Abstract: As artificial intelligence (AI) outpaces the human brain, it is invoking wide-spread fear that men and machines are moving into a conflicting zone. Some even suspect that AI machines may one day consider human beings as slow and sloppy, and thus worthy of subordination or elimination. A growing challenge to mitigate the looming crisis requires science to expand its artificially augmented intelligence by incorporating elements from the ethical–spiritual and human universe. Our endeavor to bridge the prevailing gap between science and spirituality focuses on Buddhism, which stands out in its ability to achieve a rare fusion between natural, spiritual and human worlds. This unique synthesis is specifically mediated by Buddhist ‘causality’, where one aspect explains reality based on a scientifically proven cause and effect paradigm, but the other aspect interprets it by compassionate humanism. It argues that the missing human–spiritual dimension in artificial intelligence can be remedied by the Buddhist concept of ‘causally’ linked to the idea of ‘self-enlightenment’. Being an integral part of Buddhist heritage and a leading player in cutting-edge science, Korea demonstrates abilities to emerge as a new balancer to incorporate the best of science, spiritually and humanism to build next-generation AI machines with distinct human qualities.

Keywords: science; spirituality; humanism; self-enlightenment; artificial intelligence; Buddhism; Confucianism; Korea

1. Introduction

With the advent of the fourth industrial revolution, there seems to be a new contest brewing between men endowed with compassionate wisdom and machines, driven by artificially augmented ‘machine intelligence’ or Artificial Intelligence (AI).¹ A paper published in the journal Nature by researchers at Google’s “Deep Mind” have come up with a Differential Neural Computer (DNC), which is capable of generalized learning. The advent of the DNC has the potential to fundamentally alter the AI landscape. This development possibly can ignite a kind of “Hard Take-off” scenario espoused by Nick Bostrom in his seminal work, “Superintelligence: Path, Dangers, Strategies” (Bostrom 2014). Considering the path-breaking changes in AI, British cyberneticist Kevin Warwick’s book “March of the Machines: Why the New Race of Robots Will Rule the World” (Warwick 1997) has come up with a dark vision of the future. Echoing similar concerns, four leading scientists— theoretical physicist Stephen Hawking, cosmologist Max Tegmark, physics Nobel laureate Frank Wilczek, and computer scientist Stuart Russell—published an appeal in the British newspaper The Independent on 2 May 2014 cautioning us against the complacency of writing off AI machines as mere science fiction: “Success in creating AI would be the biggest event in human history. Unfortunately, it might also be the last, unless we learn how to avoid the risks.”

Indeed, AI is growing at a rate quicker than most of us can comprehend. Whether we agree with such advances or not, the influence of AI driven technological changes in society have given birth to what some prefer to call ‘technological singularity’—a hypothetical point when technological growth becomes uncontrollable and irreversible. The question arises as to where we go from here and how we handle rapidly accelerating advances in the
field of AI. Experts agree that this sort of complex challenge cannot be handled responsibly only by AI scientists. In order to carefully manage the multi-dimensional impact of AI, leaders in science, philosophy, and religion are required to agree to coordinate and work together, possibly in a formal institutional setting.

The need to create a wider interface involving inputs from science, philosophy and spirituality pushes us to look beyond the field of secular ethics, which is significantly weaker than religious ethics. Secular ethics is largely confined to non-binding moral philosophy whereas religious ethics derives ideas from the existence of God, which has a discernable binding power of morality. The God, understood to be morally perfect, is the source of moral values. If there were no God, then those moral values would have no force. In order to bring science, philosophy and spirituality into a common platform, we need the binding power of religious spirituality that can have effective moral constraints on increasingly uncontrolled AI.

Our search to find an appropriate spiritual-religious ethics having closer ties with scientific principles stops at Buddhism, particularly with its ideas of causal correlations explaining inter-relations between science, spirituality and humanity. These ideas are connected to numerous tasks performed by machine intelligence including learning, deducting, and reasoning. In fact, Buddhism has a lot to say about the nature of the human mind and human intelligence. “Buddhism conceives that the body is bigger than the brain and intelligence has wider application than just focusing on the problem-solving activities of the human brain.” (See Promta and Himma 2008, p. 117). It stresses on correlations between brain, mind and intelligence and believes there are sources of intelligence other than the human brain. According to a famous Thai scientist Ravi Bhavilai, human emotions must have a source other than the brain, and this source must be the mind. Buddhism teaches us that every human mind contains universal part that functions, so to speak, as a moral ‘spy’, which has the potential to guide AI in a more humanely acceptable path.

Having deep-seated roots in socially embedded Buddhism and being at the forefront of ongoing fourth industrial era changes, South Korea (hereafter Korea) is demonstrating unique potential to emerge as a key balancer in resolving growing apprehensions between men and machines. By looking back to its rich depository of Buddhist philosophical thought, specifically its emphasis on self-enlightenment embedded firmly in compassionate humanism, Korea seems to be successful in resurrecting individual consciousness, which was seriously undermined by the dominant Confucian collective solidarity. These enlightened and empowered individual scientists have acquired the vision to possibly produce cutting-edge AI powered machines with inbuilt human emotions such as compassion, mercy or guilt. A world leader in the semiconductor industry, Korea indeed has the potential to gain comparative advantage in the unfolding fourth industrial era by assimilating key tenets of the Buddhist configuration of the brain, mind and intelligence, and the core functions of artificial intelligence into one composite whole.

