v1: 31 May 2026

Commentary

Can There Be Meaning Without Conscious Experience? Why Embodiment May Not Suffice for AGI

Preprinted: 21 December 2024
Peer-approved: 31 May 2026

Qeios, Vol. 8 (2026)
ISSN: 2632-3834

Marco Masi[1]

1. Independent researcher

The recent developments in artificial intelligence (AI), particularly in light of the impressive capabilities of transformer-based Large Language Models (LLMs), have reignited the discussion in cognitive science regarding whether computational devices could possess semantic understanding or whether they are merely mimicking human intelligence. Recent research has highlighted limitations in LLMs' reasoning, suggesting that the gap between mere symbol manipulation (syntax) and deeper understanding (semantics) remains wide open. While LLMs overcome certain aspects of the symbol grounding problem through human feedback, they still lack true semantic understanding, struggling with common-sense reasoning and abstract thinking. This paper argues that current debates about LLM grounding often conflate operational semantic competence with intrinsic semantic understanding. While embodiment, multimodality, human feedback, and world-model learning may improve functional grounding, they do not by themselves explain why symbols or vectors should become meaningful for a subject. True meaning-making also may demand a connection to subjective experience, which current AI lacks. The path to artificial general intelligence (AGI) must address the fundamental relationship between symbol manipulation, data processing, pattern matching, and probabilistic best guesses, on the one hand, and true knowledge that requires conscious experience, on the other. I therefore defend a conscious-semantics thesis: intrinsically grounded meaning plausibly requires phenomenal consciousness. A transition from AI to AGI could necessitate semantic understanding, which is closely tied to subjective experience. This is an invitation to take the phenomenological first-person perspective seriously and to realize that intrinsically grounded meaning, as opposed to derived, relational, or operational semantics, is embedded in qualia. Recognition of this connection could furnish new insights into longstanding practical and philosophical questions for theories in biology and cognitive science and provide more meaningful tests of intelligence than the Turing test.

Correspondence: papers@team.qeios.com — Qeios will forward to the authors

 doi.org/10.32388/DN232Y.7

Introduction

Theories of language and meaning are not new. One could begin with Ferdinand de Saussure's fundamental insights about the meaning of signs and words as arising from their relationships with other signs and words. The structure of language is based on these relationships and differences, without which meaning could not emerge. This approach laid the foundation for modern structural linguistics and semiotics. Later, Ludwig Wittgenstein, in his masterpiece of logical positivism, *Tractatus Logico-Philosophicus*, emphasized that meaningful statements must be rooted in the clarity of concepts within a logical system of propositions. He described language as mirroring the world of facts (not things), developing a "picture theory" of language as a description of our experiences of these facts. The later Wittgenstein departed from the Tractatus in his *Philosophical Investigations*, realizing that language derives its meaning not only from logical structures but also from its contextual dependency. Language mirroring the world and facts is specific to social and practical contexts, and the meanings it conveys must have inherent flexibility and vagueness.

Edmund Husserl, the German philosopher who is known as the founder of 'phenomenology' (that is, the study of appearances and how they express themselves to our conscious experience), was instead more interested in getting rid of all mental presuppositions by positing phenomenal statements as the grounding basis, including all the observed physical phenomena, the sensations of our bodies, and mental events as we actually experience them and how they present themselves to us. To do this, according to Husserl, we must resort to what he called the 'phenomenological reduction' method—that is, we should refrain from trying to explain the phenomena by something which is supposed to be non-phenomenal or beyond-phenomena, as all that we know for sure is only the experience of phenomenal consciousness. For example, we should adopt the 'epoché' attitude—that is, an attitude that refrains from conjecturing if there is a world existing independently from our consciousness. This epoché helps us to reduce the world to pure phenomena, abstract from extra-mental things supposedly explaining things by non-phenomenal entities.

In Husserl's phenomenology, "noesis" is the subjective act of consciousness (e.g., perceiving, judging, imagining), while the "noema" is the intentional content or object as it is experienced in that act. The distinction marks the correlation between the act (how consciousness is directed) and its object-as-meant (what is given in that directedness), rather than a separation between mind and an external thing. This mental act has the characteristic property of 'intending' or 'being about' an object in its mental dimensionality, something he called 'intentionality', borrowing the term from the German philosopher, psychologist, and Catholic priest Franz Brentano, whose theories influenced Husserl.

The noesis–noema correlation maps onto a core linguistic distinction: the noesis parallels the speaker's act of meaning, while the noema corresponds to the meaning-content as structured and identifiable in what is said and its propositional content. Saussure's signifier–signified relation resonates with Husserl's idea that meaning is not just an external object but something constituted in consciousness-related intentional acts.

Others, like Martin Heidegger, Hannah Arendt, Jean-Paul Sartre, Maurice Merleau-Ponty, Jacques Derrida, Hubert Dreyfus, and others followed in Husserl's footsteps, emphasizing how meaning arises within the very structure of

doi.org/10.32388/DN232Y.7

conscious experience. There is no "meaning" floating in the world independently.

In particular, Hubert Dreyfus, one of the later phenomenologists witnessing the rise of IT, was among the most prominent philosophical critics of AI. He challenged the fundamental assumptions of mainstream AI research, arguing that human intelligence is essentially embodied and situated within a meaningful world, and thus cannot be replicated by machines operating solely on formal rules and symbolic manipulation[1]. Although some of his predictions proved incorrect (e.g., that computers would never master chess or facial recognition), as we shall see, his broader critique articulated in terms of embodiment and what he referred to as "contact theory" with the world, anticipated later developments in embodied and enactive approaches to cognition. What is now often described as "embodied cognition" or "sensorimotor embodiment" reflects a line of thought that has become increasingly central in contemporary debates.

While phenomenology grounds meaning in embodied, lived experience and intersubjectivity, Noam Chomsky's theory of "transformational-generative grammar" championed the view that language is grounded in an innate, formal computational system. Language is a product of the mind, an idea based on a biolinguistic conception, in which the human ability to use complex forms of language is pre-wired in the brain, in a "language acquisition device." This premise led him to hypothesize the existence of universal grammar, with core syntactic linguistic knowledge being genetically inherited.

Saussure's, Husserl's, Dreyfus', Wittgenstein's, and Chomsky's theories are just a few examples of the many competing theories that have emerged over the years. However, a definitive answer to the question of the nature of language and meaning remains elusive.

