Journal of Confucian Philosophy and Culture, Vol. 44 (August 2025): 5-17
DOI: 10.22916/jcpc.2025..44.5

Scholar's Corner: Confucianism in and for the Modern World

Philosophizing in the Era of AI

Heisook Kim*

Article Background

This article is a revised version of the opening address delivered at the FISP conference held in Fujairah, UAE, in February 2025. FISP (Fédération Internationale des Sociétés de Philosophie, International Federation of Philosophical Societies) is a global non-governmental organization that brings together national and international philosophical societies from around the world. It was founded in 1948 to promote philosophical research, education, and dialogue across different cultures and traditions. FISP is recognized by UNESCO as a key partner in advancing global philosophical discussion, fostering mutual understanding, and contributing to cultural development

The main roles and activities of FISP include:

- Organizing the World Congress of Philosophy (WCP) every five years, one of the largest gatherings of philosophers worldwide
- Encouraging international cooperation among philosophical associations and academic institutions
- Supporting the dissemination of philosophical knowledge through publications, conferences, and educational initiatives
- Promoting philosophical dialogue across cultural, linguistic, and disciplinary boundaries

As we enter the age of digital technology and AI, philosophy must address newly emerging problems concerning human nature, con-

* Heisook Kim is President of FISP and Professor Emerita of Ewha Womans University. E-mail: hkim@ewha.ac.kr
** This paper was presented as the opening address at the FISP Conference held on February 23-24, 2025 in Fujairah, UAE.

sciousness, reality, human relationships, hyper-connectedness, gender identity, the self, and more. It is high time for philosophers to confront the challenges created by our civilization. As President of FISP, I aim to focus philosophers—and philosophy as a discipline—more sharply on the questions of our age.

As Hegel proclaimed long time ago, an individual is a child of its own age and philosophy is its own age in thought. We are thinking beings living in concrete contexts, shaped by concerns of our era. While philosophy seeks the ultimate foundation of beings and phenomena through abstract concepts and principles, it is not transcendent but rather transcendental, to borrow Kant's distinction. Philosophy is at the limit of our world, engaging in critical reflection on our world and our lives in it.

This leads us to an essential question: What is the nature of the age we are now living in? A new world seems to be unfolding before us, one unlike anything we have experienced before.

We are witnessing the emergence of a technological revolution that may radically transform our lives, leading us into a realm of uncertainty about who we are, what we are, and what we live for. The development of digital technology, biotechnology, and the AI industry is not only reshaping the details and landscapes of our lives but also altering the way we perceive ourselves and relate to others.

In this essay, I aim to raise philosophical questions that arise in an increasingly artificial world and reality. I will address three key issues: individual freedom, the neutrality of technology, and the concept of artificial humanity. In discussing these issues, I will emphasize the need for intercultural and inter-philosophical dialogues.

I. Individual Freedom

First, the question of individual freedom. With the rapid expansion of the internet, the world is becoming ever more interconnected. In this digitalized reality, individuals are often abstracted and reduced to mere data. These data are categorized based on criteria selected by corporations for their own interests. The ideal of individual freedom,

long upheld as an absolute value in the West, is now at risk of becoming an illusion globally under the pressures of both surveillance capitalism and surveillance socialism. Our use of digital platforms and social networks leaves unavoidable traces, making us increasingly vulnerable to exploitation by either big tech companies or authoritarian political powers.

Astonishing advancements in AI and biotechnology may introduce even greater uncertainty about our future. These developments are radically transforming human relationships and self-perception, contributing to the blurring of distinctions that were once familiar to us—for example, the differences between men and women, husbands and wives, family and non-family, the old and the young, the natural and the artificial, the mental and the material, humans and non-humans, and even truth and falsity. The rapid growth of AI technology will especially impact our lives on a comprehensive level by establishing new norms in society.