Ideas and arguments in this paper are categorized in the following manner. Section 2 explores the inter-linkages between science, spirituality, and humanity to examine how and why Buddhism works as a bridge among the three. Section 3 outlines Buddhism’s acceptance of scientifically acclaimed causal correlations, encouraging empirically validated investigations, which at the same time delineates its ability to maintain human emotions by stressing on ‘self-enlightenment’ and human compassion. This particular trait of Buddhism, combining intelligence and compassion, has the power to defuse the looming crisis between emotionally intelligent human beings and artificially intelligent machines. Section 4 analyzes how Buddhist ‘self-enlightenment’ that combines brain and mind can help AI to remove its stigma of ‘black box’, a syndrome where complex and layered calculations are inaccessible to their programmers, let alone end users. Section 5 shows how Korea, having developed its own brand of Buddhism, is likely to emerge as a ‘new balancer’ between ‘self-enlightenment’ and ‘artificial intelligence’. Section 6 concludes that Buddhist ‘self-enlightenment’ exhibits the promise to help build ethical–moral codes to restrain the unfettered drive of artificially intelligent machines.
2. ‘Science–Spirituality–Humanity’ Interface in Buddhism

A carefully organized fusion between spirituality, humanity and science places Buddhism in a distinct category of philosophical thought. More precisely, the Buddhist ‘principle of causality’ establishes a strong linkage between spiritual, human and scientific spheres. Uniquely placed Buddhist ‘dependent origination’ is based upon the principle of cause and effect, which sets forth a realistic framework for understanding the universe. It states that everything in the universe is conditioned, relative and interdependent. In other words, nothing occurs independently or by itself; all phenomena in the universe are in the conditioned state which only exist due to their supportive components. Without these inter-linked support systems, the phenomena will come to an end. Therefore, nothing in the universe is permanent and everything is in a constant state of change. As such the Law of Dependent Origination and the Law of Impermanence are intertwined and each supports one another. The ‘causal correlations’ encourage empirically validated investigations of reality pertinently expressed by Buddhism’s ‘twelve fold formula of causation’ embedded in the ‘wheel of life’ concept.

Since the 19th century, a number of scholars have argued that Buddhism is rational and uniquely compatible with science (McMahan 2004). Buddhist modernism is defined by figures such as Migettuwatte Gunananda, Anagarika Dharmapala, Paul Carus, Shaku Soen, D.T. Suzuki, Henry Steel Olcott and Edwin Arnold. These scholars have stressed that Buddhism accepts the scientific ideal of the universe ordered by natural laws. They see Buddhism mainly as a science of the subjective world, while science is mainly concerned with the external and material world. Some of Buddhist doctrines such as impermanence and emptiness have been compared to the scientific understanding of the natural world. According to the Encyclopedia of Buddhism, “What Gotama did was not to devise a law or formulate a system, but to discover a law, to preserve a system. His understanding may be compared to that of Copernicus or Galileo, Newton or Harvey, in physical science. Buddhism extends the natural laws, the laws of causality to the mental or psychic domain, or more exactly, perceives their operation in this sphere, and thereby disposes the idea of supernatural or transcendental agencies working independent of or in contravention to the natural laws of the universe (Singh 1996, pp. 47–48).” Stressing on the scientific reasoning embedded in the Buddhist methods of inquiry, the Dalai Lama points to causality and empiricism as common philosophical principles shared between Buddhism and science (Gyatso 2005). According to Jean Piaget’s ‘theory of cognitive development’ Buddhism is a fourth mode of thought beyond magic, science and religion (see Kress 1993; Tambiah 1990). A distinct scientific nature of Buddhist philosophy is revealed by the fact that Buddha kept his silence when confronted with ‘trans-empirical nature of things and confined himself to what is empirically given.

However, in the other religious traditions, spirituality and science remain at a distance. In Christianity and Islam, the first cause is divine, which is not related to any cause. Spiritualism has created its own faith-based inward-looking boundaries and sees scientific causality as an absolutely alien concept. Spirituality without science has embarked on a singular path promoting mysticism, superstitions, and inculcating anti-science attitudes. Though, on several occasions, faith-based religious explanations were challenged by the idea of causality. Similarly, science maintains a clear distance from non-scientific faith-based concepts. On this account, Buddhism differs sharply with other religions that exclusively follow faith-based explanatory verticals. In fact, causality articulated by Buddhism remains an exception. It can be said that there exists a wide gap between the reality perceived by Buddhism and the other religions.

A relentless progression in science has reached new highs with arguments supporting ‘autonomous machines’ that can make decisions and reach conclusions of their own, without waiting for any human inputs. At this juncture, a pertinent question comes: ‘how can Buddhist understanding of science particularly its ideas explaining brain, mind and intelligence enrich AI?’ Indeed, the Buddhist tradition of investigating the mind for more than 2500 years can do a lot to enhance the performance of AI. The real challenge is not if
machines can think and learn but whether they can think with compassion? The answer to this challenge lies in the age-old tradition of Buddhism to investigate the function of the mind and its inter-linkages with the brain. Despite extensive search by neuroscientists on the anatomy and functioning of the brain, there is very little understanding of the mind, thought, consciousness, intuition and intelligence. The core issue is that Artificial Neural Networks miss out on brain and mind translation. Buddhist learning can help to restore the missing link between the brain and mind. This way, the performance of Deep Learning AI, which mimics the biological brain, can be enhanced.

First time, a new AI, dubbed as a Differential Neural Computer (DNC), has achieved generalized learning and the ramifications could be truly seismic. Sensing the potential of AI in mimicking the biological brain, governments and businesses around the world are going to spend more than US $500 billion in 2023. Deep Learning was conceived in the 1940s but it was only a few years back that the intuitive cognitive (learning through experience) system using continuous mathematics complimented the classical AI based on rational (programmed) cognitive systems. This significant transition from the intuitive to the rational is the source of AI’s noteworthy achievement in the last half-decade. There are superhuman feats of AI in the field, such as board and strategy games, image recognition, natural language processing and the automation of cars.

Besides path-breaking changes in AI, some disillusionment is coming from the fact that Deep Learning employs heavy mathematics, thus only part of its high-performance solution can be explained rationally; the rest is instinctual, a diverging path from human intelligence. Due to AI’s monstrous data crunching, inscrutable black-box operations, its inability to apply abstract concepts to different situations, and difficulty in building models, there are growing apprehensions that Deep Learning is nearing a plateau. Despite Deep Learning algorithms requiring millions of training examples, using vast amounts of quality data, it is still far-short of human capabilities of reasoning, understanding and common-sense.

