Journal of Inclusive Methodology and Technology in Learning and Teaching
ISSN 2785-5104
Anno 6 n. 4 (2026)

EDIZIONI UNIVERSITARIE ROMANE
editoria scientifica dal 1952

Beyond the Boundaries of the Mind? Artificial Intelligence, subjectivity, and the Externalization of Human Faculties

Marco Piccinno [1]

[1] Università Telematica Pegqaso; marco.piccinno@unipegaso.it
[2]

Abstract: The personal and social dimensions of the contemporary world are increasingly called upon to engage with the emergence of artificial intelligence.
From a strictly epistemic perspective, there thus emerges a need to inquire into the "nature" of artificial intelligence, with the aim of identifying possible interpretive frameworks capable of clarifying the significance it may assume for personal identity.
In other words—and this is the focus of the present reflection—the meaning of artificial intelligence cannot be separated from the definition of what the mind is. Consequently, any inquiry seeking to address this issue cannot avoid a deeper examination of the relationship between the two.
From this perspective, the possible trajectories along which the issue may be developed are certainly manifold. Nevertheless, it is possible to identify a number of conceptual cores around which the discussion may be organized: a) Is the mind exclusively a correlate of the central neurocerebral structure, or can it also be "embedded" in structures other than neuronal substrates? B) Is it still plausible to interpret the mind–brain connection on the basis of the hardware–software analogy? c) How, and through what processes, can material neural connections generate immaterial contents such as thoughts, feelings, beliefs, and so forth (the problem of *qualia*)? d) Is the capacity to "intend" the objects of the world—that is, semantic capacity—specific only to the human mind, or can it also be extended to the digital devices that support AI?

Keywords: Artificial intelligence; extended mind; cognitive externalization; intentionality; consciousness

Copyright: © 2026 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Neurons, qualia, and functional organization

The problem of consciousness concerns the capacity of a human subject to be aware of personal states and of the surrounding environment. Its central difficulty arises from the relationship between two different descriptions: neural processes, which operate without being conscious in themselves, and the mental representations through which perceptions become objects of awareness. External stimuli reach sensory receptors, travel through peripheral and central nervous pathways, and are processed by activities that are not experienced directly. Nevertheless, their outcome may be a conscious image, sensation, thought, or feeling. The question is how such a transition can be understood without confusing the physiological process with the experience to which it is related.

This difficulty is expressed by the problem of qualia a term used to describe the subjective reverberations elicited by sensory stimuli. Particularly significant in this regard are Jackson's thought experiments, which bring into focus what a physical account of sensory processing appears to leave unexplained. A stimulus can be described in terms of neural receptor and transmitter, yet the pain of a sting, the taste of food, or the distinct appearance of a colour is experienced from a first-person perspective. Similarly, a landscape may be accompanied by joy, melancholy, serenity, or unease. The emphasis falls on the qualitative and individual form in which a content is present to a subject, notwithstanding the common physiological structures involved in perception.

Dennett's account (1981, 1991) challenges the assumption that qualia constitute independent mental entities exceeding the processes that make behaviour possible. In his approach, stimuli are processed through multiple interacting pathways, and their effects are linked to memories, words, decisions, and movements. Consciousness is not explained by postulating a single cerebral location at which all information is finally displayed. Rather, the significance of a content depends, according to the author, on the connections it establishes with different operations. A sound, for example, can turn the head, prompt a verbal description, revive the memory of an earlier event, and support the decision to move towards or away [from the source]. From this perspective, conscious content is explained in terms of this complex network of dispositions and effects.

In particular, Dennett's model, which he calls *Multiple Drafts* (1991), is based on the parallel processing of sensory information. Stimuli from the external world enter consciousness, but not because they are channelled into a particular structure within the central nervous system that unifies them and makes them conscious. In this sense, the procedural components of the process can be identified in the pathways that shape the conscious perception of external objects, such as an apple. Sensory activity discriminates shape, colour, size, and position; further processing connects the object with recognition and previous experiences of taste or smell; linguistic, decisional, and motor activities may then lead a person to name the apple, choose to eat it, and reach for it. The explanatory emphasis is placed on processes and on the availability of their results for other activities, rather than on a mental content that is separate from and independent of that process. In this account, Dennett also considers the importance of cultural contents, which he terms memes and which, in his view, can be compared to biological genes. They consist of expressions, customs, gestures, proverbs, melodies, and patterns of behaviour shared by individuals within the same cultural and social context, which can be preserved, replicated, and modified through exchanges among those individuals.

