FILOZOFIA I NAUKA
Studia filozoficzne i interdyscyplinarne
Tom 13, 2025

Mariusz Mazurek

LIMITATIONS OF ARTIFICIAL INTELLIGENCE. WHY ARTIFICIAL INTELLIGENCE CANNOT REPLACE THE HUMAN MIND

https://doi.org/10.37240/FiN.2025.13.1.6

ABSTRACT

This paper examines the historical and contemporary challenges in the development of artificial intelligence. Replicating the human mind—with its intentionality, self-awareness, and creativity—has been, from its inception, a major challenge in artificial intelligence (AI) research. A fundamental issue remains the inability of AI systems to address "why?" and "for what purpose?" questions, underscoring the distinction between machines and humans in terms of meaning-making and contextual interpretation.

Contemporary technologies, such as neural networks and deep learning, aim to emulate cognitive mechanisms observed in biological systems. Nonetheless, limitations such as the "frame problem" and the inability to simulate intentional states persist as significant barriers in the development of these systems. These challenges have prompted an interdisciplinary approach, integrating engineering, philosophy, psychology, and biology, and have led to the emergence of the concept of "naturalness" in AI design. This concept emphasizes mimicking not only cognitive functions but also the adaptive and goal-directed processes characteristic of the human mind.

Keywords: Artificial intelligence, intentionality, AI limitations, human vs. machine thinking.

THE HISTORY AND CHALLENGES IN THE DEVELOPMENT OF ARTIFICIAL INTELLIGENCE

The history of AI, though relatively short—spanning just a few decades—is marked by contrasts, alternating between periods of enthusiasm, high expectations, and disappointment.[1] This is understandable, given that the field tackles one of the most complex and enigmatic issues: the attempt to

[1] S. Russell, P. Norvig, *Artificial Intelligence: A Modern Approach*, Pearson, New Jersey 2021.

understand the human mind and create a machine capable of emulating it. It is an ambitious endeavor that has captivated philosophers, inventors, and dreamers for centuries, from alchemical attempts to create a golem to literary visions of artificial humans, such as Karl Čapek's robots or Isaac Asimov's androids.[2]

In the realm of AI, bold declarations about the imminent replication of human intellect by machines have never been in short supply.[3] During the periods of greatest optimism, researchers promised rapid breakthroughs poised to transform science and technology. However, each wave of optimism was followed by failures and unforeseen difficulties that necessitated a reevaluation of foundational assumptions. While pioneers often acknowledged the challenges of replicating human cognition, some philosophical perspectives suggest alternative frameworks where intelligence emerges through interaction rather than predefined computations. These perspectives challenge the traditional computational metaphor of the mind and offer new insights into the evolution of AI.

From its inception, the field of AI has also attracted its share of critics—both scientists and philosophers—who have challenged prevailing methods and assumptions.[4] Critics have highlighted the oversimplifications inherent in approaches to human intelligence, warning against excessive reductionism that could lead to flawed conclusions. It is important to distinguish between different levels of AI: Narrow AI (ANI) is designed for specific tasks, such as image recognition or language translation, while Artificial General Intelligence (AGI) would require the ability to generalize knowledge across different domains, similar to human cognition. Many skeptics argue that achieving AGI is unlikely due to fundamental differences between biological cognition and machine-based operations.

Despite these challenges, AI continues to evolve, reshaping reality and sparking new waves of hope and apprehension. Each era in its history, from early models based on simple algorithms to contemporary neural networks, has brought both significant progress and unresolved questions. These struggles demonstrate that AI is not merely a technological pursuit but also a profound reflection on human intelligence, its uniqueness, and the limits of our understanding.[5]

---

[2] K. Capek, *R.U.R. (Rossum's Universal Robots)*, Aventinum, Prague 1921.

[3] R. Kurzweil, *The Singularity is Near: When Humans Transcend Biology*, Viking, New York 2005; G. Marcus, E. Davis, *Rebooting AI: Building Artificial Intelligence We Can Trust*, Pantheon, New York 2019.

[4] H. L. Dreyfus, *What Computers Can't Do: A Critique of Artificial Reason*, MIT Press, London 1992.