In order to surmount these growing pains, Deep Learning can take a leaf out of Buddhist wisdom (Hershock 2021; Hongladarom 2020). The practice of Deep Learning, comprising of inputs, assigning weights and biases, optimization and outputs, has a similarity to Buddhist wisdom tools of hearing, contemplation, stabilization (samatha), meditation and insight (vipassana). By integrating Buddhist learning, AI can use less data, minimize large mathematical operations, and combine the algorithms with human experience and intuition to solve problems like climate change and incurable diseases that humanity often confronts. However, the assimilation of Buddhist learnings into AI is a complex task that would require coordinated efforts of Buddhist scholars, neuroscientists and AI specialists to identify methodologies that can help Buddhist insights to be systematically incorporated into AI models.

3. Buddhist ‘Principle of Causality’ with a Sense of Compassion

The ‘principle of causality’ in Buddhism has evolved from its inception. Early Buddhist theory goes beyond the commonsense notion of causation, and recognizes a system whose parts are mutually dependent. This dependence has been termed as ‘dependent origination’ (paticcasamuppāda), which conforms to the definition given by Buddhaghosa, a fifth-century Indian Theravada Buddhist commentator, translator and philosopher. Eventually, Buddhist began to look into the several factors that cause an event to change. In the Abhidharmika schools, notably the sarvāstivāda, hetu and pratyaya became salient aspects of the rise (Karundasa 1996).

Buddhism has never promoted explanations exclusively based on faith or belief. Whenever an explanation of reality lacked valid empirical support, Buddha maintained total silence. By not responding to the queries raised regarding the ‘trans-empirical nature of things’, Buddhism has shown its adherence to causality and negation to any divine intervention as the first cause. Indeed, in a major departure from the faith-based understanding of the universe, Buddhism scrutinizes spirituality in an empirically sound manner and thus
makes its basic assumptions much more compatible with science. Buddhism denies the existence of a first cause in any form and rejects the existence of a permanent self (atman). Here, Buddhism breaks away from other religions, particularly Hinduism. The nature of existence in Buddhism is interpreted in terms of the twin truths: conventional and ultimate. Conventionally, there exist beings and things, but only as conceptual entities (prajñaptisat). Ultimately, they do not exist, because they have no permanent core. So how does the inexistent world function? Conceptual entities are dissected into impersonal phenomena or dharmas as ultimate units.

There are two important aspects of the Buddha’s discovery of the nature of dhamma or dharma; (1) ‘causation’ (paticasamuppāda); (2) ‘causally produced dhamma’ (Kalupahana [1975] 1986, p. 68). It is a distinction between the causal relation and the causally related. The problem of causation, therefore, involves two aspects: the rule or pattern according to which things change, and the things themselves that are subject to change. The nature of dhamma is closely connected with the theory of impermanence (anicca in Pali and anitya in Sanskrit) and momentariness or Ksanikavāda (Khanikavāda in Pali) originated from the Sanskrit words ‘Ksana’ and ‘vada’; when both appear together they can be interpreted as “theory of the momentariness”, that is, the manner in which a dhamma presents itself or has existential status (Kalupahana [1975] 1986, pp. 74–75). It is here that the concept of self-existence or own-nature (svabhava) enters. The sarvāstivāda postulated the four ways in which change takes place with respect to the dharmic analysis. This analysis is criticized by the Madhyamika as being untenable or inapplicable to the very concept of causality.

Where the Sarvāstivāda postulated the svabhava, the Saṅgrāntika rejected it and did not recognize a static moment; dharmas are constantly arising and disappearing. But they were still left with the problem of explaining the connection between two successive moments. The conclusion derived by Kalupahana ([1975] 1986) stresses that the causally produced dhamma (paticasamuppāda-dhamma in Pali or pratiṭṭhya samutpāda in Sanskrit) is an empirical phenomenon which includes the mental concepts (dhamma) as well. According to early Buddhism, there are no accidental occurrences because everything is causally conditioned or produced (paticasamuppānapptam). Due to Buddha’s discovery of the conditioned nature of things, Buddhist texts are replete with assertions to the effect that he who perceives the causally conditioned nature of things also perceives the truth (dhamma), and he who perceives the truth also perceives the conditioned nature.

By making an analysis of paticasamuppāda, as described in the Samyukta Nikāya (II, 26), it is synonymous with what he refers to as the causal nexus, which has four main characteristics: ‘objectivity’ (tathāta), ‘necessity’ (avitathāta), ‘invariability’ (anavihihāti), and ‘conditionality’ (idappacayatā). By understanding the characteristics of the causal nexus, one is able to understand the causal nature of the concept of paticasamuppāda or popularly referred to as the ‘Wheel of Life’. According to Kalupahana ([1975] 1986), early Buddhism did not simply accept the mere constant conjunction of two things or constant association of succession. The early Buddhists looked upon the occurrence of events by virtue of a plurality of causes. The later developments are on the sarvāstivāda and saṅgrāntaka views on causality, where the former reasserts a form of Satkāryavāda or the ‘identity theory’ subscribed to by the Samkhya system, and the latter, a form of asatkāryavāda or the ‘nonidentity theory’, is subscribed to by the Vaiśeṣika system. Finally, it can be said that in early Buddhist thought, the theory of causation was employed to explain all types of causation available in the world of experience, including nirvāṇa, however, in Madhyamika thought causality was employed to explain only the relativity of the phenomenal, the theory itself being considered transcendent.