Similar questions arose in the field of IT. In the midst of phenomenologists' reflections, and shortly before Wittgenstein published his Investigations, Claude Shannon, the modern father of information theory, was contemplating the notion of information. What does the word "information" mean? The etymology of words is often more insightful than our contemporary understanding of them. The word "information" derives from the Latin "informare," which suggests that something has been formed or shaped–by molding, carving, or puncturing an object into a pattern or by modifying its physical, internal, or external state. It is about forming or changing something, making it a medium for symbols that convey a message, expressing our thoughts to someone who can understand those symbols and refer to them meaningfully. Whether it is letters and words on paper or bits in a computer, information is always about forming patterns or modifying internal states in objects.

On the other hand, the information we receive also "forms" and "shapes" ideas in our minds. There is, therefore, an important distinction to be made between physical information and semantic information. Physical information conveys a message in the form of symbols, signs, and tokens, and in communication theory, it is quantified by the Shannon information measure. However, a sequence of symbols and its quantification in bits are, in and of themselves, not meaningful unless a mind apprehends them and "collapses" them into a meaningful semantic whole. Physical information has no meaning whatsoever if there is no mind receiving the message and translating it into coherent thoughts.

Shannon was acutely aware of this distinction. In his seminal 1948 paper, "A Mathematical Theory of Communication," he pointed out: *"The fundamental*

doi.org/10.32388/DN232Y.7

problem of communication is that of reproducing at one point either exactly or approximately a message [a sequence of discrete symbols] selected at another point. Frequently the messages have meaning; that is they refer to or are correlated according to some system with certain physical or conceptual entities. These semantic aspects of communication are irrelevant to the engineering problem"[2].

In physics and statistical theory, several other definitions of information exist. However, most of these information measures are closely related to Shannon information, and none of them quantifies semantic content.

Unlike a mathematical or physical theory of communication, semantic information concerns the meaning of a message. The idea that meaning cannot be reduced to pattern recognition alone and that a physical representation of something is fundamentally different from the thing being represented becomes intuitively evident with the famous Gestalt figures. For example, the popular Rubin's vase-face figure, where our mind switches between the visual interpretation of a vase and two faces, shows how patterns *informed* into material structures (the figure on a piece of paper) have no meaning in and of themselves. A meaning-making mind must convert physical information into a coherent semantic whole and eventually even make a choice between mutually exclusive ones. It is a cognitive process that highlights how the significance of things is not inherent in the things themselves "out there" but rather emerges as a perceived content in us.

For about three decades, these questions were largely ignored as irrelevant philosophical subtleties. Unfortunately, physical information was often conflated with semantic information, and both were simply referred to as "information" without further distinctions. However, the progress of IT, the ever-increasing (physical) information processing power of computers, the advent of AI, and new findings in neuroscience have forced us to reconsider these simplistic assumptions. Questions about the nature of consciousness and the mind have resurfaced, and the relationship between computational and mental states has become a matter of debate.

However, what are consciousness and semantics?

Consciousness in biological systems is widely believed to emerge from complex, integrated activity within the brain. Rather than residing in a single structure, it is thought to arise from the dynamic coordination of sensory processing, memory, attention, and self-representation across distributed neural circuits. Theories such as Integrated Information Theory[3] and Global Workspace Theory[4] propose that consciousness depends on both the quantity and organization of information processing—where integration and accessibility of information are key. This idea relies on the assumption that in biological organisms, the neural architecture enables a unified experience of perception and thought, giving rise to the subjective awareness we call consciousness. Others embrace metaphysical ontologies like dualism, panpsychism, or idealism. However, the true origin and nature of consciousness remain largely debated. (For comprehensive reviews on the subject, see: [5][6][7]).

Thus, there is no universally accepted definition of consciousness. Nonetheless, in the present context, the term 'phenomenal consciousness' suffices to illustrate our point. Phenomenal consciousness refers to the subjective, qualitative aspects of experience—often described as 'qualia'—which include raw sensory perception like the redness of blood, the pain of a toothache, and the sweetness of sugar. Phenomenal consciousness encompasses more than just sensory

qualia; it includes all conceptual and nonconceptual experiences related to time, space, causality, the body, the self, and the world. Or, to frame it in Thomas Nagel's famous "what it is like to be" formulation: an entity is conscious if there is a subjective experience associated with being it[8]. From now on, we will simply refer to phenomenal consciousness as consciousness.

It is important to note that consciousness is exclusively a first-person subjective experience that cannot be fully captured through third-person empirical investigation. While I cannot demonstrate that others are conscious, I am aware that I am conscious, and therefore, I have no reason to doubt that other human beings like me are also conscious.

What is semantics? Here, we adopt a general understanding. Even though we will focus on the linguistic dimension of semantics due to the disruptive impact of LLMs, we refer not only to formal linguistic competence, which encompasses lexical semantics (the retrieval of individual word meanings) and compositional semantics (the construction of meaning from multi-word utterances), but also to a type of 'non-verbal semantics' grounded in world knowledge and functional linguistic competence, applicable in non-verbal contexts (such as extracting meaning from a picture). This can be described as 'general conceptual knowledge,' which does not necessarily rely on linguistic inputs but is crucial for fluent language use. While LLMs excel in formal competence, they often struggle with the functional aspects of language[9]. In the following, we will focus on the complex reasoning that relies on non-linguistic or 'pre-linguistic' world knowledge and that determines functional linguistic skills, subsuming it under the label of 'true semantic understanding' or just 'semantics.'

The Explanatory Gap Between Symbols and Semantics

In 1980, one of the most debated arguments against computationalism came from John Searle, who formulated his celebrated "Chinese Room Argument." This thought experiment was designed to challenge the notion that a Turing machine–and, by extension, any (non-quantum) computer or AI system running an algorithm–can truly understand something simply by processing symbols[10]. Searle, who does not understand Chinese, imagined himself locked in a room with a set of rules (in English) for manipulating Chinese symbols. By receiving Chinese characters as input and following the rules to produce the correct output in Chinese, he could mimic a native Chinese speaker, though without understanding what the Chinese symbols actually referred to–that is, what they meant. This thought experiment was designed to show that, no matter how perfect a simulation of understanding might be, it does not imply genuine comprehension of the language. One is merely following the syntactic rules of a program without any grasp of meaning (semantics). It highlights the distinction between syntax (symbol manipulation) and semantics (meaning), and the possibility of correctly processing symbols without understanding what they mean.