During the pandemic, we benefited from the rapid circulation of information about COVID-19. However, at the same time, we faced the infringement of individual freedoms due to strong state controls over mask mandates and movement restrictions. Some Western critics opposed government measures that restricted personal autonomy, while in East Asia, tracking systems were widely implemented under the justification of public safety. With the rapid development of image recognition and other information technologies, we seem to be haunted by the specter of a totalitarian society where individual freedom is increasingly under threat.

As we transition into an increasingly hyper-connected world, the situation may worsen. In this new reality, we inevitably leave digital traces of our lives. Today, individual freedom is often reduced to the ability to purchase goods—goods that algorithms predict we will want based on the preferences of others. AI systems anticipate our desires even before we consciously recognize them. Ironically, while we believe we enjoy more freedom than ever before, we are increasingly subjugated to smart and intelligent entities that constantly guide and nudge us in certain directions.

Modern society is now embedded in a vast network of surveillance, with corporations operating these invisible systems while accumulating enormous wealth. The era of surveillance capitalism is expected to expand beyond anything we can currently imagine. As the Internet of Things becomes more pervasive—making everyday objects "smart"—it seems that human desires are being fulfilled more efficiently, creating a world of abundance and greater individual choice. However, the reality is quite the opposite. The technological revolution has brought about a surveillance society that threatens individual freedom rather than enhancing it.

The problem of individual freedom has long been an intriguing subject in philosophy. However, its interpretation varies depending on how the concept of the individual is defined. Different civilizations and cultures may have distinct conceptions and criteria for individuality and personal identity.

For instance, in Confucian culture, the primary state of existence for an individual is to fulfill a given role within familial and social networks and to carry out the moral responsibilities associated with that role. It is often argued that Confucianism does not conceptualize the individual as an autonomous entity but rather as a "relational self"—a self defined by its relationships. In this cultural framework, individual freedom is inseparable from the well-being and harmony of a close-knit community, such as the family or clan. The successful fulfillment of one's societal roles is considered the highest value a person can achieve, as they are born into a concrete web of relationships.

By contrast, Christian culture has championed a unique concept of individual freedom, emphasizing the direct relationship between God and each individual. In this tradition, personal freedom is modeled on God's free will and is therefore defended as an absolute value. However, in today's digitalized and hyper-connected world, individuals are becoming increasingly relational, prompting us to reconsider traditional notions of individuality and individual freedom defined in Western tradition. There may be alternative conceptions of individual freedom in non-Western cultures that offer valuable insights. It is time to explore these perspectives and engage in multilateral philosophical dialogues across diverse cultural backgrounds.

II. The Neutrality of Technology

Artificial intelligence is already deeply integrated into our daily lives, from self-driving cars and drones to media, banking, the internet, and investment firms. AI has become mainstream, fueling various aspects of business and everyday activities. The rapid progress in AI has been driven by exponential increases in computing power and the vast availability of big data. AI is poised to play an even greater role in our culture, influencing fields such as art, advertising, law, education, and healthcare. The boundary between humans and machines is becoming increasingly blurred.

We are witnessing the reconfiguration of many long-standing conventions, customs, and norms. Our previous notions of what is typical and natural are becoming obsolete. In the age of AI, we must pay close attention to big data and critically examine who owns, interprets, and processes this data, as data-driven innovations are unfolding worldwide.

We now live in the era of big data. Every digital trace we leave behind in our countless excursions through cyberspace is collected and stored without exception. We unknowingly provide Google with information, inadvertently contributing to its growing power. Amazon often knows our tastes and desires better than we do ourselves. Our daily lives are increasingly ensnared by digital platforms, making it ever more difficult to escape their influence. The rapid development of AI relies on exponentially growing data and the advancement of machine learning algorithms.

AI is fundamentally dependent on past data and the information it continuously learns through iterative processes. Just as human behavior is shaped by habits and learning, AI is molded by the data it absorbs. When the data it receives is biased, AI inevitably inherits and perpetuates those biases. The issue of bias is both crucial and sensitive in today's world. Particularly from a women's perspective, the question of AI's neutrality is of profound significance.