Dharmas are momentary; they arise and vanish in space and time in conformity with definite principles that regulate their flow and interdependence: Karma and Pratītya Samutpāda. The term Karma, literally means “action” or “deed”, and denotes the principle of ethical causation: there are no agents, but there are actions and their consequences. Such actions accumulate and yield their fruits: particular body–mind configurations evolving in cyclic rebirths (samsāra). The principle of pratītya samutpāda (dependent origination),
denotes the conditionality or interdependence of existential phenomena. Essentially, it accounts for the conditioned flux of phenomenal existence, in particular the interdependent flow of the five aggregates with no ontological substratum.

The terms *hetu* (cause) and *pratyaya* (condition), occur as a compound or separately. As a compound, *hetupratyaya* denotes the principle of causes and conditions applicable to all aspects of existence. When included in lists of conditions, it denotes the first condition, the condition qua cause. Individually, they are virtually synonymous, or form either separate or correlated models of causality. In terms of soteriology, causality is integrated into the four noble truths. The first truth is *duḥkha*, or life is suffering. The second truth teaches the origin of suffering, identified by the Buddha as craving. Otherwise, the origin is interpreted in terms of *karma* and dependent origination. The third noble truth, the cessation of suffering, teaches the eradication of *karma*, leading to rebirths, and the cessation of suffering: appeasement of dependent origination. The fourth truth identifies the path which leads one away from craving and suffering.

Causality in Buddhism has twin dimensions—one explains reality scientifically based on ‘causal correlations’; the other aspect is deeply linked with compassion to others. Buddhism is known for its compassion—a human capacity to comprehend others’ grief and misery. Avalokitesvara is a bodhisattva who embodies *karunā*, which is important in all schools of Buddhism. For Theravada Buddhists, dwelling in *karunā* is a means for attaining a happy present life and heavenly rebirth. For Mahayana Buddhists, *karunā* is a co-requisite for becoming a Bodhisattva. Buddha’s journey to enlightenment begins when he sees an old man, a sick person and the people carrying a human dead body. He was able to comprehend others’ misery or *duḥkha*. The multidimensional causality, having scientific cause and effect mixed with compassion, has the potential to add a human dimension to artificial intelligence.

The missing human dimension of the AI is important in order to understand human imperfections, largely driven by the irrationality of the psycho-emotional world. Human emotions, such as compassion, guilt, and anger, often guide human beings to go against the so-called empirically validated decisions. It reminds us that human imperfections breed contradictory solutions leading to the problem of ‘moral dilemma’: whether to make decisions based on pure intelligence or listen to human emotions such as forgiveness, compassion, remorse, etc. The Buddhist understanding of perfection is linked to the brain and imperfections tied to the mind can help AI to resolve long-standing moral dilemmas between intelligence and emotions.

### 4. Self-Enlightenment, Machine Intelligence and the Moral Dilemma of AI Fraternity

A Buddhist ‘principle of causality’ provides an empirical as well as emotional foundation following which one can attain ‘self-enlightenment’, a higher state of awakening. The idea of self-enlightenment in Buddhism incorporates elements of intelligence and compassion (Fussell 1995). Bodhi or ‘wisdom’ (*prajñā* or *jñāna*) in Buddhism are different from ‘knowledge’ which is much more related to information, but ‘wisdom’ is deeply linked to intelligence with human compassion. Buddhism has accorded the highest priority to ‘compassion’, which has been a major theme throughout much of Buddhist history. Compassion originates from a sort of self-love which develops to a love towards others. In fact, *karunā* is usually translated as compassion—a wish for others to be free from suffering and the causes of suffering. In Theravada Buddhism, *karunā* is one of the four “divine abodes” (*brahmavihara*), along with loving kindness (Pali: *metta*), sympathetic joy (*mudita*) and equanimity (*upekka*). Buddhist karunā even argues that the feeling of compassion should not be limited to human beings alone but it should encompass the animal world too. This overpowering idea of compassion to others has led Buddhism to denounce violence in all forms. The Buddhist emphasis on human compassion has the potential to complement the missing human dimension in the current generation of AI machines.

Buddhist philosophy teaches us that there are at least three major sources of intelligence: man, community, and the world as a whole. It believes that the multifaceted
intelligence of the mind directs us to be something more than biological robots. Considering the multidimensionality of the mind, it is unclear whether AI machines with social awareness and intelligence would care about the well-being of others, both machines and human beings, and if this is even theoretically possible. Nevertheless, Buddhism raises the possibility of helping AI in creating a machine with the intelligence of the universe. Among other things, Buddhism is a study of mind. Scholarly work on developing a framework of mind derived from Buddhist philosophy can be used for building an AI model with a human face. The five-aggregate models in Buddhism—where activities of a subject with a mind are classified into five aggregates: physical matter and body (Rūpa), feeling or affect valence (Vedanā), perception or cognition of conceptions (Saññā), volition (Saṅkhāra in Pali and Saṁskāra in Sanskrit) and phenomenal consciousness (viññāna in Pali and Vijñāna in Sanskrit)—can serve as a guiding framework of AI studies.

In addition, a ‘strong AI’ debate on consciousness can further gain insight from Buddhist philosophy. “Buddhism as a religion is unique in its ability to consider life as not uniquely human but also animal, and even machine” (Kaul 2022). When asked if robots could ever become sentient beings, the Dalai Lama replied: ‘if the physical basis of computer acquires the potential or the ability to serve as basis for a continuum of consciousness a stream of consciousness might actually enter into a computer’ (quoted in Kaul 2022). Clearly extending the possibility of life beyond humans, the Dalai Lama envisions the prospect of machine cognition and machine life. The Buddhist philosophical approach seems to be useful in creating compassionate AI and selfless robots. In order to model emotions, models of the mind are essential and the Buddhist traditions distinguish different versions of compassion through four sublime states of mind: metta (love and loving-kindness, karuna (compassion), mudita (sympathetic joy), uppekkha (equanimity) and virtues (in Sanskrit: Pāramītā and in Pali: Pārami), which can assist in building robotic character and its capacity to learn through interactions. As Buddhism imagines humans to be more than mindless machines, this can be a signpost for machines that may have a mind.