Thus, it raises the question of whether machines, while capable of the former, are also capable of the latter. There is only the signifier without the signified, in the sense that a Turing machine manipulates symbols according to syntactic rules (an algorithm) without having either an abstract conceptual designation or any real-world referents.[1]

Based on these arguments, Searle introduced the distinction between “strong AI”—nowadays referred to as “Artificial General Intelligence” (AGI)—and “weak AI”[11]. There is no consensus on what exactly “general intelligence” means.

There are a variety of definitions, but broadly speaking, AGI refers to a hypothetical form of AI capable of performing any intellectual task that a human can. It is commonly understood as a system that can learn and apply its intelligence, exhibiting broad and flexible generalization across multiple domains and real-world task competence, with human-like cognitive abilities. However, most current conceptions of AGI treat intelligence in functional and computational terms, without any commitment to subjective experience.

Here, as will become clear throughout this essay, general intelligence refers to a system possessing what I like to call “semantic awareness”—that is, a form of cognition capable of reasoning and generalization abilities based on a genuine semantic comprehension (in a sense that will be clarified later) of language, data, and, eventually, sensory inputs. Such a system would go beyond the mere emulation currently achieved by symbol manipulation, best-fit algorithms, or next-token predictions.

It is therefore useful to distinguish between “functional AGI,” which exhibits sophisticated information processing and behaviorally indistinguishable performance, and “phenomenal AGI,” which would possess genuine semantic awareness grounded in subjective experience rather than mere syntactic manipulation.

Nobel laureate Roger Penrose also expressed his doubts in his seminal book *“The Emperor's New Mind: Concerning Computers, Minds, and the Laws of Physics,”* where he laid out his argument that the mind can’t be equated with a Turing machine and emphasizes the role of non-algorithmic processes. His ideas were influenced by Gödel's incompleteness theorems, which demonstrate the existence of true mathematical statements that cannot be proven algorithmically within any formal system. While he does not provide a singular, definitive explanation, Penrose contends that conscious understanding isn’t computational. The quality of understanding requires awareness and arises from the conscious mind's ability to perceive and interpret reality in ways that transcend purely mechanistic or computational systems, suggesting that consciousness involves non-algorithmic processes that are tied in particular to quantum mechanics[12].

In the 1990s, Searle’s thought experiment and Penrose’s argument were reinforced by the so-called “symbol grounding problem,” discussed by Harnad[13]. This issue highlights the difficulty of explaining how symbols in computational systems—whether words, numbers, streams of bits, signals, or more complex representations—can carry meaning without being grounded in sensory experiences or real-world interactions. A gap remains between functional symbols and number-crunching processes on the one hand and meaningful mental states on the other. The divide between syntax and semantics persists.

The question becomes even more complex when we consider that symbols do not always refer to concrete objects or real-world phenomena but can also represent abstract and intangible concepts like “beauty,” “justice,” or even the concepts of “abstraction” and “meaning” themselves.

Harnad raises the issue that we do not really know what “meaning” itself is. We cannot simply assume that semantics can be reduced to a form of computation, as computation follows specific syntactic, logical, and mathematical rules based

on symbol manipulation—not their meanings. He referred to this as the "symbol/symbol merry-go-round"—the idea that symbols can refer only to other meaningless symbols. Avoiding this infinite regress requires a bridge between those representations and the things or concepts to which they refer.

Even for a Turing machine that convincingly passes the Turing test—that is, its cognitive processes are indistinguishable from those of a human—the question would remain: Does it meaningfully connect its internal representations to their referents and their causal relationships, or is it a mere Searle's Chinese room mimicry, without any real semantic awareness of what they mean? Does it truly understand? What does it mean to "truly understand"? Harnad realized that *"there is a difference between inert words on a page and consciously meaningful words in our heads"*[14].

Thus, following mechanical rules in symbol manipulation is one thing, and what humans mean by "understanding," "knowing," or "comprehension" is another. Semantics remains outside the formal description of any computational model.

The question, then, is: If symbols are meaningless in and of themselves, what generates meaning? While Shannon's notion of physical information is conceptually and mathematically well-defined, it is unclear how to define semantic information and even less clear how minds convert the former into the latter. How do neural activity patterns supposedly constitute and implement meaning? How do symbols and signs encoded in language elicit meaning in the brain through their reciprocal exchange? It seems that a mind is required as an interface between Shannon information (which refers to physical patterns or states in a medium) and semantic information (which meaningfully grounds the symbol to its referent in a much less tangible mental domain).

The most promising approach to answering these questions and aiming at "naturalizing meaning" seemed to be representationalism—an attempt to explain mental states and their contents in terms of representations and intentionality. However, decades of research in cognitive science and philosophical debates aimed at connecting mental processes that manipulate representations of external reality with cognition, perception, and consciousness have led to little substantial progress.[2] Currently, non-representational theories emphasizing cognition as deeply rooted in bodily experiences and actively engaged with the environment are receiving more attention. (We will address these later.) The advent of Large Language Models (LLMs) and their limitations suggests that this must be the case. However, I will argue that while embodied cognition—that is, cognitive processes rooted in interactions with the environment—might provide valuable insights into the relationship between computation, symbols, representations, and sense-making semantics, it can, at best, be only a necessary but not sufficient condition. Phenomenal consciousness and its whole psychological dimension cannot be sidestepped, as it is inextricably intertwined with meaning-making.

Meaning in a Time of Large Language Models

With the advent of transformer-based LLMs, particularly in their popular form as ChatGPT, these questions elicited renewed interest. LLMs are designed to generate human-like text based on vast amounts of data. Using deep learning neural networks, they predict and generate word sequences by learning patterns, grammar, context, and semantics from large datasets. Trained on diverse sources, including books, websites, and conversations, LLMs can answer

questions, complete texts, translate languages, summarize information, and engage in natural dialogue. We have all been impressed by their abilities, which seem to suggest that they can reason effectively and possess a degree of semantic understanding.[3] The deep learning architecture underpinning these models is the transformer, whose key feature is its ability to effectively process and uncover long-range dependencies in input data. In simple terms, transformers in LLMs process and generate text by learning the probability distribution of long-range sequences of words and predicting the next word based on those previously generated. The transformer architecture has proven to be exceptionally powerful, especially in tasks that seem to require some level of semantic understanding. This fact challenges the notion that symbol manipulation alone is insufficient for cognitive states involving deep understanding.