Una diversa concezione è proposta da Chalmers (1996, 2010), che distingue la descrizione fisica o funzionale di un sistema dal suo versante fenomenico. I processi funzionali possono essere descritti mediante relazioni fra input, transizioni interne e output. L'esperienza cosciente, tuttavia, non viene semplicemente identificata con tali operazioni. Chalmers propone principi che collegano organizzazione funzionale ed esperienza, considerandole come aspetti distinti. Il punto fondamentale non è che una sequenza di operazioni spieghi meccanicamente come si presenta l'esperienza, bensì che debba essere considerata una relazione sistematica fra organizzazione funzionale e stati coscienti. Questa posizione si distingue sia dalla completa

separazione di mente e corpo sia dalla riduzione dell'esperienza soggettiva al comportamento osservabile.

The principle of organizational invariance occupies a central place in this argument. If the fine-grained causal organization of a system remains unchanged, Chalmers argues that corresponding experiences should also remain unchanged. The imagined progressive replacement of neurons with silicon elements is used to discuss fading qualia: a subject might retain the same functional capacities while an alleged alteration of subjective experience went unnoticed. The related example of dancing qualia imagines switching between functionally equivalent neural and silicon arrangements. If experiences changed radically despite every relevant functional relation remaining unchanged, the subject would be unable to register the change through ordinary cognitive processes.These thought experiments are employed to defend a link between organization and experience, not to demonstrate that a replacement has actually been performed.

Information plays another significant role in Chalmers's analysis. It concerns the differentiation of possible states within a system and the way in which such differentiation can be linked to physical processes and phenomenal properties. The example of an electrical switch illustrates how the selection of one state rather than another has distinct effects on a circuit. Yet the discussion also confronts a difficulty: how can a general principle connecting information and experience avoid indiscriminately attributing consciousness to every elementary device? The possibility of exceptionally simple forms of experience and the criteria for relating information to consciousness remain questions raised by the theory. The account therefore makes explicit both the proposed functional correspondence and its unresolved boundaries.

2. Beyond and between neurons: computation, body, and intentionality

The search for an account of consciousness that cannot be exhausted by neural descriptions is on the basis of Penrose's work (1989, 1994). His argument considers the limits of computation through questions associated to Gödel's incompleteness theorem and Turing's analysis of computability. The contrast concerns the execution of formal rules and the understanding of what they can guarantee. A machine performs the procedures specified for it; human mathematical understanding, in Penrose's view, cannot be fully represented by by adopting essentially algorithmic criteria. The philosophical claim is that conscious insight cannot simply be equated with the mechanical application of a fixed set of instructions.

The significance of these limitations becomes particularly apparent in what is known as the "halting problem". Penrose proposes assigning a machine a task that, although it falls within a computational domain, is inherently unsolvable: finding an odd number that is the sum of two even numbers. He observes that the machine uses its computational capabilities to solve this problem, yet remains unable, even after countless attempts, to understand that it is unsolvable. The machine thus repeats its computational operations indefinitely because, despite being able to execute them, it cannot understand their limitations. Penrose, by contrast, regards this capacity as distinctive of the human mind and, in his view, as marking an insurmountable difference between the two. The example allows Penrose to further

clarify the nature of this difference, which essentially comes down to the gap between applying a rule and understanding its limitations.

At the same time, this thought experiment gives Penrose an opportunity to assert the non-computational nature of the human mind, precisely because of its capacity to recognise the limits of computation. Knowledge of these limits does not arise from computational processes themselves but transcends their capabilities in a fundamental and irreducible way. In other words, according to Penrose, the human mind cannot be equated with that of a machine because the former is aware of itself and its potential, whereas the latter is not.

Penrose extends such considerations to the more general question whether conscious mathematical understanding is computational. His negative answer leads him to seek processes of a different kind, not reducible to conventional algorithmic organization.

Building on these premises, Penrose seeks the physiological foundations of consciousness, which he believes can be found in neural processes associated with quantum physics. He therefore focuses his research on structures within neurons, particularly the subneuronal structures of the cytoskeleton and microtubules, rather than solely on communication between neurons.. The proposed significance of such structures lies in the attempt to locate a physical basis for activity that, in his view, cannot be adequately describedby by neuron-based account. The theory therefore does not detach consciousness from the material world; it looks for material processes of a kind different from those ordinarily invoked in accounts of neural information processing.

Searle's account (1980, 1998) gives further prominence to the first-person and qualitative character of consciousness. Conscious states are not an independent substance separated from the brain; they belong to the biological life of the subject and arise through neural processes. At the same time, Searle refuses to eliminate their subjective nature from the account. Consciousness is an objective feature of biological reality in the sense that it exists, yet it has a subjective mode of existence because it is experienced by someone. The explanatory problem concerns the emergence of first-person phenomena from neurobiological activity. Thus the material origin of consciousness and the irreducibility of its subjective description are held together within the same approach.

This position is linked to the critique of the analogy between the mind and a computer program. In the Chinese-room thought experiment, a person who does not know Chinese receives Chinese symbols and uses formal instructions to combine them into apparently appropriate answers. Correct manipulation of the symbols does not mean that the person understands their meaning. Searle uses the example to question whether the execution of syntactic rules is sufficient for semantic understanding and reaches the conclusion that computation, understood as rule-governed manipulation of formal elements, does not by itself establish intentionality: the capacity of mental states to be directed towards things, events, or states of affairs. The difference between syntax and semantics therefore occupies a central place in his argument.

