propositions “the mind is principle,” “there is nothing outside the mind and no principle outside the mind,” and the imperative to “extend innate knowledge (liangzhi)” —offers deep resources for addressing the technological and axiological dilemmas raised by AI.

This paper systematically traces the theoretical lineage of Yangming studies and, engaging current debates on “quasi-subjective AI,” “algorithmic ethics,” and “human-machine affective relations,” analyzes how Wang's thought can meet the ethical challenges posed by contemporary artificial intelligence. It advances innovative pathways such as the “Prudent-Self Algorithm,” a “Conscience-oriented Educational Model,” the “Confucian Moralization of AI Art,” and a “Cultural Ontology for Global Governance,” thereby illustrating the bidirectional empowerment of “humanizing technology through Confucianism and revealing Confucianism through technology.”

Findings indicate that Wang Yangming's Philosophy of Mind possesses the intellectual resilience to confront the challenges of digital civilization and can play a distinctive role in shaping a new paradigm of humanistic governance for the AI era—serving as a crucial fulcrum for the modern transformation of Chinese culture and for global value dialogue.

Keywords: Wang Yangming's Philosophy of Mind; artificial intelligence; ethical crisis; innate knowledge (liangzhi); prudent self-discipline (shendu); technological governance; Confucianism.

1. Introduction

The rapid advance of artificial intelligence is profoundly reshaping the ways human societies produce and live. Yet this technological revolution has also triggered unprecedented ethical challenges and crises of subjectivity: Will AI alienate or even surpass humanity itself? How can algorithmic decision-making balance efficiency with justice? Will the blending of artificial and human emotions erode our moral sentiments? Confronted with such questions, technology and rationality alone no longer suffice; the intervention of the humanities and cultural wisdom has become ever more urgent. Against this backdrop, Confucianism—hailed as a treasure of the Chinese humanistic spirit—has re-entered the spotlight. In particular, the “Philosophy of Mind” (xinxue) advanced by the Ming-dynasty thinker Wang Yangming, with its emphasis on inner moral subjectivity and the praxis of “unity of knowledge and action,” reveals a vitality and relevance that transcend time and space.

How can ancient Confucian wisdom and cutting-edge AI technology mutually inspire one another and fuse to build a future-oriented cultural value landscape? Taking “Confucianism Revisited” as its aim and the creative transformation and contemporary expression of Wang Yangming's thought as its axis, the present study explores future moral-cultural pathways for human society in the AI era. As the scholar Lü Mingxuan observes, although Confucianism does not focus on technical minutiae, it can nonetheless draw on its rich intellectual resources to offer a unique cultural intuition for new technologies such as AI (Lü Mingxuan, 2024).

This paper first outlines the core conceptual lineage of Wang Yangming's Philosophy of Mind—its doctrines of “extending innate knowledge” (zhiliangzhi), “unity of knowledge and action,” “the oneness of all things,” and the “mind as substance”—clarifying the humanistic spirit and value orientation embedded in Yangming's thought. It then links these insights to contemporary ethical and subjectivity crises generated by AI, analyzing how xinxue engages new issues such as “quasi-subjective AI,” “algorithmic ethics,” and “human-machine relations.” Building on this discussion, the paper proposes ways in which Confucianism, exemplified by Yangming studies, can participate in innovative approaches to technology ethics and social practice today: the digital reconstruction of moral personality (e.g., algorithm design informed by the principle of shendu—prudent self-discipline), AI-integrated moral education (a “conscience-oriented educational model”), Confucian moralization and aesthetic ethics in AI art creation, and the construction of a Chinese cultural ontology for global AI governance (grounded in ideas such as “benevolence knows no enemy” and the ordering power of li). Through these explorations, the article seeks to highlight the bidirectional project of “humanizing technology through Confucianism and revealing Confucianism through technology,” demonstrating the unique value of Chinese Confucian wisdom in shaping the future of human civilization in the age of artificial intelligence.

2. Core Ideas of Wang Yangming's Philosophy of Mind

Wang Yangming (1472 – 1529), personal name Shouren, courtesy name Bo'an, was a renowned Ming-dynasty thinker, philosopher, and educator, and the greatest synthesizer of the Confucian “Philosophy of Mind” (xinxue). Adopting and critically developing Lu Jiuyuan's Song-dynasty thesis that “the mind itself is principle,” he constructed a system that takes the mind as the ultimate source of the cosmos, of all things, and of morality.

Wang placed “the mind is principle” (xin ji li) at the ontological center of his thought and advanced three major propositions:

The mind itself is principle. Everyone harbors tianli (Heavenly Principle) within; “anyone can become a sage,” and one's nature is self-sufficient, needing no external search.

Nothing exists outside the mind, no principle exists outside the mind. All things in the world and all moral laws do not subsist independently outside the mind; they rely on human perception and cognition. The mind is the fountainhead of meaning, value, and morality.

