Pegem Journal of Education and Instruction, Vol. 15, No. 4, 2025 (pp. 3844-3858)

RESEARCH ARTICLE WWW.PEGEGOG.NET

The Philosophical and Ethical Foundations of Artificial Intelligence

Akram Belkhiri [1], Iliyes Lahri [2]

[1] Mohamed El Bachir El Ibrahimi University, Bordj Bou Arreridj, Algeria
Email: akram.belkhiri_1@univ-bba.dz ; https://orcid.org/0009-0005-1562-796X

[2] Mohamed El Bachir El Ibrahimi University, Bordj Bou Arreridj, Algeria
Email: iliyes.lahri@univ-bba.dz ; https://orcid.org/0009-0000-8455-3179

Received: 12/05/2025 ; Accepted: 24/10/2025 ; Published: 13/12/2025

Abstract:
Philosophy plays an important role in the development of artificial intelligence through the nature of the questions it raises about the achievements it has realized and those it may realize in the future, as well as through the issues and trends that have emerged as a result of differing viewpoints regarding the nature of the thinking involved in artificial intelligence and the extent to which this field can advance. This study aimed to define the conceptual framework of artificial intelligence, as it is a newly emerging concept that has required tracing the stages of the emergence of this specialization, which is witnessing rapid developments and innovations at all levels. This has led to the formation of a dialectical tendency between supporters of this emerging technology and opponents who fear the stripping of human beings of their humanity.
The study also revealed the necessity of subjecting machines to ethical constraints and equipping them with ethical algorithms, through addressing machine ethics and artificial intelligence ethics, as well as human–machine interaction, on the grounds that the world is living a new phase of scientific progress referred to as the fifth scientific revolution. After the emergence of the metaverse, humanity is on the verge of the emergence of the post-human, and the effects these transformations raise on human moral and psychological behavior.
Keywords: philosophy, artificial intelligence, consciousness, ethics, ethical risks.

Introduction:

In recent times, the world has witnessed enormous revolutions in the field of technological development. These revolutions become even more intense and dangerous when they are related to emerging technological advances, especially those associated with artificial intelligence technologies and digitalization. Human beings thus find themselves facing a transformation and a shift from the traditional concept of technology—summarized by scholars as an attempt to make the world a small village—to a modern concept that further miniaturizes the world and challenges the initial definition. Contemporary humans have come to view technology and digital intelligence as having turned the world into a digital unit—far removed from the notion of a village—and transformed data and information into condensed digital units capable of competing with humans in intelligence and characteristics. There is no disagreement that artificial intelligence today plays an effective role across all disciplines, possessing numerous capabilities such as distinguishing between ideas, recognizing individuals' states, facial recognition, and an enormous capacity for prediction and knowledge.
All these modern characteristics that artificial intelligence has come to possess were initially intended to make human life easier and more convenient across various fields. However,

there are those who take an opposing stance, viewing these developments as a source of fear and serious risk regarding the ability of artificial intelligence to violate human privacy and strip humans of their humanity through breaches of personal data and rights related to voice and image. This development goes even further by comparing artificial intelligence with human intelligence and raising the possibility of AI surpassing human intelligence. This has caused some to feel anxious about the future of humanity, leading to the question of whether the spell could turn against the sorcerer and artificial intelligence might escape human control, especially given that there are many fundamental processes that indicate the possibility of such an outcome.

In this era, humans have moved from imitating and simulating other creatures to imitating themselves and simulating their own faculties. After once seeking knowledge for its own sake, humans now pursue what is practical and utilitarian, with little concern for ethical or moral values and little interest in highly abstract theories, instead relying on realism in every cognitive endeavor. As a result, humans have become subordinate to the machine, deprived of freedom in decision-making. At this precise moment, it has become necessary to activate the critical role of philosophy, based on the premise that every science has its own philosophy. Accordingly, this paper raises questions about the most important philosophical foundations upon which artificial intelligence is based, the role of contemporary humans in the face of these developments, and whether they can preserve their moral values. Can artificial intelligence possess all human faculties and perform their functions with the same effectiveness and mastery?

First: Definition of Artificial Intelligence:

Until recently, the theories of relativity and quantum mechanics represented a comprehensive philosophical revolution, revising all previous concepts and breaking the framework of Newtonian physics. However, these theories have now become conventional, having fulfilled their historical role. What humans once considered imaginary and impossible to achieve has now become a reality. For example, thinking about flight during the medieval period was considered mere fantasy, yet today it has become a primary means of transportation and entertainment. Moreover, humans have become increasingly attached to matter and machines, employing them to perform all their tasks. Machines have mastered every craft and industry, and they are now capable of rivaling their original creators, especially since they have begun to emulate human mental and intellectual abilities. This necessitated, in this research paper, presenting a concept of artificial intelligence.[1]

Artificial intelligence is defined as "that branch of computer science concerned with creating machines capable of performing intelligent tasks." Intelligence here refers to any activity that requires humans to use their mental abilities, such as translating a text from one language to another, conducting a medical diagnosis, or solving a specific problem. From this perspective, the role of researchers in the field of artificial intelligence is to simulate human intelligence by designing and developing software that enables machines—such as computers—to handle various situations intelligently.[2]

Artificial intelligence is also defined as the science that seeks to understand the nature of human intelligence through the creation of computer programs capable of simulating human behavior; that is, solving problems and making decisions in diverse situations. The program can determine the appropriate solution by relying on multiple pre-programmed reasoning processes, in addition to its ability to learn from previous mistakes. From this definition, it can be understood that artificial intelligence relies on entering specific data into the computer in the form of

[1] Badawi, Abdel Fattah Mohamed. Philosophy of Science: Science and the Future of Humanity—Where to? Quba Publishing House, Saudi Arabia, 1st edition, 2007, p. 11.

