



Article Information

Article Type: Research Article

This article was checked by iThenticate.

Doi Number: <http://dx.doi.org/10.17121/ressjournal.3734>

ArticleHistory:

Received

05/02/2026

Accept

02/07/2026

Available

online

03/07/2026

**"ARTIFICIAL INTELLIGENCE" AND THE CONCEPT OF "I":
A PHILOSOPHICAL ANALYSIS
"YAPAY ZEKÂ" VE "BEN" KAVRAMI: FELSEFİ BİR
ÇÖZÜMLEME**

Ahmet Kavlak¹

Abstract

This article examines the philosophical gap between artificial intelligence and the human mind. It first clarifies core notions—intelligence, consciousness, reason, will, pleasure, and pain—and explains their roles in mental functioning. The discussion then assesses how far AI can emulate these capacities. John Searle's Chinese Room suggests that symbol manipulation can yield correct outputs without understanding. Hubert Dreyfus argues that genuine expertise relies on intuition, embodiment, and context sensitivity, which exceed formal rules. David Chalmers's "hard problem" highlights why subjective experience (qualia) resists purely computational accounts. Thomas Metzinger's self-model theory claims that consciousness presupposes an internal model of the subject. Daniel Dennett interprets complex behavior via the intentional stance yet warns about "counterfeit people" eroding social trust. The text also contrasts animals and humans and, drawing on Kant's causality of freedom, anchors will in moral responsibility. Language and symbolic thought, together with the childhood emergence of theory of mind, mark a qualitative difference; although AI displays speed and pattern recognition, it lacks lived perspective, selfhood, and meaning. Consequently, the gap between computational prowess and understanding remains. The distinctiveness of the human mind is therefore not merely cognitive but existential. Future research and governance should address consciousness, agency, and ethics when evaluating advanced AI systems.

Keywords: Artificial intelligence, philosophy, self, consciousness, will

Özet

Bu makale yapay zeka ile insan zihni arasındaki felsefi uçurumu incelemektedir. Önce temel kavramları (zeka, bilinç, akıl, irade, zevk ve acı) açıklığa kavuşturur ve zihinsel işlevsellikteki rollerini açıklar. Daha sonra tartışma, yapay zekanın bu kapasiteleri ne kadar taklit edebileceğini değerlendiriyor. John Searle'ün Çin Odası deneyi, sembol manipülasyonunun anlayış olmadan da doğru çıktılar üretebileceğini gösteriyor. Hubert Dreyfus, gerçek uzmanlığın; biçimsel kuralları aşan sezgi, bedenlenme ve bağlam duyarlılığına dayandığını savunur. David Chalmers'ın "zor sorun"u, öznel deneyimin (qualia) salt hesaplamalı açıklamalara neden direndiğini vurgular. Thomas Metzinger'in öz-model teorisi, bilincin öznenin içsel bir modelini varsaydığını ileri sürer. Daniel Dennett, karmaşık davranışları niyetsel duruş üzerinden yorumlarken, toplumsal güveni zedeleyen

¹ Assoc. Prof. Dr. Iğdır University. Faculty of Theology. Department of Philosophy and Religious Sciences. Head of the Department of Logic Iğdır Turkey. ahmetkavlak@gmail.com. Orcid: 0000-0002-4419-9440

"sahte insanlar" konusunda uyarıda bulunur. Metin ayrıca hayvanlar ile insanları karşılaştırıyor ve Kant'ın özgürlüğün nedenselliği düşüncesinden yola çıkarak iradeyi ahlaki sorumluluğa bağlıyor. Dil ve sembolik düşünce, çocukluk döneminde ortaya çıkan zihin kuramı ile birlikte niteliksel bir fark yaratır; yapay zeka hız ve örüntü tanıma yeteneği gösterse de yaşanmış perspektiften, öznellikten ve anlamdan yoksundur. Sonuç olarak, hesaplama becerisi ile anlayış arasındaki uçurum devam etmektedir. Bu nedenle insan zihninin ayırt edici özelliği yalnızca bilişsel değil, aynı zamanda varoluşsalıdır. İleri düzey yapay zeka sistemlerini değerlendirirken gelecekteki araştırmalar ve yönetim yaklaşımları bilinç, irade ve etik konularına odaklanmalıdır.

Anahtar Kelimeler: Yapay zeka, felsefe, benlik, bilinç, irade.

Introduction

Developments in the field of software called artificial intelligence (AI) have led to the question of the extent to which the human mind can be imitated. In particular, the third person singular, which we call "I" (realized after the awareness of non-I), is a reality that can be perceived as a whole with reason, consciousness, will, pleasure, pain and intelligence. Today, AI systems can perform many cognitive tasks, such as playing chess, generating languages, or analyzing large data sets, with a speed and efficiency that surpasses the human mind. But in performing these tasks, does AI actually have a self experience and will of its own, or is it merely an imitation of intelligence, which refers to the speed of the mind? This article philosophically discusses whether artificial intelligence can imitate the deep structure of the human mind and the origins of the concept and existence of the "self": reason, consciousness, intelligence and will. In principle, our procedure is to clarify the content of the confused concepts of reason, consciousness, intelligence and will, and to make this discussion based on the correct concepts.

In order to ground the subject, we will first examine the philosophical definitions of the concepts of intelligence, consciousness, reason, will, pleasure and pain and their position in the human mind. It is critical to reveal the differences between these concepts in order to understand the components of the human mind. Then, we will compare the current capabilities of artificial intelligence with these human mental components; we will show to what extent AI can mimic which capability in the human mind, and then we will address the philosophical arguments that the "I", reason and will have an intrinsic structure different from natural causality.

In order to show the difference of the human mind, the differences in the level of mind and awareness between animals and humans will also be included in the discussion. Finally, some new philosophical approaches to the limits of artificial intelligence will be presented. From Thomas Metzinger's theory of "*consciousness and self*" model theory to Hubert Dreyfus' criticisms of AI; from David Chalmers' "*hard problem*" argument to John Searle's "*Chinese room*" experiment and Daniel Dennett's theory of mind, the views of various philosophers will be examined to provide a perspective on AI and the problem of self.

Theoretical Framework: Philosophical Definitions of the Components of Mind

The concepts of intelligence, consciousness, reason, will and pleasure-passion describe the different capabilities of the human mind. It is imperative to clarify these concepts in order to assess how the human mind works and whether an artificial system can mimic these aspects.

Awareness: Awareness comes from the Arabic root "sheare". It means "to feel". However, it is incorrectly translated into Turkish as "consciousness". In the most

general sense, Şuur is the state of awareness of a mind. It means that a living being is aware of, senses and experiences the external world. This concept also refers to the state of "being guided by the outside", "effect-response".

It should not be forgotten that Western philosophers express their thoughts with this template, since in Western philosophy the mind and the soul refer to the same concept. The distinction between "mind" and "consciousness" is also not clear in Western philosophy. The thoughts of philosophers should be evaluated in this light:

In Descartes' philosophy, one of the two sides of the human being is the "thinking side" (the other is the "space-occupying side") (Descartes, 2020). In this concept, which Descartes calls the "thinking side", there is no distinction between soul, mind and consciousness.