[5] K. Stephan, G. Klima, *Artificial Intelligence and its Natural Limits*, AI & Society, 36(1), 2021, pp. 9-18. https://doi.org/10.1007/s00146-020-00995-z

KEY BARRIERS IN AI DEVELOPMENT AND THEIR IMPLICATIONS

Every form of intelligence, whether human or artificial, is subject to fundamental limitations stemming from its structure and operational mechanisms. In the case of the human mind, these include constraints such as information processing speed, memory capacity, and the efficiency of signal transmission. Despite these barriers, humans utilize their intelligence in a uniquely adaptive manner—adjusting to changing conditions, solving novel and complex problems, and creating innovative solutions. This ability to adapt and create represents the essence of human intelligence, which remains challenging to fully replicate in artificial systems.

Technological advancements have shown that many obstacles to building advanced AI models have been overcome, yet new challenges continually emerge. Some limitations once considered insurmountable have become obsolete due to progress in areas such as quantum computing and advanced neural networks.[6] For example, the increase in computational power now allows for the real-time processing of vast amounts of data, a feat deemed unattainable just a decade ago.

However, more fundamental issues remain unresolved. Many of these pertain to areas such as nonlinear thinking, intuition, and effective decision-making under uncertainty.[7] Despite advancements in deep learning, AI systems often fall short. Modern AI models, such as generative neural networks and reinforcement learning, have demonstrated the ability to solve complex problems in specific domains, although their flexibility remains limited.

Today's AI is continually being developed to address these challenges. New technologies, such as transfer learning and hybrid approaches that combine various model types, are being explored to overcome the constraints of current systems. This dynamic field of research compels engineers to seek not only more efficient algorithms but also methods for simulating adaptive and creative capabilities, which remain predominantly human domains.

These ongoing efforts underscore that while AI has made remarkable strides, its development is far from complete. The pursuit of overcoming these barriers not only propels technological innovation but also deepens our understanding of intelligence itself, both artificial and human.

[6] R. Penrose, *The Emperor's New Mind: Concerning Computers, Minds, and the Laws of Physics*, Oxford University Press, Oxford 1989.

[7] J. McCarthy, *Making Robots Conscious of Their Mental States*, w: K. Furukawa, D. Michie, S. Muggleton (eds.), Machine Intelligence 15, Intelligent Agents, Oxford University, Oxford 1999, pp. 3-17.

THE LIMITS OF MACHINE THINKING AND THE PROBLEM OF INTENTIONALITY

Classical arguments suggest that creations are inherently dependent on their creators, limiting their capabilities.[8] From this perspective, machines can operate only within the boundaries set by their programmer, and their intelligence, even in its most advanced form, will never surpass the creative level of their creator. The results of machine operations, while sometimes unpredictable, are ultimately constrained by the rules and data implemented within them.

At the early stages of AI development, Alan Turing compiled a comprehensive catalog of arguments challenging the possibility of creating truly thinking machines.[9] These arguments encompass various perspectives, including theological, logical, and intuitive ones. For example, the so-called theological objection posits that thinking is an inherent attribute of the soul, which is a divine gift that no machine can ever receive. Meanwhile, the "head-in-the-sand" objection reflects a more emotional standpoint, assuming that accepting thinking machines would be too terrifying for humans to acknowledge as reality.

Turing also highlighted the absence of conscious experience in machines, which remains a key distinction between human and artificial intelligence. However, some researchers propose that AI could simulate aspects of intentional behavior through predictive modeling and reinforcement learning. While such simulations do not equate to genuine consciousness, they challenge simplistic assumptions about the rigidness of AI decision-making. While machines can "simulate" emotional responses—signaling success or failure, expressing joy at success, or showing embarrassment at a mistake—these are merely programmed reactions devoid of genuine consciousness. This lack of feeling and subjective experience fundamentally differentiates them from the human mind, posing a significant barrier to creating truly thinking and conscious machines.

Despite their increasing sophistication, computational machines – by which we mean both the physical hardware and the abstract algorithms and mathematical models they execute – remain far from replicating the nuanced and multidimensional capabilities of human intelligence.[10] Traits such as politeness, ingenuity, beauty in interpersonal relationships, friendliness, or the ability to take initiative are deeply rooted in human experience and cultural context, making them exceptionally challenging to replicate in

---

[8] L. Floridi, *The Philosophy of Information*, Oxford University Press, London 2013.