[2] Thagard, Paul. Computational Philosophy of Science. The MIT Press, London, 1988, p. 2.

commands processed with simple responses such as "yes" or "no," which makes some of its operations patterned and automated in nature. [3]

Intelligence is defined as "the science that studies how to make computers perform tasks that humans do, but in a better way than humans." [4]

The previous definitions agree that artificial intelligence represents a simulation of human thinking methods and the way humans utilize their acquired experiences in various fields, in addition to mimicking their abilities in understanding different languages and recognizing images. Technological developments have led to the creation of methods and programs that make the computer a machine capable of performing tasks characterized by intelligence and human expertise. All artificial intelligence systems rely on patterns of logical thinking similar to those used by humans, although they may differ in degree or method.

Today, artificial intelligence is considered the latest achievement of the human mind in the field of computer science, as it aims to enable machines to think in a way similar to human thought and solve problems with capabilities that may exceed human ability, through a set of pre-programmed reasoning processes

Second: The Origin of Artificial Intelligence:

Machines have demonstrated their ability and superiority over humans in performing tasks that require significant physical effort, such as lifting heavy objects or moving from one place to another. Nevertheless, many tasks remain exclusive to humans due to their possession of intelligence and reasoning abilities. From this perspective arose the challenge embraced by artificial intelligence researchers, which is the attempt to replace humans with intelligent machines in fields that require mental and cognitive abilities, just as machines previously replaced humans in areas requiring physical strength. [5]

The idea of creating an intelligent being or machine capable of performing human mental tasks has existed since ancient times. However, John McCarthy believes that serious scientific research in this field actually began after World War II, when several researchers independently started working on developing intelligent machines. Alan Turing (1912–1954) and the American mathematician Claude Shannon (1916–2001) are among the earliest researchers in this field. McCarthy points out that Turing published an article titled *Computing Machinery and Intelligence* in 1950, while Shannon discussed the idea of programming a mechanical man to play chess in the same year. However, McCarthy believes that Turing was ahead in proposing the idea of artificial intelligence, as Turing had delivered a lecture on artificial intelligence in 1947. [6]

In the late 1950s, several researchers emerged in the field of artificial intelligence, among the most prominent of whom were John McCarthy, Marvin Minsky, Allen Newell, and Herbert Simon. Shapiro referred to them as the "Fathers of Artificial Intelligence" in recognition of their pioneering contributions. McCarthy and Minsky established an artificial intelligence laboratory at the Massachusetts Institute of Technology, while Newell and Simon created another laboratory at Carnegie Mellon University. Later, McCarthy moved to Stanford University, where he established a new laboratory. According to Shapiro, these three universities, along with the University of Edinburgh in the UK, where Mitchie established a dedicated AI department, represent the leading centers in this field.[7]

However, historians agree on dividing the development of artificial intelligence into four stages:

---

[3] Bonnet, Alain. Artificial Intelligence: Its Reality and Future. Translated by Ali Sabry Farghly et al., Alam Al-Ma'rifa, Kuwait, 1993, p. 11.

[4] Belkacemi, Manal. Artificial Intelligence: Shaping the Future, Vol. 1. Dar Al-Ta'leem Al-Jam'a'i, Alexandria, 2019, p. 13.

[5] Mccarthy, J., What is Artificial Intelligence?avaliable at URL: (2007) http://www-formal.stanford.edu/jmc/whatisai.html

[6] Mccarthy, J, Some Philosophical Problems from the standpoint of Artificial intelligence, avaliable at URL: (1969) http://www-formal.stanford.edu/jmc/mcchay69.html, 2002.

[7] Shapiro, S.C., (ed.), (1992) Artificial intelligence, In: Encyclopedia of Artificial intelligence, Vol.1.John-Wiley& Sons, New York

1. The First Stage: This stage can be described as the formation and foundation phase, typically starting from the end of World War II, with the Dartmouth Conference considered the official starting point of artificial intelligence. During this period, thinkers and scientists sought to develop machines capable of imitating human intelligence. [8]

The year 1940 marked the first attempts to build a system capable of thinking and using logic in its operations, rather than relying on fixed relationships between symbols and responses. These efforts led to the creation of neural networks in an attempt to simulate the functioning of cells in the human nervous system. Research also began on developing a machine capable of playing chess intelligently, competing with humans, as pioneered by Shannon. This stage is characterized by finding solutions to games and solving puzzles using computers. [9]

In this stage, artificial intelligence focused on using the social and human sciences to develop AI models and systems. Concepts from cognitive psychology and computational linguistics were employed to develop knowledge-based systems and problem-solving approaches.

2. The Second Stage: This stage can be described as the period of stagnation and decline, spanning from the 1980s to the early 1990s. The cause of this slowdown was primarily due to funding shortages and a lack of interest, as expectations had not been met. This situation persisted until the early 1990s, when interest in AI applications was renewed. During this stage, knowledge-based AI systems were further developed, benefiting from insights from the social and human sciences. Concepts from sociology, economics, and anthropology were applied to develop expert systems and improve their performance in applications such as medical diagnosis and strategic planning.

3. The Third Stage: This stage extends from the mid-1990s to the end of the first decade of the 2000s and is called the "Golden Age," during which artificial intelligence experienced significant flourishing, leading to the emergence of numerous AI systems. Advanced computers appeared that were capable of defeating humans in chess, and the first speech recognition and voice-command systems were developed. [10]

During this stage, AI techniques crystallized to include symbolic translation and processing as much data as possible. Research even extended further to understanding how the biological brain functions through its fundamental operations and attempting to simulate human intelligence. Additionally, spoken language commands were utilized to create an industrial robot that imitates certain human behaviors and actions.

4. The Fourth Stage: This stage extends from the 2000s to the present and is referred to as the era of contemporary artificial intelligence. During these years, there has been a strong surge in AI applications across various fields, characterized by their tremendous ability to compete with the human mind, as they rely on extremely extensive inputs, in contrast to the limited human senses. [11]

Continuous developments have led to the emergence of robots, which embodied some of the most prominent and exciting advancements in the field of artificial intelligence through the creation of a new science called robotics, considered a manifestation of artificial intelligence. During this stage, there was a strong focus on ethical and social implications, especially with the expansion of AI applications in the social and human sciences. This led to the emergence of ethical considerations in studies and research,

[8] McCarthy, J., Minsky, M. L.Rochester, N., & Shannon, C.E. (1955). A proposal for the Dartmouth summer research project on artificial intelligence. http://www-formal.stanford. edu/jmc/history/dartmouth/ dartmouth.html

[9] Ibrahim, Abdullah Al-Faqi. Artificial Intelligence and Expert Systems. Dar Al-Thaqafa for Publishing and Distribution, Jordan, 2012, p. 170.