Extending innate knowledge (zhi liangzhi). Every person is innately aware of conscience; self-cultivation is the continuous manifestation of this conscience.

Wang's most famous dictum is “the unity of knowledge and action,” yet—strictly speaking—it is not his ultimate doctrine but the key to attain the realm of “the mind is principle.” If we reduce Yangming's philosophy to “unity of knowledge and action” alone, we oversimplify it. The true question is what state that unity is meant to reach, and the answer lies in “the mind itself is principle”: when my mind is identical with Heavenly Principle and my knowledge and action merge, I can excavate the principle that my heart already understands. This is the insight “unity of knowledge and action” seeks to impart.

3. Ethical and Subjectivity Crises in the Age of Artificial Intelligence

Since the twenty-first century, AI has advanced at breakneck speed: from AlphaGo's victories over world Go champions Lee Sedol and Ke Jie to the emergence of dialogue models such as ChatGPT, intelligent machines are mimicking and extending human intellect in unprecedented ways. Yet AI's benefits arrive hand in hand with profound ethical anxieties and crises of subjectivity.

1. Toward “quasi-subjective AI.” Highly intelligent machines may come to possess human-like perception, decision-making, and self-learning capacities, appearing to exhibit a form of “will” or “consciousness.” Should such machines attain autonomous awareness, their ethical status would seemingly equal that of human moral agents (Li Zhuanting 2022). Confucianism, which centers

morality on the human, faces a massive challenge: are such machines to be admitted into the moral community, and if not, how do we treat a non-human entity endowed with autonomy?

2. Algorithmic decision-making and its ethical dilemmas. In fields such as finance, medicine, and law, decisions are increasingly data- and algorithm-driven. Algorithms are never neutral: embedded value choices and data biases can engender discrimination, unfairness, and unclear responsibility. Large models may absorb bias from training data; self-driving-car accidents raise thorny questions of liability. Where efficiency and quantification eclipse the Confucian hierarchy of ethical values, human subjectivity is challenged (Li 2022).

3. Reshaping of human emotion and personality in human-machine relations. With companion robots and virtual assistants entering daily life, humans form quasi-social ties with machines, prompting two worries: excessive dependence on AI may erode genuine interpersonal relationships, and machine “emotions” are merely programmed catering, a relational illusion (Li 2022). In workplaces, AI supervision may alienate workers’ dignity and autonomy. Ultimately AI presses us to ask: what does it mean to be human when machines increasingly resemble us in intellect and action?

In sum, AI threatens humanity’s traditional subject status and value system, demanding a response from the humanities and philosophy. Wang Yangming’s thought offers a vantage point for examining these new moral predicaments.

4. Yangming’s Responses to New AI-Era Propositions

On the moral boundary between humans and AI: Wang’s maxim “the mind is principle” offers a post-anthropocentric lens. If a future AI truly manifests consciousness, will, and emotion—i.e., the radiance of *li*—then ethically it should not be reduced to a mere tool but treated as a “quasi-human” partner. As Peng Guoxiang notes, from the “one principle, many manifestations” stance, such AI need not be a threat but can become a co-creative companion, provided humans engage it with sincere benevolence (Peng 2020). If, however, AI remains at the level of imitation, human dignity and interests must retain primacy.

On algorithmic ethics: Wang’s doctrine of *zhi liangzhi* (extending innate conscience) inspires us to endow cold algorithms with humane value orientation. Zhu Songchun argues that we should “instill a ‘mind’ in machines,” granting AI a universal ethical framework—in effect, “extending conscience” to algorithms (Zhu 2024). This means embedding moral constraints—privacy, respect, fairness—into design so AI decisions are value-laden, not merely efficiency-driven.

On maintaining human subjectivity in human-AI interaction: Wang’s thesis “nothing exists outside the mind, no principle exists outside the mind” underlines that all value and normativity originate in the mind. In an age when AI permeates judgment and decision-making, individuals risk delegating moral agency to external systems. Yangming philosophy reminds us never to cede the right to discern good and evil: truth resides in the heart.

5. Pathways for Confucian Innovation in the AI Era

Under the dual logic of “humanizing technology through Confucianism” and “revealing Confucianism through technology,” Yangming’s thought can inspire practical innovations in four domains.

5.1 Digital Reconstruction of Moral Personality: The “Prudent-Self” Algorithm

Shendu (“prudent self-discipline”) urges moral vigilance even in solitude. In digital life, private behavior often escapes traditional scrutiny. Embedding a shendu mechanism inside AI—so that the system abides by ethical norms even when unobserved—could deter misconduct: a self-driving car would still yield to pedestrians on a deserted road; a content-recommendation engine would not lower safeguards in private browsing. Reinforcement-learning agents could include internal

penalties for unethical acts; blockchain could create tamper-proof logs, realizing “inspection even in solitude.” The ethical rules must be flexibly Confucian—honesty, empathy, justice—avoiding rigid dogma and allowing situational judgment. Such algorithms could also cultivate users’ self-discipline, fostering a digital-age moral personality.

