ISSN 2965-1557
http://doi.org/10.1590/2965-1557.038.e202635053

Original Article

Revista de Filosofia Aurora

PUCPRESS

periodicos.pucpr.br/aurora

What Is It Like *to Be* a Machine? Bats, Intelligent Design, and the Problem of Other Minds

Como é ser uma Máquina? Morcegos, Design inteligente e o Problema das Outras Mentes

Léo Peruzzo Júnior [a]
Curitiba, PR, Brasil
[a] Pontifícia Universidade Católica do Paraná (PUCPR)

Como citar: PERUZZO JÚNIOR, Léo. What Is It Like to Be a Machine? Bats, Intelligent Design, and the Problem of Other Minds. *Revista de Filosofia Aurora*, Curitiba, v. 38, e202635053, 2026. DOI: https://doi.org/10.1590/2965-1557.038.e202635053

Abstract

Starting from Thomas Nagel's classic essay *What Is It Like to Be a Bat?*, this article investigates the philosophical problem of consciousness and its implications for Artificial Intelligence. It argues that an objective description of physical and computational processes is insufficient to explain the subjective character of experience, that is, the fact that there is something it is like to be a particular organism. Drawing on Nagel's analysis of bats, the article discusses the persistence of the so-called explanatory gap between the physical correlates of consciousness and phenomenal experience. The discussion is then extended to the domain of intelligent machines, asking what it would mean to be a machine and whether artificial systems could possess their own forms of experience or understanding of the world. The article critically examines the contemporary imaginary surrounding Artificial General Intelligence,

[a] Postdoctoral Researcher at Università Ca' Foscari Venezia (Visiting Scholar). Professor in the Graduate Program in Philosophy (Master's and PhD) at the Pontifical Catholic University of Paraná (PUCPR), at the Centro Universitário Franciscano do Paraná (UNIFAE/PR), and at the Faculdade Vicentina (FAVI). PhD in Philosophy from the Federal University of Santa Catarina (UFSC), e-mail: leo.junior@pucpr.br

highlighting the limitations of approaches that reduce intelligence to information processing and algorithmic optimization, while arguing that intelligence also involves pragmatic engagement, meaning-making, and structural coupling with the environment. Finally, it draws on autopoietic enactivism to defend the view that different forms of life and cognition generate different worlds of meaning, suggesting that the real challenge is not whether machines can think like us, but whether they might come to inhabit a world of meanings of their own.

Keywords: Consciousness. Artificial Intelligence. Problem of Other Minds. AI Imaginaries. Enactivism.

Resumo

Partindo do clássico ensaio What Is It Like to Be a Bat?, *de Thomas Nagel, este artigo investiga o problema filosófico da consciência e suas implicações para a Inteligência Artificial. Argumenta-se que a descrição objetiva dos processos físicos e computacionais não é suficiente para explicar o caráter subjetivo da experiência, isto é, o fato de haver algo que é como ser um determinado organismo. A partir da análise dos morcegos de Nagel, discute-se a persistência do chamado gap explicativo entre os correlatos físicos da consciência e a experiência fenomenológica. Em seguida, o problema é deslocado para o domínio das máquinas inteligentes, questionando-se o que significaria ser uma máquina e se sistemas artificiais poderiam possuir formas próprias de experiência ou compreensão do mundo. O artigo examina criticamente o imaginário contemporâneo da Inteligência Artificial Geral, destacando os limites das abordagens que reduzem a inteligência ao processamento de informações e à otimização algorítmica e sustentando, ainda, que inteligência envolve inserção pragmática, produção de significado e acoplamento com o ambiente. Por fim, recorre-se ao enativismo autopoiético para defender que diferentes formas de vida e cognição produzem diferentes mundos de sentido, sugerindo que o desafio não é saber se as máquinas pensam como nós, mas se podem vir a habitar um mundo de significados próprio.*

***Palavras-chave**: Consciência. Inteligência Artificial. Problema das Outras Mentes. Imaginários da IA. Enativismo.*

Introduction

As computational systems increasingly perform tasks traditionally associated with human cognition, there is a growing temptation to understand intelligence as a phenomenon reducible to information processing itself. Accordingly, the hypothesis that an intelligent machine should be modeled on the design of human intelligence encounters, on the one hand, the assumption that thinking is essentially a form of computation and, on the other, the claim that such projects can succeed under denaturalized conditions, that is, without a living body or an environment. For this reason, thinking about an intelligent machine does not merely involve discussing its computational performance. Rather, it requires redefining what it means to exist as a machine, as well as what it means to exist as any other kind of being.

Some of the questions raised by Artificial Intelligence can be traced back to the debate initiated by Thomas Nagel in his celebrated essay *What Is It Like to Be a Bat?* (1974). By arguing that every conscious state possesses an irreducibly subjective character, Nagel maintains that no objective description of physical processes can fully capture what is experienced from the first-person perspective. His analysis seeks to expose the limits of reductionist approaches to the mind, introducing a problem that remains unresolved: how can the objectivity of scientific explanations be reconciled with the constitutive subjectivity of conscious experience? Put differently, is there a fundamental difference between accurately describing the functional capacities of a machine and understanding what it is like to be such a system?