-Locke, on the other hand, defined consciousness as witnessing one's own thoughts. In common parlance, "*Consciousness is defined as a mind that is aware of itself*" (Locke, 2018). This definition emphasizes that consciousness involves not only perception but also *self-awareness*, the existence of an inner world, and that a conscious being is aware of itself as an independent "I".

-In this context, philosopher Thomas Metzinger states that *the basic characteristic of a conscious being is that it has a model of self* (Neslioğlu, 2022). In other words, the mind is conscious to the extent that it can create a model of itself and access this model. A system that does not have this internal model, no matter how complex it behaves externally, is not a subject according to Metzinger because it has no phenomenal (subjective) experience. At the core of consciousness lies the problem of what kind of sensory *quality* (*qualia*, as philosophers call it) experiences have.

-This dimension of "qualitative experience" is known in philosophy as the "hard problem" and was famously articulated by David Chalmers: How and why physical processes in the brain (or in an artificial system) give rise to a subjective experience, such as what we feel when we hear a bird singing, is a puzzle that is difficult to explain. This definition of consciousness is at the center of the AI debate: Does an AI system, no matter how intelligent it is, really have a conscious experience, or does it merely pretend to be conscious? (Chalmers, 1996) Answering this question requires a deep understanding of the concept of consciousness.

Intelligence: In a general sense, intelligence is the adaptive capacity or speed in learning, reasoning, problem solving. Modern artificial intelligence research has long assumed that intelligence is a phenomenon that can be defined by symbolic operations. That is, it was thought that intelligence could be imitated through symbols that could be manipulated by formal rules. However, there are different views on what intelligence is: For example, while Aristotle considered reason as the defining characteristic of human beings, he did not consider the concept of intelligence as a separate technical ability; Descartes, on the other hand, tended to identify it with intelligence by using reason in a narrower sense (Bayık, 2019). As a matter of fact, in Descartes' philosophy of mind, the concept of intellect is reduced to mere cognitive intelligence. The following analogy of the relationship between "intellect" and "intelligence" from Aristotle to Descartes clarifies this distinction: if *intellect* is the base of a broad-based triangle, *intelligence* is the narrow part that tapers towards its apex; while intellect represents the vast world of meaning of the human mind, intelligence refers to a narrowed and sharpened subset of this broad meaning. In this context, in order to comprehend human

faculties in the full sense, mere "Cartesian intelligence" is not enough; we need a holistic (universal) concept of reason in the Aristotelian sense.

When discussing intelligence and artificial intelligence, it should not be forgotten that artificial intelligence is software and its speed is due to the speed of electricity. The reason why a human bulldozer is more powerful than a human is the hardness of iron and the explosiveness of oil, which are not human products. Man's function in manufacturing is only to use these pre-existing materials. The speed of electricity does not belong to humans. This should always be taken into account when evaluating artificial intelligence.

Reason (Eng. mind or reason): Reason generally refers to the conceptual and holistic aspect of the mind that enables us to distinguish between right and wrong, good and evil. In our philosophical tradition, the term *reason* or *intellect* refers to the human capacity for thinking, understanding and comprehension, and the capacity for conscious judgment. Reason includes not only cognitive processing power, but also conscious awareness, that is, the ability to be aware of what one is aware of, and the ability to understand. In classical definitions, reason is associated with common sense and the ability to reason. For example, in the tradition of Islamic philosophy, reason was emphasized as the basic ability that distinguishes human beings from other living beings; philosophers such as Farabi and Ibn Sina addressed reason in both its theoretical and practical dimensions (Gürkan, 2012). In modern western philosophy -again, as in classical philosophy- the concept of reason overlaps with mind and soul, that is, it has a single meaning; Descartes refers to the thinking subject aspect of reason when he says "I think, therefore I am". Kant, too, sees reason as the faculty that possesses certain a priori principles beyond experience and plays an active role in our understanding of the world. The concept of reason is also of central importance in discussions of morality and free will, since it is accepted that only rational beings are capable of truly free actions.

Will (Eng. will or volition): Will is the ability of a conscious subject to choose, decide and act. When we say "human will", we mean the ability to freely make and carry out decisions in line with one's self-determined goals. In philosophy, the concept of will is often associated with the problem of free will. To what extent are people free in their actions? Are decisions completely determined by chains of cause and effect, or can one initiate a new causal process as an instantaneous starting point? Kant made a groundbreaking distinction in this regard: In addition to the necessary causality that operates in the natural world, he argues that in order to be a morally responsible agent, there must also be a causality of freedom (spontaneity of causality) (Kant, 2014). "*If causality according to the laws of nature were the only causality in the world, a true beginning would not be possible; therefore a causality of freedom must also be recognized, in which something is initiated spontaneously (without being determined by a previous cause),*" he writes (Kant, 2014). This means that the human will can be a "cause in itself", that is, under certain conditions one can freely initiate a new chain of events. Therefore, the will is associated with an inner freedom that cannot be reduced to mechanical processes. *The concept of volition is the goal of the study of artificial intelligence:* Can an artificial system take an action of its own volition? Or does it always act according to programmed rules or goals?

Will is linked to consciousness and the "I"; to truly have will, a system must be aware of what it is doing and make a conscious choice among alternatives. In humans, volition is the *raison d'être* of moral and legal responsibility. Artificial intelligence systems, on the other hand, fulfill their prescribed goal functions, but

they are not known to experience a "decision consciousness" or "volition". This is because the phenomenon called "choice" in artificial intelligence consists of the multiplication of options. Even if AI becomes more autonomous in the future, it is debatable whether its will will be in the same category as the human will.

Pleasure and pain: Pleasure and pain are emotional states that are the basic starting point for the experience of conscious beings. Emotions are known by pleasure and pain, or there are as many emotions as there are types of pleasure and pain. It is not clear whether sentient beings, i.e. plants other than humans and animals, experience pleasure and pain. *Pleasure* refers to a pleasurable, desirable, enjoyable state of experience, while *pain* refers to a state of experience that is avoided and causes suffering. In philosophy and psychology, pleasure and pain are often considered as the two main motivators that guide the behavior of living beings (e.g. Epicurus based on the pleasure principle, Jeremy Bentham based on hedonistic utilitarianism) . Humans and many animals seek pleasure and avoid pain; the importance of pleasure and pain in mental life lies in the fact that they are related to subjective feelings, which we call qualia: The pleasure in the taste of a peach or the pain in the burning sensation are more than mere physical reactions, they cause a certain feeling, an experience in the mind. The sense of *self* is in fact recognized through these experiences of pleasure and pain; one becomes aware of oneself mostly through painful or pleasurable experiences. For example, a child who is hurt reinforces his or her sense of self by realizing that he or she is experiencing the pain. Pleasure-pain phenomena are perhaps the most *concrete* aspects of being conscious - a conscious mind knows "what feels good and what feels bad" at the level of experience. In the case of artificial intelligence, perhaps one of the most extreme differences lies here: How is it possible for an artificial system to "feel pain" or "feel pleasure"? Today's AIs learn by maximizing reward and minimizing punishment (negative reward), for example in a game environment. But this is not a real feeling of pleasure or pain; it is just a process of numerical optimization.