[10] Laham, Mohammed. Artificial Intelligence: Stages of Initiation, Development, and Foundational Principles. Computer Academy, 2020, p. 73.

[11] Warwick, Kevin. Fundamentals of Artificial Intelligence. Translated by Hisham Ahmed Mohamed, General Egyptian Book Organization, Cairo, 2013, p. 122.

particularly regarding issues such as privacy, algorithmic bias, and the responsible use of AI in sensitive fields, which have become critical issues requiring careful attention.[12] At this stage, artificial intelligence began to be viewed with caution and apprehension due to its alarming developments, which were perceived as threatening human existence, ethics, and intrinsic value as a whole.

Today, artificial intelligence is considered the latest achievement of human science in the contemporary world. Scientists classify it as the fourth scientific revolution, with such influence and significance that it surpasses previous scientific revolutions since the beginning of the modern era. The sequence of these revolutions begins with the first industrial scientific revolution in 1760, marked by the invention of the first steam engine and its use in factory machinery, ships, and trains.

The second industrial scientific revolution occurred around 1900, with the discovery of the internal combustion engine and the use of electrical power in automated production, facilitating the development of factories, automobile manufacturing, and other technologies.

The third scientific revolution spans from 1920 to 2000, characterized by the invention of electronic technologies and information technology. Robots and computers played a central role in what is now called the digital revolution, which transformed economic patterns and daily life in societies. This stage also saw the integration of computers and robots into most manufacturing processes, communications, education, and other sectors.

The fourth stage is represented by artificial intelligence technology, beginning in the new millennium and continuing until 2020. Scientists agree that this stage combines the technologies and efforts of previous scientific revolutions. Artificial intelligence has become a revolutionary invention with diverse applications across various fields, acting as a technology that simulates human abilities and unites humans and machines. Some contemporary thinkers even suggest that AI may be humanity's final achievement, potentially displacing humans from their central role in life, marking the beginning of a new era that some call the "post-human stage."[13]

This movement focuses on modern technologies such as genetic engineering, information technology, and nanotechnology and molecular technology in artificial intelligence. It is a stage aimed at improving human life through reliance on technological advancements, with Max More being one of the most prominent advocates of transhumanism.

Some scholars also propose a fifth stage in the development of artificial intelligence, referred to as the fifth scientific revolution, characterized by the emergence of a new type of innovation known as the Metaverse. The term "Meta" means "beyond" or "after," while "Verse" means "world" or "universe." When combined, the meaning becomes "beyond the world," or a virtual world parallel to the real world we live in. The difference between the two is that the virtual world is based on computational assumptions resembling real science, allowing users to move, play, meet friends, shake hands, and talk to them as if they were physically present, through the use of Metaverse glasses. Mark Zuckerberg is considered the first to announce the renaming of Facebook to Meta in October 2021.[14] He points to the advent of a new phase in artificial intelligence, a technological revolution in which real life merges with virtual life, enabling humans to interact and meet each other as if they were in a real environment, while all of this takes place within a virtual

[12] Mittelstadt, B. D., Allo,P., Taddeo, M., Wachter, S., &Floridi, L. (2016). The ethicsof algorithms: Mapping thedebate. Big Data & Society, p3

[13] Basyuni, Amal Abdel Monem Abdel Fattah. The Dialectics of the Relationship Between Science and Philosophy in the Contemporary Era: Artificial Intelligence as a Model. School of Creed and Philosophy, Faculty of Islamic and Arabic Studies for Girls, Mansoura, Al-Azhar University, [Year], p. [page number].

[14] **Zaatar, N.** (2022). The virtual world "Metaverse": A psychological perspective. **Journal of Human Sciences**, **Vol. 9, No. 2**, pp. 1016–1029, University of Larbi Ben M'hidi – Oum El Bouaghi, Algeria.

world that simulates the real world. It is as if we are referring here to Plato's ideal world, which imitates the real world and has been criticized on many occasions by numerous philosophers. This idea, however, has been embodied in reality through concepts of artificial intelligence, and more precisely, in the Metaverse.

Third: The Relationship between Philosophy and Artificial Intelligence:

The roots of the relationship between philosophy and artificial intelligence trace back to the classical problem concerning the source of knowledge, raised by modern philosophers, foremost among them René Descartes in the rationalist approach[1]. He questioned the origin of knowledge and the role of reason in its formation, making reason the source of all intellectual production. Additionally, reason represents a faculty that distinguishes humans from other creatures, namely, intelligence . However, this faculty will become shared with machines once artificial intelligence is given the opportunity to present new ideas that completely contradict all previously established norms[3]—ideas that claim the possibility of machines acquiring knowledge, which implies acquiring intelligence similar to that of humans At this very point, it became necessary to activate the role of philosophy, which has long borne the responsibility of defending human principles and ethics, raising the questions: Can a machine be conscious? And can it compete with humans in the faculty of intelligence ?

Philosophy is linked to artificial intelligence when the latter relies on rationalist philosophy and computational capabilities, functional philosophy, scientific philosophy, and the study of ethics and values that philosophy evokes whenever human ethical principles are challenged[6]. Two-valued formal logic contributed to the creation of innovative artificial language, where a neuron composed of two characters acts as a model for an artificial neural cell that simulates the neurons in the brain, taking the value 1 or 0 depending on the values of neighboring neurons, just as biological neurons operate[15].