Thinking about an intelligent machine requires more than simply equating cognition with computation. In this respect, it is difficult to agree with the claim that "the inspiration guiding cognitive science is that, at an appropriate level of abstraction, a theory of natural intelligence should have the same basic form as theories explaining sophisticated computational systems" (Haugeland, 1981). This assumption presupposes that cognitive processes can be exhaustively described by computational models, relegating to the background the constitutive role of the body, action, the environment, and the forms of life within which intelligence emerges and acquires meaning. Legg and Hutter (2007, p. 15), for example, argue that a general definition of artificial intelligence is that "intelligence measures an agent's ability to achieve goals in a wide range of environments." Consequently, they maintain that intelligence should not primarily be understood as memory, abstract reasoning, language, or accumulated knowledge. These capacities may contribute to intelligence, but they do not define it. Rather, intelligence is defined by an agent's overall performance in achieving goals across a variety of environments and, consequently, by its capacity to adapt its behavior in potentially unfamiliar contexts.

Against this background, the problem of Artificial Intelligence cannot be reduced to the technical question of whether machines are capable of performing increasingly complex tasks (Hopster, 2021; Ulnicane; Erkkilä, 2023; Kudina; van de Poel, 2024). What is ultimately at stake is whether intelligence consists merely in the capacity to achieve goals across different environments or whether it also presupposes a particular mode of being situated in the world, especially given AI's disruptive potential to transform existing practices such as writing and education. According to Kudina and Van de Poel (2024, p. 4), for example, "AI needs to be developed and designed with consideration for the

sociotechnical systems in which it will be embedded. For example, how we should understand values such as fairness and explainability in the design of AI systems will partly depend on the specific context and the concrete sociotechnical systems in which they will be embedded." They therefore argue that designers should not rely solely on general design principles or abstract metrics of fairness but must instead adapt these concerns to specific contexts. Moreover, if intelligence cannot be abstracted from its biological conditions and reconstructed within computational systems, then much of the contemporary imaginary surrounding AI appears less as a description of the actual capabilities of machines than as a projection of anthropocentric expectations concerning the nature of cognition.

Drawing on a critical analysis of the imaginaries associated with Artificial General Intelligence, this paper argues that understanding the world involves specific forms of pragmatic engagement, meaning production, and coupling with the environment. For this reason, the decisive question is not whether machines can reproduce isolated human capacities, but whether artificial systems could develop their own modes of relating to the world independently of a centralizing system such as the brain. Ultimately, the challenge is to understand what it means to conceive of AI as a sociotechnical system, that is, one "constituted by diverse social, political, economic, cultural, and technological factors" (European Commission, 2023, p. 10).

Nagel´s Bats

In 1974, Thomas Nagel published the essay *What Is It Like to Be a Bat?*, one of the earliest philosophical attacks on materialist reductionism, which conflates the explanation of matter with the organization capable of generating qualitatively new properties. The relationship between water and the formula H_2O, the problem of the gene and DNA, and the relationship between an oak tree and a heap of hydrocarbons, which contain the same chemical substances but lack biological organization, are ultimately examples showing that a physicochemical description of components is insufficient to explain the emergence of complex properties, especially subjective conscious experience.

The central point of Nagel's argument is the thesis that every conscious state possesses a subjective character. That is, there is something it is like to be the organism undergoing that state: "An organism has conscious mental states if and only if there is something that it is like to be that organism" (Nagel, 1974; Portuguese edition, 2013, p. 109). According to Nagel, an organism has conscious mental states if and only if there is something it is like to be that organism, which may be described as the subjective character of experience. This implies that the explanatory gap between the phenomenological experience of being conscious and the physical and causal powers that produce such an experience raises highly specific philosophical problems. After all, when we refer to the physicochemical functions responsible for instantiating or producing conscious sensations, we are not yet explaining the phenomenal character of experience. Rather, we are merely describing its objective correlates, while the explanatory gap between physical processes and the way those processes are manifested to a subject, or indeed to any other animal, remains intact.

At the same time, Nagel chose bats rather than wasps or flatfish because bats are mammals, thereby preserving our intuition that conscious experience occurs in them. Bats, however, possess a perceptual apparatus radically different from our own, centered on echolocation, making their mode

of perception a fundamentally alien form of life. More than one thousand bat species, for example, employ echolocation, using sounds produced through vocalizations or wingbeats to locate food in environments with little or no light (Vater; Kössl, 2004; Voigt; Lewanzik, 2012). What Nagel ultimately intends to demonstrate is that even if we possessed a complete understanding of the bat's nervous system, or even if we could hang upside down and imitate its behavior, this would tell us only what it is like for us to behave as bats, not what it is like, from the bat's own internal point of view, to be a bat. The same analogy applies to any other being, since our imagination depends upon our own experience and is therefore inherently limited.