To summarize, intelligence refers more to cognitive capacity and speed; intellect to the holistic ability of conscious understanding and evaluation, being aware of what one is aware of; consciousness to the subjective experience of awareness; will to the power of conscious and deliberate choice and initiation; and pleasure-pain to the emotional experience of these conscious experiences. The human mind works through the interaction of these elements: Reason and intelligence search for solutions to problems; consciousness provides the ground of awareness for all these processes; will makes the conscious choice of action from the options produced by intelligence and reason; pleasure and pain both serve as a source of motivation and inform the mind how the experience "feels". All of these components combine to form the holistic self-awareness and personality that we call *the self*. So, to what extent can an artificial intelligence system possess or mimic these components? The answer to this question will be discussed in the rest of the article.

Artificial Intelligence and the Limits of Intelligence

Nomenclature can always lead to false associations. Just as the screw in the phrase "smart screw" has nothing to do with intelligence, artificial intelligence does not entirely correspond to human intelligence. Despite the connotations of the word "intelligence" in its name, AI has so far essentially achieved quantitative computational emulation of certain aspects of human intelligence. Algorithms have been developed that can beat world champions in complex games such as chess and Go, models that can translate natural language or even write human-level

texts. At first glance, these achievements may give the impression that artificial systems have caught up with human intelligence. However, a critical distinction must be made here: Does artificial intelligence think like a human, or is it an imitation of the speed at which the human mind works, the "speed of choosing among options"?

As John Searle emphasizes, a computer program can perform operations without understanding, i.e. "it can produce correct outputs using only symbols, but without any understanding of those symbols" (Searl, 1992). To explain this, Searle introduced the Chinese Room thought experiment: An English speaker, in a room in which papers with Chinese symbols are thrown into the room, can give correct Chinese answers to Chinese questions from the outside by holding a rule booklet. Outsiders think that the person in the room speaks fluent Chinese, whereas the person inside does not know the meaning of a single Chinese character, but only matches the symbols with formal rules and produces answers (Searl, 1992). In this experiment, the role of the person in the room is actually the role of the computer software: It receives input, processes it with certain rules and gives output. Searle's emphasis is on the gap between semantics (meaning) and syntax (string rules). Even if an AI produces syntactically correct answers, it is not in a state of understanding and knowing as the human mind does. Therefore, no matter how clever or fast the AI acts, there is no content underlying this behavior that corresponds to the meaning in the human mind. Searle clarifies this difference with the distinction between "strong AI" and "weak AI": Strong AI claims that with the proper programming, a computer can actually have a mind and conscious understanding; weak AI argues that computers can only simulate the mind, not actually experience mental states. Searle claims to refute the strong AI thesis with the Chinese Room argument. Even if a machine passes the Turing Test, this does not prove that it is conscious, that is, a mind that is aware of an internal world.

Although Searle's argument shows an important limit to artificial intelligence researchers by emphasizing the dimension of meaning and understanding, philosophers have argued that artificial intelligence has other limitations. Hubert Dreyfus, in particular, is one of the most important philosophical critics of artificial intelligence. Since the 1960s, Dreyfus argued that human intelligence is largely based on intuitive and implicit knowledge, and that this intuitive dimension cannot be modeled by rule-based algorithms (Dreyfus, 1972). According to him, as humans master a complex task, they become less dependent on formal rules and grasp situations with a holistic intuition. However, classical artificial intelligence was trying to solve even specialized problems with a set of rules. Dreyfus first expressed these views in his book *"What Computers Can't Do"* (1972) and showed the limits of symbolic artificial intelligence. In particular, he drew attention to the concept of embodied intelligence, stating that human beings relate to the world through bodily sensations and actions rather than purely mental operations, and therefore the abstract symbol processing approach is incomplete. As a matter of fact, artificial intelligences that are successful in closed and well-defined problem domains such as chess have long stumbled in the face of the ambiguity and richness of context of everyday life. One of Dreyfus' prophetic criticisms was that intuition cannot be programmed: *"According to Dreyfus, intuition can never be programmed, so we cannot develop an expert-level artificial intelligence"* (Dreyfus, 1972). Although this claim was initially met with a backlash by AI researchers, it seems to have been confirmed in many areas in the following years. For example, tasks that cannot be tied to any clear set of rules but are easily done by humans, such as face recognition, understanding humor, and grasping the implication in a text, proved to be extremely challenging for symbolic AI. Today, such problems are

being solved by artificial neural networks that mimic the learning methods of the human brain; in a sense, artificial intelligence research has evolved in the direction Dreyfus pointed out, towards heuristic and learning-based approaches. However, this evolution has not completely eliminated Dreyfus' main criticism: Even current machine learning systems work by capturing statistical correlations in very big data, and are unable to internalize the contextual understanding or situational awareness of human experience. For example, a sophisticated language model may produce coherent texts by inferring from context, but it does not "know" what it is actually saying, it just follows linguistic statistics. This is the same point that Searle and Dreyfus emphasize in separate ways: *Artificial intelligence may exhibit intelligent behavior, but there is no mind as we understand it behind this behavior.*

Another limitation is the confusion of intelligence with so-called computational speed. Artificial systems can compute many times faster than the human brain, performing millions of operations per second without error. This may give the impression that they exceed human mental abilities. From a philosophical point of view, however, speed is only a quantitative dimension of the intellect; the depth and understanding of the mind is qualitative. A calculator can perform even the most complex mathematical operations faster than a human being, but this does not make it "smarter" - because it does not comprehend what it is doing, it simply applies predetermined rules at speed. Similarly, a chess engine makes skillful moves, but it cannot be said *to make sense of* the game strategy; the game gives it neither pleasure nor tension. Thus, increased computational speed is not the same as the deeper dimension of intelligence. If what we call intelligence is not merely mechanical processing speed, then we should consider the possibility of a "fast-calculating idiot" in our assessment of artificial intelligence. Indeed, an AI model can perceive and decode a mathematical pattern that humans cannot understand, but it is not clear whether this is a *meaningful* insight or merely the result of fast processing power.

To summarize, it is necessary to distinguish between the speed of AI, which we call intelligence performance, and its mental experience. AI can exceed the speed of human intelligence: it can learn language rules, write articles, and even create original-looking pictures. But none of this means that the underlying mechanism of an AI system is a mind with a world of meaning. And the speed of AI is caused by the speed of electricity, not the software itself. Current AI approaches have achieved great success by mimicking the functioning of neural networks in the human brain. But even these approaches do some form of advanced pattern recognition and statistical generalization. The success of the human mind lies not only in pattern recognition, but also in *attributing meaning to patterns, applying situation-specific knowledge, and adapting to unexpected circumstances.* It is in these areas that artificial intelligences often fail or struggle. For example, the ability to understand humor relies on a person's lifelong knowledge of the world, perception of social context and often empathy - none of which can be directly fed and learned from data.