Do LLMs truly understand? Do they possess a semantic understanding beyond imitation? Do they ground symbols as humans do?

While LLMs perform surprisingly well, nobody knows exactly why. Their internal complexity is such that they must be regarded as a black box. The issue is further complicated by the possibility of distinguishing different forms of "grounding." Modern LLMs, as artificial neural networks, compute over high-dimensional vectors rather than discrete symbols, making it more accurate to speak of a "vector grounding problem"[15]. Millière and Mollo distinguish five notions of grounding: referential (how linguistic items refer to real-world objects), sensory-motor (such as linking textual and visual representations), relational (how words relate to other words through definitions), communicative (the establishment of an intersubjective, rule-based language to communicate meaning and ensure mutual understanding), and epistemic (the relationship between linguistic expressions and factual knowledge-based data). Using an analogy with Chalmers' distinction between the easy and hard problems of consciousness[16], the hard problem of symbol grounding pertains to referential grounding. The other forms of grounding are, so to speak, "easy" in the sense that they can be addressed through a more or less sophisticated representational theory. Harnad's "symbol/symbol merry-go-round" relates to referential grounding.

However, Millière and Mollo argue that LLMs can partially overcome the referential grounding problem, thereby acquiring more than interlinguistic functions through human fine-tuning. This fine-tuning endows LLMs with normative, world-involving functions that allow them to develop intrinsically meaningful representations of the world—going beyond mere word associations by adding an epistemic layer that connects language to world references. Through human feedback, LLMs gain knowledge about how the world operates, which enables them to form world-model representations rather than purely linguistic ones, even without embodied cognition with sensorimotor skills. In this way, LLMs at least partially address the referential grounding problem.

Yet, these world-involving functions are ultimately human-induced; the human factor is the bridge between language and the world. Human feedback via supervised learning provides modern LLMs with a form of referential grounding that aligns vector representations of words with real-world referents, endowing them with intrinsic meaning that, once fine-tuned, may no longer rely directly on human input. The only way for LLMs to achieve any form of referential grounding is through the intervention of a semantically grounded human agent who sets the normative framework, thereby guiding the non-grounded agent in

establishing a world model—that is, a world-system of referents. Thus, this merely shifts the symbol grounding problem from the machine to the human agent: If an artificial neural network cannot achieve symbol (or vector) referential grounding, how are those biological neural networks inside our skulls able to do so?

Moreover, empirical evidence suggests that there is not anything or "anyone" in an LLM that genuinely "understands" beyond best-guessing and mimicry. Recent research has shown that LLMs can deviate in their responses when presented with irrelevant information, with performance declining when an unrelated sentence is added to a problem. This does not alter the problem's semantic content but appears to "distract" the model[17]. The illusion of "reasoning" displayed by LLMs relies largely on recognizing patterns with a strong token bias rather than true comprehension; this also accounts for their limited generalization ability[18]. Further studies reveal that LLMs lack genuine mathematical reasoning; instead, they replicate logical steps seen in training data without understanding. For instance, rephrasing the same question can yield different answers, and adding redundant information irrelevant to the solution significantly lowers their mathematical reasoning performance. A slight increase in task difficulty can similarly lead to notable performance drops[19]. LLMs tend to translate statements into operations (e.g., interpreting "discount" as "multiplication") without understanding the underlying semantics. This behavior aligns more closely with probabilistic pattern matching than with actual logical reasoning. Such limitations are evident even in basic grade-school math problems and are expected to become more apparent in complex mathematical assessments. That LLMs do not go much beyond sophisticated forms of "semantic-free" imitation is also suggested by the fact that they are inherently limited by their ability to solve problems that significantly differ from those they saw during their training session.[4] Additionally, the idea that LLMs exhibit "emergent abilities"—skills that are absent in smaller models and that appear only in larger ones—is also under scrutiny. These abilities may be fictitious and not an inherent result of scaling models up in complexity[20]. Further evidence suggests that larger and more instructable language models become less reliable[21].

Other investigations have highlighted AI's limitations in reasoning, as demonstrated by LLMs' failures in seemingly simple tasks such as counting words or reversing a list[22]. This raises the question of why LLMs are generally effective in multi-step reasoning yet struggle with surprisingly trivial problems. Chain-of-thought prompting, which involves generating intermediate reasoning steps before arriving at a final output, exhibits characteristics of both memorization reasoning (in which the model mimics patterns learned from training data) and probabilistic reasoning (in which the model selects the most probable output based on the input)[23]. Further studies have examined how transformer LLMs tackle compositional tasks that require breaking problems into sub-steps and synthesizing those steps into precise answers. It turns out that they often simplify these tasks by matching linearized subgraphs—training examples that mirror the computations needed to solve test examples—without developing systematic problem-solving skills. Moreover, abstract multi-step reasoning problems based on autoregressive generation (the statistical prediction of the next sequence value based on previous values) tend to deteriorate rapidly as task complexity increases[24].

A study conducted by Apple[25] with Large Reasoning Models (LRMs)—a new type of AI designed to excel at complex problem-solving by breaking tasks into smaller, logical steps—revealed that while these models perform better on reasoning benchmarks, they experience a complete collapse in accuracy beyond certain complexities. Their reasoning effort increases with problem complexity up to a point, after which it declines. In high-complexity tasks, they fail to develop generalizable reasoning capabilities, experiencing collapse. Additionally, their limitations in performing exact computations became apparent; they might succeed in one task but repeatedly fail to provide correct answers in others. Apple's take is that these models create an illusion of thinking but lack true reasoning. They are essentially large pattern matchers that falter when they encounter data outside their training set, leading to breakdowns in generalization. Before we can consider them capable of genuine thinking processes with true reasoning abilities, we must be aware that something crucial is still missing.

This suggests that LLMs' abilities scale asymptotically: Beyond a certain threshold, adding more neural layers, faster processing, increasing training data, enhancing supervised reinforcement learning, and further fine-tuning or prompt engineering do not significantly improve their performance.[5]

Additionally, unlike humans, LLMs show unreliable intrinsic self-correction without external feedback[26][27]. This aspect is crucial, as one might argue that humans are prone to similar flaws. However, humans can recognize their logical, inferential, and deductive mistakes, which allows them to restart the problem-solving process with different premises or methodological approaches. This capacity to overcome biases and reach the desired result is possible only because of a semantic awareness that extends beyond mere symbol manipulation, enabling individuals to identify incorrect or nonsensical outcomes even if they do not know the correct answer.