It is well known that Nagel's primary target is physicalist theories that seek to reduce mental states to physical or neurophysiological states, particularly the work of J. J. C. Smart (Philosophy and Scientific Realism), David K. Lewis (An Argument for the Identity Theory), Hilary Putnam ("Psychological Predicates"), D. M. Armstrong (A Materialist Theory of the Mind), and D. C. Dennett (Content and Consciousness). Nagel does not deny that the mind may, in some sense, be physical. Rather, he argues that we lack even a conceptual understanding of how a purely objective description could capture the subjective character of experience, that is, what it is like to have the experience of being conscious.

Unlike bats, or indeed any other animal that may be conscious in the way bats or humans are, scientific objectivity requires an impersonal perspective that is accessible from any point of view. Consciousness, however, always involves a unique point of view, even if one reduces it to a neurophysiological instantiation. It is precisely here that traditional arguments appear to talk past one another. On the one hand, they explain conscious processes as supervenient upon physical processes. On the other hand, they fail to preserve the subjective dimension of the experience of being conscious. A satisfactory theory of consciousness, therefore, would be one capable of preserving the phenomenal character of experience while also explaining, in objective physical terms, why and how certain neurophysiological processes give rise to an intrinsically subjective mode of existence (Peruzzo Junior, 2025). In other words, such a theory would have to demonstrate, not merely descriptively but explanatorily, why, whenever particular physical and neural conditions obtain, there is something it is like to be the subject in question, thereby preserving the singular perspective of experience while simultaneously situating it within the objective order of nature.

In light of the problem formulated by Thomas Nagel, three questions place pressure on the limits of an objective explanation of consciousness. The first asks whether it is possible to be conscious without the experience of being conscious. If accepted, this hypothesis would reduce consciousness to a purely functional description while eliminating its phenomenal character. The second asks whether it would be possible to be conscious while perceiving the world through sonar. This highlights the fact that even with a complete physical and neurophysiological description, the problem remains of how to integrate the qualitative aspect of experience into an impersonal perspective, whether in our own species or in any other. The third asks whether it would be possible to be conscious of the world after removing experiential content altogether, thereby leading to an empty form of consciousness that appears to dissolve the very idea of experience.

Taken together, these three questions demonstrate that any attempt to fully objectify consciousness faces the difficulty of preserving the subjective point of view without which there is no genuine conscious phenomenon. Thus, even if I could perceive and behave like a wasp or a bat without altering my fundamental constitution, my experiences would bear little resemblance to those of these other animals. It is precisely at this point that many neurophysiological theories encounter their greatest difficulty, since they conflate the description of consciousness with the experience of consciousness itself. Moreover, while changes in the physicochemical constitution of a human being under otherwise ordinary environmental conditions may profoundly disrupt the experience of being conscious, even if we were gradually transformed into a bat, nothing in our present constitution would enable us to imagine what the experiences of such a creature are like (Nagel, 1974).

According to Nagel (1974), the fact that we cannot indefinitely postpone developing a language capable of accommodating a detailed phenomenological description of a Martian or a bat should not lead us to dismiss as insignificant the claim that bats and Martians may have experiences fully comparable to our own. Nevertheless, an epistemological gulf remains between the possibility of such experiences and our actual capacity to understand them, since our conceptual schemes are shaped by particular forms of life. In other words, the obstacle is not merely physicochemical in nature, but also stems from an incompatibility between the conceptual frameworks we employ and those that would structure the experience of other forms of life.

The same limitation reappears when the question is transferred to the artificial domain. If, in the case of bats, we encounter the impossibility of accessing the subjective character of a radically different form of existence, something analogous occurs when we ask what it would be like to be an intelligent machine. Even if we could describe with complete precision its computational architecture, algorithms, and functional capacities, such a description would tell us only how those systems operate from our own perspective, not whether, or in what sense, there is anything it is like to be such a system. In this context, the difficulty lies not merely in the absence of empirical data but in the possible inadequacy of our own conceptual schemes for grasping a mode of existence that does not share the biological and phenomenological conditions structuring human experience. It is precisely this question, namely, what it would mean to be an intelligent machine, that constitutes a philosophical problem of the first order.

What Is It Like to Be an Intelligent Machine?

Irish writer Mark O'Connell's book To Be a Machine: Adventures Among Cyborgs, Utopians, Hackers, and the Futurists Solving the Modest Problem of Death (2017) offers a distinctive portrait of the contemporary transhumanist imaginary. By following scientists, engineers, hackers, and advocates of the technological transcendence of the human condition, the book reveals a common assumption underlying many narratives about artificial intelligence: the conviction that intelligence can be detached from biological life and artificially reconstructed within technical systems of potentially unlimited capacity. As Ray Kurzweil (2005, p. 45) argues, "the Singularity will allow us to transcend these limitations of our biological bodies and brains. We will gain power over our fates. Our mortality will be in our own hands." Within this horizon, thinking about an intelligent machine no longer means

merely discussing computational performance. Rather, it requires redefining what it means to exist as a machine.

The question, therefore, is not simply what an intelligent machine is, as though it were sufficient to enumerate its functional properties or algorithmic capacities. The philosophical formulation asks instead what it is like to be an intelligent machine, thereby shifting the problem from the technological to the phenomenological domain. To ask what it is like to be an intelligent machine is to investigate whether, beyond functional simulation, there exists any form of experience, intentionality, or relation to the world that could be attributed to artificial systems, or whether such categories remain inseparable from the biological and historical conditions of human existence.