This creative scientific project was based on a philosophical-scientific integration, combining three main foundations: the scientific knowledge of physiology and the functionality of brain neurons[1], the formal analysis from the philosophy of logic developed by Bertrand Russell, Whitehead, and other contemporary logicians[2], and the scientific technology derived from the philosophy of computer construction established by the English mathematician and philosopher Alan Turing[16]

One of the reasons that make artificial intelligence a philosophical subject is the ambiguity and lack of clarity regarding the foundations upon which this intelligence is based[1]. As Michio Kaku states in his book Physics of the Impossible: "The fundamental laws underlying artificial intelligence are still not well understood"[17].

One of the reasons that make artificial intelligence a philosophical subject is the ambiguity and lack of clarity regarding the foundations upon which this intelligence is based. As Michio Kaku states in his book Physics of the Impossible: "The fundamental laws underlying artificial intelligence are still not well understood." Moreover, if we consider the fact that since 1950, in just half a century, artificial intelligence has developed and its achievements in terms of cognitive capacity have multiplied by approximately ten billion times, as indicated by scientific statistics.[18] We find ourselves facing a problem concerning how to reconcile this enormous development, which has permeated most aspects of our lives, with the fact that this field suffers from a lack

[15] Russel , Norvig : Artificial Intelligemce A modern Approach , second edition , Pearson Education , New Jersey 2003 , P. 16

[16] Izonar , Amit : Artificial Intelligence and Soft Computing , CRC Press , London-New york-Washigton, 2000 , eBook , P. 25

[17] **Kaku, M.** (2013). Physics of the Impossible (S. D. Kharfan, Trans.). Kuwait: World of Knowledge Series, Vol. 399, p. 130.

[18] **Kaku, M.** (2001). Future Visions (S. D. Kharfan, Trans.). Kuwait: World of Knowledge Series, Vol. 270, p. 41.

of clearly defined fundamental laws. This constitutes, from the perspective of the philosophy of science, an invitation to find its path in this domain, which directly evokes philosophical curiosity, at least to understand the reasons behind this ambiguity and to formulate them philosophically. Furthermore, investigating the intellectual structure underlying any activity or phenomenon—represented by the fundamental laws governing artificial intelligence—is a distinctly philosophical requirement.

Furthermore, the field of artificial intelligence is closely linked to the philosophy of mind, as it represents a continuous process of replicating the functions of the human mind in electronic circuits. In other words, artificial intelligence serves as an innovative experimental laboratory to study human intellectual activity beyond the complexities of the human self, aiming to understand how thought arises, its mechanisms, drivers, and how its processes are executed[1]. The development of artificial intelligence is, at the same time, a development in understanding how the human mind operates.

Artificial intelligence is also connected to one of the most important branches of philosophy, namely the theory of knowledge, which has occupied philosophers for many centuries. AI and its development can serve as an experimental laboratory for philosophy, where an artificial entity simulates the human mind and senses, acquires knowledge from the external world, issues commands, and executes procedures according to what it knows and how it interacts with the external environment[4]. Thus, artificial intelligence inevitably remains within the scope of philosophical questions raised in epistemology, which must be confronted. These questions include: What does a computer or robot know? Does the nature of instrumental/artificial thought resemble the nature of human organic thought? Can a robot reach human-level analytical ability, self-reliance in solving cognitive problems, and possess emotions and feelings? And what are the ethical implications of interacting with artificial entities imbued by humans with thoughts and feelingsAnd can a robot [19]. distinguish between right and wrong actions, both moral and immoral, in addition to all the questions posed by epistemology.

From the above, it becomes clear that philosophy is connected to artificial intelligence through several fundamental topics that form the foundation upon which it is built. These are abstract subjects that philosophy has concerned itself with since ancient times, particularly regarding language, the mind, intelligence, emotions, ethics, and others.

Fourth: The Nature of Artificial Intelligence and Philosophy's Position on It:

The scientific birth of artificial intelligence was not as straightforward as we might imagine; rather, it went through various conflicts[1]. Biologists were wary of this new entrant, leading them to call for its rejection on the grounds that the human brain is an extremely complex organ that is impossible to replicate[2]. Linguists criticized it, arguing that language generation requires mental and psychological abilities that are difficult to simulate[3]. Psychologists rejected it on the basis that knowledge cannot be separated from the accompanying psychological states[4]. Even computer scientists themselves objected, claiming that a machine cannot produce creative knowledge beyond what it has been programmed to know.[20] However, these fears and apprehensions were alleviated when proponents of artificial intelligence realized that it was impossible to exaggerate in making this creation compete with the human who conceived it[1]. This understanding restored peace of mind to the skeptics and transformed them from opponents into participants in the

[19] **Mousa Hussein, K.** (n.d.). Artificial Intelligence from the Perspective of the Philosophy of Mind: The Journey of the Mind to Silicon. Baghdad: University of Baghdad, College of Arts, Department of Philosophy, p. 9.

[20] **Al, N.** (1994). Arabs and the Information Age. World of Knowledge Series, Kuwait, April Issue, pp. 144–146.

development of the advantages of artificial intelligence .

In the field of philosophy, as in other scientific disciplines, artificial intelligence has sparked widespread debate and intense discussion among contemporary philosophers. This debate resembled the major philosophical problems known throughout the history of philosophy. However, its characteristics, components, and applications have widened the gap of disagreement among thinkers regarding its evaluation, as well as determining its intellectual references and the domain it belongs to—whether scientific, mechanical, technical, or philosophical—since it addresses the mind, a uniquely human property that depends on understanding language and various human actions.

The question posed by Turing in his article *"Computing Machinery and Intelligence"*, although framed in mathematical and computational terms, appeared primarily as a philosophical question: Can a machine think? And is the nature of thought in a machine the same as human thought To answer this, Turing presented arguments regarding the possibility of endowing a machine with mental abilities, arguments that sparked a wave of objections. Ringel summarized this by stating: "The problem began when some attempted to blur the distinction between human traits and abilities and those of the machine"[21].

This question divided the community of philosophers into two groups. The first group adopted Turing's ideas, arguing that a machine could be endowed with mental abilities enabling it to act intelligently and even surpass human intelligence . In contrast, supporters of the second group rejected the notion that a machine could succeed humans in their greatest distinguishing faculty. They argued that it is absolutely impossible to equate or compare natural human intelligence with artificial intelligence, which can only imitate human thought and mind[3]. Human intellect and emotions cannot be possessed by any other being.