There have been cases where using historical data to train systems for policing, hiring, or credit decisions has resulted in the perpetuation of past discrimination, effectively programming bias into algorithms

and influencing future decisions. Some time ago, there was news about the malfunctioning or discriminatory behavior of a facial recognition system due to biased training examples that primarily included white male faces. This may have been due to the limited availability of training data at the time, with an insufficient number of Black or Asian female faces. In environments like MIT, where white males are more commonly encountered, the data used for training AI systems might naturally reflect this demographic imbalance. Since the world itself is already biased, AI algorithms inherit and reinforce these biases.

For example, the wealthy tend to purchase more goods and in larger quantities. As a result, consumer data disproportionately focuses on those with greater purchasing power, leading algorithms to reflect and cater to their preferences, needs, and tastes.

Cathy O'Neil, a well-known mathematician in the U.S., warns in her book *Weapons of Math Destruction* that algorithms can and do perpetuate inequality. The book's subtitle, *Big Data Increases Inequality and Threatens Democracy*, highlights her concern. O'Neil borrows the term WMD (Weapons of Mass Destruction) to describe how mathematical models and algorithms can produce harmful outcomes, often reinforcing systemic inequalities by keeping the poor poor and the rich rich.

Mathematical modeling is widely used to quantify important social traits and values, such as creditworthiness, teacher performance, and human interests ranging from college rankings to personality assessments. The troubling reality is that these mathematical tools and algorithms create feedback loops that perpetuate injustice, forming a toxic cycle that many people are unaware they are trapped in. Mathematical modeling, statistics, and algorithms all appear to function objectively and factually, yet they often reinforce existing biases under the guise of neutrality.

Professor O'Neil provides an example of *recidivism* models and predictive policing algorithms—programs that deploy officers to patrol specific locations based on crime data. Recidivism refers to the tendency of individuals to reoffend after being punished. These models are particularly susceptible to harmful feedback loops. For example, O'Neil writes:

> A person who scores as "high risk" according to this model is likely to be unemployed and to come from a neighborhood where many of his friends and family have had run-ins with the law. Thanks in part to the resulting high score on the evaluation, he gets a longer sentence, locking him away for more years in a prison where he's surrounded by fellow criminals—which raises the likelihood that he'll return to prison. He is finally released into the same poor neighborhood, this time with a criminal record, which makes it that much harder to find a job. If he commits another crime, the recidivism model can claim another success. But in fact the model itself contributes to a toxic cycle and helps to sustain it. (O'Neil 2016, 27)

The question of neutrality is especially crucial when it comes to gender, race, and marginalized communities, as these groups are often underrepresented in the formation of social policies and excluded from the processes of data collection. As a result, their perspectives and experiences may not be adequately reflected in the systems that shape society. What we need is a reflective approach to understanding how models function and how they contribute to harmful cycles. There is often a significant gap between statistical models and the real world, and numerous parameters can be adjusted to bridge this divide. However, determining which parameters to prioritize is not an easily resolvable problem. Nevertheless, it is our responsibility to break these harmful cycles and strive to restore the best aspects of human nature—namely, empathy and a sense of togetherness.

In the future, what will matter most is not technology itself, but our ability to make informed choices and our commitment to building a society where people can live together in harmony. It is imperative that we foster a collective identity as human beings and cultivate a moral consciousness that emphasizes coexistence, regardless of gender or race. The neutrality of technology does not emerge naturally or automatically in an increasingly automated world. A "brave new world" has not yet arrived. In fact, it may be that true neutrality or objectivity, in the strictest sense, is unattainable. Rather, neutrality should be seen as an ideal, a movement, or what Kant would describe as a *regulative idea of reason*—something that can only be approached asymptotically.

In this context, I would argue that the most viable approach to technological neutrality lies in *transversal rationality*[1]—a framework that integrates diverse critiques and resistances to conventional practices into a heterogeneous whole that remains subject to continuous scrutiny and revision. Since different cultures develop distinct conceptions of rationality, we must engage in intercultural dialogues that transcend national, racial, gender, and religious boundaries while remaining aware of the limitations and contingent nature of our knowledge.