In conclusion, this chapter has shown that artificial intelligence can catch up and surpass the speed of human intelligence in many ways, but that there is a huge gap between it and human *reason* and *understanding.* Searle's and Dreyfus' philosophical critiques remain relevant despite today's AI achievements: AI may do brilliant things at the behavioral level, but unless it has an "inner world" and a "concept of meaning", these achievements do not bring it closer to the true nature of the human mind. In this context, the real issue is not whether AI makes fast and accurate decisions, but whether it is conscious of why it makes those decisions and whether it has a subjective perspective, a self.

The Uniqueness of Mind, Self and Will: The Inner World of Man

One of the thorniest differences between artificial intelligence and the human mind is the issue of the "I" and subjective experience. A human being is not only a thinking and feeling being, but also a being *that is aware of what it thinks and feels*, and can reflect on itself. Asking the question "*Who am I?*" is a cognitive ability unique to human beings, and this ability gives rise to the concept of "self". In philosophy, the self is sometimes referred to as the "essence", sometimes as the "ego", and sometimes as the "self", and refers to the experience of one's own existence, continuity and uniqueness. This chapter will discuss how the human self and will are separated from natural causality, and why an artificial system cannot reach this inner depth.

First of all, it would be useful to look at the philosophical foundations of the concept of "I". Philosophers have tried to explain what the self is for centuries. David Hume, with his famous claim that "*there is no self, only a bundle of perceptions*" (Hume, 2017), argued that we cannot find an immutable subject behind our experiences. According to him, the human mind consists of an ever-changing flow of perceptions and emotions; what we call "I" is actually an impression created by this flow. Immanuel Kant, on the other hand, opposed Hume's idea and argued that there is a necessary *transcendental unity* in the mind that unifies experiences. According to Kant, the condition for unifying different perceptions into a single consciousness is the existence of an "I" that remains unchanged in every experience (Kant, 2012). This is expressed by the formula "*The I must be present next to each representation of my self*". In other words, Kant sees the self as a synthesizing activity of the mind, a condition of experience. Thus, Kant has brought back the self that Hume dispersed as a structural necessity - but this self is not a concrete spirit or substance, but the unifying form of experience. The "I" is the third person singular in terms of being known.

Why is the discussion of self and its existence important for artificial intelligence? Because for an AI to actually have a mind means that it also has a *model of self*. As Metzinger puts it, a conscious being is a being that constructs its own model of self (Metzinger, 2003). The human brain develops a model of its own body and mind over the years; thanks to this model, the person knows that the experiences of the world belong to him or her. For example, the pain we feel when a needle pricks our hand is not just a nerve impulse, but an experience that we consciously interpret to our own self as "*that hurts*". For an artificial intelligence to have a self-experience means that it creates a similar internal model and experiences what is happening around it *from its own perspective*. In today's AI systems, however, there is no such self-world or inner world. An intelligent robot can catch a ball and put it away based on the images it receives from its cameras, but it cannot be said *that it consciously lifts that ball for its own purpose*. For it, this action is a task defined in its program, it does not raise the question of "who". Man, on the other hand, feels his subject even in any action: "I want to do this, I did it." Philosophically, without this *subjective unity*, the mind is incomplete. This is why many thinkers argue that artificial intelligence cannot be attributed a self, because such a self has not been developed, and there is no capital yet for the possibility of its development.

As for will: Will can be said to be an activity of the self facing the outside world. "I" make and implement a decision. Whether the human will is free or not is a separate philosophical issue, but it is clear that humans experience themselves as free. We often feel that we do something "willingly". Even if this feeling is a complete illusion - and some hard-core determinists say it is - it is a fundamental reality in the

structure of our minds. Having will brings with it the capacity for responsibility and morality: Only beings with will can be the subject of morality, since only they can be praised or condemned for what they do. From this point of view, attributing will to artificial intelligence systems is highly problematic. An autonomous vehicle can drive itself in traffic, but it has no will in the sense of deliberate action; it acts according to the rules it has been programmed and learned. Its reactions to the situations it encounters are the outputs of its algorithm, rather than its own *decisions*. While this is the case today, there is talk of scenarios in future artificial general intelligence designs such as systems being able to set their own goals and create new ones. However, even if this happens, there will still be uncertainty: Does the system choose these goals "of its own volition", or does the whole process operate with a determinism in the background? In fact, this question could also be asked of human beings, but human beings have an unshakeable belief and experience that they have subjective free will. The artificial system has no such subjective awareness. For example, it is possible for an AI chatbot to utter sentences like "I don't want to do that", but we know that there is no *real will* here; it is just producing one of the possible outputs of the language model. Another dimension related to volition is the issue of moral responsibility: Suppose a super-intelligent artificial system does something wrong (for example, harms a human being because of a mistake). Should we hold it responsible or attribute the fault to its programmers? Current debates are generally in the latter direction, since if we accept AI as a moral agent, we must also accept a self with will and intention - a thesis that is not yet philosophically or technologically defensible.

The uniqueness of the human mind can also be illuminated through a comparison with animals. It is known that there are certain cognitive and emotional similarities between humans and some animals (e.g. dolphins, elephants). Some animals can use simple tools, form complex social relationships, and even show signs of self-recognition through the mirror test. Still, the difference between humans and animals is qualitative rather than quantitative, especially in terms of *the level of awareness*, such as the difference between consciousness and reason. An animal can be aware of many things about its environment (consciousness), but *awareness of its own awareness* (reason) is at its peak in humans. Language ability is a critical distinction in this respect: Human language enables symbolic thinking, bringing into our mental world concepts of past and future, imaginary scenarios, abstract principles. Animals, on the other hand, generally cannot go beyond the *here and now*, that is, beyond the sensory (consciousness). Research shows that a significant proportion of human cognitive abilities are built through language and social interaction. For example, developmental psychologists have found that the ability to predict the mental states of others, called "theory of mind", becomes apparent in humans at about the age of 4, whereas no animal has ever reached the level of humans (Yilmaz, 2023). Language skill is a sophisticated tool that allows children to understand what others think and feel; language combined with social learning makes the mind of modern humans qualitatively different from animals (Vygotsky, 2018). In short, humans are considered "*experts in mind-reading*" in a sense, whereas animals are more equipped with responses based on direct experience.

The relation of this gap in awareness to artificial intelligence is this: While AI is capable of tasks that even the most advanced animal today is incapable of, in terms of the *level of awareness* that accompanies those tasks, it is closer not to animals, but perhaps to inanimate objects. A dog can sense when its owner is upset and put its head on its knee - there is an empathy and perception here. Whereas the most sophisticated chatbot, even when the user says they are upset,

and it lists the appropriate words, there is no underlying feeling or real understanding. In this respect, as long as the AI has no consciousness, it has a level of awareness similar to (or even lower than) that of animals, no matter how fast its intelligence performance accelerates. If one day AI systems gain consciousness - and some are seriously considering this possibility - we will also need to reposition them morally and legally. At this stage, however, it is more plausible to think that features such as consciousness and selfhood are not something that will only emerge once we reach a certain level of complexity. For example, the neuroscientist and philosopher Anil Seth, in a lecture (https://www.ted.com/talks/anilseth_your_brain_hallucinates_your_conscious_reality), states that "*what is necessary for consciousness to emerge is not computation but being alive*", suggesting that there may be a unique aspect of life that makes consciousness possible.