1- Position of Artificial Intelligence:

Supporters of this approach base their stance on the central idea of artificial intelligence, which claims that AI is a genuine intellectual activity . In other words, an AI machine thinks as a human being does. If the development of artificial intelligence continues to progress, this trajectory may culminate in the creation of purely material entities possessing intellectual and even emotional capabilities comparable to, and possibly surpassing, the sophistication that distinguishes humans from other known beings . Proponents of this approach adopt a philosophical tendency rooted in contemporary material monism, unifying the mind, the brain, and the computer . This perspective is summarized by the American mathematician Claude Shannon (1916–2001), who once asked, "Can machines think " His answer was, "Yes, certainly." When asked to clarify, he said, "I think, don't I"[22] This implies a complete correspondence between human thought and artificial thought, or that human thinking functions exactly like a computer, establishing a parallel between the mind and the brain, which is considered an advanced computer, as inferred from Shannon's response[1]. All defenders of this view maintain that a computer capable of passing the Turing Test possesses the same nature of intelligence and human-like thinking.

Artificial intelligence is the construction of intelligent software that enables a computer to surpass its human counterparts in the ability to think and solve problems[1]. Moreover, it also possesses the capacity for understanding and perception.[23]

[21] Ringle, M. (ed.), (1999), Philosophy and Artificial Intelligence, In: Philosophical Perspectives in Artificial Intelligence, Humanities Press, Brighton, England: Harvester, p 1

[22] **Kaku, M.** (2013). Physics of the Impossible (S. D. Kharfan, Trans.). Kuwait: World of Knowledge Series, Vol. 399, p. 134.

[23] Oussayef, K. (2002), Practical Application of Philosophy in Artificial Intelligence, avaliable at URL : http://www.Cs.rochester.edu/u/brown/242/assts/temprojs/phil.pdf

The Turing Test involves placing a mentally sound human in an isolated room and the computer to be tested in another room. Then, a tester or investigator is seated in a third room . The investigator conducts a written conversation via a typewriter connected to the computer, and the same type of interaction occurs with the human . Both sets of responses are delivered to the investigator through the typewriter without revealing whether they come from the computer or the human . The investigator is required to possess a high level of intelligence and carefully formulate questions capable of distinguishing between the human's and the computer's answers . Turing specified that the conversation should last five minutes. If 30% of the time elapses and the investigator cannot differentiate the computer's answers from those of the human, the computer program is considered to have successfully deceived the investigator and passed the test . This implies that the behavior of the computer's intelligence mirrors that of human intelligence. Turing predicted that by the year 2000, it would be possible to build a computer with one billion bytes capable of passing the Turing Test successfully. However, opinions on the accuracy of this prediction varied depending on the skill of the investigator conducting the test, as computers succeeded in deceiving some investigators but failed with others[24].

In Turing's experiment, the investigator was unable to distinguish between the human and the computer due to the mental similarity exhibited by the computer and its different logical capabilities . Computer science and artificial intelligence specialists justified their position by arguing that a thinking machine can solve problems and adopt a logical approach in dealing with matters. It is also capable of solving functions considered unsolvable according to conventional mathematical proof patterns, by applying quantum arrangements to its physical structures and equipping it with special electronic chips resembling the activity of neural cells, enabling it to learn through experience based on prior knowledge.[25] Supporters of this approach challenged the assumptions of traditional philosophy regarding human uniqueness and the possession of the faculty of reason. This led advocates to take a further step in developing a philosophical view supported by scientific arguments, asserting that the brain is merely analogous to the electronic components and all the material elements of a computer. The mind, in this context, is the program or the set of programs that run on this computer.[26]

From the foregoing, it becomes evident that machines and robots have entered virtually every field, presenting themselves as the optimal alternative to humans due to the valuable services they provide, saving both money and effort, and excelling in precision thanks to their tremendous intellectual capabilities. Their services have even surpassed human activities, as reflected, for example, in designing architectural forms for residences and resorts that even the most skilled engineers cannot create². Likewise, exploratory machines provide precise assessments of subterranean resources, accurately identifying the type of mineral and its degree of purity³. These exceptional intellectual abilities render humans incapable of competing with the intelligence of machines.

2- Position of Human Thought:

Supporters of this approach, who reject the possibility of artificial intelligence surpassing human intelligence, argue that AI differs fundamentally from human reasoning. They emphasize that it is impossible for AI to achieve software comparable to or approaching the human level of intellect. Some proponents of this view have gone so far as to call for the complete cessation of AI projects, considering them a waste of time and money.

[24] Russel , Norvig : Artificial Intelligemce A modern Approach , op.cit. , P.948

[25] **Al-Doudi, M. L.** (n.d.). An Exploratory Attempt into the Nature of Artificial Intelligence from a Qur'anic Perspective [Electronic edition], p. 686.

[26] **Searle, J.** (2007). The Mind (M. Hanna, Trans.). Kuwait: World of Knowledge Series, Vol. 343, p. 58.

The contemporary American philosopher Geoffrey Sayre (1956), specializing in epistemology and ethics, sarcastically remarked on the computer theory of mind: "The pursuit of artificial intelligence based on the fashion of the computer theory of the mind does not merely chase the shadow of achieving satisfactory results; it is time to redirect the efforts of AI researchers … and the enormous funds spent on their attempts to tree-lined streets instead of what the computer theory aspires to"[27].

They justify their position by arguing that humans are the only beings whom God endowed with the faculty of reason, favoring them over other creatures for a wisdom and purpose determined by divine will. These faculties are unique to humans alone, as consciousness is linked to biological characteristics and cannot be transferred to a machine. Thus, the human mind achieves a significant superiority over all developments in artificial intelligence, which lacks the distinctive features of consciousness and cannot possess any degree of freedom to initiate or abandon actions outside the framework of software algorithms that dictate the course of every operation. In this sense, artificial intelligence is a "being in itself" rather than a "being for itself," in Sartre's terms, as the nature of any action is predetermined and precedes its existence[28].