[9] A. M. Turing, *Computing Machinery and Intelligence*, Mind, 59(236), 1950, pp. 433-460.

[10] B. M. Lake, T. D. Ullman, J. B. Tenenbaum, S. J. Gershman, *Building Machines That Learn and Think Like People*, Behavioral and Brain Sciences, 40, 2017. https://doi.org/10.1017/s0140525x16001837.

AI systems. Even more abstract aspects, such as a sense of humor, the ability to distinguish right from wrong, moral decision-making, or everyday phenomena like taste and emotional preferences—such as enjoying strawberries with cream or experiencing love—remain beyond the reach of current computational approaches.

As early as the mid-19th century, Lady Ada Lovelace highlighted the limitations of computational machines. Collaborating with Charles Babbage on the first designs of calculating machines, she astutely observed that mathematical machines are inherently constrained in their creativity.[11] They can only perform tasks that have been predefined and clearly described by humans. She noted that for a machine to think in a truly original way, it would require profound knowledge of the mechanisms of human thought. This insight, articulated nearly two centuries ago, remains relevant and aptly illustrates the fundamental limitations of contemporary AI models.

Limitations in human cognition are replicated in limitations in artificial models. Humans' ability to act outside established rules, to go beyond algorithms and patterns, is closely tied to their emotionality, intuition, and capacity to create something entirely new. Although AI systems exhibit increasing levels of adaptability, their ability to generate novel, autonomous goals remains fundamentally limited. While they can optimize within predefined frameworks, the capacity for open-ended, self-directed thinking—characteristic of human cognition—remains elusive.[12]

Since the early days of AI research, scientists have noted fundamental differences between the biological structure of the human nervous system and the digital systems of machines. The human nervous system operates in a continuous, dynamic manner, relying on real-time interactions, while traditional digital machines function on discrete states and precise algorithms. The contemporary development of connectionist networks and neuromorphic systems represents an attempt to mimic this biological complexity. Through modern technologies such as deep neural networks, AI developers strive to replicate the architecture and functioning of the human brain, aiming to bring artificial systems closer to natural intelligence.

For many years, it was widely believed that human behavior could be compared to the operation of an efficient computer program; this belief led to the simplified assumption that thought processes are based on the same mechanisms as those found in digital machines. AI creators assumed that all human actions could be described and formalized using a set of mechanistic rules and principles. This approach inspired computer scientists and engi-

[11] A. Lovelace, *Notes by the Translator on Menabrea's 'Sketch of the Analytical Engine'*, Taylor's Scientific Memoirs, 3, 1843, pp. 666-731.

[12] Cf. D. Hassabis, D. Kumaran, C. Summerfield, M. Botvinick, *Neuroscience-Inspired Artificial Intelligence*, Neuron, 95(2), 2017, pp. 245-258.

neers for decades, who saw it as an opportunity to develop comprehensive models of AI.

Over time, biologists, psychologists, and philosophers adopted computational models to study the brain. The so-called "computer metaphor" became widespread, suggesting that the human mind (reduced to the human mind) should be viewed as a kind of complex computational system. Cognitive phenomena were believed to be reducible to a simple set of components, such as basic actions, data structures, rules, or representations.

Although this approach has led to many successes and allowed for a better understanding of certain aspects of the human mind, over time, researchers began to recognize that the computer metaphor might be overly reductionist. The human brain operates on principles that are not always easily formalized within rigid logical frameworks. This challenge prompted researchers to seek new ways of modeling cognitive processes that account for the complexity and flexibility of human intelligence. As a result, modern AI — referring to the state-of-the-art systems and methodologies developed in recent years — is moving further away from simple computer analogies and toward more biologically inspired designs.

In contrast to the simplified approach that reduces mental processes to a set of fixed, formalizable rules—a view famously criticized by philosophers such as Hubert Dreyfus—there has emerged a growing emphasis on the inherent, "natural limitations" of AI. In this context, the term "mechanical rules" does not suggest that mental processes function according to the laws of physical mechanics; rather, it refers to the artificial confinement of complex cognitive phenomena into discrete, formal constructs. Conversely, "natural limitations" highlights the intrinsic, context-dependent constraints of human cognition that resist such reductionism. Those approaching AI from a more engineering-focused perspective often concentrate on developing intelligent expert systems. While such systems are impressive in terms of their scope of knowledge and capabilities, they remain narrowly specialized, which clearly distinguishes them from the flexibility and versatility of human intelligence.