Currently, AI technologies are dominated by intelligent machines capable of thinking, learning and attempting autonomy in decision making. However, in order to achieve superior performance, AI should make machines that can mimic human intelligence. Indeed, the first phase of AI was symbolist and relied on rules, logics and symbols and found solutions that humans could understand. However, in the past few years unsupervised machine learning has taken precedence. In other words, in contrast to earlier era ‘weak or narrow AI’ that has been focusing on specific tasks, we see the emergence of ‘strong AI’ capable of thinking and acting like humans, but the coming ‘super AI’ may even exceed human intelligence. In the newest and remarkable development, Deep Neural Networks (DNN) use massive amounts of data to learn and make predictions, and thus inching quite close to the human brain.

However, the over-dependence on data raises two significant issues related to biases and explanations: one, a system trained on human data AI can import human prejudices and biases, particularly in the context of race and gender; two, the reasoning of DNN is in ‘black-box’—based on complex and layered calculations that are inaccessible to humans, so their own programmers let alone the end users are unable to have an explanation for such automated decision-making. A question here arises—‘when something does not go right, who is responsible? This has led to the need of developing ‘explainable AI or XAI’, since explaining decisions is critical for accountability and trust.

It is crucial for the AI fraternity to tackle the moral dilemma originating from the emotionless side of the fast-approaching machine age. There is a danger that AI machines can ultimately prove human beings as incompetent and inefficient and thus worthy of only subordinate status. In the absence of any agreeable resolution to the emotionlessness of AI machines, public opinion would not support building an emotionless world controlled by only artificially intelligent machines. For instance, the control of nuclear weapons requires not only an efficient and intelligent decision to launch an attack on the opponent but the capacity to foresee the possibility of impending destruction of humanity from the planet.
earth; hence, the willingness to reconsider decisions by factoring human emotions and opening various channels of dialogue is essential.

The future willingness of humanity to allow unrestricted growth of AI technologies would depend on how efficiently it combines emotions and artificial intelligence into a composite new approach. The larger question is how Buddhist spirituality can assist in integrating the twin spheres of emotion and intelligence in the AI framework. Analysis here emphasizes that Buddhism is the fourth thought beyond magic, religion and science, which has the potential to induce human compassion and create a dedicated moral–ethical compass for AI machines. By adding a Buddhist ethical–moral dimension to AI machines, it is possible to develop a kind of ‘guilt sensor’ that could possibly turn off machines facing moral dilemmas. With the right mix of intelligence and compassion these sensors can provide a human face to artificially augmented machines. The multi-dimensionality of Buddhist causality can help in conceptualizing content for these sensors to guard against the possible brutality committed by emotionless but artificially intelligent machines.

Moreover, Buddhist causality is important not only to enlighten AI machines but the wider scientific community that is at the helm of promoting cutting-edge AI. The majority of research on AI is funded by military-related establishments which are focused on designing and manufacturing AI machines mainly for their use in battlefields (Scharre 2018). Researchers involved in these high-tech projects are mostly blind-folded by certain biases and preconceived notions that total military conquest over the enemy is the only viable option available. They are not aware of the conquest by other means such as ideas, social values and cultural norms. These researchers have hardly any time or capacity to contemplate with compassion about human suffering and misery that conflict zones produce. In this context, the Buddhist ideas of compassion based on multi-dimensional causality can teach these scientists a few lessons about the moral–ethical side of their decisions. This understanding can be crucial in conceptualizing next generation AI machines.

5. Between ‘Human’ and ‘Machine’ Intelligence’: Korea Emerging as a New Balancer

How can the challenge to embed two distinctly different dimensions of intelligence—one, ‘human’ rooted in the spiritual-religious ethics and the other, ‘machine’ embedded in the next-generation AI—be tested in a real country case? In order to test such a complex inter-relationship between human and machine intelligence, we need a country that has deep-seated spiritual religious ethics and also a key player in the emerging field of artificial intelligence. Given the legacy of socially embedded Buddhist philosophical tradition and having a widely recognized leadership in high-technology sectors like robotics, automation and AI, we have selected Korea as an appropriate case to study mutual exchange between human and machine dimension of intelligence.

A justification for not selecting Asia’s other Buddhist countries with acknowledged AI prowess such as Japan or China as a case study, instead solely focusing on Korea, lies in the varying levels of social embeddedness that Buddhism in Asia has accomplished. Due to distinct historical circumstances, unlike Japan and China, Korean Buddhism stood up twice to national cause, once during the Japanese invasion and again when Japanese colonialism attempted to undermine Korean identity. In order to preserve national identity, Korean Buddhism tried to create doctrinal harmonization by laying the foundation of ‘inter-penetrating Buddhism’ or Tongbulgyo, taught by the Korean monk Wŏnhyo (Ch˘o 1989). Walking along with the collective resolve of society, Korea witnessed the rise of a distinct brand of Buddhism, which is called Minjung Bulgyo or ‘Buddhism for people’ (Mun 2014). This distinct connect with the wider society gave rise to ‘doing Buddhism’ in Korea, which gradually transformed into ‘socially engaged Buddhism.’ A distinct social embeddedness of Korean Buddhism helped the nation to make a smooth transition from the ‘collective consciousness’ skillfully fermented by age-old Confucianism to the ‘individual consciousness’ that Buddhist philosophy clearly articulated. The Buddhist ‘self-enlightenment’ helped Korea to restore ‘individual conscious’, which is powering its transition to the fourth industrial era. On the contrary, Buddhism in Japan and China
could not develop a Korea-style social connect and thus failed to deconstruct their age-old collective consciousness.

