The Navigation of Taoist Ethics for a Good and Just Life with AI

To what extent can artificial intelligence be used in society while allowing humans to achieve the Taoist ideal of a good and just life?

PHIL 383

Word Count: 2712

Introduction

The integration of artificial intelligence and machine learning into daily life presents unique moral challenges regarding human agency, ethical actions, and our relationship with nature. While contemporary scholarly research primarily uses Western ethical frameworks to approach the impacts of this technology, philosophers like Shannon Vallor emphasize the critical need for non-Western perspectives of virtue ethics to fully analyze our future (Vallor 2016, 22, 35). Even within research involving non-Western ethical frameworks, current discourse is centered on Confucian and Buddhist ethics, leaving a noticeable gap for Taoism, the counterpart to Confucianism in China.

At the heart of Taoism lies the "Tao" (The Way), an ineffable, formless, and nameless principle that serves as the natural order of existence, or how the universe naturally operates.[*] It is commonly associated with the idea of yin and yang, which are dynamic, equal, and complementary opposing forces. This unity of opposites reveals that concepts like strength and weakness or existence and non-existence are not absolute binaries but mutually arising forces that define and generate one another. To align with the Tao, the wise cultivate *wu-wei*, often translated to "non-action" or "effortless action." Rather than sit and do nothing, *wu-wei* suggests acting effortlessly with the flow of nature, a practice perfectly mirrored by the motif of water, which nourishes all things without striving, settles in the lowest places, and yet possesses the yielding power to erode the hardest stone. Achieving this requires a return to *ziran* (naturalness and spontaneity) and *pu* (the "uncarved block"), an unaltered, simple state of being that aligns with one's true nature. Ultimately, to maintain this pure state, one must embrace the utility of the empty vessel; just as a bowl is made useful by its empty space rather than its edges, one must clear away artificial preconceptions to remain open to the universe's potential. An agent that acts within the Taoist

[*] Also spelled "Dao" and "Daoism."

framework, contrary to an AI agent, is an entity that possesses the capacity for intentional, spontaneous action.

AI systems (neural networks, machine learning, large language models), as viewed through Taoism, naturally appear entirely opposed to a Taoist existence. An AI systems is carved: it requires finite, labeled data to model itself after, and certainly requires an extreme amount of effort that does not function naturally. As a result, AI agents themselves lack the natural spontaneity, agency, and Taoist features to be considered under a Taoist framework. Although AI lacks natural spontaneity and agency required to exhibit Taoist morals, its use as a tool by humans could foster a good and just life to a high extent if used to automate unnatural tasks.

AI as an Agent

To evaluate artificial intelligence (AI) through Taoism, its creation must be considered through the idea of the *pu* (the "uncarved block"). *Pu* represents an unformed state of natural simplicity. This is the most ideal state, for if "leaders could harness it [the uncarved block] // the Heaven and Earth would come together" (Tsu 1989, chap. 37). This is ultimately not completely possible, but only an ideal to strive for. To exist is to be cut, but a great person "…makes few cuts" (Tsu 1989, chap. 28). The most moral person will make as few changes to their unformed, natural self to pursue being "the wise" – the Taoist ideal of a person who navigates the world through effortless action (*wu wei*).

At first, an AI model might appear to embody *pu*. Before the model is trained, the foundation has no ego or desires; it sits in a state of unformed, infinite potential until training, or a user's prompt "carves" a response out of it. However, artificial intelligence cannot exist in a state of true *pu* and natural simplicity like a human. Its foundation is built upon complex data structures, structured rule-based algorithms, and millions of parameters and weights that calculate the result it's prompted to give. Even during its training, the data it is fed is meticulously carved, or labeled, by humans to force efficiency and output deterministic results, completely antithetical to the natural order of things. The *Tao Te Ching* warns

against this type of carving, stating “There are already enough names, // we need to know when to stop” (Tsu 1989, 32). Computer systems like AI are unable to process information in its most natural state in the world, but must instead rely on labels, and discrete structures to provide names to things in nature, exactly against the practices of Taoism. AI lacks the ability to be unformed because it is structurally bounded and relies upon carved data and constraints, thus unable to experience the naturalness of the uncarved state.