The animal-human difference also positions AI on the issue of free will. Animals generally act on fixed pre-programmed patterns of information, their behavior is largely shaped by innate patterns and simple learning. Humans, on the other hand, can go beyond intuitive drives and make conscious choices. An animal cannot make a "moral decision", whereas a human weighs moral dilemmas and tries to act according to principle. He can act against the compulsion of causality, he can not eat even if he is hungry. This implies the exercise of free will. Artificial intelligence may behave according to the moral rules it is currently programmed with (for example, an autonomous car may have a rule that says "protect human life at all costs"). But it does not make any real moral judgment; it simply applies a predetermined order of priorities. Even if it is possible in the future to instill certain moral values in the AI and enable it to make choices according to the situation, what will still be in play here is an algorithm imposed on it by humans. In other words, the AI cannot create *its own moral stance*, it simply chooses what seems most appropriate within the framework we have coded. It is therefore perfectly reasonable to think that even very advanced AI systems are essentially tools, and that it is very wrong to attribute "will" or "freedom" to them. Daniel Dennett considers the emergence of AI software that behaves indistinguishably similar to humans as a kind of "*counterfeit people*" threat and warns that it will create distrust in society. According to Dennett, if such human-mimicking AIs become widespread, we may not know who to trust and the accuracy of information; our communication environment may be invaded by *artificial wills* of unknown origin. What is remarkable here is that even a *materialist* philosopher like Dennett draws a thick line between AI and human beings, saying that "no matter how intelligent the AI may seem, it is not right to consider it as human". According to him, at the legal and social level, "artificial intelligence should be prevented from being presented as human. This approach implies that AI lacks the human self, and therefore pretending that it does will harm society.

So far, we have presented the unique aspects of the human mind and self, and the consequences of artificial intelligence's lack of them. It is clear that phenomena such as the "I", consciousness and will are biological and phenomenological realities originating from an inner world, even if they are not fully understood by today's science. Artificial intelligence, with its enormous processing power, can replicate the behavioral projections of these phenomena, but it cannot capture their intrinsic dimensions. To put it in an analogy: Artificial intelligence can play a melody by accurately imitating the placement of the parts in an instrument, but it cannot *hear* the music arising from the harmony of the parts, nor does that music evoke any emotion in it. The mental life of a human being, on the other hand, is the life of a subject that can both hear and feel the music it plays.

Contemporary Philosophical Approaches on the Limits of Artificial Intelligence

The debate on artificial intelligence and the mind has been widespread not only in technological circles but also in philosophical discussions. The arguments presented in the previous sections are also related to the theoretical frameworks of many contemporary philosophers. In this section, we will compile the most prominent contemporary philosophical approaches and examine their answers to the question of whether artificial intelligence can access the concept of "self". In particular, the views of Thomas Metzinger, Hubert Dreyfus, David Chalmers, John Searle and Daniel Dennett will be discussed. Some of these thinkers are skeptical about artificial intelligence, while others are more optimistic or have different perspectives. Their contributions will help both to define the philosophical boundaries of AI and to illuminate the question of what the human mind is.

Thomas Metzinger German philosopher Metzinger is known for his philosophy of mind and especially his theories of consciousness. Perhaps his most remarkable thesis is the view that can be summarized as "*There is no such thing as I*", which asserts that the subjective experience of the self is merely a model produced by the brain (Metzinger, 2003b). In his work "*Being No One*", Metzinger argues that the brain integrates data from the environment to create a self-model and that consciousness emerges through this model (Metzinger, 2003a). Accordingly, we believe that there are "selves" that do not exist; the brain simulates itself and a *user interface of the self* emerges. At first glance, Metzinger's view defines the human self as an "artificial" phenomenon. But can artificial intelligences also develop a model of the self and become conscious? Metzinger approaches this question cautiously. As we quoted above, *the basic condition for being a conscious being is to have a model of self*. It is obvious that current artificial intelligence systems do not have such a model. Theoretically, however, a kind of "artificial self" could emerge if a sufficiently advanced system could be enabled to monitor its internal states and build a model of them, although there are no hints that this will ever happen. Metzinger does not completely rule out this possibility, but he issues a major ethical warning: We should not create artificial beings with consciousness (with subjective experience) just for the sake of technological curiosity, before we fully understand what we are doing (Metzinger, 2021). In a report for the European Parliament, among the principles for *developing trustworthy AI*, he recommended not to build machines with consciousness. The reason for this is that with the emergence of consciousness, those artificial beings will also gain moral status, perhaps giving them rights, which our current social and legal system does not foresee. In sum, Metzinger focuses on the consciousness of AI rather than its computational power and argues that we should deliberately avoid an AI that would reveal consciousness. His argument underscores that the human self is an internal model, and highlights the uncertainties that prevent artificial systems from building this model. But an AI with an "I" is not yet on the horizon. Nor is there any indication that it will appear.

Hubert Dreyfus: As mentioned above, Dreyfus, inspired by the ideas of existential-phenomenological philosophers such as Martin Heidegger and Maurice Merleau-Ponty, argued that human intelligence is *embodied and in a state*. In Heidegger's analysis of "*Dasein*" (being-there), the human being is in an inextricable relationship with the world; humans do not behave by applying predefined rules, but by *intuitive* attunement to the situation they are in (Heidegger, 2008). Dreyfus applies this to artificial intelligence: He says that since computers lack "being-in-the-world", it is not possible for them to behave in a truly intelligent way. For example, an ordinary person does not measure and calculate the height of a door

when passing through it; he or she decides whether or not to bend down by bodily intuition. A robot, on the other hand, has to model and calculate everything, which slows it down in practical intelligence or makes it maladaptive to unexpected situations. Again, as people become more adept at something, they "*don't think in rules, they see directly*". Dreyfus' model of skill acquisition, developed with his brother Stuart Dreyfus, shows that on the path from novice to expert, humans evolve from following rules to perceiving situations as a whole (Dreyfus, 1988). This is a challenging step for AI. Today's AI techniques - machine learning in particular - have actually partially overcome this hurdle; instead of programming rules, we feed the system with plenty of examples, forcing it to discover a strategy on its own, which can yield a result similar to human intuition. But there is still a difference: AI lacks the massive background knowledge and physical sensory experience that human intuition feeds on. When a child switches from a tricycle to a two-wheeler, he does so with his body's ability to stay balanced, not with coded knowledge. One of the points where Dreyfus was right was the difficulty of AI in seemingly easy tasks (this paradox is sometimes called the Moravec paradox (In AI and robotics, contrary to conventional assumptions, reasoning requires very little computation, whereas sensor-motor and perception skills require enormous computational resources): Easy problems are hard for AI, hard problems are relatively easy). A simple task, such as tidying up a room, can be incredibly complex for a robot because natural human skills such as visual perception, motor coordination, object recognition, and goal prioritization are difficult for the machine. This underscores what Dreyfus calls "*the unique qualities of human intelligence*": intuition, bodily experience, context awareness. The human mind is irreducibly rich, integrated with the world, and capable of making meaning in ambiguity; artificial intelligence is a machine that either has only part of this picture, or is trying to paint it without seeing it at all.