The experiences of Korea in utilizing a Confucian collective consciousness based on conformity and hierarchy to succeed in late-late industrialization and then attempting to re-embark Buddhist individual consciousness are crucial to test the core premise of this study—‘How to embed Buddhist idea of compassionate intelligence in building next-generation AI machines?’ In order to upgrade its competitiveness in the fourth industrial era, Korea is trying to conceptualize and build AI machines that may not only perform the complex tasks of learning, thinking and executing but that also carry out a given task in a compassionate manner. Given the US technological dominance and China’s aggressive AI strategy, many experts see the possibility that Korea can use its ‘socially engaged Buddhism’ to leapfrog in the ever intensifying AI race.

A careful analysis of Korea’s experiences in its first industrial transformation with relative efficiency points to a broad base transition from the industrial to post-industrial era. In the newest techno-economic shift, Korea’s age-old Confucius ‘collective conformity’ that once powered the country’s mass-manufacturing-led industrial juggernaut is now facing stiff resistance. Scholars even recommended the ‘creative destruction’ of Korea’s ‘collective consciousness’ (Uttam 2011). With the ongoing paradigm shift in the patterns of industrialization, Korea needs to master knowledge-intensive high-tech industries requiring individual ingenuity and creativity. Advances in fourth industrial era technologies such as robotics, automation and AI are deeply embedded in scientific inquiry and experimentation, based on the inputs provided by the creative impulse of individual scientists.

In order to assess Korea’s potential to surmount twin transformations from an agrarian to industrial and an industrial to post-industrial era, we need to first look into Confucian collective conformity that played a key role in mobilizing and disciplining its industrial workforce. Hardworking and disciplined human resource has been credited with powering Korea’s first transformation from an agrarian to industrial era. It has been argued that during the era of fast-pace industrialization, Korea concentrated on labor-intensive, mass-manufacturing industries based on the imitation of foreign technologies through a well-thought policy of ‘reverse engineering’ (L. Kim 1997; Uttam 2006). In order to mass-produce goods and services for its export-oriented industries, Korea was able to form a ‘developmental alliance’ that facilitated a wider interest convergence between and among state, business and conservative political classes (E. M. Kim 1998; Hundt 2009; Uttam 2014). This alliance was able to utilize the nation’s collective consciousness flowing from Confucian thought to propel the wheels of industrialization.

By invoking Confucian ideas of collective consciousness under rigid hierarchy, Korea constructed a unique but informal managerial hierarchy that successfully organized teams accomplishing collective duties. In this collective social vision, decision-making was only done through achieving wider consensus. With the authorization of collective consensus, Korean management ensured employees worked exceptionally long hours. This informality in the corporate governance made Korea retain the number one position among the OECD countries when it comes to office hours devoted by employees. “Each employed Korean worked 1908 h a year in 2020, 221 h more than the OECD average of 1687 h. Korea’s annual working hours was 141 h longer than in the US, which came at 1767 h per annum, and 310 h longer than Japan’s at 1598 h” (Korea Herald, 17 August 2021) (Y.-s. Kim 2021). The new recipe, fermented by low wages and long working hours, immensely contributed to producing a “Miracle on Han River.”

Korea’s rapid economic development does not have only a glittering side but hides a dark side as well. The devastating financial crisis of 1997 came as a grim reminder that conservative ‘developmental alliance’, which used Confucian collective consciousness to discipline the workforce and suppress wages, may no longer bring prosperity to Korea. It was clear that Korea needed to move out from decades of ‘reverse engineering’ of foreign technologies and lay the foundation of its real techno-scientific advancement. The
debates following the 1997 financial crisis generated an extraordinary consensus seeking the ‘creative destruction’ of hierarchy and its associated collective consciousness. Korea launched a multi-billion dollar ‘Brain Korea’ project to promote and consolidate individual excellence and creativity (Uttam 2012). It sought the empowerment of the individual self and the protection of its autonomy to augment Korea’s dormant entrepreneurial spirit and creativity, in order to support its knowledge economy increasingly relying on automation, robotics and AI powered machines.

In the post-1997 financial crisis, Korea’s three ‘progressive’ administrations under the leadership of Kim Dae-Jung, Roh Moo-Hyun and Moon Jae-in worked hard to unravel the basic tenets of country’s old political economic order based on collective solidarity to mobilize cheap labor for its sprawling mass-manufacturing industries. The ‘progressive’ administrations introduced policies aimed at recasting their developmental priorities. Responding to the swift turn of events, the Korean state tried to promote ‘bottom-up capitalism’ led by SMEs and tech-intensive start-ups. It was the first time in Korean developmental history when large Chaebol groups were not the key priority of its powerful state system. Instead, a new ecosystem focused on individual entrepreneurship and tech-heavy start-ups became the new mantra for the post-1997 crisis Korea.

Sensing the momentum building up towards the fourth industrial era, even conservative Korean president Park Geun-hye launched initiatives aimed at expanding ‘creative industries’ by promoting the individual self and discouraging old-era collective conformity. In a sensitive political environment in Korea, no one openly talks about ending Confucian collective consciousness and adopting the Buddhist enlightenment of the individual self but in reality, a process of ‘creative destruction’ of collective consciousness and restoration of individual consciousness is well underway in Korea.

Increasingly, it is becoming obvious that Korea’s push towards an AI-powered fourth industrial era requires a new social contract that can unchain individual consciousness from the rigid Confucian conformity and restore its autonomy. The changing needs of the Korean techno-economics demand a careful utilization of creativity and imagination of individual self to excel in the fundamental sciences aimed at building a firm base in the cutting-edge technological areas. Against this backdrop, Buddhist emphasis on self-enlightenment emerges as a new mantra for revitalizing Korea’s long-dormant individual consciousness, which could empower a generation of scientists willing to infuse self-enlightenment in scientific experimentation aimed at developing fourth industrial era technologies.