At this point, two fundamental problems emerge: the concept of intelligence and the design of an intelligent machine. The first concerns the way artificial intelligence, or more precisely computational systems capable of performing specific tasks, frequently reduces intelligence to information processing, pattern recognition, and probabilistic computation. Human intelligence, however, cannot be adequately understood in these terms alone. It also involves embodiment, historicity, affectivity, intentionality, and a dynamic process of coupling with the world. By separating intelligence from nature, AI promotes a process of denaturalizing intelligence, transforming it into an abstract and operational structure that can, in principle, be artificially reproduced. This gives rise to the image of artificial intelligence as being free, precisely because it is artificial, from the constitutive limitations of biological life, thereby fostering a powerful fictional imaginary in which every physical boundary appears capable of being overcome.

The concept of intelligence, therefore, cannot be understood as an abstract, universal capacity independent of the historical conditions under which meaning is produced. Rather, intelligence should be conceived as a cultural process of semantic articulation between agents and worlds through which particular forms of perception, interpretation, and action become possible. From this perspective, intelligence designates specific modes of semantic coupling with reality. Consequently, there is no single homogeneous form of intelligence, but rather multiple forms of intelligence embedded within culturally distinct symbolic horizons. The semantic space within which an agent operates determines what can be recognized as relevant or meaningful, thereby both expanding and constraining that agent's image of the world. It is precisely this dimension that logical and computational languages appear unable to reproduce.

The second problem concerns the design of the intelligent machine itself. Can one meaningfully speak of a general design for machines? In other words, is there a normative definition of what a machine ought to be, how it ought to function, and which purposes it ought to fulfill? The design of a machine refers to the structural and functional project underlying systems capable of autonomously or semi-autonomously performing specific operations. Yet why should the design of an intelligent machine necessarily reproduce the morphology of the human body? To what extent should the design of an intelligent machine depend upon replicating the anatomical structure of human beings? Once the question is considered from the standpoint of non-human or extraterrestrial life, it becomes evident that there is no logical necessity for cognition, perception, or agency to be tied to the same anatomical and sensorimotor conditions that characterize human experience.

The contemporary super-industrial imaginary has suggested that design extends far beyond the material construction of computational devices. As Crawford (2025, p. 63) reminds us, "AI is much more than databases and algorithms, machine learning models, and linear algebra. It is metaphorical: it depends on manufacturing, transportation, and physical labor; on data centers and underground cables stretching across continents; on personal devices and the raw materials from which they are made [...]." Designing a machine has therefore come to mean prefiguring the conditions under which it operates and the modes through which it interacts with its environment. Paradoxically, although the technological imaginary frequently presents machines as abstract and deterritorialized systems, visual representations almost invariably return to anthropomorphic forms, as though the intelligibility of technical artifacts depended upon reproducing the morphology of the human body.

Within this super-industrial imaginary, the design of machines has also come to imply the disappearance of any clear distinction between humans and computers. Kurzweil (2007), for example, envisions a future in which the boundaries separating humanity and technology gradually dissolve, so that dimensions traditionally associated with human interiority, including consciousness, subjectivity, and even spirituality, become integrated into the technological domain itself. Yet what kind of imaginary presupposes the complete erasure of the boundaries between human beings and technology to the point of conceiving the fusion of human consciousness with silicon chips?

Once the design of an intelligent machine has been established, intelligence itself is expected to assume the form prescribed by that design. This is the paradox underlying the hypothesis that if the machine possesses a design, intelligence must likewise possess a design encapsulated within it. Consequently, the design of intelligence becomes even more fundamental than the design of the machine itself, for it constitutes the primary condition that guides technological development. Why, after all, should the design of an intelligent machine be modeled according to human parameters if its encapsulated functions instead simulated the design of an animal intelligence? Humanoid robots endowed with ultraviolet vision, for example, could operate according to a perceptual regime analogous to that of bees, identifying patterns invisible to the human eye and thereby radically transforming the very concept of perception. It is precisely the super-industrial imaginary surrounding Artificial Intelligence that, while promising to expand the capacities of human intelligence, ultimately reduces human experience itself to the mimetic reproduction of nature.

The Super-Imaginary of Intelligent Machines

The problem of super-imaginary shifts the question from how to build machines to how to design what will count as intelligence. Designing intelligence entails selecting in advance which aspects of human cognition will be translated into computational models and which will be excluded as irrelevant or inaccessible because they do not serve the purposes of the machine itself. In many cases, the concept of intelligence comes to signify exclusively the capacity for computation, optimization, statistical prediction, or functional adaptation. Intelligence is thereby reduced to the ability to extract results from vast amounts of information in the shortest possible time.

One aspect of this imaginary lies in the assumption that artificial intelligence is not subject to the laws of biological evolution in the same way as natural organisms. Whereas human and animal intelligence emerge from slow, contingent, and embodied evolutionary processes shaped by the demands of survival in unstable environments, artificial systems can be modified abruptly and reorganized without metabolic or temporal constraints. This assumption creates the impression of a potentially unlimited intelligence detached from the constitutive forces of organic life.