Reflections on the limitations of AI have become a significant current in this field, particularly with the growing awareness that artificial systems struggle to overcome boundaries that are naturally surpassed by the human mind. Human intelligence is extraordinarily complex and relies on abilities such as intuition, adaptability, and creativity, which are difficult to describe within strict models. In practice, many contemporary AI systems achieve excellent results in specific domains but are unable to operate beyond their narrow areas of specialization.

These considerations have influenced the development of a more realistic approach to the capabilities of AI. Instead of striving to create machines with general intelligence resembling human cognition, engineers and scientists increasingly focus on designing solutions that operate within clearly defined tasks and conditions. This approach allows for the effective use of AI in practical applications, such as expert systems in medicine, data analytics, or robotics, while acknowledging that certain aspects of human intelligence remain unattainable for machines. Recognizing AI's limitations has become a critical reference point for the further development of the field.

In everyday reasoning, we inevitably rely on an implicit background of assumptions—comprising fundamental ontological commitments and general natural laws—that remains largely unconscious and resists precise articulation. This form of tacit knowledge, which provides the contextual framework for our understanding of the world, is distinct from the intentional states that drive goal-directed thought. This knowledge, which resists easy formalization within clearly defined rules, poses a significant challenge for AI systems. Machines are unable to replicate intentional states that define the human ability to direct thoughts and actions toward external objects. A crucial obstacle is the absence of "causal powers" inherent to the human brain—capabilities that enable not only data processing but also the imbuing of that data with meaning in the context of the external world.

Human mental processes are not limited to computational or syntactic operations. If they were, cognitive abilities could easily be transferred to any computational environment, regardless of its specific characteristics. However, research shows that such a functionalist approach to human cognition is insufficient. Evidence for this can be found in both historical and contemporary discussions, which emphasize the incommensurability of certain aspects of human intelligence with mechanistic models of AI.

Another significant limitation of AI is the lack of self-awareness in machines. A computer does not possess a "self," the capacity for subjective perception of the world, which is an inherent part of human consciousness. The knowledge processed by machines is always provided externally, excluding the possibility of an internal, self-aware reception of reality. This lack of subjectivity represents one of the greatest barriers to creating systems capable of fully imitating the human mind. Machines can process information and respond to stimuli, but they cannot experience, feel, or consciously reflect on their actions in the same way humans do.

One of the most interesting objections to the capabilities of AI is the reference to the Turing Test as a method of assessing whether a given system can be considered intelligent. This test is based on comparing the responses of a machine and a human to any questions. If the person conducting the conversation cannot distinguish the machine's responses from human ones, the test is considered passed. Although the Turing Test is one of the oldest and simplest tools for evaluating AI, its usefulness in fully understanding the capabilities of machines remains limited.[13] It focuses solely on external manifestations of intelligence, ignoring internal cognitive and intentional processes.

Criticism of AI's potential also often refers to psychological functionalism, the belief that mental states can be described solely based on their functions and interactions. However, this approach does not account for the feature of intentionality, which is a key element of the human mind. Intentionality refers to the ability to direct thoughts and actions toward a specific object, state of affairs, or goal. This property is typical of human beliefs, needs, desires, or emotions such as fear or intentions. This ability requires not only information processing but also attributing meaning to it in the context of its relationship with the surrounding world.

Human mental processes are not limited to computational or syntactic operations. If they were, cognitive abilities could easily be transferred to any computational environment, regardless of its specific characteristics. However, research shows that such a functionalist approach to human cognition is insufficient. Evidence for this can be found in both historical and contemporary discussions, which emphasize the incommensurability of certain aspects of human intelligence with mechanistic models of AI.

Another significant limitation of AI is the lack of self-awareness in machines. A computer does not possess a "self," the capacity for subjective perception of the world, which is an inherent part of human consciousness. The knowledge processed by machines is always provided externally, excluding the possibility of an internal, self-aware reception of reality. This lack of subjectivity represents one of the greatest barriers to creating systems capable of fully imitating the human mind. Machines can process information and respond to stimuli, but they cannot experience, feel, or consciously reflect on their actions in the same way humans do.