Carefully resurrected individual consciousness in Korea helped its tech industry, led by giants such as Samsung, Hyundai, and LG, to experiment with AI; nonetheless, it has demonstrated only mixed commercial success. However, a sense of urgency took hold in 2016 when AlphaGo, an AI program developed by Google’s DeepMind Technologies, defeated a Korean professional Go player with a 9 Dan rank, Lee Sae-dol, at a traditional Chinese board game in Seoul (Xu 2019). This event came as a shock to Korea and thus prompted government to set-up AI research institutes and outline an ambitious national AI strategy (Ministry of Science and ICT 2019). In addition to state support, corporate Korea greatly expanded its in-house AI divisions. These efforts received further encouragement by the country’s empowered scientific community. The Korean government’s concerted attempts culminated into conceptualizing and producing ExoBrain, a Korean equivalent to IBM’s Watson, which was developed by the Electronics and Telecommunications Research Institute (ETRI), a government-funded research institute located in Daedeok Science Town in Daejeon city. A few years back, ExoBrain was able to beat two quiz champions and two collegiums. The victory of ExoBrain is a milestone in Korea’s AI research and development. Not only has the state been involved, but Korea’s corporate sector has come forward to enhance its AI capabilities. Samsung is working to create AI semiconductors by 2030.

Korean society’s profound knowledge of Buddhist self-enlightenment emanates from wide-spread acceptance that Buddhism is primarily a philosophy, not a religion in its truest sense. Buddhism classified as a philosophy allows the possibility of philosophical criticism without unconditionally believing the truth that Gautama Buddha discovered.
The belief that Buddhism is a philosophy and it has promoted the concept of ‘doing philosophy’ which involves wider society to continuously engage with Buddha’s teachings. ‘Doing philosophy’ in Korea is rooted in the society in the same way as the practice of Buddhist meditation (Pak 2017). Due to wider social acceptance, Buddhism has been incorporated into Korean traditional culture. As a result, many people outside the practicing population are deeply influenced by Buddhist teachings (Sungtaek 2002). The Korean scientific community as well is exposed to ‘doing philosophy’ and learned ‘practicing philosophy’ by participating in popular Buddhist meditation drills. Additionally, there have been attempts to make Buddhism ‘socially engaged’ by tackling problems of the environment and deeper involvement in social welfare activities (King 1994). Historically, under the intense pressure from Confucianism and rapid expansion of Christianity, Korean Buddhism successfully embedded its teachings in the wider society. To this end, Korean Buddhism initiated the harmonization of different strands within Buddhist philosophy into ‘inter-penetrating Buddhism’ that provided a unified spiritual-philosophical interface to tackle problems that Korea has been facing.

The meeting of autonomous individuals with Buddhism’s scientific outlook unleashed Korea’s creative energy. A well-thought creative freedom empowered by Buddhist enlightenment has contributed to the research of a new class of AI machines that possibly can think rationally and act compassionately. The above discussion points to the fact that Korea’s success in the fourth industrial era lies in its capacity to re-embrace Buddhist ideas of self-enlightenment. Individual scientists enlightened by Buddhist philosophy have acquired the potential to unravel brain-mind configurations focusing on compassionate intelligence. How much Korea becomes successful in restoring individual consciousness is seemingly commensurate to its comparative advantage in creating a new generation of AI machines.

6. Conclusions

The findings of this study confirm how the uniqueness of Buddhism that carefully integrates spiritual, human and scientific spheres through its ‘principle of causality’ helps in acquiring itself a central role in the discourse regarding next-generation AI machines. The Buddhist ideas are impacting AI in two ways: one, through brain and mind-induced human intelligence, which is incorporated into DNN-assisted generalized learning; two, ‘causal correlations’ defining ‘self-enlightenment’, which is responsible for the shaping of individual personalities. The enlightened individuals are potentially becoming scientists capable of conceptualizing and designing AI machines with human qualities that possibly could comprehend others’ suffering and pain.

Given Buddhism’s unique mediating position between spirituality, humanity and science, it is safe to argue that any satisfactory resolution to the emotionlessness of science lies in its concept of ‘enlightenment’ that combines both ‘intelligence’ and ‘compassion’ in an empirically sound manner. This study validates our basic premise that Buddhist spirituality deeply integrated with the scientific temper has the potential to contribute in developing an appropriate moral–ethical compass in the form of ‘guilt sensors’ that can create effective checks and balances in the functioning of AI machines. Moreover, Buddhism’s moral–ethical insight can also be introduced to the larger community of scientists to make them aware of the gravity that the new machine age can impose on humanity.

A thorough investigation of the Korean case confirms how its scientific community, well-acquainted with Buddhist philosophical tradition, is deriving an added edge to meticulously tackle the challenges emerging out of dichotomous relations between artificial intelligence and human compassion. The Korean society’s accumulated knowledge of Buddhist thought, operationalized by ‘doing philosophy’, and its wide-spread practice of mediation is empowering its scientific community to embed human compassion with AI research. It is helping Korea to assume leadership in building AI machines that can potentially have, for instance, a ‘Buddha Eye’, capable of sensing other’s misery. This unique
advantage in building emotionally compatible AI machines can place Korea to lead the fourth industrial revolution from the front. In order to achieve a pole position in the rapidly evolving AI research, Korea to some extent has accomplished the ‘creative destruction’ of its own Confucian collective consciousness, which was perfected to promote ‘group logic’ to power its mass-manufacturing industries. On the ruins of collective consciousness, Korea has promoted autonomous and enlightened individual self that Buddhist philosophy always argued for. The birth of an empowered, autonomous and compassionate individual in Korea has the potential to easily embrace scientific temper with human face.

An empowered and enlightened individual is Korea’s new link to rapidly evolving technologies critical for the success in the fourth industrial era. Certainly, many believe that Korea’s first transition from an agrarian to industrial era was mediated by society’s collective consciousness that Confucian values taught; however, the nation’s ongoing second transition from an industrial to post-industrial or fourth industrial era requires a new intellectual framework that scientifically validated, unorthodox ideas of Buddhism can provide. A significant move from Confucian ‘collectivism’ to Buddhist ‘individualism’ is helping Korea to lay the foundation of a new understanding of human intelligence. In addition, the Buddhist conception of human intelligence is also supporting a community of scientists that can infuse human compassion into scientific reasoning. A paradigmatic shift to re-embrace Buddhist teachings focused on shaping autonomous and enlightened individual is empowering Korea to strike a new balance between spirituality, humanity and science. This rebalance in Korea can positively contribute to the development of AI with a human face.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data sharing not applicable.