According to Korteling *et al.* (2021), although Artificial Intelligence systems have achieved significant advances in autonomy and computational capacity, they remain structured according to an operational logic fundamentally different from that characterizing human cognition. While such systems are capable of performing complex reasoning and problem-solving tasks, they do not share, at least within the current technological horizon, the conscious, intentional, and experiential dimensions that characterize biological intelligence. In this context, Korteling *et al.* (2021) argue that the similarities between artificial and human intelligence are primarily functional rather than ontological, making it necessary to develop a critical understanding of the cognitive limits of artificial agents. As AI systems increasingly occupy autonomous decision-making roles within institutional, political, and military settings, it becomes essential that those responsible for their deployment possess adequate conceptual frameworks for distinguishing the specific characteristics of artificial cognition from those of human experience.

Korteling *et al.* (2021) further observe that an influential imaginary underlying the development of Artificial Intelligence presupposes the possibility of Artificial General Intelligence (AGI) capable of reproducing, or even surpassing, the totality of human intelligence. This imaginary assumes that human intelligence constitutes the ultimate and authentic paradigm of intelligence, serving as the benchmark against which all other cognitive forms are evaluated. The very expression Artificial Intelligence, for example, reveals a conceptual and epistemological tension, as though artificially produced intelligence necessarily occupied a derivative or secondary position in relation to biological intelligence. The aspiration to create systems endowed with universally human-like capacities also reflects an anthropocentric self-understanding according to which human beings are essentially rational entities capable of solving complex problems through experience, intuition, and logical reflection. Yet if intelligence is ultimately a matter of information and computation rather than flesh, blood, and carbon atoms, is there any insurmountable physical or philosophical reason why machines could not become more intelligent than human beings in every conceivable respect?

The argument advanced by Roli, Jaeger, and Kauffman (2022), in turn, begins from the premise that general intelligence depends upon the capacity to identify and exploit new affordances, that is, opportunities for action that emerge contingently and unpredictably through the relationship between an agent and its environment. They argue that such possibilities cannot be fully predicted, enumerated, or formalized in advance because the potential uses of objects and situations remain indefinitely open to innovation. Since algorithmic systems operate upon predefined sets of rules and possible states, they are incapable of genuinely discovering new affordances: "However, a number of serious problems arise when considering the higher-level integration of task-solving capabilities. All of these problems are massively confounded by the fact that real-world situations often involve information that is irrelevant,

incomplete, ambiguous, and/or contradictory" (Roli; Jaeger; Kauffman, 2022, p. 2). Accordingly, they argue that biological cognition possesses creative and adaptive capacities that transcend the limits of algorithmic computation, making the achievement of genuine Artificial General Intelligence impossible within the currently dominant computational paradigm.

Roli, Jaeger, and Kauffman (2022) further emphasize three fundamental difficulties. First, there is the formal problem of determining which criteria should be used in any given situation to evaluate the success of an action or decision. Second, there is the difficulty of selecting, among countless available possibilities, which tasks genuinely deserve attention and which aspects of a context are actually relevant. Finally, there is an even more fundamental question concerning how to determine what ought to count as valuable or desirable. A truly general artificial intelligence would need to establish, revise, and refine its own goals autonomously, without depending upon human guidance. In other words, it would need to know what it seeks to achieve, which presupposes the capacity to possess interests, motivations, or purposes of its own. What is ultimately at stake, therefore, is the claim that only genuine agents are capable of exploiting the power of affordances and thereby generating genuinely open-ended evolutionary dynamics, unlike algorithmic attempts to reproduce such dynamics within the domain of artificial life.

Although artificial systems are trained on enormous volumes of data and interact dynamically with users and digital environments, their operation remains constrained by predefined architectures, bounded parameter spaces, and optimization structures oriented toward specific objectives. Consequently, it is not sufficient to claim that AI algorithms generate representations of knowledge and that such knowledge therefore corresponds to some state of affairs in the world. Rather, it is essential to recognize the difference between generating representations of knowledge and producing representations that are relevant, intelligible, and useful for some cognitive agent. Intelligent machines do not inhabit the world, nor are they capable of representing or reasoning metaphorically. Instead, they remain unable to cope with the uncertainty arising from contextual circumstances or, as Buijsman and Veluwenkamp (2022) argue, they introduce a form of epistemic dependency that prevents users from making genuinely critical decisions.

In this sense, the question what is it like *to be an* intelligent machine? encounters a semantic difficulty arising from the very idea of Artificial General Intelligence. What is conventionally called artificial intelligence does not constitute a form of subjective experience capable of understanding diverse contexts, but rather a collection of computational operations directed toward problem solving. If, as Roli, Jaeger, and Kauffman (2022) argue, the innovation produced by algorithmic systems remains restricted to recombinations of previously defined elements within an already delimited space of possibilities, then the machine is incapable of assigning genuinely new meanings to the world, being limited instead to reorganizing meanings that have already been established. Within algorithmic contexts, novelty can only be represented combinatorially, since an algorithm cannot discover or generate properties or relations that were absent from its original ontology. Because it operates deductively, an algorithm cannot genuinely improvise, for it is incapable of discovering new causal properties of an object that are not already implicit in its logical premises.