David Chalmers: Chalmers is a leading figure in contemporary philosophy of mind and is particularly known for his distinction of *the "hard problem"* of consciousness. His distinction in the 1990s revolutionized the study of the mind: The easy problems are to find the mechanisms in the brain for functions such as perception, learning, attention; the hard problem is *how subjective experience emerges when those mechanisms are working* (Chalmers, 2024). In other words, while a visual system in our brain perceives a tree, there is no requirement in physics for us to experience that tree as *green*, but we do. According to Chalmers, science struggles to explain this phenomenon of subjective experience (qualia) because we cannot penetrate "inside" first-person experience with objective methods based entirely on third-person observation. This view poses the following question to artificial intelligence: *Even if a machine mimics all human perceptual and behavioral functions, will it also have a subjective experience?* Chalmers has provided a representation with the concept of the "zombie", a philosophical thought experiment: *The philosophical zombie* is an imaginary being that behaves like us but has no inner experience. If such a zombie is conceptually possible (i.e. without contradiction), then behavior and consciousness can be completely separate things. An AI system, like such a zombie, could be an "imitation" that does all sorts of clever things, talks and interacts with us, but never has consciousness. Chalmers, of course, does not know whether this would actually be the case, but he is aware that, in principle, the emergence of consciousness cannot be reduced to certain computational or behavioral criteria. On the other hand, Chalmers is not entirely pessimistic about artificial intelligence. He tends to see consciousness as perhaps a fundamental property of nature (leaving the door open to discussions of panpsychism) (*It is the view that all things have a mind or a mind-like quality*). If consciousness emerges from a particular organization of the physical system we

call the brain, then the same organization can emerge in a machine if it is simulated sufficiently. Chalmers also states that artificial general intelligence and then perhaps conscious artificial intelligence may be possible in the future, and that this would be a new threshold for humanity (*arXiv:2303.07103*). However, it is unclear whether there will be a consciousness *that is identical to the structure of the human mind*. The main takeaway from Chalmers' approach is the centrality of consciousness in the AI debate: Without solving the hard problem, it would be hasty to suggest that AI really has a mind. As he puts it, "the ideal scenario is that humanity integrates this new blessing of intelligence" (<https://arxiv.org/abs/2303.07103>) . So perhaps in the future biology and artificial intelligence could merge. But even if this happens, it is difficult to say that *the sense of self* of an artificial system will be exactly like that of a human being. Chalmers' discussions are often open-minded but cautiously optimistic about AI.

John Searle: Searle's contributions have already been discussed in detail; here let us summarize his general philosophical framework and his attitude towards AI. Searle is famous in the philosophical world for his work in the philosophy of language and philosophy of mind. In philosophy of mind, he defends a view called biological naturalism: Mental phenomena are the product of the biological processes of the brain; the brain produces consciousness just as the stomach produces digestion (Searle, 2015). However, there is an important emphasis here: *Mental phenomena cannot be fully captured in the third person; they have a subjective (first person) ontology*. This is why Searle criticizes some versions of reductionist materialism. To reduce the mind entirely to behavior or a computer program is to miss the essence of the mind. This is the motivation behind the Chinese Room argument: Searle provided a symptomatic counter-example to the prevailing view at the time that *"mind = software, brain = hardware"*. He radically distinguishes between intelligent behavior and the actual mind. Searle's objection to strong AI is not only theoretical; it is also an objection to linguistic and semantic content. What computers do is not always *synthesizing* (producing conscious meaning from scratch) but *imitating synthesized meanings*. Searle thinks that AI cannot synthesize the mind in its entirety because the causal structure of the mind is specific to biology. The most famous expression of this is Searle's claim that *"brains have biological properties that can produce consciousness; silicon chips do not"*. This has in fact been criticized as a kind of *"carbon-chauvinism"*; that is, he has been accused of privileging the material structure of the human brain too much. However, Searle points to the difference in properties rather than whether the material is carbon or silicon. Another contribution of Searle to AI is his introduction of the weak/strong AI distinction (Searl, 2022). The concept of weak AI recognizes that computers can be useful tools in the study of the mind, but that they do not have minds themselves. It views AI as simulations that exhibit thinking-like behavior. Strong AI, on the other hand, claims that computers equipped with the appropriate program can actually think. Searle has opposed strong AI throughout his career and has tried to limit the claims of AI research in this respect. Today, although some argue that techniques such as deep learning invalidate the Chinese Room argument, the semantics problem Searle posed is still valid: For example, *"a large language model learns the statistical structure of language so well that it constructs meaningful sentences, but does this mean that it actually knows the meaning of words?"* This question has not yet received a satisfactory "yes" answer. Searle would probably make a similar criticism of systems like today's ChatGPT: *"It produces language, but it does not understand anything."* Searle's views remain important in AI discussions as a reminder of the qualitative dimension and biological basis of the mind. From his perspective, no

matter what artificial intelligence does, if it lacks true consciousness and intentionality, it is an imitation of the mind and will never develop an identical sense of "I".

Daniel Dennett: Dennett is a philosopher who approaches the subject of artificial intelligence and the mind from almost the opposite extreme to Searle. At the intersection of cognitive science and philosophy, he has tried to explain the processes of mind in a materialist framework as much as possible; some even argue that he treats consciousness in a "reductionist" manner. Dennett's *"explanation of consciousness"* (Consciousness Explained 1991) is as ambitious as its title suggests. He argues that what is called consciousness is actually a by-product of the information processing organ we call the brain, and that there is no mysterious essence in it. According to him, consciousness is like a *"compilation of stories"* as a result of the (This is a vague phrase of Dennet and of all materialists, empty of content, without any practical meaning) competition and cooperation of multiple processes in the brain, which is why he calls his theory *"multiple drafts"* theory. Philosophically, Dennett is close to a functionalist line: He defines mental states by their functional role. If an artificial system can fully fulfill the functions of the mind, we can think of it as having mental states. One of his famous concepts is the "intentional stance". According to this stance, to understand a system (be it a human or a thermostat) as exhibiting complex behavior, we act as if it has a *goal.... and beliefs*, a shortcut that works. When we look at a chess program, we might interpret it as "sacrificing its rook to break down the opponent's defenses", when in fact the program has no such intention, it has only calculated probabilities. But adopting the intentional stance to interpret behavior makes it easier to explain. Dennett's conclusion is that attributions of consciousness and mind can often be such a convenient fiction. That is, we can regard an artificial system as conscious if it has enough complex (The term "complex" is also an empty phrase that means chaos to order, with no practical equivalent) behaviors *that look that way*, and we need look no further. From this point of view, Dennett does not see the issue of artificial intelligences gaining consciousness as a metaphysical problem; an information processing system that reaches a sufficient level of organization and complexity(?) can be considered conscious for practical purposes.