One of the most interesting objections to the capabilities of AI is the reference to the Turing Test as a method of assessing whether a given system can be considered intelligent. This test is based on comparing the responses of a machine and a human to any questions. If the person conducting the conversation cannot distinguish the machine's responses from human ones, the test is considered passed. Although the Turing Test is one of the oldest and simplest tools for evaluating AI, its usefulness in fully understanding the capabilities of machines remains limited. It focuses solely on external manifestations of intelligence, ignoring internal cognitive and intentional processes.

The criticism of AI's potential also often refers to psychological functionalism, the belief that mental states can be described by solely basing on their functions and interactions. However, this approach does not account for the

[13] M. Minsky, *The Society of Mind*, Picador, London 1988.

feature of intentionality, which is a key element of the human mind. Intentionality refers to the ability to direct thoughts and actions toward a specific object, state of affairs, or goal. This property is typical of human beliefs, needs, desires, or emotions such as fear or intentions. This ability requires not only information processing but also attributing meaning to it in the context of its relationship with the surrounding world.

Not all mental states are intentional—for instance, neck pain or a feeling of boredom do not refer to any specific object. Conversely, not all intentional states must be conscious. A person may hold beliefs or needs that are not expressed as conscious thoughts yet still influence their behavior. This complexity of the human mind, encompassing both conscious and unconscious intentional states, presents a significant challenge for AI developers striving to replicate these processes in machines. Ultimately, it is the absence of intentionality and the capacity for conscious direction of actions that distinguishes AI systems from human thinking.

Intentional states are complex and multidimensional phenomena that can be analyzed from various perspectives. From a psychological standpoint, they relate to processes such as attention focus, orienting responses, or the ability to act purposefully. From a physiological perspective, these states are associated with the activation of specific brain regions and the nervous system, measurable with tools like magnetic resonance imaging or electroencephalography. This allows scientists to track which brain regions are activated in response to different intentional thoughts and actions, helping to better understand their biological foundations.

For John Searle and many linguists[14], intentional states are also closely linked to speech acts.[15] They are analogous to the so-called illocutionary force—the intention accompanying an utterance—and its propositional content, the message conveyed by the words. In this context, intentionality becomes a crucial element of communication, enabling not only the exchange of information but also the expression of intentions, emotions, and goals.

Logical considerations of intentional states suggest that they are a particular type of intentional sentences. Their meaning and truth value are not determined solely by the value of logical components but by their reference to context and possible worlds. Jaakko Hintikka, in his theory, connects intentionality with various scenarios of events that could occur in so-called "possible worlds."[16] This multifaceted approach highlights the complexity and multidimensionality of intentional processes.

---

[14] N. Chomsky, *Aspects of the Theory of Syntax*, MIT Press, London 1969.

[15] J. R. Searle, *Minds, Brains, and Programs*, Behavioral and Brain Sciences, 3(3), 1980, pp. 417-424. https://doi.org/10.1017/s0140525x00005756

[16] J. Hintikka, *The Intentions of Intentionality and Other New Models for Modalities*, Springer, Dordrecht 1975; L. McIntyre, *Post-Truth*, MIT Press Essential Knowledge Series, London 2018.

In the field of AI, researchers such as Roger Schank, Robert Abelson, and David Rumelhart have attempted to create models that partially reflect human intentionality. One approach involved introducing belief systems as components of an internal representation of knowledge in computers. This allows these systems to better simulate human thinking by taking into account context, goals, and relationships between objects.

Daniel Dennett, a philosopher specializing in consciousness and AI, proposed the concept of the "intentional stance."[17] According to this idea, an artificial system can be perceived as "intentional" if we interpret its actions as if they were the result of conscious decisions and goals. However, this is more of an interpretive tool than an actual replication of human intentionality. In practice, despite attempts to create systems that simulate beliefs and goals, artificial models remain a mere approximation of true intentionality, which arises from the biological and emotional complexity of the human mind.

Research on AI, especially in the context of modern connectionist methods that attempt to mimic the functioning of the human nervous system, has significantly expanded discussions about the boundaries between human and artificial thinking capabilities. One of the key areas of these debates is the issue of intentionality—the ability to direct thoughts and actions toward specific goals or objects—and the question of whether it can be reduced to purely physical processes.