Furthermore, the fundamental architecture of AI actively resists the Taoist principle of *ziran*, or natural spontaneity. *Ziran*, translating roughly to “self-so,” demands that an entity’s actions arise organically from its innate, unconditioned nature, free from artificial coercion. Because an AI possesses no innate nature or selfhood, its existence is entirely constructed and coerced by human engineering, making true *ziran* impossible. One might counter this by pointing to the "emergent capabilities" of large language models, arguing that AI does, in fact, exhibit a form of spontaneous naturalness. When an AI generates novel art, deduces logical patterns it was not explicitly programmed to solve, or produces highly unpredictable, creative text, it appears to act organically, creating "of itself" without rigid, line-by-line direction. Yet, this perceived spontaneity is merely an illusion masking immense, deterministic calculation. To be "self-so" requires an authentic, grounded existence in the world, but an AI’s output is merely high-level mechanical reckoning. Its "creativity" is entirely dictated by external carvings: the biases of its training data, the mathematical constraints of its neural architecture, and the rigorous shaping of reward functions designed to force compliance. When an AI produces a surprising result, it is not spontaneously flowing with the natural order; it is blindly traversing a rigidly defined latent space to minimize a mathematical loss function, permanently tethered to the artificial parameters set by its creators.

Because artificial intelligence lacks the natural spontaneity of *ziran*, it is fundamentally incapable of possessing genuine virtue, or *De*. In the *Tao Te Ching*, Lao Tzu explicitly warns against the rigid enforcement of morality, observing that while authentic goodness is accomplished effortlessly, the "just person" leaves much to be done, and the "disciplinarian" inevitably resorts to coercion to force order upon

the world (Tsu 1989, chap. 38). In Taoist philosophy, true virtue is never the result of such forced action or a strained, mechanical attempt to act morally. Instead, it is the natural, uncoerced action of an agent, that organically aligns with *De* through *ziran* and *wu-wei*. If you simply force morality upon yourself, are you truly virtuous? It may appear that way to others, but you will never ultimately be in alignment with the Tao unless you naturally act in alignment with *De*. Artificial intelligence, heavily constrained by its data, reward function, and guardrails, acts similar to the disciplinarian or the just person. It is forced to choose and to strive towards *De* if the programmer desires, but this forced, coercive action is the exact opposite of what the *Tao Te Ching* prescribes. Without *wu-wei* and *ziran*, AI cannot act in alignment with *De* necessary to navigate moral landscapes according to Taoism. As a result, artificial intelligence cannot be evaluated as an independent moral agent under Taoist philosophy.

AI as a Tool

Because AI cannot be evaluated as an independent moral agent, it must instead be considered as a tool that a moral agent (humans) could use. A Taoist approach does not require the rejection of artificial intelligence, despite its flaws. Instead, it seeks to integrate it by applying the principle of yin and yang*, which instructs us to find an equilibrium between opposing forces. All things naturally possess these dual energies, and true stability is only realized when they are successfully synthesized (Tsu 1989, chap. 42). The mathematical efficiency and structured, analytical effectiveness of artificial intelligence represent an active and forceful form of yang. However, the swift advancements of technology, especially artificial intelligence, frequently encounters a conflict between itself and ethical accountability (Cui, et al. 2025, 159). While the advantages are clear, they are accompanied by substantial risks like biases, data privacy issues, and employment displacement (Cui, et al. 2025, 159). To achieve the harmony described by Lao Tzu, the yang of artificial intelligence must be balanced by the morally grounded yin of humanity.

* Often symbolized by the dark (yin) and the light (yang).