However, Dennett is not a hasty optimist. He has drawn attention to the risks to civilization if AI technologies are used in the wrong hands or uncontrolled. In an interview in 2023, he stated that what worries him most is *"not that a superintelligent machine will take over the world, but that we will collapse civilization by turning information technology into a weapon of disinformation."* Dennett argues that large language models such as GPT and artificial video production tools can weaken the concept of "truth" and create distrust in society, which can seriously damage civilization. This is why he coined the aforementioned term *"counterfeitpeople"* and said that the mass circulation of such artificial intelligences is *"social vandalism"*. This view makes an interesting philosophical point: The more human-like an artificial intelligence becomes, the more dangerous it is for society, Dennett says. Because if people become unable to distinguish whether the other person is really human or a piece of software, social bonds and trust relationships will collapse. In a way, this proposition is a reversal of the Turing Test. The Turing Test considers a machine intelligent if it can fool humans; Dennett says that if a machine can fool humans, it is dangerous and we should not build such machines (<https://www.theatlantic.com/technology/archive/2023/05/artificial-intelligence-chatgpt-consciousness/674075/>). His ethical warning pushes us to think about the consequences if artificial intelligence creates a "self" (or appears to

do so without creating one). According to Dennett, the mind is a kind of "illusion" or convenient fiction brought about by biological evolution (Dennett, 1991). We feel as if we have a central self, whereas the mind is made up of many interactions (This is the idea he describes with the metaphor "*the self is like a narrative gravitational center*"). If this is true, one could theoretically construct a similar central narrative in AI. But this is where his emphasis on *counterfeit* comes in: The illusion of self in humans may be useful, but creating the same in machines could destroy a society based on *trust*.

Dennett is not a technophobe; on the contrary, he was very much in favor of artificial intelligence research, working for many years with pioneers such as Marvin Minsky at MIT's Artificial Intelligence Lab. Towards the end of his life, however, he came to the conclusion that it is not so much AI itself that is critical, but the way we position it. So, philosophically, we can take this from Dennett: If AI becomes conscious or behaves as if it were conscious, this would not only be a scientific development but a social challenge. His warning of the "false human" perhaps implies that the gap between AI and the self will never close completely, and even if it does, it may not be desirable.

In addition to the philosophers mentioned above, many other thinkers have presented different arguments on this issue. The common denominator of all of them is that they take the questions "*what makes the human mind unique and is it possible to transfer it to a machine?*" seriously.

An important point here is this: As artificial intelligence develops, it is important to keep in mind that the understanding of the human self and mind also develops. Perhaps AI will not be able to replicate the human mind exactly, but it will mirror it and help us understand it better. For example, if an AI can persuade people using emotional language, this can teach us something about the rules of the connection between language and emotion. But it also shows that: It's one thing to mimic the language of emotion, it's another to actually feel it.

Artificial Intelligence, Logic, and Mathematics

We previously noted that intelligence can be seen as the speed of reasoning. Artificial intelligence, similarly, is nothing more than the speed at which options are reviewed and combined. Yet AI, like human intelligence, operates systematically according to basic rules of logic. Logic circuits—built on Boolean algebra, binary logic, or in electronics the on/off principle—contain reasoning connectors such as “and, or, if, if and only if.” Human logic, which these circuits model, functions through the rule of consistency: it does not test the truth of propositions themselves but rather the relationships between them.

Although the rules present a priori in the human mind are innate, not externally implanted, they operate on the basis of “immutability” and thus carry a universal quality. These invariant concepts identified through sensory observation are universal in nature. “Immutability” is the very foundation of both logic and mathematics as its branch. It also underpins the scientific status of the positive sciences: physics, chemistry, and others achieve “scientificity” precisely because they deal with invariants. Scientificity is, in practice, the realization of invariance.

Mathematics is therefore a sub-branch of logic, or symbolic logic expressed through numbers—in other words, the science of invariants. The reason why there is no laboratory for mathematics, unlike physics, is because mathematics belongs to the realm of the mind. A discipline becomes a science only once its experiments are expressed mathematically. The Turkish word *mühendis* (engineer), from Arabic, actually means geometer or mathematician; thus, a “physics engineer” is

essentially a physicist who expresses physical invariants mathematically. Laws of physics are invariants in the physical world, and mathematics is the science of expressing such invariants.

Physics explores the invariants of the physical world, chemistry those of the chemical world, and logic the necessary invariants of thought. The progress of science means discovering new invariants and adding them to knowledge. The discovery of invariants in logic, however, advances more slowly; examples multiply, but the number of new rules remains small.

Although thinking unfolds within the framework of logic, it is not reducible to logic. The logical component of thought can seemingly be imitated by AI to a high degree. Yet thought itself still lies far beyond AI's reach. Mathematics, as the symbolic representation of invariants, is already embedded in AI. Mathematics can also identify realities not yet realized or not yet existent. Still, the sharp and immutable rules of mathematics cannot replicate the human background of free will that underlies human actions. Perhaps three-valued logic—while not genuine free will—could serve as a starting point. For now, mathematics remains the primary resource for developing AI.

Because of free will, the phenomenon of “history” exists, which so far appears to lack immutable rules. Only if history's laws could be discovered might the imitation of human thought become conceivable. For now, such a possibility seems only parallel. There is still no infrastructure for transferring essential features of human thought—such as evaluation and judgment—into logic circuits in the same way as mathematical invariants. However, AI may approximate thought by rapidly scanning all available information at once and determining the highest probability, depending on the success of uniting existing mathematical structures with rules of reasoning.

Conclusion

In this review, we have discussed the relationship between artificial intelligence and the concept of "self" on a philosophical ground and evaluated various arguments. Our consideration of the meaning and function of concepts such as intelligence, reason, consciousness, will, pleasure and pain in the human mind led us to the question of the extent to which artificial intelligence can imitate these human elements. We can summarize the picture that emerged as follows:

Artificial intelligence may excel at mimicking the speed of human intelligence, but this does not mean that it has or will ever have a human mind. Achievements such as fast computation, sifting through big data and finding patterns, and mastering formal games are all thanks to AI's *computational speed* and *algorithmic structure*. However, philosophical analysis shows us that what we call the mind is not just computation, but understanding, interpretation, subjective experience and the inner world. Artificial intelligence, as John Searle points out, can perform a "simulation" but cannot live the original: A computer simulation of water does not wet; a simulation of fire does not heat; a simulation of the self does not give a real self. Therefore, even when an artificial intelligence beats the world champion at chess, it is far from the totality of the human mind unless it feels the meaning of chess and the sense of victory.