Patricia Churchland presented an argument[18] illustrating these challenges. She claimed that if intentionality is a physical phenomenon, it should be possible to build a computer capable of experiencing it. However, to date, no computer exhibits intentional states similar to those of humans, leading to the conclusion that intentionality is not merely a physical trait. This antireductionist stance opens the door to further analyses and provokes questions about the nature of human consciousness.

Many researchers challenge the second point of this reasoning, arguing that the firm assertion of machines' inability to possess intentionality may be premature and insufficiently justified. There is also criticism of the first assumption, notably expressed by John Searle. In his view, intentionality results from the biological structure of the brain and the unique processes that the human organism employs to create meanings and goals. These biological properties are not only difficult to replicate but may be entirely unattainable in artificial systems.

Thus, although contemporary technologies enable increasingly advanced simulations of cognitive processes, the difference between human and artificial thinking seems fundamental. Research suggests that intentionality is

[17] D. C. Dennett, *The Intentional Stance*, Bradford Books, Cambridge 1987.

[18] P. S. Churchland, *Neurophilosophy: Toward a Unified Science of the Mind-Brain*, MIT Press, Cambridge 1988.

not a simple function of physical processes but emerges from more complex biological interactions, whose understanding and replication remain among the greatest challenges of modern science.

Intentionality remains one of the most difficult barriers in the development of AI, limiting machines' ability to fully mimic human thought. This limitation arises because intentionality touches on the boundaries of our knowledge about ourselves—how human minds assign meaning to the surrounding world and direct actions toward specific goals. Despite advances, AI still does not match the complexity and flexibility of the human mind.

The inability of machines to replicate intentionality is a fundamental limitation that does not seem likely to be overcome in the near future. Intentionality requires not only information processing but also the ability to interpret it in the context of goals and meanings—something that, in the human mind, arises from extraordinarily complex biological and psychological processes. Without this ability, artificial systems remain tools rather than autonomous entities capable of conscious thought.

Despite its limitations, research in AI has opened new perspectives in the study of cognitive processes, or cognitive science.[19] This field, inspired by efforts to understand the human mind, analyzes processes such as perception, learning, memory, and problem-solving. Many of these issues intersect with the philosophy of mind, which raises fundamental questions about what it means to think, to be conscious, and to act intentionally.

AI, initially focused on creating machines capable of performing human-like tasks, has become a key to foster interdisciplinary research on the nature of human intelligence. As researchers continue to explore these topics, AI not only drives technological advancements but also provokes new questions about what makes humans unique in a world of intelligent systems.

In my view, AI should not be confined solely to an engineering discipline; rather, its goal ought to include the pursuit of naturalness—aiming to replicate the adaptive, context-sensitive processes inherent to human cognition. AI is not limited to developing systems and devices for practical use; it is also a theoretical science aimed at creating models that explain human behavior. For these models to be effective, they should be based on interdisciplinary research on human nature. It is crucial for AI projects to integrate the knowledge of engineers, psychologists, biologists, linguists, philosophers, and logicians—an approach that is thankfully becoming more common.

Naturalness in AI should involve not only mimicking the mechanisms of human thought but also adopting an appropriate methodological approach. First, it is necessary to find a balance between the reductionist simplification

[19] D. J. Chalmers, *The Conscious Mind: In Search of a Fundamental Theory*, Oxford University Press, New York 1996.

of reality and a holistic perspective that accounts for the complexity of cognitive processes. A historical example of this debate was the 1970s controversy between frame-based methods and semantic decomposition approaches. Today, research on connectionist networks is particularly promising, emphasizing the importance of relationships and interactions among system components, which allows for a more comprehensive understanding of AI functioning.

Second, AI must address the challenge of transitioning from posed problems to effective solutions. For humans, this process is often intuitive and unconscious—they can extract the key elements of a situation while ignoring irrelevant details. Machines, unlike humans, struggle to disregard unnecessary information and tend to consider all possibilities, often inefficiently. This is especially evident in robotics, where rapid responses to dynamic environmental changes are required.

Third, it is worth considering the principle of "subsidiarity," which suggests tailoring the level of detail in AI systems to the actual cognitive processes they aim to simulate. Higher levels of system organization exhibit features that cannot be fully reduced to their lower-level elements. Therefore, it is unrealistic to expect that the microstructures of machines will fully explain the complexity of the human psyche.