Consciousness is explained through its essential properties: it is qualitative, subjective, and unified. **Qualitativeness** means that every mental process is accompanied by a specific type of feeling, such as pain, joy, or satisfaction. This qualitative aspect elevates consciousness to the level of subjectivity, meaning that consciousness is characterized by a kind of ontological subjectivity belonging to the first-person perspective of the conscious agent. As for the **unity of consciousness**, its basis is that no matter how many types of feelings a single self experiences, distributed across different parts of the body, they remain operative within the unified field of consciousness and form parts of it. In summary, this analysis shows that the three properties of consciousness are inseparable and belong to a single phenomenon: consciousness itself.

The contemporary American philosopher John Searle was among the first to discuss this claim, criticizing the lack of boundaries for machine intelligence and attempting to demonstrate the impossibility of achieving this type of artificial intelligence. To this end, Searle, in the course of refuting the idea of artificial intelligence, proposed a thought experiment called the **Chinese Room**, as opposed to the **Turing Test**, which is used to prove the possibility of machine intelligence. To understand Searle's argument against artificial intelligence surpassing human intelligence, it is necessary to present the idea of the Chinese Room.

Searle asked us to imagine him inside a room with a single entrance and exit, containing only a bundle of papers and a set of symbols and signs in the Chinese language. He speaks only English and knows nothing of Chinese. However, a Chinese speaker passes a paper with a question written in Chinese through the entrance, and Searle is supposed to respond to the question in Chinese as well, passing the answer out through the exit.

Searle cannot even distinguish between Chinese and Japanese writing, yet he possesses in the room a "dictionary" in English—a set of instructions in command form, such as "if... then...," alongside Chinese symbols whose meaning he does not understand. The function of this instruction manual is to provide a set of rules that enable Searle to manipulate Chinese symbols purely formally. According to Searle, since a computer performs calculations, it similarly manipulates a set of symbols and instructions formally.

Assuming Searle follows the instructions in the manual without making a single mistake, the output answers will match the input questions.

[27] Russel , Norvig : Artificial Intelligemce A modern Approach , op.cit. , P. 947

[28] **Hussein, K. M.** (n.d.). Artificial Intelligence from the Perspective of the Philosophy of Mind: The Mind's Journey to Silicon [Previously cited reference], p. 21.

Thus, to outside observers, Searle appears to be proficient in Chinese, and in this way, he would pass the Turing Test successfully.

Searle then asks: "Does this mean that I truly understand Chinese?" The answer is, of course, no. What Searle (or the computer in the case of artificial intelligence) does is apply a set of logical operations dictated by the instruction manual (program) to shuffle Chinese symbols in the room (a database). He does not know their meaning, yet produces answers (outputs) to a set of questions (inputs) without any understanding. Searle himself admits he knows nothing about Chinese. However, if the Chinese Room were transformed into an English room, only then would the process be based on meaning, reference, and understanding, rather than on mere blind, formal symbols.

As a result, the computer understands nothing; it does not comprehend stories or poems. It merely performs calculations, and all it "knows" is computation. It does not even realize that it is carrying out these operations, which are executed automatically through electrical fields and physical phenomena within the components of the device.[29]

Searle's purpose in these criticisms was to challenge two fundamental ideas. The first is the claim that a machine can, or might be able to, truly understand—such as understanding short stories or producing poetry. The second is the notion that the way a machine behaves and its relationship to the program it runs can explain how humans understand things (the idea that the mind is a computer program).

Through the Chinese Room, Searle illustrates the reasons for rejecting these two ideas. He builds his logical argument, based on the thought experiment, from three premises leading to a conclusion. These premises are:

1. The programs on which artificial intelligence systems operate are **formal (syntactic)** in nature.
2. Minds have **conceptual, semantic mental content**, which carries meaning.
3. **Formal syntactic structures** alone are not sufficient to serve as the basis for the meaning or significance that characterizes mental content.[30]

What Searle seeks to demonstrate through this experiment is the impossibility of the claim made by artificial intelligence that the type of thought involved in AI is identical to human thought. This is because AI thought is limited solely to the **formal, syntactic dimension** of human cognition, without attaining the semantic content of the structures it manipulates. Human thought, by contrast, possesses both **syntactic structure and semantic content** simultaneously.

The conception Searle defends is that machines are incapable of attaining **full consciousness** as humans do—a form of consciousness that enables adaptation to circumstances, events, and situations. This adaptation reflects the presence of **intuition and intelligence**, faculties unique to humans, which no machine, regardless of its sophistication, can replicate.

Proponents of this view agree that artificial intelligence can never constitute a truly conscious activity. At best, what can be achieved in robotics is the **imprinting or copying of certain commands** onto machines, enabling them to repeat these commands in a stereotypical manner. Such repetition is devoid of genuine intelligence or intuition, operating solely at a formal level without any real understanding. It may superficially resemble human behavior but does not match the capabilities or skills of human intelligence.

Fifth: Ethics of Artificial Intelligence

With the significant technological advances brought about by AI in human life, philosophical and ethical issues have come to the forefront. AI outputs have caused widespread transformations in all fields of contemporary life, extending to the behavior of

[29] https://manshoor.com/society/artificial-intelligence-philosophy-chinese-room/2024/16/0305:20،

[30] Searle, John. **The Mind**, op. cit., pp. 105–108.

individuals in society, even affecting social norms, rules, and values. This has occurred thanks to its diverse technologies at the intellectual and cultural levels, transforming the world into a small village without boundaries or barriers, breaking social constraints, and prompting a reconsideration of socialization according to new educational concepts of social behavior.

These changes highlight that humans, as dual beings, learn not only through reason but also through emotion. AI, combined with the physical and digital environment, now provides an abundant source of **cultural and intellectual production**, targeting all segments of society. It has even begun to compete with schools and families in the process of socialization, raising concerns about its potential negative effects on psychological and social life.