Intelligent Machines and Extended Functions

In discussing the philosophical foundations of Artificial Intelligence, Russell and Norvig (2010) revisit the assertion that machines may perhaps act intelligently or, indeed, may even do what they are actually thinking. Within this context, they examine three objections to the possibility of intelligent machines: the argument from disability, the mathematical objection, and the argument from informality, all of which are presented as attempts to establish limits on the capacity of machines to reproduce forms of behavior regarded as intelligent only in human beings.

The argument from disability maintains that there are various human capacities that could never be reproduced by machines. Among these are creativity, humor, moral consciousness, intuition, aesthetic experience, and the capacity to love. According to this perspective, regardless of how sophisticated a computational system may become, it would remain incapable of performing certain activities associated with the uniqueness of the human condition. Russell and Norvig (2010, p. 1177) observe, however, that such claims usually take the form of mere intuitive assertions, rarely accompanied by rigorous demonstrations: "Of course computers can do many things as well as or better than humans, including those that people believe require great insight and human understanding. However, this does not mean that computers use insight and understanding in performing these tasks." The argument from disability frequently expresses nothing more than a lack of understanding concerning the mechanisms involved in these capacities, especially the fact that computers do not perform particularly well in sustaining unrestricted conversation, as required by the Turing Test.

The mathematical objection, in turn, relies primarily on the results obtained by Gödel and Turing. The incompleteness theorem demonstrates that, within any formal system sufficiently powerful to represent arithmetic, there exist true propositions that cannot be proved within the system itself. Likewise, Turing's work reveals the existence of computationally undecidable problems for which no algorithm can provide a general solution. On the basis of these results, some philosophers concluded that the human mind could not be reduced to a computational procedure, since human beings would be capable of recognizing the truth of certain propositions that escape the limitations of formal systems. Furthermore, computers constitute finite-state systems, since they possess only a limited number of possible internal configurations. As a consequence of this structural limitation, their operation can be understood as a succession of transitions between determined states. Thus, if a computer has a limited amount of physical memory, it can store only a limited amount of information, since there are only a finite number of possible configurations for all of these components.

Finally, the argument from informality maintains that human behavior cannot be entirely captured by formal rules: "Essentially, this is the claim that human behavior is too complex to be captured by any simple set of rules and that, since computers can do nothing more than follow a set of rules, they cannot generate behavior as intelligent as that of human beings" (Russell; Norvig, 2010, pp. 1177-1178). The informality of human behavior requires constant adaptation to unpredictable contexts, linguistic ambiguities, and novel situations, characteristics that appear to elude any finite set of instructions. In response, Russell and Norvig (2010) observe that rationality does not necessarily require the explicit formulation of every possible rule, since intelligent systems could operate through adaptive

mechanisms, learning, and probabilistic strategies. Thus, the inability to make explicit all the rules governing human behavior would not imply that such behavior is, in principle, non-computable.

John Haugeland, in Artificial Intelligence: The Very Idea (1985, p. 214), in turn, argues that the competence of an agent consists in the capacity to achieve goals through rational decisions grounded in knowledge and contextual awareness, such that the attribution of beliefs, intentions, and goals must render the observed behavior intelligible. In this sense, intelligence cannot be understood solely on the basis of the semantic correctness of linguistic expressions, but requires the articulation of symbolic meaning with the pragmatic meaning of actions. Since human beings both speak and act, words and behaviors must be interpreted as coherent manifestations of the same mental structure. According to Haugeland, this requirement applies equally to artificial systems: a genuinely intelligent robot cannot limit itself to producing linguistically appropriate utterances, but must act consistently with what it expresses.

What is at stake, therefore, is that language itself constitutes a mode of action or, as Haugeland states, "the answer lies in a fact about language so fundamental that it is easy to overlook: speaking is, in itself, a form of acting" (Haugeland, 1985, p. 215). A robot, or an intelligent machine, must master the pragmatic conditions governing its use of language as well as the effects that particular utterances produce in concrete contexts. For this reason, Haugeland concludes that no system neglecting the pragmatic dimension of language could genuinely be said to speak a language, since linguistic competence presupposes not only meaning but also rational action situated in the world.

We thus return to the indicative argument: can we speak of a profound difference between acting intelligently and knowing that one is acting in such a manner without resorting to a phenomenological explanation? A genuinely intelligent machine cannot merely produce semantically appropriate utterances, but should integrate meaning and behavior, articulating its linguistic manifestations with the pragmatic demands of the world in which it operates. This means that, regardless of its internal states, it should be capable of possessing a unified understanding of the world so as to extend itself into it. Although an intelligent machine may exhibit outstanding performance in particular tasks, it remains incapable of predicting the environment and the pragmatic conditions of its interlocutors. This is so not because the machine cannot establish rules to grasp human behavior, but because social dynamics are open, contingent, and culturally situated.