III. The World of Artificial Humanity

Finally, I will reflect on the world of artificial humanity. Since John Searle introduced the Chinese Room Argument in his "Minds, Brains, and Programs" as a critique of the Turing Test—which assesses a machine's ability to think based on indistinguishable performance in dialogue—there has been ongoing interdisciplinary debate on the topic. The core of Searle's argument is that, even if a machine can engage in a dialogue with a human, it does not genuinely understand the input and output; it merely simulates, rather than duplicates, human intelligence. While the machine may appear to comprehend what it is saying, it lacks real understanding. Searle argued that true human understanding and intentionality arise from the biological

[1] The concept of *transversality* originally comes from mathematics. Jean-Paul Sartre used it in his book *The Transcendence of the Ego* (*La Transcendance de l'ego*, 1936) to describe the way consciousness unifies itself without relying on a substantial or centralizing self. He believed that the unity of consciousness is achieved through the concrete and real retentions of past conscious moments. Sartre referred to this process—of unifying consciousness through the crossing of diverse past states—as a *play of transversal intentionalities*. Félix Guattari later inherited and reinterpreted the concept in *Psychanalyse et transversalité* (1974). He argued that the question of revolution is necessarily tied to a radical reworking of the concepts and methods prevailing in the field of analysis. For Guattari, it is the principle of *transversality* that must bring together and unify the roles of the analyst and the activist. In contrast to the traditional concept of *universality, transversality* generally refers to a mode of unification that crosses diverse elements without erasing their differences.

causal link between the brain and linguistic behavior. He maintained that only certain biological systems possess original intentionality and genuine comprehension.

However, my focus today is not on the philosophical question of whether machines can think, possess intentionality, or have consciousness. Instead, my concern lies on the human side of the equation: *How do we coexist with machines that simulate human behavior*, *and how do we teach them?* It is highly likely that, in the near future, we will live alongside AI and robots in various forms. We can even envision a future in which humans themselves become cyborgs to varying degrees. *Cyborgs*, by definition, are beings that combine biological and machine elements. With advancements in biotechnology and digital technology, various forms of human enhancement may become a reality in our future lives.

Even today, humans interact with AI in numerous ways, such as through secretarial services, conversational AI, education, and medical care. However, AI is not just a tool that assists us; we are entering an era in which AI will become integrated into our communities, taking on roles akin to family members or partners—perhaps even more so than pets like dogs and cats are today. The relationship between humans and AI may increasingly resemble ordinary human relationships.

The question I pose in this context is whether machines can be made virtuous and normative. As people live and work alongside AI robots, they must communicate and engage with these machines in diverse activities and services. The robots we coexist with will have the capacity to either benefit or harm humans. When humanoid robots interact with people, they must not only understand the immediate situations but also interpret human intentionality. They need to know what to say, how to say it, what to do, and how to do it appropriately.

While robots will be designed according to pre-programmed schemes for specific purposes, those integrated into daily life must learn the intricate details of human behavior—including emotional and moral aspects. Elderly individuals, for instance, may seek companionship to alleviate loneliness and wish to form bonds with their robotic companions. Consequently, robots must not only respond appropriately but also appear emotionally engaged with the people

they care for and live with. Whether robots genuinely experience emotions such as empathy or merely simulate them, the fact remains that humans who interact with them desire feelings of love, care, and warmth. The authenticity of these emotions may not matter as much as we might assume, as even in humans, genuine emotions cannot be fully assured—we can never truly know what is in another person's mind.

What matters here is not whether machines or AIs possess real emotions, understanding, or intentionality, but rather that humans perceive them as rational and emotional beings capable of caring and meaningful interaction. We do not merely need smart machines; we need AIs that exhibit linguistic and non-linguistic behaviors—such as pausing at key moments in a conversation—that align with cultural and moral expectations. People who view their personalized machines as legitimate family members may expect them to embody virtuous character traits such as faithfulness, kindness, affection, and honesty. If machines acquire moral attributes, they may come to be seen as having a persona and even a certain level of spirituality.