Moreover, AI has deprived humanity of one of its most essential values since its inception: **freedom**. Humans have become dependent on pre-programmed systems in nearly all aspects of life, unable to perform actions without consulting or relying on machines that shape their patterns of thinking. This has resulted in the loss of independence as a natural activity, transforming human action into a pre-structured, mechanized process governed by fixed mechanisms. Despite human progress in understanding and solving problems, achieving what once seemed impossible, and overcoming distances and barriers, humans have not been able to preserve the freedom they have long celebrated.

This underscores the necessity of establishing an **ethical framework** to govern interactions with artificial intelligence. Accordingly, the ethical dimension becomes a genuine concern in determining how to handle these technologies and how to transform them into a means of understanding human behavior, regulating societal dynamics, and guiding them toward virtue. It involves building values, cultivating human faculties, promoting justice and equality, and safeguarding privacy and human dignity.

This is particularly important given issues such as **data collection without the knowledge or consent of individuals**, manipulation of images and videos, inciting tensions between races and nations, distortion of facts, spreading chaos, manipulating minds, and falsifying evidence.[31]

The **ethics of science and technology** is not merely a philosophical subject; rather, it is **an interdisciplinary field**, meaning that addressing its problems requires the input of multiple disciplines. One could even say it may require contributions from all the humanities. Science explains the ongoing scientific and technological developments, specialists in **philosophy of ethics** examine the resulting ethical implications, legal experts propose legislation, and religious scholars provide guidance in religious societies, and so on.[32]

Machine ethics is considered a branch of **artificial intelligence ethics** concerned with incorporating or ensuring ethical behavior in human-made machines. Its goal is to design a machine that follows an ideal ethical principle, or a set of ethical principles, that the machine must adhere to while performing its functions and interacting with humans. Machine ethics is thus a new branch of **technical ethics** that focuses on ensuring that the behavior of machines toward human users—and even toward other machines—is ethically acceptable.[33]

The study of **artificial intelligence ethics** is based on two main domains: the relationship of the machine to humans, and the relationship of humans to the machine. The first domain focuses on questions such as: In what ways can machines be beneficial or harmful? Can

[31] Al-Asad, Saleh. "Artificial Intelligence between Fears and Aspirations." **Journal of Economy and Development**, vol. 9, 2021, p. 2.

[32] Rania, Atef. "The Impact of Technological Advancement on Ethics: An Analytical and Critical Study." **Journal of the Egyptian Philosophical Society**, vol. 31, no. 31, Cairo, 2022, p. 308.

[33] Jamal, Ali Khalil Al-Dahshan. "Ethical Dilemmas of Fourth Industrial Revolution Applications." **International Journal of Research in Educational Sciences**, vol. 3, no. 3, p. 67. http://search.mandumah.com/Record?1053557

machines possess ethics? What ethical behavior should a machine follow when a problem arises? The second domain concerns the human-machine relationship, addressing questions about how machines are used, how they are designed, and how they are interacted with. Do machines have rights as well as duties? All these questions are addressed by the applied branch of **machine ethics**.

Take, for example, the **autonomous car**. Despite the comfort it provides, reducing traffic congestion and pollution, and assisting sick or disabled individuals, it raises numerous ethical issues identified by designers and users through a series of moral questions:

- How should an autonomous car act when faced with a situation where it must choose between hitting a group of people or an individual
- How should it save its passengers if the brakes fail
- What should it do if it must choose between the safety of its passengers and the safety of pedestrians?
- How should it respond to sudden situations, such as an animal crossing the road, requiring sudden braking that might cause a rear-end collision?
- Who should be held accountable if the car causes the death of a person: the manufacturer or the license holder?

Therefore, the algorithms of autonomous cars must include **ethical principles** to guide their decisions in situations where harm is unavoidable. What applies to autonomous cars also applies to machines used in medical fields, drones, military robots, and all machines that can provide services to humans while potentially causing harm.

The field of **machine ethics** emphasizes the creation of machines that protect humans from unethical use. If a person operates a program that would cause harm to humans, the program itself should refuse to execute the harmful action. Machines should not simply perform every command they are given; instead, they should be able to refuse to carry out unethical actions, saying: "I will not perform this unethical act," rather than: "I did it because I was instructed to do so.[34]

The **virtual life project, the Metaverse**, announced by Mark Zuckerberg, has sparked significant concerns in scientific and media circles due to its potential risk of transforming our reality into an ethically corrupt environment. It is well known that a large proportion of virtual electronic games involve **violence and killing**, in which children and adolescents assume the roles of characters—not as mere spectators, but as active participants. The more skilled they are in inflicting harm on their virtual opponents, the more points they earn, allowing them to assume increasingly aggressive roles, all under the supervision of the game administrators.[35]

All these ethical issues make it necessary to provide solutions through the **design of machines that respect human rights and ethics**, and by programming algorithms that address ethical considerations. This requires the involvement of **psychologists, philosophers, and sociologists** in formulating and determining what constitutes ethical or unethical actions.

Conclusion

It can be said that **artificial intelligence**, despite its remarkable achievements that have transformed contemporary life, remains a phenomenon with a dual impact. On one hand, it represents the pinnacle of human scientific and intellectual progress, and on the other, it raises profound philosophical and ethical challenges that touch the very essence of humanity. While some supporters see it as a new horizon for thought and creativity, opponents fear the domination of machines over human consciousness. The real challenge lies in **balancing technological advancement with the preservation of human values**. Thus, the future of artificial intelligence will not be determined solely by its technical power but also by how successfully humans can

[34] Liao, S. M. (2020): Ethics of artificial intelligence, Oxford University Press, p. 450

[35] **Zaatar, Nour El-Din. "The Virtual World: The Metaverse from a Psychological Perspective." Op. cit., p. 1025.**

guide it ethically and philosophically toward serving the common good and protecting human dignity.

Artificial intelligence encompasses **deep philosophical dimensions**; it is not merely a set of mathematical algorithms and does not represent an independent cognitive system. It has multiple cognitive dimensions: in addition to the **philosophy of mind**, AI is connected to **computer science, binary logic, mathematics, psychology, human cognition, neuroscience, and cognitive engineering**, all of which contribute to its formation and are bound to respect human ethical considerations.