For this reason, although artificial systems may simulate aspects of behavior regarded as intelligent with remarkable efficiency, the question remains whether such performance amounts to genuine understanding or merely to the sophisticated reproduction of patterns. The problem becomes even more complex when we consider biological forms of cognition radically different from human experience. Cnidarians, such as Pelagia noctiluca, possess neither a centralized brain nor a nervous system organized around a hierarchical processing center. Their interaction with the environment emerges from the coordination of epithelial cells and diffuse neural networks composed of multipolar neurons distributed throughout their bodies, exhibiting quiescent states during the night (Nath *et al*., 2017). Even so, they are capable of perceiving stimuli, directing movement, capturing prey, and adapting to environmental conditions. In the case of cnidarians, asking "what is it like to be intelligent?" does not mean investigating the existence of a unified representation of the world, but rather understanding the very bodily dynamics of coupling with the environment and the capacity to know

it through a particular mode of existence. Cnidarians perceive the world without eyes and are capable of sensing it without requiring a central processor, as is the case with human beings.

The common imaginary has anthropomorphized the concept of the "intelligent machine" by associating it exclusively with human conscious capacities. Pennartz, Farisco, and Evers (2019) propose an inside-out approach according to which internal characteristics of conscious experience can be inferred through observable indicators such as goal-directed behavior, model-based learning, episodic memory, susceptibility to illusions, visuospatial abilities, and specific neuroanatomical and physiological structures, none of which can be observed in current Artificial Intelligence systems.

The fact is that the design of intelligence does not appear necessarily to depend on a central processing unit or reflective consciousness, but rather on the capacity to respond effectively to the demands of the world. Cnidarians, like Nagel's bats, demonstrate how our conception of knowledge can be entirely illusory when we regard ourselves as its sole observers. The philosophical problem of Artificial Intelligence therefore seems to extend beyond biology itself by raising the possibility that computational processes may achieve this kind of environmental embeddedness and, consequently, cease to be confined to the formal manipulation of symbols.

An intelligent system, unlike a machine, does not necessarily require a centralized brain, just as numerous living organisms perform complex activities through distributed architectures. It is important to recall here that the autopoietic enactivism of Maturana and Varela constitutes a profound shift in the traditional explanation of life (Maturana; Varela, 1997): understanding the world is not reducible to information processing but consists in the production of meaning through the structural coupling between organism and environment, rooted in specific modes of perception, action, and embodiment. The echolocation of bats and the perceptual capacities of cnidarians illustrate this point in an exemplary manner, demonstrating that knowing the world also means inhabiting it from a singular perspective inaccessible to those who do not share the same embodied conditions of existence.

Accordingly, any account of intelligent machines, or "artificial intelligences," requires that we abandon excessively anthropocentric and centralized models of cognition, recognizing that different forms of life may produce different worlds of meaning, even in the absence of a central processor (brain). The philosophical question, therefore, is not whether a machine thinks as we do or correctly simulates thought, as Turing envisioned, but whether it may come to constitute its own horizon of meaning from which its relationship with the world ceases to be the mere execution of rules and instead becomes an authentic way of inhabiting it — just as occurs with bats, cnidarians, and many other forms of life.

Conclusions

The problem formulated by Thomas Nagel continues to provide an important point of reference for the philosophy of Artificial Intelligence. The question "what is it like to be a bat?" constitutes, on the one hand, an objection to the materialist reductionism of consciousness and, on the other, highlights the limitations of theories that identify intelligence with information processing or functional performance. An objective description of physical, biological, or computational systems does not eliminate the question of how a given being inhabits the world. From this perspective, asking "what is

it like to be an intelligent machine?" does not merely mean investigating its capacity to solve problems or perform complex tasks. Rather, it means examining whether there could be some form of pragmatic embeddedness, meaning production, and a distinctive perspective from which an artificial system might constitute its own world of meaning.

This discussion also suggests the need to revise part of the anthropocentric imaginary that shapes expectations surrounding Artificial Intelligence, particularly those that regard the brain-centered model as a kind of archetype for the construction of intelligent systems. Intelligence does not appear to constitute a homogeneous, universal property independent of the material and historical conditions in which it emerges, but rather different modes of coupling between organisms, or eventually artifacts, and their respective environments (Peruzzo Junior; Karasinski, 2023). Bats, cnidarians, and countless other forms of life indicate that knowing does not require reproducing the human cognitive architecture, but establishing singular relationships with the world. Likewise, if artificial systems eventually develop autonomous forms of cognition, this will probably not result from the mere replication of human intelligence, but from the constitution of their own modes of organization, perception, and meaning production. In other words, the technology that emulates Artificial Intelligence systems may ultimately advance its performance much further by replicating extra-human forms of life than by replicating the human brain.

The philosophical investigation of Artificial Intelligence, therefore, should shift its focus from comparing machines with human beings to examining the neurobiological, environmental, and cultural conditions that make the emergence of intelligent organisms possible. The first epistemic consequence, therefore, is to understand whether artificial systems could eventually come to inhabit a world as we inhabit it, constitute semantic horizons, and establish their own relationships with what surrounds them, or whether they will remain limited to the formal manipulation of symbols and the execution of previously structured functions. As long as these questions remain open, Artificial Intelligence programs will continue to give rise to imaginaries that extend beyond the domain of computational engineering and reach central issues in the philosophy of mind, epistemology, and ontology.