For Searle, mental states can be about objects in the world, while the status of a physical configuration as a computational symbol depends upon an interpretation. Electrical activity in a machine is not, simply by virtue of being electrical activity, a meaningful sentence or a mathematical assertion. The interpretation of a

configuration as a symbol is assigned within a system of meaning attribution. The account of Author accordingly distinguishes the causal operations of a machine from the intentional content of a conscious state. Searle does not deny that a brain can be described computationally for certain purposes; he denies that such a description, on its own, explains consciousness or supplies the semantic character of mental life.

3. Mental representations and the conditions of their validity

The discussion of consciousness leads to the further question of the validity of mental representations. A subject may have a representation of an object, event, or state of affairs, but the fact that the representation is consciously held does not by itself show that it is true. The epistemological problem concerns the criteria by which beliefs and propositions can be justified. In this regard, reflection on the subject has identified four distinct criteria that the mind uses to validate its constructs: correspondence, justification, coherence, and practical consequences These approaches are not concerned with establishing how mental contents arise; rather, they examine the conditions under which those contents can be validated.. Distinguishing the origin of a representation from its justification is therefore indispensable to the issue under examination.

The correspondence account relates the truth of a proposition to the state of affairs to which it refers. A statement about the world is considered true when things are as the statement says they are. This formulation gives the represented object or event a determining role in the appraisal of a belief. It also raises difficulties concerning the relationship between linguistic propositions and the reality described by them. A representation is not identical with its object: it employs particular expressions, classifications, and points of view. The problem is thus not simply the existence of an external world but the specification of the relation that makes a proposition adequate to it..

Foundationalism is based on the idea that mental contents, which it identifies as "beliefs", fall into two categories: basic beliefs and "derived" beliefs. Basic beliefs consist of self-evident mental contents that are immediately apparent as such, rather than established through reasoning. Within this conception, these beliefs provide the foundation for further propositions, giving rise to a structured system of knowledge. independently of argumentative processes. This position also fails to consider that what seems immediate to one person may be shaped by habit, expectation, or common sense. Even beliefs accepted because they have been handed down through tradition may appear unquestionable in the absence of independent justification. The search for a secure foundation must therefore clarify what distinguishes genuinely basic justification from the mere persistence over time or psychological force of a conviction.

Coherentist theories shift the criterion of justification from basic beliefs to the relations among beliefs within a system (Hempel, 1950). A proposition gains epistemic standing according to the way it fits with other propositions, whether it avoids contradiction and whether it contributes to the intelligibility of the whole. This does not require the identification of one proposition that justifies all others from outside the network. Instead, justification is examined through the reciprocal

relations of the elements composing a corpus of knowledge. The same belief may be assessed differently when the contexs is changed, since, from this perspective, what matters is the relationship each proposition bears to all the others. The coherence of a system thus becomes a decisive feature of epistemic evaluation.

Pragmatism, finally, grounds the validity of mental representations in the observable consequences they produce in the contexts in which they are adopted (Peirce, 1905; Dewey, 2014). From this perspective, the meaning of a proposition is related to the particular outcomes it generates in experience. The basic criterion, therefore, lies neither in subjective preference, nor in coherence or self-evidence, but in what is experienced in practice as a consequence of the mental representation. The practical criterion also has epistemic limitations. Clearly, a belief can be useful without being strictly true. Moreover, the criterion of utility proposed by pragmatism is decidedly contextual. An idea that produces favourable consequences in certain circumstances may prove inappropriate in others. Likewise, a false statement may yield practical benefits, while a true one may lack immediate utility.

References

Chalmers, D. J. (1995). Facing up to the problem of consciousness. Journal of Consciousness Studies, 2(3), 200–219.

Clark, A. (2025). Extending minds with generative AI. Nature Communications, 16, Article 4627.

Clark, A., & Chalmers, D. J. (1998). The extended mind. Analysis, 58(1), 7–19.

Dennett, D. C. (1991). Consciousness explained. Little, Brown and Company.

Floridi, L. (2023). Etica dell'intelligenza artificiale: Sviluppi, opportunità, sfide. Raffaello Cortina Editore.

Floridi, L. (2025). AI as agency without intelligence: On artificial intelligence as a new form of artificial agency and the multiple realisability of agency thesis. Philosophy & Technology, 38, Article 30.

Piccinno, M. (2026). Che cosa ci siamo messi in testa. Le rappresentazioni mentali tra pedagogia, filosofia e neurosceinze. Pensa

Searle, J. R. (1980). Minds, brains, and programs. Behavioral and Brain Sciences, 3(3), 417–457.