While the development and application of artificial intelligence aim to enhance human comfort and welfare, this does not prevent us from emphasizing the need to **avoid the risks and threats** arising from increased reliance on this technology. This can be achieved by establishing **regulatory and ethical mechanisms** that complement AI, assist in its development, and mitigate its negative effects through the creation of **ethical and legal frameworks** that safeguard fundamental human rights and establish a value system governing relationships in an era where machines may surpass humans.

We indeed need **ethical policies** that regulate robot behavior and address the ethical risks of automated systems. AI systems must be subject to **legal and ethical oversight**, as any use of AI technology outside these rules could inevitably lead to catastrophic consequences. Therefore, it is crucial to develop **carefully designed legal regulations** to manage AI risks, prevent it from becoming a threat to humanity, and establish the **ethical and legal foundations** for governing AI resea

Bibliography :

1. Al-Asad, Saleh. "Artificial Intelligence between Fears and Aspirations." *Journal of Economy and Development*, vol. 9, 2021.
2. Bonnier, Alain. *Artificial Intelligence: Its Reality and Future*. Translated by Ali Sabry Farghali et al., *Alam Al-Ma'refa*, Kuwait, 1993.
3. Abdel-Moneim Abdel-Fattah Basiony, Amal. "The Dialectic of the Relationship Between Science and Philosophy in the Present Era: Artificial Intelligence as a Model." School of Theology and Philosophy, Faculty of Islamic and Arabic Studies for Girls, Mansoura University, Egypt.
4. Badawi, Abdel-Fattah Muhammad. *Philosophy of Science: Science and the Future of Humanity—Where To?* Dar Quba for Printing and Publishing, Saudi Arabia, 1st edition, 2007.
5. Jamal, Ali Khalil Al-Dahshan. "Ethical Dilemmas of Fourth Industrial Revolution Applications." *International Journal of Research in Educational Sciences*, vol. 3, no. 3, p. 67. http://search.mandumah.com/Record?1053557
6. Rania, Atef. "The Impact of Technological Advancement on Ethics: An Analytical and Critical Study." *Journal of the Egyptian Philosophical Society*, vol. 31, no. 31, Cairo, 2022.
7. Searle, John. *The Mind*. Translated by Michel Hanna, *Alam Al-Ma'refa Series*, vol. 343, Kuwait, 2007.
8. Abdullah, Ibrahim Al-Faqi. *Artificial Intelligence and Expert Systems*. Dar Al-Thaqafa for Publishing and Distribution, Jordan, 2012.
9. Ali, Nabil. *Arabs and the Information Age*. *Alam Al-Ma'refa Series*, Kuwait, April 1994.
10. Kaku, Michio. *Physics of the Impossible*. Translated by Saad Al-Din Kharfan, *Alam Al-Ma'refa Series*, vol. 399, Kuwait, 2013.
11. Kaku, Michio. *Future Visions*. Translated by Saad Al-Din Kharfan, *Alam Al-Ma'refa Series*, vol. 270, Kuwait, 2001.
12. Karim, Moussa Hussein. *Artificial Intelligence from the Perspective of Philosophy of Mind: The Journey of the Mind to Silicon*. University of

Baghdad, College of Arts, Department of Philosophy.
13. Warwick, Kevin. *Fundamentals of Artificial Intelligence*. Translated by Hisham Ahmed Muhammad, Egyptian General Book Authority, Cairo, 2013.
14. Muhammad, Laham. *Artificial Intelligence: Stages of Initiation, Development, and Foundational Principles*. Academy of Computing, 2020, p. 73.
15. Muhammad, Lahbib Al-Doudi. *An Exploratory Attempt at the Nature of Artificial Intelligence from a Qur'anic Perspective*. Electronic Edition.
16. Manal, Al-Balqasmi. *Artificial Intelligence: Building the Future*, vol. 1, Dar Al-Ta'leem Al-Jamei, Alexandria, 2019.
17. Nour El-Din, Zaatar. "The Virtual World: The Metaverse from a Psychological Perspective." *Journal of Humanities*, no. 3, 2022. http://search.mandumah.com/Record/1285730
18. Carthy, J., Minsky, M. L.Rochester, N., & Shannon, C.E. (1955). A proposal for the Dartmouth summer research project on artificial intelligence. http://www-formal.stanford. edu/jmc/history/dartmouth/ dartmouth.html
19. Artificial Intelligence, Humanities Press, Brighton, England: Harvester.
20. https://manshoor.com/society/artificial-intelligence-philosophy-chinese-room/2024/16/03
21. Izonar , Amit : Artificial Intelligence and Soft Computing , CRC Press , London-New york-Washigton, 2000 , eBook .
22. Liao, S. M. (2020): Ethics of artificial intelligence, Oxford University Press.
23. Mccarthy, J, Some Philosophical Problems from the standpoint of Artificial intelligence,
24. Mittelstadt, B. D., Allo,P., Taddeo, M., Wachter, S., &Floridi, L. (2016). The ethicsof algorithms: Mapping thede-bate. Big Data & Society, p3
25. Oussayef, K. (2002), Practical Application of Philosophy in Artificial Intelligence, avaliable at
26. Ringle, M. (ed.), (1999), Philosophy and Artificial Intelligence, In: Philosophical Perspectives in
27. Russel , Norvig : Artificial Intelligemce A modern Approach , op.cit. , P.948
28. Russel , Norvig : Artificial Intelligemce A modern Approach , second edition , Pearson Education , New Jersey 2003 .
29. Shapiro, S.C., (ed.), (1992) Artificial intelligence, In: Encyclopedia of Artificial intelligence,
30. Thagard, Computational Philosophy of Science, The MIT Press, London. P. (1988).
31. URL : http://www.Cs.rochester.edu/u/brown/ 242/assts/temprojs/phil.pdf
32. Vol.1.John-Wiley& Sons, New York