After all, it does not require complex scientific research to recognize the shortcomings of LLMs in reasoning, semantic competence, and the understanding of inherent meaning. Sustained use and interaction often reveal interesting insights. For example, if one were to ask how many U.S. states have names beginning with the letter K, one might be told that there are four: Kansas, Kentucky, Kansas, and Kentucky. These issues similarly affect generative AI. If instructed to create a picture with no elephant in the room, it might repeatedly generate images featuring an elephant in a room, thereby demonstrating a failure to grasp negation (the "not-questions"). When asked to depict someone writing with their left hand, the model may refuse to comply, always producing an image of a right-handed person instead. This makes sense if we keep in mind that the model makes probabilistic guesses rather than semantic ones, as it has been trained predominantly on a large dataset where the proverbial sentence 'elephant in the room' and images of right-handed individuals are quite common.[6]

These failures do not prove that LLMs lack consciousness or intrinsic semantics. They are, however, consistent with the hypothesis that current systems rely heavily on relational, distributional, and scaffolded competence rather than intrinsically grounded semantic awareness. Overall, we are beginning to realize that despite billions being invested in research and development, LLMs continue to hallucinate, fabricate information, and lack a genuine understanding of the world. They excel in deep learning but struggle with generalized abstract

reasoning, and they have difficulty understanding wholes in terms of their parts.[7]

Though we are dealing with a black box, we know it is ultimately a Turing machine working with symbols according to logical, syntactic, context-based, or probabilistic rules in someone else's natural language. These symbols stand for external referents grounded in the subjective experiences of someone else—experiences that the symbols themselves cannot convey to others. While it would be unfair to label LLMs as mere "stochastic parrots," they still fall short of human capabilities in natural language inference; their dialog is limited in information and can be unreliable, frequently failing in differentiating between fact and fiction and generating inaccurate claims (for a good summary of the strengths and weaknesses of LLMs see [28]). Grounding LLMs in human reasoning and common sense through humans-in-the-loop will remain essential[29].

That LLMs associate patterns relationally without understanding them—that is, that they learn correlations between words rather than concepts, thereby reducing their internal processes to statistical regularities without comprehension—has also been highlighted in recent work by Gonen et al.[30]. They illustrate this with vivid examples: when asked who likes the color yellow and what such people do for a living, an LLM may respond that they work as school bus drivers. Its "knowledge" consists solely in the co-occurrence patterns linking words related to yellow with those related to school buses. It does not apprehend the phenomenal quality of "yellowness" but only uncovers correlations within its text-based item representations.

Similarly, self-driving cars' ability to classify objects and people is impressive, yet countless instances reveal their lack of comprehension regarding what they are "seeing" (e.g., distinguishing between real people on the street and fictional figures on a billboard). Classifying and categorizing are not the same thing as understanding. The critical questions are whether a car can effectively navigate without a semantic understanding of its environment—recognizing the street, cyclists, traffic lights, and so on—and whether a level V self-driving car could achieve comprehension that even humans might struggle to acquire without conscious experience.

It could be argued that further fine-tuning and additional training sessions might resolve these issues and reduce the error rate. After all, humans also make mistakes. However, this is not the decisive point. The crux of the matter is that when the model fails, it can do so spectacularly, sometimes producing absurd answers that clearly illustrate a lack of genuine understanding.

On the other hand, it is also true that humans are not entirely immune to similar senseless perceptual hallucinations or ludicrous cognitive failures, particularly in individuals with neurological disorders. Oliver Sacks's bestselling book "*The Man Who Mistook His Wife for a Hat*" compiles case studies of such rare but real phenomena[31]. However, the question remains whether such pathologies can be addressed without relating them to a phenomenological content. Neural networks fit data and classify information, whereas humans integrate sensory data into a unified experience—what we might call a "perception of meaning"—that represents a semantic whole, no matter how senseless it may appear. Humans can overcome these hallucinations precisely because of conscious experience, while machines are confined to sense-making that cannot go beyond relational systems and statistical inferences based on abstract tokens.

Current systems show impressive reasoning-like competence, but their generalization, robustness, and exact algorithmic control remain contested. While LLMs' cognitive skills are undoubtedly impressive and surprising, it remains unclear whether there is any true reasoning or common-sense knowledge that extends beyond sophisticated Chinese room machinery. There is no "ghost in the machine" with semantic awareness; fundamentally, the machine does not understand a thing. We must seriously consider whether we are falling prey to what Floridi terms "semantic pareidolia": the attribution of intentionality to statistical pattern-matching systems, and the perception of meaning in correlations where none exists, much like seeing faces in clouds[32]. Machine cognition certainly doesn't align with the human understanding of 'cognition.' When we look at a green apple or a yellow pear, we do not perceive vectors or matrices.

However, one might argue that nonetheless, the machine does possess a form of cognition as well. We should not assume that human-like cognition is the only possible one. In fact, representing reality as a complex relationship of multidimensional vector space enables AI to engage in meaningful conversations with us. Yet, without succumbing to anthropocentrism and placing our cognition at the center of the universe, we need to acknowledge the differences between human and machine meaning-making. If there is none, the question arises how the internal workings, ultimately represented as digits in a memory chip, translate into the rich experience of perceiving the 'greenness' of an apple, feeling its shape and size, and even tasting it. A memory cell labeled "apple" or "pear" is merely a symbol encoded in the physical state of a circuit. There exists a subjective and experiential dimension that cannot be encapsulated by a symbol alone, regardless of how complex the computations leading to its identification may be. This issue is not exclusive to machines; it also applies to humans. If you have never tasted an apple, no amount of description—be it through language, science, mathematics, computation, or neurological explanation—can truly convey what it is like tasting it. If even the most advanced AI can pass a molecular biology exam yet is unable to learn what for humans are much simpler tasks, such as driving a car, and remains far from achieving human intelligence —or even the intelligence of a cat or dog—something might be fundamentally flawed in our understanding of what intelligence truly is. Alternatively, we might be overlooking a crucial aspect of our nature as living beings.