Fourth, naturalness should also relate to simulating the evolutionary nature of human intelligence. The human mind has developed over millennia, adapting its capabilities to changing environments and goals. Similarly, AI should be designed to emulate adaptive thinking and action.

Implementing these principles can bring AI closer to human cognitive abilities while defining the boundaries that technology may never surpass. These natural limitations not only present challenges but also serve as inspiration for further research and innovation.

PROSPECTS FOR THE DEVELOPMENT OF ARTIFICIAL INTELLIGENCE

AI continues to evolve not only as an attempt to replicate human cognition but also as a complementary tool augmenting human intelligence. From personalized education systems to medical diagnostics, AI is increasingly designed to support, rather than replace, human expertise. Modern AI systems have achieved remarkable results in various fields, such as natural language processing, data analysis, and robotics. However, they remain far from replicating the complexity of the human mind.

A key challenge facing AI is intentionality—the ability to direct thoughts and actions toward specific goals and to imbue information with meaning in the context of the external world. While machines can answer questions like "what?," "where?," or "how?," they struggle to fully comprehend "why?" and "for what purpose?"—questions requiring deep understanding and awareness. The inability to feel, experience, and consciously direct actions sets AI apart from human thinking.

The present era reveals new perspectives in AI research. Technologies such as neural networks, deep learning, and language models are becoming increasingly advanced, allowing for the simulation of some aspects of human thought.[20] Additionally, there is a growing emphasis on striving for naturalness in AI—not merely mimicking human cognitive mechanisms but adopting a more interdisciplinary approach.[21] Combining knowledge from engineering, psychology, biology, and philosophy enables a deeper understanding of both the limits of machines and the nature of human intelligence.

Today's applications of AI in fields such as medicine, education, and transportation highlight its value as a tool that supports humans rather than replaces their abilities.[22] At the same time, challenges like the "frame problem" in robotics serve as reminders that humans still surpass machines in cognitive flexibility and the ability to solve problems intuitively.

Although contemporary AI is far from achieving human-level consciousness, its development provides not only practical benefits but also prompts profound reflections on the nature of intelligence and the boundaries of technology. In an era where AI plays an increasingly prominent role in everyday life, these questions are not only challenges but also critical components of the future.

REFERENCES

K. Capek, *R.U.R. (Rossum's Universal Robots)*, Aventinum, Prague 1921.

D. J. Chalmers, *The Conscious Mind: In Search of a Fundamental Theory*, Oxford University Press, New York 1996.

N. Chomsky, *Aspects of the Theory of Syntax*, MIT Press, London 1969.

P. S. Churchland, *Neurophilosophy: Toward a Unified Science of the Mind-Brain*, MIT Press, Cambridge 1988.

D. C. Dennett, *The Intentional Stance*, Bradford Books, Cambridge 1987.

H. L. Dreyfus, *What Computers Can't Do: A Critique of Artificial Reason*, MIT Press, London 1992.

L. Floridi, *The Philosophy of Information*, Oxford University Press, London 2013.

B. Goertzel, *Artificial General Intelligence*, Springer, Berlin 2007. https://doi.org/10.1007/978-3-540-68677-4

D. Hassabis, D. Kumaran, C. Summerfield, M. Botvinick, *Neuroscience-Inspired Artificial Intelligence*, Neuron, 95(2), 2017, pp. 245-258.

J. Hintikka, *The Intentions of Intentionality and Other New Models for Modalities*, Springer, Dordrecht 1975.

R. Kurzweil, *The Singularity is Near: When Humans Transcend Biology*, Viking, New York 2005.

[20] B. Goertzel, *Artificial General Intelligence*, Springer, Berlin 2007. https://doi.org/10.1007/978-3-540-68677-4

[21] I. Rahwan et al., *Machine Behaviour*, Nature, 568(7753), 2019, pp. 477-486. https://doi.org/10.1038/s41586-019-1138-y

[22] S. Russell, P. Norvig, *Artificial Intelligence*, op. cit., p. 28.