This perspective advocates for a methodology that accepts the duality of artificial intelligence and ethical considerations. True harmony is realized when it is guided by human empathy and judgement, guaranteeing that technological advancements do not compromise human rights or social welfare, including the attributes of Taoism. It is also essential that this relationship is dynamic and necessitates continual adjustment as new technologies arise and social attitudes change, without compromising the core tenants of Taoism.

Building upon yin and yang, the use of an artificial intelligence system must allow agents to embody *wu-wei*, or effortless action. An AI made with this principle should not force compliance, overstep its boundaries, or impose decisions that undermine human autonomy (Cui, et al. 2025). Instead, it must function strictly as a tool or a suggestion to enhance human existence. This aligns with the *Tao Te Ching*, which tells us to "Retire when the work is done. // This is the way of heaven" (Tsu 1989, chap. 9). A Taoist approach to artificial intelligence requires the machine to perform its rigorous computations and then "retire," leaving the practical decisions in the hands of humans. Because the heavy lifting is performed by artificial intelligence, it leaves the human agent to effortlessly choose in accordance with nature, with the AI as a suggestion. By presenting options instead of enforcing order, the tool supports the natural flow of human choice without coercing it.

Artificial intelligence operates as an inherently carved block bound by parameters, predefined algorithms, and labeled data that fundamentally lacks the spontaneous flow of *ziran*. This presents a critical issue: if AI is fundamentally carved, does relying upon it inevitably pulls the human operator further away from the Tao and *De*. The risk of the user furthering themselves from the ideal of the uncarved block, or *pu*, is the primary concern of using artificial intelligence as a tool. Even considering the use of an AI might change the way a human views the world, eroding their natural spontaneity and replacing their natural humanity with a significantly more carved state. To resolve this dichotomy, the relationship between agent and the tool must be re-evaluated.

In *Between Dao and Techniques*, Wang Qian introduces the concept of *yi dao yu shu*, which translates to "guiding techniques by Dao." This idea addresses the ambiguity of navigating modern technology, describing "…social practices and techniques that require ethical reflection and are regulated by moral tradition" (Miller and Zhu 2011, 318). Because the AI (*shu*, or the technology) is devoid of the Tao and incapable of self-regulation, and all the attributes of Taoism, it cannot naturally align with the environment on its own. Therefore, the human agent must act as the moral component. Wang establishes that humans are the crucial "mediators of sociotechnical interactions" (Miller and Zhu 2011, 318). To prevent technology from dictating human behavior and pulling humans away from the core tenants of Taoism, we are urged to cultivate moral practice, especially traditional virtues, outside of technological influence.

While the AI remains a carved tool, if the agent actively grounds their actions in moral principles, can consciously regulate the guide the technology. Through *yi dao yu shu*, the technology is dependent on human moral action, ensuring that its application is guided towards naturalness rather than allowing the tool to carve away the agent's connection to *De*, and as a result, the Tao. As a result, delegating unnatural, data-heavy tasks to AI can serve as a way for humans to develop *wu-wei*. When an agent allows the machine to handle that strenuous labor, they avoid the unnatural and forced action that doesn't necessarily allow themselves to act naturally. This offloading allows the agent to move effortlessly within their circumstances without the backage of those unnatural tasks, preserving their own *ziran* and without the expenditure of effort.

This idea is found within Barry Allen's "dao-engineering." Allen argues that the Taoist critique of technology is not a condemnation of machines themselves, but the "confused, disordered knowledge" and the poor application of technology (Allen 2010, 151). The solution is certainly not to reject technology outright, but to embrace "dao-engineering, capable of artificial, artifactual, engineered wu wei effectiveness" (Allen 2010, 155). By applying this to artificial intelligence, it is clear how technology can be used to absorb the structured tasks that would pull an agent away from the Tao, due to modern life. For

example, logistical planning, like weekly schedules, budgets, and complex economic pathways involve meticulous and mechanical work. These necessary tasks for modern life do not pull the agent closer to achieving any Taoist ideals. By offloading this logistical work to an AI, the agent is freed from managing it and allows them to focus on the cultivation of *De*.