Meaning and Qualia

Wittgenstein might not have been entirely off: Language, like mathematics, is not merely a human cognitive creation unrelated to the organization of the world but, rather, might reflect its structure. We might conjecture that the statistical distribution of words is not solely tied to human cognition but also offers a window into the conditional structure of the world as mediated through human language. The patterns embedded in language might reflect the patterns inherent in the world. Linguistic patterns track real-world patterns, and which representational structure may be captured within LLMs[33]. If so, given that LLMs are pattern matchers trained to "autocomplete" based on complex relationships between vast numbers of tokens representing properties of the world and their interrelations, their effectiveness might be less surprising.

Nonetheless, while any mathematical or formal proposition acquires meaning only when it is communicated from grounded semantic agents to other grounded semantic agents, one could go a step further and suggest that all of

what we perceive and conceptualize about the world is ultimately a symbol, sign, token, or image within our cognitive awareness. This might be even less surprising to the philosophical idealist, who posits that the world itself is an "idea" or symbolic expression. Such a view aligns with certain Eastern philosophical and mystical theories in which the universe represents the expression of a creative, transcendental "real-idea," forming the foundation for all meanings, signs, words, and human language[34].

Be that as it may, it is not necessary to venture into metaphysical speculations to capture some essential aspects of how our sense-making cognitive processes emerge. The primary challenge here is not conceptual or theoretical but, rather, the shift from a third- to a phenomenological first-person perspective.

In this respect, it might also be interesting to consider Harnad's recent follow-up, which he formulated in the form of a dialogue with ChatGPT-4[35]. According to Harnad, what ChatGPT lacks is *"a direct sensorimotor grounding to connect its words to their referents and its propositions to their meanings."* Machines operate within the realm of symbol manipulation without any grounding in real-world physical interactions and experiences; they lack the subjective and conscious dimensions of understanding. He further suggests that the only solution is to embody AI models—that is, to ground them—with sensory and motor outputs that allow them to interact with and learn from their environment, thereby building a world model. An ″embodied AI″ is a system that goes beyond mere abstract symbol manipulation found in traditional software. It is physically linked to the real world through sensors that detect environmental conditions and actuators that allow it to interact with that environment. This integration of perception and action creates a continuous feedback loop that enhances learning, as opposed to depending solely on abstract data. The idea of embodiment is crucial in fields such as robotics, virtual agents, and cognitive science, where it is seen as vital for fostering more adaptive, context-sensitive, and human-like intelligence.

However, Harnad carefully distinguishes the symbol grounding problem from any allusion to the hard problem of consciousness (or the "problem of qualia" or the "explanatory gap"[16]). The philosophical issue of qualia involves how neural processes give rise to the subjective experiences of "what it is like" to be in a qualitative state of consciousness. The symbol grounding problem is about representation and understanding in artificial systems, not about why the complex machinery in our brain supposedly gives rise to subjective experiences.

Qualia—the introspectively accessible, subjective, phenomenal aspects of our conscious lives, such as the bitterness and warmth of coffee, the redness of a tomato, the smell of freshly cut grass, or the sharp pain from a paper cut—are not reducible to mere information processing. Our private experiences of pleasure, pain, feelings, and emotions, along with sensory perceptions like seeing, hearing, touching, smelling, and tasting, convey qualities of the world such as colors, sizes, and shapes. These experiences are not merely concepts or abstract symbols in a quantitative database; they are grounded in a qualitative experiential dimension.

Harnad's insights are inspiring; however, in my view, he fails to close the circle. He begins by associating meaning with *"the subjective, felt experience of understanding,"* the *"phenomenological aspect of what it feels like to mean or understand something"*—elements that AI lacks, as it has no *"direct experiential grounding"* or *"direct sensorimotor experiences."* Yet, somewhat surprisingly, he

stops short of explicitly concluding that, for a machine to achieve what he terms *"direct symbol grounding"* (and, ultimately, a state of AGI), it must be conscious.

Disentangling the symbol grounding problem from the hard problem of consciousness and then reiterating that natural language must be grounded in real-world experiences is a misstep. This is because the former is an aspect of the latter. The symbol grounding problem is itself "grounded" in the problem of qualia and the explanatory gap.

I submit that the gap between syntax and semantics is qualitatively much deeper than previously assumed. Scaling alone will not bridge this gap. While embodied architectures could extend text-based information to include sensory-based data with feature-detecting and abstracting capacities from real-world engagement, this would not transform Shannon-type sensory information into semantic information, thereby grounding the ungrounded in any miraculous way. There is no reason to believe that replacing or augmenting the linguistic symbols in an LLM super-dictionary database with a sensorimotor super-dictionary of pixels and mathematical functions representing environmental sensory signals could provide genuine grounding beyond syntactic computational understanding within an abstract categorical space. Names and verbal descriptions require not only a sensorimotor embodiment but also a conscious experience of the features they represent to be grounded in a semantic space. Vast data sets (textual, numerical, and/or sensory), immense computing power, and any form of embodiment may be necessary to build meaningful world models. However, it is questionable whether these elements are automatically sufficient to allow an AI system to gain a true understanding of the properties of the real world it is engaging with, without having a subjective experience of those very same properties.

After all, one can teach children how to roll a ball without giving them millions of examples, because a child does not see the world through numerical or symbolic representations but through a sense-making mind that relies on lived experiences. Biological organisms' meaning-making has an essential component that machines fundamentally lack—a gap that information processing alone cannot bridge. To put it in Vallor's words: *"We are more than efficient mathematical optimizers and probable next-token generators"* [36].

Meanwhile, taking a first-person perspective reveals that our semantics-based cognition is inherently tied to conscious experience—always. We cannot truly understand colors, sounds, tastes, smells, temperature, touch, or sight without having directly experienced seeing a color, hearing a sound, tasting chocolate, smelling an odor, or touching an object. Is there a difference in "meaning" between, say, listening to a Tchaikovsky symphony and "knowing" it only in the form of a digital transcription? An LLM might "know" everything textually about apples or the Fourier transform of a sound signal, but there is no semantics to extract, as all this remains an abstract, multidimensional vector representation defined by weighted numerical parameters. Without experiencing "what it is like" to taste an apple or feel its texture or to be immersed in the chills and thrills of musical rapture, it cannot achieve any form of symbol—or vector—grounding. Similarly, someone might explain all the chemistry and physics of an H2O molecule, but you cannot grasp the essence of wetness until you have felt the sensation of water yourself.