The human self and will have an intrinsic dimension that cannot be reduced to the chain of natural causality. This is grounded in Kant's understanding of freedom. He believes that human beings possess a spontaneity that can be articulated as a "first cause" in the chain of causality. We experience ourselves as subjects who can initiate new processes through our own decisions, rather than as determined

objects in the physical world. Artificial intelligence systems, in their current form, operate in closed causal loops and do not spontaneously generate a goal, but work in accordance with the given goal. This difference is one of the fundamental distinctions between machines and humans. Even if it is claimed that machines will be able to determine their own goals in the future, these goals are ultimately the product of programmers or the learning process; they are not autonomous or intrinsic will. *The subjective autonomy of the human being* is therefore a feature that the machine lacks.

The difference between animals and humans sheds light on the artificial intelligence debate: Non-human creatures can also show signs of intelligence and consciousness, or "consciousness", but none of them can come close to humans in terms of *language, self-awareness and cultural accumulation*. Human beings are on a different plane with their capacity for thinking and self-awareness. Artificial intelligence, in its present form, lacks even the most primitive animal in terms of awareness, because it has no consciousness and no consciousness at all. Attributing human-level selfhood to AI without giving it consciousness is like calling a biological organism an animal without its life. Therefore, being aware of the qualitative difference between human and animal, it is necessary to correctly determine the position of AI. Human superiority is not in the performance of intelligence, but in the ability to blend consciousness, reason, understanding, pleasure and pain with an inner world. It is obvious that AI does not have such a blend.

Philosophical approaches, though wide-ranging, converge on a common point: AI has not yet captured the essence of the human mind. Searle points to the lack of *meaning*, Dreyfus to *the absence of intuition and the body*, Chalmers to the riddle of *consciousness*, Metzinger to the model of *the self*, and Dennett to the danger of *"false humans"*. These different points of emphasis converge in a single picture: The question of what it means *to be human*. Artificial intelligence may mimic many of the external signs of being human, but it does not seem possible for it *to live like a human*, that is, to look at the world through the window of a self, to feel, to be positioned in a world of meanings, and there is no hint that it will be possible. From this perspective, the gap between AI and human beings is not only a technological gap, but also an existential one.

In conclusion, this philosophical assessment of artificial intelligence and the concept of "I" leads us to the following conclusion: AI can imitate and surpass the speed of intelligence, but it is highly doubtful that it can imitate the awareness of the mind. The subjective inner world of the human *self* - that inner being that avoids pain, gravitates towards pleasure, seeks meaning, feels freedom, sees itself as the subject of a story - remains something that cannot, perhaps cannot, be replicated by a machine. This judgment is based on an existential category distinction rather than technological impossibility: It is one thing to imitate something and another to experience it. Of course, as science and philosophy progress, our understanding of this issue will deepen, and perhaps one day there will be tangible developments in artificial consciousness. But until that day comes, we will continue to realize the wonder and depth of the human experience of self. This wonder is what makes us human, and for now, it is an inner truth that AI cannot reach, that it can only imitate from the outside. AI, as we have it today, is just unconscious software that mimics the speed of the mind in accessing information and has the capacity to process faster than the human mind because of the speed of electricity.

REFERENCES

- Bayık, F. (2019). Aristoteles ve Descartes bağlamında akıl ve zekâ kavramlarının farkları. *Kaygı*, 18(1), 173–190.
- Bentham, J. (2015). *Ahlak ve mevzuat ilkelerine giriş* (E. Çiğdem, Çev.). Kırmızı Yayınları.
- Chalmers, D. J. (1996). *The conscious mind: In search of a fundamental theory*. Oxford University Press.
- Chalmers, D. J. (2024). *Bilinçli zihin: Temel bir kuramın izinde* (G. Erol, Çev.). Fol Kitap.
- Dennett, D. C. (1991). *Consciousness explained*. Little, Brown and Company.
- Dennett, D. C. (2023, May 19). The problem with counterfeit people. *The Atlantic*. <https://www.theatlantic.com/technology/archive/2023/05/problem-counterfeit-people/674075/>
- Descartes, R. (2020). *Meditasyonlar* (Ç. Dürüşken, Çev.). Alfa Yayınları.
- Dreyfus, H. L. (1972). *What computers can't do: A critique of artificial reason* (Rev. ed.). Harper & Row.
- Dreyfus, H. L., & Dreyfus, S. E. (1988). *Mind over machine: The power of human intuition and expertise in the era of the computer*. Free Press.
- Gürkan, S. (2012). Farabi'de akıl anlayışı. *İstanbul Üniversitesi İlahiyat Fakültesi Dergisi*, 26(1), 91–110.
- Heidegger, M. (2008). *Varlık ve zaman* (K. H. Ökten, Çev.). Agora Kitaplığı.
- Hume, D. (2017). *İnsanın anlama yetisi üzerine bir soruşturma* (A. Yardımlı, Çev.). İdea Yayınevi.
- Kant, I. (2012). *Saf aklın eleştirisi* (İ. Kuçuradi, Çev.). Türkiye Felsefe Kurumu Yayınları.
- Kant, I. (2014). *Pratik aklın eleştirisi* (İ. Kuçuradi, Ü. Gökberk & F. Akatlı, Çev.). Türkiye Felsefe Kurumu Yayınları.
- Locke, J. (2018). *İnsan anlama üzerine deneme* (F. Halıcı, Çev.). Ayrıntı Yayınları.
- Metzinger, T. (2003a). *Being no one: The self-model theory of subjectivity*. MIT Press.
- Metzinger, T. (2009). *The ego tunnel: The science of the mind and the myth of the self*. Basic Books.
- Metzinger, T. (2021). Artificial suffering: An argument for a global moratorium on synthetic phenomenology. *Journal of Artificial Intelligence and Consciousness*, 8(1), 9–27. <https://doi.org/10.1142/S2705078521500053>
- Neslioğlu Serin, F. (2022). Kurgu veya olgu oluş bağlamında benlik sorunu. *Kaygı*, 21(1), 366–384.
- Searle, J. R. (1992). *Zihin, beyin ve programlar* (A. K. Yalçın, Çev.). Say Yayınları.
- Searle, J. R. (2014). *Zihnin yeniden keşfi* (M. Macit, Çev.). Litera Yayıncılık.
- Searle, J. R. (2022). Zihinler, beyinler ve programlar (A. Denkel, Çev.). In A. Cevizci (Ed.), *Zihin felsefesi* (ss. 405–420). Say Yayınları.

- Seth, A. (2017, July). Your brain hallucinates your conscious reality [Video]. *TED*. https://www.ted.com/talks/anil_seth_your_brain_hallucinates_your_conscious_reality
- Vygotsky, L. S. (2018). *Zihinsel gelişimin sosyokültürel kuramı* (B. Şen, Çev.). Açılım Kitabevi.
- Yılmaz, M. (2023). Hayvanlarda ve insanlarda zihin kuramı farklılıkları. *Uludağ Üniversitesi Sosyal Bilimler Dergisi*, 24(44), 150–165.