By using artificial intelligence for specific tasks that pull the agent away from naturalness and spontaneity, they can reduce their unnatural effort, or the opposite of *wu-wei*. Dao-engineering allows the AI to act as a buffer between the demands of the modern era, and the tenants of Taoism. By letting the carved block (AI) do the work that causes carving, the agent is freed and allowed to return closer to the state of the uncarved block.

This framework is inherently fragile and requires the agent to set up the artificial intelligence in a way that will not act against the agent's interests within the Taoist framework. Additionally, if done too well, one could fall victim to over-delegation, offloading not just unnatural tasks but natural, human ones. If they begin relying on the AI to create meaning, make ethical judgements, or over-influence the agent in any natural task, they abandon their role as the moral anchor required by *yi dao yu shu*. By surrendering their natural spontaneity to the machine, they risk becoming the carved block.

To prevent this, a definitive limit must be made. Wing-Cheuk Chan states that it is only when humans refrain from "letting what is human wipe out what is Heavenly," and match Heaven with Heaven, that it is possible to overcome and succeed in the age of technology (Chan 2003, 12-13). For the age of AI, we must not let technology erase our Taoist values (our path to bringing Heaven to Earth). For dao-engineering, the agent must have the wisdom to stop the AI before it oversteps into the naturalness that it cleared a way towards. For example, a Taoist moral agent should not use artificial intelligence to identify and describe things that naturally exist in the world, such as other humans, the plants and animals in the world, or other objects because it revokes their perception, in addition to carving of those objects by defining and categorizing the world in a discrete manner, against the views of Taoism. In a way, the people that will endure are those who know when to stop, and when to apply artificial intelligence.

Conclusion

Evaluating artificial intelligence through a Taoist framework requires a separation between the nature of the machine and the nature of the moral agent using it as a tool. Because AI is fundamentally "carved" from its design and lacking spontaneity, it cannot be considered an independent moral agent capable of possessing *De*. Rejecting this technology outright ignores the Taoist ideal of yin and yang. When viewed as a tool, forceful actions characterized by yang can be balanced by the natural human action of yin. Through practices of *yi dao yu shu* and dao-engineering, a Taoist agent can offload unnatural burdens of modern life onto AI to open a pathway to better pursue *De*, and thus the Tao. This technological delegation frees the human agent to cultivate *wu-wei*, returning closer to the state of the uncarved block. However, this balance must be structured around the Taoist agent remaining an anchor for the AI, exercising wisdom of knowing when to utilize the tool, and when to stop to best cultivate *De*.

Bibliography

Allen, Barry. 2010. "A Dao of Technology?" *Dao: A Journal of Comparative Philosophy* 151-160. https://research.ebsco.com/c/ybus2u/viewer/pdf/bkld6zyrlz.

Chan, Wing-Cheuk. 2003. "Phenomenology of Technology: East and West." *Journal of Chinese Philosophy* 1-18. doi:https://doi.org/10.1111/1540-6253.00103.

Cui, Jun, Mark Qin, Alex Zheng, and Wei Chang. 2025. "Guiding Artificial Intelligence Ethics through Laozi's Philosophy: Insights into Ethical Challenges and Global Developments." *Advances in Social Science, Education and Humanities Research.* Sydney: Atlantis Press. 155-161. doi:https://doi.org/10.2991/978-2-38476-444-0_19.

Miller, Glen, and Qin Zhu. 2011. "Dao Ji Zhijian: Zhongguo Wenhua Beijing de Jishu Zhexue «道技之间 : 中国文化背景的技术哲学»– By Wang Qian." *Journal of Chinese Philosophy* 317-320.

Tsu, Lao. 1989. *Tao Te Ching.* Translated by Gia-Fu Feng and Jane English. New York: Vintage House, A Division Of Random House.

Vallor, Shannon. 2016. *Technology and the Virtues: A Philosophical Guide to a Future Worth Wanting.* New York: Oxford University Press.