This leads us to Jackson's famous knowledge argument[37]: Neurophysiologist Mary may have all the physical information about color vision, but she still lacks knowledge if she has not experienced the qualia of colors.

The question is: Can any physical information be linked to semantic information without the presence of subjective experience?

One could reverse the argument and ask, "What kind of 'knowledge' could a purely phenomenological experience without physical information or conceptualization convey?" It is unlikely that this would lead to semantic comprehension. I might experience the blueness of the ocean, the chill of an environment, the sourness of a lemon, or its shape in my hands without having any concept or understanding of what the ocean, the environment, or a lemon truly represents. Experience, representation, and conceptualization are intimately intertwined.

This does not mean, however, that congenitally blind individuals cannot understand others discussing spatial properties, geometric forms, light, colors, or other visual sensations. They have an indirect understanding of these concepts, achieving referential grounding much like an LLM—not through sentience but via information mediated by other sentient agents. And it is known that blind subjects are not defective in their learning of language, space, and objects. Blind learners not only learn the forms of language but their meanings as well[38]. The key difference between even the most advanced AI and a person with sensory impairments is that the absence of sight in humans is compensated for by other subjective experiences, like sound, touch, taste, and smell. These experiences enable the indirect semantic comprehension of visual concepts even if direct visual perception is lacking, as the representation is supported by other forms of conscious experience—something a machine cannot replicate.[8]

What kind of comprehension do blind individuals have of the visible world, and to what extent can they eventually recover an ordinary understanding of it? This is not a new question. For instance, in the 17th century, William Molyneux proposed a thought experiment involving a congenitally blind person who has learned to recognize objects solely through touch. Imagine this person can distinguish a sphere from a cube by touch but has never seen them. Now, suppose this person could suddenly see. Molyneux questioned whether, if the sphere and cube were placed on a table, this person would be able to identify which was which *without* touching them.

The question was debated by philosophers like Locke and Berkeley, who essentially agreed that if our mind cannot establish a relationship between the sensations of the tactile and visual worlds, the answer to Molyneux's question must be negative. The connection between these two realms—specifically, grounding visual experience in meaningful semantic content known only through tactile experience—cannot be established through mere intellectual inference. To create this connection, one must link real-world experiences with previously grounded concepts derived from other kinds of lived experiences into some world-model.

Interestingly, the Molyneux problem can now receive a scientific answer. It is possible to gain some insights into the state of "meaningless awareness" from individuals affected by congenital cataracts that are later treated. A congenital cataract is an organic anomaly present at birth that clouds the eye's natural lens and can result in amblyopia—a disorder in which the brain fails to process visual stimuli. In the 1930s, Marius von Senden became the first to describe the perception of space and shape in congenitally blind individuals before and after surgery[39]. When the sight of previously blind patients was restored and their bandages removed, they did not see the world as one might assume. Instead,

they experienced a blotch of chaotic colored patches that meant nothing to them. They required time and practice to make sense of what they saw.

In 2011, neuroscience shed further light on this topic. An Indian research project, "Project Prakash," aimed at treating blind children while also seeking answers to scientific questions about how the brain develops and learns to see, provided evidence supporting Locke and Berkeley's views[40]. In the treatment of congenitally blind children aged 8 to 17, researchers found that, upon gaining sight, these children struggled to visually match objects they had previously known only through somatosensory information. However, this capacity developed quite rapidly; their skills in relating visual perception to somatosensory sensation improved within a few days of sight restoration and were nearly fully present within a few months. Physical information could finally be grounded in semantic information through sensory tactile experiences, as the latter was already rooted in a meaningful conceptual unity.

This indicates that linking tactile knowledge to visual knowledge is not an innate ability. Furthermore, it demonstrates that meaning emerges not only from the relationships among words, symbols, or vectors but also through the contextualization of different forms of qualitative experiences. Only after this experiential stage can the mind connect a symbol (the signifier) to its referent in the real world (the signified).[9]

One thing that must be noted is that semantic content does not require an experiential link to the physical world—that is, direct sensorimotor engagement. Most notably, abstract and intangible concepts like "freedom," "courage," "truth," "beauty," "justice," and "wisdom" lack objective referents in the physical world. Nonetheless, they are deeply meaningful to us and, in a sense, real and concrete. Their concreteness arises from subjective feelings and a "perception of meaning" rooted in the experiential dimension of the mind. These concepts are thoughts, often accompanied by an inner emotional state, that also have a phenomenologically meaningful aspect. Even purely abstract entities, such as numbers, lack physical reality outside our minds (or, as Platonists would argue, exist only in a world of perfect Forms), and the symbols for numbers hold no inherent significance unless they are associated with a "number sense"—an experiential link between symbol and quantity mediated by a subjective mental experience of extension or quantity.[10]

Qualia are not limited to sensory experiences; thoughts are "mental qualia," and emotions are "emotional qualia." The "perception of meaning" is, above all, a subjective "mental quale" that emerges from and is intertwined with other qualitative experiences, whether somatosensory, affective, or otherwise. In this way, there exists a phenomenon of what "it is like to be" in each of these internal states—an aspect that any symbol-to-symbol or representation-to-representation relationship cannot fully capture.

Moreover, semantics has a holistic character. The perception of meaning is always associated with an integration of information that neuroscience still struggles to understand[41]. How the brain integrates information from various sensory inputs (sight, sound, touch, etc.[11]) and distinct neural processes into a unified, coherent experience is unclear. For example, when we look at an apple, we do not perceive separate visual features like color, shape, and size individually; rather, we see a single, whole apple. It is what I referred to as a semantic whole that suddenly appears in our awareness by looking at a figure made of pixels, parts, structures, boundaries, colors, etc. The question of how

different parts of the brain "bind" features together into one experience, despite being processed in different brain regions, is commonly known as the "binding problem" and is intimately related to the unity of consciousness.