Data availability statement

The main focus of this article is contributions of a theoretical or methodological nature, without the use of empirical data sets. Therefore, in accordance with the journal's editorial guidelines, the article is exempt from being deposited in SciELO Data.

References

BUIJSMAN, S.; VELUWENKAMP, H. Spotting When Algorithms Are Wrong. *Minds and Machines*, v. 33, n. 4, p. 541-562, 2022. DOI: 10.1007/s11023-022-09591-0.

CRAWFORD, K. *Atlas da IA*: poder, política e custos planetários da inteligência artificial. São Paulo: Sesc Edições, 2025.

EUROPEAN COMMISSION. EU-U.S. *Terminology and Taxonomy for Artificial Intelligence*. 2023. Disponível em: <https://digital-strategy.ec.europa.eu/en/library/eu-us-terminology-and-taxonomy-artificial-intelligence>. Acesso em: 4 maio 2026.

HAUGELAND, J. *Mind Design*: Philosophy, Psychology, and Artificial Intelligence. Bradford Books, 1981.

HAUGELAND, J. *Artificial Intelligence*: the Very Idea. Cambridge: MIT Press, 1985.

HOPSTER, J. What are socially disruptive technologies? *Technology in Society*, v. 67, p. 101750, 2021. DOI: 10.1016/j.techsoc.2021.101750.

KORTELING, J. E. H.; et al. Human versus Artificial Intelligence. *Frontiers in Artificial Intelligence*, v. 4, art. 622364, p. 1-13, 2021. DOI: 10.3389/frai.2021.622364.

KUDINA, O.; VAN DE POEL, I. A sociotechnical system perspective on IA. *Minds and Machines*, v. 34, art. 21, p. 1-9, 2024. DOI: 10.1007/s11023-024-09680-2.

KURZWEIL, R. *The Singularity Is Near*: When Humans Transcend Biology. London: Penguin Books, 2005.

KURZWEIL, R. *A era das máquinas espirituais*. São Paulo: Aleph, 2007.

LEGG, S.; HUTTER, M. Universal Intelligence: A Definition of Machine Intelligence. *Minds & Machines*, v. 17, p. 391-444, 2007. DOI: 10.1007/s11023-007-9079-x.

MATURANA, H.; VARELA, F. *De máquinas e seres vivos*: autopoiese; a organização do vivo. São Paulo: Artes Médicas, 1997.

NATH, R. D. *et al*. The Jellyfish Cassiopea Exhibits a Sleep-like State. *Current Biology*, v. 27, p. 2984-2990, 2017. DOI: 10.1016/j.cub.2017.08.014.

NAGEL, T. What Is It Like to Be a Bat? *The Philosophical Review*, v. 83, n. 4, p. 435-450, 1974.

PENNARTZ, C. M. A.; FARISCO, M.; EVERS, K. Indicators and Criteria of Consciousness in Animals and Intelligent Machines: An Inside-Out Approach. *Frontiers in Systems Neuroscience*, v. 13, art. 25, p. 1-23, 2019. DOI: 10.3389/fnsys.2019.00025.

PERUZZO JUNIOR, L.; KARASINSKI, M. Cognitive Artifacts and Human Enhancement. *Ethics in Science and Environmental Politics*, v. 23, p. 45-52, 2023. DOI: 10.3354/esep00208.

PERUZZO JUNIOR, L. Artificial intelligence, extended cognition, and the narratives of cyberimmortality. *Unisinos Journal of Philosophy*, v. 26, n. 1, p. 1-13, 2025.

ROLI, A.; JAEGER, J.; KAUFFMAN, S. How Organisms Come to Know the World: Fundamental Limits on Artificial General Intelligence. *Frontiers in Ecology and Evolution*, v. 9, art. 806283, p. 1-14, 2022. DOI: 10.3389/fevo.2021.806283.

RUSSELL, S.; NORVIG, P. *Artificial Intelligence*: A Modern Approach. New Jersey: Prentice Hall, 2010.

ULNICANE, I.; ERKKILÄ, T. Politics and Policy of Artificial Intelligence. *Review of Policy Research*, v. 40, n. 5, p. 612-625, 2023. DOI: 10.1111/ropr.12574.

VATER, M.; KÖSSL, M. The ears of whales and bats. *In*: THOMAS, J. A.; MOSS, C. F.; VATER, M. (eds.). *Echolocation in Bats and Dolphins*. Chicago: Chicago University Press, 2004.

VOIGT, C.; LEWANZIK, D. No cost of echolocation for flying bats revisited. *Journal of Comparative Physiology B: Biochemical, Systemic, and Environmental Physiology*, v. 182, p. 831-840, 2012. DOI: 10.1007/s00360-012-0663-x.

Editores responsáveis: Léo Peruzzo Júnior e Jelson Oliveira.

RECEBIDO: 20/07/2026
APROVADO: 07/09/2026

RECEIVED: 07/20/2026
APPROVED: 09/07/2026