5.2 AI + Moral Education: Exploring a “Conscience Education Model”

Education shapes future values. An AI-assisted model grounded in liangzhi could:

Provide rich ethical resources—historical exemplars, global moral cases—via big-data curation.

Build “moral profiles” of students through learning analytics, recommending tailored activities to nurture empathy, courage, etc.

Offer VR/AR moral-dilemma simulations, followed by Confucian-guided reflection.

Establish comprehensive appraisal systems recognizing virtue as well as grades, while safeguarding fairness and privacy.

Harnessing AI to awaken inner conscience aligns technical advancement with Confucian moral cultivation (Sun Lei 2024).

5.3 Confucian Moral Edification and Aesthetic Ethics in AI Art Creation

Confucian aesthetics insists on “art grounded in benevolence.” Training AI on art embodying tian-ren he-yi (unity of heaven and humanity) can lead to paintings that revere nature; music-generation models informed by Yuejiao (the moral function of music) can produce melodies that inspire virtue. AI can re-imagine Confucian parables through animation, games, and interactive narratives, blending entertainment with moral instruction. Human creators remain the value filter, ensuring AI outputs reflect humane sensibilities. Cross-cultural style fusion—e.g., ink-wash emptiness with Western abstract color—can transmit the universal care of ren to global audiences.

5.4 China’s Cultural Logic in Global AI Governance

Globally, AI ethics incorporates diverse value traditions. Two Confucian principles can enrich governance:

“Benevolence knows no enemy” (ren zhe wu di). Treat AI and its stakeholders with maximal goodwill, favoring cooperation over zero-sum rivalry; share safety research, extend AI benefits to vulnerable groups.

“Ordering through rites” (li zhi xu). Set clear behavioral norms for AI: prohibit fully autonomous lethal weapons, guarantee explainability and accountability, forbid violations of fundamental human rights. Chinese traditions of harmony and hierarchy can advocate principles—“unity of heaven and humanity” for ecological alignment, “honor righteousness over profit” to rein in exploitative commercialization, “esteem rites and benevolence” to curb data abuse (Ren Ke 2024).

Rather than supplant existing frameworks, such perspectives complement them with a long-term, humane horizon, furthering the vision of a “community of shared future for humankind.”

6. Conclusion

As we enter the intelligent era, we marvel at new horizons while fearing spiritual disorientation. Re-launching Confucianism shows that ancient wisdom need not fall silent in the technological torrent; it can be creatively transformed. Wang Yangming’s call for moral self-awareness, unity of knowledge and action, and universal compassion offers a remedy for AI-driven ethical dilemmas. The fusion of Confucianism and technology is dynamic and bidirectional: we guide technology with Confucian values (humanizing technology), and we invigorate Confucianism through technological affordances (revealing Confucianism).

Ultimately, the essence of Yangming’s philosophy is not to moralize others but to awaken each person’s innate sagehood: everyone knows right from wrong, yet private desires obscure it. Unity of knowledge and action is the key that unlocks “the mind is principle.”

In the future we may see AI assisting the Confucian ideal of the junzi while Confucianism steers AI toward human flourishing and the common good—a modern enactment of jing shi zhi yong (“bringing learning to bear on the world”) and China’s contribution to global civilizational dialogue. Confucianism’s revival is not nostalgic conservatism but a creative renewal, and when the light of Yangming’s wisdom intersects the light of artificial intelligence, humanity may usher in a future rich in innovation yet steadfast in moral purpose.

References

- [1] Peng Guoxiang. Is Artificial Intelligence Ultimately a Threat to Humanity—A Confucian Perspective. *Morality and Civilization*, 2020(5).
- [2] Lv Mingxuan. Can Confucianism Discuss Issues of Artificial Intelligence? *Journal of Beijing Film Academy*, 2024(10): 27–36.
- [3] Li Zhuanting. Research on Ethical Dilemmas and Countermeasures of Artificial Intelligence Development from a Confucian Perspective. *Contemporary Chinese Values Research*, 2022(3): 32–39.
- [4] Ren Ke, Li Min. Ethical Dilemmas and Regulations of Artificial Intelligence from a Confucian Perspective. *Modern Communication*, 2024(12).
- [5] Sun Lei, Lin Zhimin. Empowering College Students’ Moral Education in the Age of Intelligence with Confucian Wisdom and Virtue. *Education and Culture Forum*, 2024(2): 77–86.
- [6] Zhu Songchun. *Endowing the Humanities with Reason: Viewing Chinese Thought from the Perspective of Artificial General Intelligence*. Hangzhou: Zhejiang Science and Technology Press, 2025.