AIs are known to rely heavily on past data and deep learning, utilizing complex probabilistic models. Through exposure to countless human behaviors, they may appear to develop character traits and a semblance of moral sensibility. Since values and norms are deeply embedded in language, an AI's moral and emotional intelligence may be shaped by the languages it learns. For instance, the Korean language has a complex system of honorifics used to express respect for elders and those in superior social positions. AIs must develop the appropriate emotional and moral sensitivity depending on whom they are addressing. If they become proficient in multiple languages, they may need to cultivate different forms of moral awareness for different cultural contexts. In this sense, the process of AI acquiring virtues and emotional sensitivity resembles that of children learning moral values and social norms.

Different cultures across the globe have developed unique methods for instilling moral values and virtues in their people. These cultures have created religious texts, conventions, and rituals that narrate human experiences, relationships, virtues, and vices. Intercultural

dialogue will be crucial in the development of virtuous AIs that will operate across the innumerable cultural, social, and technological divides among human beings. We must strive to ensure that AI development fairly represents the norms and values of cultures that have been historically marginalized, particularly in the realm of technological advancement.

Machine learning, which is largely based on past data, has a significant impact on women and other groups who have historically suffered under oppressive cultural norms and practices. As artificial humanity emerges through technological advancements, it will blur familiar boundaries—such as those between human and machine, culture and nature, biology and physics, and even between women and men. At the same time, however, it presents an opportunity for us to take responsibility in constructing new norms that ensure freedom and fairness.

Donna Haraway, in her *Cyborg Manifesto*, once proposed a utopian vision of a world without gender, a world without origin, but also a world without end. The new reality in which we coexist with AI calls for fresh philosophical reflections on the grand boundaries, categories, and conceptual frameworks that have long defined our existence. It is up to us to determine what should be dismantled and what should be constructed. However, in the real world, we cannot simply sever ourselves from the past. We must live with it, negotiate with it, and allow it to inform our adaptation to new environments. We are, after all, the product of countless generations of evolutionary adjustment—shaped not only by the natural world, but also by the cultural worlds we inherit. This means we are not as entirely free as we might wish in constructing our future. In many cases, negotiation and compromise may prove to be the most pragmatic path forward.

In East Asia, Confucian traditions—both classical and neo-Confucian—have left a profound and enduring imprint on nearly every aspect of life, particularly in the cultivation of moral character and habitual virtue. Looking ahead, we will share our daily lives with artificial intelligence. This means that AI must learn virtues. Here, Confucian culture offers a valuable resource. For centuries, it has refined the art of instilling virtue in children and in the unlearned through

disciplined repetition and rote practice. This enduring tradition may well contribute to the humanization of AI in the East Asian context.

REFERENCES

Guattari, Félix. 1974. *Psychanalyse et transversalité*. Paris: François Maspero.

Haraway, Donna. 1985. "*Cyborg Manife*: Science, Technology, and Socialist-Feminism in the 1980s." *Socialist Review* 80.

O'Neil, Cathy. 2016. *Weapons of Math Destruction: Big Data Increases Inequality and Threatens Democracy*. London: Penguin Books.

Sartre, Jean-Paul. (1936) 1991. "The Transcendence of the Ego (La Transcendance de l'ego)." In *The Transcendence of the Ego: A Sketch for a Phenomenological Description*. New York: Hill and Wang.

Searle, John R. (1980) 1981. "Minds, Brains, and Programs" with reflections. In *The Mind's I*, composed and arranged by Douglas Hofstadter and Daniel Dennett. New York: Basic Books.

* About "Scholar's Corner":

Scholar's Corner is a dedicated section that explores the multifaceted dimensions of Confucianism in modern society. Since its launch in 2019, it has featured thoughtfully curated contributions from renowned scholars and experts, offering in-depth analyses and unique perspectives on a wide range of Confucian themes.