Conflicts of Interest: The author declares no conflict of interest.

Notes
1 Artificial intelligence is a type of technology that performs tasks related to human-intelligence including learning, deducting and reasoning. This branch of technology is programmed to continuously self-improve by generalized learning through crunching massive amounts of data and acquiring knowledge through experience of what works.

2 A direct email communication with author Somparn Promta revealed that Professor Ravi Bhavilai was a physicist teaching at the Department of Physics at Thailand’s Chulalongkorn University. Quoted in (Promta and Himma 2008, p. 178).

3 Causation is regarded as fundamental to science. Hence, Buddhist ‘principle of causality’ which refers to the ‘twelvefold formula of causation’ embedded in the concept of ‘wheel of life’ creates a close proximity with science. Discarding faith-based analysis of the questions pertaining to reality around us, Buddhism took a step forward from its age and made it compatible with today’s science age.

4 In Buddhist thought ‘Wheel of life’ concept explains the movement in life in the form of ‘causal correlation’ producing ‘constant association of succession’.

5 Several contemporary scholars of Buddhism have argued against the idea that Buddhism is compatible to science. Lopez argues that attempts to make Buddhism compatible with science severely restrict Buddhism. Evan Thompson has also criticized the narrative of Buddhism as being uniquely scientific. For details, see (Lopez 2008; Thompson 2020).

6 Causality has found expression in the philosophical debates ranging from Aristotle to Hume. Aristotle is justifiably the single greatest contributor to the theory of causality in Western philosophy. His articulation of the four types of causes—material, efficient, formal and final—forms the foundation of modern conceptions of causality. These philosophical arguments were able to somewhat challenge faith-based explanations of reality. For details, see (Olen 1983; Beauchamp 1974; Wallace 1974; Hankinson 1998).

7 More on discouraging abstract theorizing, see (Organ 1954).

8 In the Mahayana tradition it is one of a suitable of terms employed to denote the Buddha-nature, such as “gotra”.

9 Causality in Buddhism refers to conditions created by a plurality of causes that necessarily co-originate phenomena within and across lifetimes, such as karma in one life creating conditions that lead to rebirth in one realm of existence for another lifetime.
 Assertions such as ‘Miracle on Han River’ indicate Korea’s rapid economic growth phase spanning from 1961 to 1997. During that period, Korea announced its National Strategy for AI in 2019 and made AI a key element of its Digital New Deal initiative. Judging from the nearly 6000 AI-related patents that have been filed by Korean companies over the past 12 years (over 3000 of which were registered by Samsung Electronics alone) the AI market in Korea is flourishing. With an investment of over $2 billion in AI research and training, South Korea’s highly publicized AI strategy aims to make the country among the top 4 contenders in AI. The Key AI players in Korea include Samsung Electronics, Hyundai Motors, SK Telecom, KT, LGU+, Naver and Kakao.

The idea of ‘collective conformity’ in Korea flows from the Confucian worldview that sees collective consciousness as a binding factor over individual consciousness. In the contest between collective and individual consciousness, Korea in recent times has witnessed the persistent waning of collective solidarity.

‘Individual consciousness’ in Korea flows from the Buddhist idea of ‘self-enlightenment’, which argues that an individual should be directed by the light or knowledge of his/her inner-self. When an individual is at the centre, its uniqueness can find ways to express creatively.

‘Developmental Alliance’ refers to convergence of interest among and between dominant actors such as state, business, conservative society, and bureaucracy to forge consensus regarding the nature & direction of national economic development.

 eats any form of ‘Ahimsa’ or non-violence is part of its five precepts. See (Harvey 2000).

Intelligence encompasses a range of abilities that are associated with theoretical and practical rationality. It has always been linked with the mind and conception. The general assumption is that we cannot understand something without first conceptualizing it. We always breakup a complex problem into sub-problems, build a model in our minds, and then use that model to solve a problem. There are various types of intelligence, including analytic intelligence and emotional intelligence etc.

Korea announced its National Strategy for AI in 2019 and made AI a key element of its Digital New Deal initiative. Judging from the nearly 6000 AI-related patents that have been filed by Korean companies over the past 12 years (over 3000 of which were registered by Samsung Electronics alone) the AI market in Korea is flourishing. With an investment of over $2 billion in AI research and training, South Korea’s highly publicized AI strategy aims to make the country among the top 4 contenders in AI. The Key AI players in Korea include Samsung Electronics, Hyundai Motors, SK Telecom, KT, LGU+, Naver and Kakao.

The idea of ‘collective conformity’ in Korea flows from the Confucian worldview that sees collective consciousness as a binding factor over individual consciousness. In the contest between collective and individual consciousness, Korea in recent times has witnessed the persistent waning of collective solidarity.

‘Individual consciousness’ in Korea flows from the Buddhist idea of ‘self-enlightenment’, which argues that an individual should be directed by the light or knowledge of his/her inner-self. When an individual is at the centre, its uniqueness can find ways to express creatively.

‘Developmental Alliance’ refers to convergence of interest among and between dominant actors such as state, business, conservative society, and bureaucracy to forge consensus regarding the nature & direction of national economic development.

Asserts such as ‘Miracle on Han River’ indicate Korea’s rapid economic growth phase spanning from 1961 to 1997. During this period, Korea registered close to double digit growth and thus the country successfully accomplished its first transformation from agrarian to industrial era.

By introducing the ‘Brain Korea’ project, the Korean state acted in the aftermath of a crippling financial crisis in 1997 to restore individual creativity and experimentation by directing substantial monetary incentives to scientists and researchers.

References


Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.