B. M. Lake, T. D. Ullman, J. B. Tenenbaum, S. J. Gershman, *Building Machines That Learn and Think Like People*, Behavioral and Brain Sciences, 40, 2017. https://doi.org/10.1017/s0140525x16001837
A. Lovelace, *Notes by the Translator on Menabrea's 'Sketch of the Analytical Engine'*, Taylor's Scientific Memoirs, 3, 1843, pp. 666-731.
G. Marcus, E. Davis, *Rebooting AI: Building Artificial Intelligence We Can Trust*, Pantheon, New York 2019.
J. McCarthy, *Making Robots Conscious of Their Mental States*, w: K. Furukawa, D. Michie, S. Muggleton (eds.), Machine Intelligence 15, Intelligent Agents, Oxford University, Oxford 1999, pp. 3-17.
L. McIntyre, *Post-Truth*, MIT Press Essential Knowledge Series, London 2018.
M. Minsky, *The Society of Mind*, Picador, London 1988.
R. Penrose, *The Emperor's New Mind: Concerning Computers, Minds, and the Laws of Physics*, Oxford University Press, Oxford 1989.
I. Rahwan et al., *Machine Behaviour*, Nature, 568(7753), 2019, pp. 477-486. https://doi.org/10.1038/s41586-019-1138-y
S. Russell, P. Norvig, *Artificial Intelligence: A Modern Approach*, Pearson, New Jersey 2021.
J. R. Searle, *Minds, Brains, and Programs*, Behavioral and Brain Sciences, 3(3), 1980, pp. 417-424. https://doi.org/10.1017/s0140525x00005756
K. Stephan, G. Klima, *Artificial Intelligence and its Natural Limits*, AI & Society, 36(1), 2021, pp. 9-18. https://doi.org/10.1007/s00146-020-00995-z
A. M. Turing, *Computing Machinery and Intelligence*, Mind, 59(236), 1950, pp. 433-460.

OGRANICZENIA SZTUCZNEJ INTELIGENCJI. DLACZEGO SZTUCZNA INTELIGENCJA NIE ZASTĄPI LUDZKIEGO UMYSŁU?

STRESZCZENIE

Sztuczna inteligencja od początku swojego rozwoju stanowi obszar intensywnych badań, cechujący się zarówno znaczącymi osiągnięciami, jak i nieuchronnymi ograniczeniami. Kluczowym wyzwaniem w tej dziedzinie jest próba odwzorowania ludzkiego umysłu, który charakteryzuje się zdolnością do intencjonalnego myślenia, samoświadomości oraz kreatywności – właściwościami trudnymi do redukcji do struktur obliczeniowych. Fundamentalnym problemem pozostaje niezdolność systemów sztucznej inteligencji do formułowania odpowiedzi na pytania „dlaczego?" oraz „po co?", co podkreśla różnicę pomiędzy maszynami a człowiekiem w aspekcie nadawania sensu i interpretowania kontekstu. Współczesne technologie, takie jak sieci neuronowe oraz uczenie głębokie, dążą do odwzorowania mechanizmów poznawczych obserwowanych w systemach biologicznych. Mimo to ograniczenia, takie jak „frame-problem" czy brak zdolności do symulowania stanów intencjonalnych, wciąż stanowią istotne bariery w rozwoju tych systemów. Refleksje nad tymi wyzwaniami zapoczątkowały interdyscyplinarne podejście, integrujące inżynierię, filozofię, psychologię oraz biologię, a także zainicjowały koncepcję „naturalności" w projektowaniu systemów AI. Koncepcja ta zakłada naśladowanie nie tylko funkcji poznawczych, ale również procesów adaptacyjnych i celowych charakterystycznych dla ludzkiego umysłu. Niniejszy artykuł analizuje historyczne oraz współczesne wyzwania związane z rozwojem sztucznej inteligencji, uwzględniając filozoficzne rozważania nad granicami technologii. Podkreśla również, że pomimo rosnącej roli AI we współczesnym świecie, jej rozwój stanowi inspirację do głębszego zrozumienia natury ludzkiej inteligencji, jej unikalnych właściwości oraz ograniczeń.

Słowa kluczowe: sztuczna inteligencja, intencjonalność, ograniczenia AI, ludzkie a maszynowe myślenie.

ABOUT THE AUTHOR – PhD, Institute of Philosophy and Sociology of the Polish Academy of Sciences, Nowy Świat 72, 00-330 Warsaw.

E-mail: mariusz.mazurek@ifispan.edu.pl