The implications for AI are that without a conscious subject (an entity or individual capable of conscious qualitative sensations of the properties of the world), a machine cannot grasp anything beyond the purely abstract "understanding" of its representations, symbols, letters, words, sentences, signs, and numbers. A self-driving car, no matter how advanced or sophisticated its neural network and information processing system might be, cannot comprehend what an image of a street, a cyclist, or a traffic light represents if it lacks the complementary subjective experience and an integrative process that binds them into a meaningful unity. True knowledge cannot be achieved through functional processes alone; understanding requires a specific type of qualitative phenomenal dimension. An image can be converted into vector spaces, matrices, relationships, curve fittings, and probability laws, beyond their complex functional relationships. However, unless the light signals hitting the pixels of the image on the CCD camera elicit a subjective lived experience in someone or something, all these mathematical abstractions, multidimensional vectors, or neural representations will remain meaningless physical information without conscious semantic grounding. While physical information can be measured by (negative) Shannon entropy, without sentience, nothing can convert it into semantic information. This is because "understanding" is an aspect of phenomenal consciousness itself and cannot be abstracted from the hard problem of consciousness.

Semantics, Life, and the Unconscious

Thus, consciousness plays a fundamental role in the transition from symbols to semantics. An information processing system may interact with the environment through sensory-motor embodiment, but that alone does not make it a true semantic agent. An embodiment allowing for the representation of a world model should not be confused with a system experiencing the world. For it to achieve this, it must possess a comprehension of the world based on experiences, not just representations, abstract symbolic descriptions, or indirect linguistic grounding—that is, "qualia-less" physical information. Transitioning from the vectorization of language to the vectorization of the environment may represent a significant advancement in AI, but this alone is unlikely to overcome its semantic deficiencies, as these will still be ungrounded quantifying numbers or symbols all the way down. Embodiment may be a necessary condition, but it is not sufficient to transform unconscious apprehension into conscious comprehension. To achieve a deeper understanding of the world, including its contents, properties, and related concepts, I must grasp it qualitatively through conscious experience—that is, with qualia in the form of sensations, feelings, and lived sensory perceptions. A tight and inextricable relationship exists between feeling and knowing, sensing and understanding, perceiving and comprehending. We cannot disconnect these aspects of cognition and treat them separately. While the hard problem of consciousness addresses how consciousness emerges from material and/or functional processes, such as neural activity in the brain, the symbol grounding problem explores how meaning arises from these same processes. Symbol grounding is impossible without experience because experiencing meaning is an inherent aspect of consciousness. To understand something implies being conscious of it. Meaning emerges not only due to a relationship between abstract tokens but also as a pre-

linguistic association of experienced qualities into a unified construct apprehended and comprehended as a singular aware sensation. Meaning *is* subjective perception; the "perception of meaning" *is* a quale in itself.[12]

There is, however, a potential objection to this viewpoint.

Many of our cognitive processes occur without our conscious awareness. Experimental psychology demonstrates that sophisticated cognitive abilities seem to occur without consciousness. For instance, individuals can process semantic content and focus on objects even when masking techniques prevent the information from reaching their reported access consciousness. In cases of blindsight, individuals who are cortically blind can respond to visual stimuli they do not report perceiving, due to lesions in the primary visual cortex. Remarkably, they can often correctly identify objects that they believe they do not see.

Does this show that understanding is possible in the absence of consciousness?

I believe this conclusion is premature. Our understanding of consciousness is largely based on a superficial subjective experience. What we refer to as 'unconscious' or 'lack of attention' may actually represent another form of conscious awareness and attention—not unconsciousness, but a subliminal, non-metacognitive—that is, non-reportable—type of awareness. We often confuse different states of consciousness with being 'unconscious' because we cannot relate them to our ordinary waking state of awareness. In these states, we may not be genuinely unconscious; instead, we may simply lack mnemonic access to those experiences.

What is it like to be in a conscious state without memory? Are we truly 'unconscious' during dreamless sleep? Are we unconscious during anesthesia? Are we unaware while in a hypnotic trance or sleepwalking?

I argue that we do not have definitive answers. Phenomenal consciousness—that is, some form of sentience—can be present but quickly fade within seconds. Consciousness-mediated semantic cognition might still function in these altered states. Therefore, we should avoid making hasty conclusions, as this remains an open question.

Revisiting Semantic Grounding Through Contemporary Scientific Perspectives

We now turn briefly to contemporary scientific theories that may further substantiate the claim that consciousness is necessary for semantic grounding. From the standpoint developed above, contemporary theories of meaning-making can be reconsidered within a broader and more encompassing framework.

For example, ***affective neuroscience*** reframes meaning as biologically instantiated valuation, rooted in the living body rather than in abstract cognition alone. It investigates the neural mechanisms underlying emotions, mood, and motivation, integrating methods and concepts from psychology, neurobiology, and cognitive science to examine how cognitive states are conditioned by affective states such as fear, pleasure, anger, and attachment. These are intrinsically bound to phenomenal consciousness, insofar as they are not merely functional or physiological processes but are essentially constituted by what it feels like to experience them. Affective neuroscience has shown that "felt significance"—that is, the sense that something matters and means something—is not a purely abstract or cognitive attribution but emerges from evolutionarily

conserved valuation systems that are tightly coupled to bodily states and subjective sensations. Work pioneered by Jaak Panksepp demonstrates that primary affective states such as seeking, fear, or care are intrinsically (positively or negatively) valenced prior to reflective thought. These states are deeply integrated with autonomic and interoceptive processes, such as heart rate, hormonal signaling, and visceral sensations, so that valuation is literally felt in the body rather than computed in a detached, symbolic manner[42]. Interoceptive states (e.g., hunger, pain, arousal) allow bodily conditions to shape subjective feelings and thereby determine motivational salience, driving goal-directed behavior. On this view, "significance" is not imposed by rational deliberation but arises from consciously felt bodily states. Higher cortical processes can refine, reinterpret, or even override these signals, but they remain anchored in this affective foundation.[13]

While affective neuroscience focuses on the body, ***enactivist theories*** examine its interaction with the environment. They likewise reject the idea that cognition primarily consists in building and manipulating internal representations of a pre-given world. Instead, they argue that meaning emerges through embodied engagement with the environment[43]. Organisms are not passive information processors but autonomous systems that actively enact a meaningful world through their ongoing interactions, guided by their physiological organization and needs. Sense-making arises through the establishment of a perspective from which certain environmental features count as relevant or irrelevant. The organism thus "enacts" a world of lived significances relative to its environmental context. Nutrients, threats, or mates are not inherently meaningful; they become meaningful insofar as they matter for the organism's continued existence. Accordingly, enactivism emphasizes that cognition does not consist in internal symbolic models or representations but arises directly from the organism's embodied condition within dynamic feedback loops of interaction with its environment. As Evan Thompson puts it, "*Living is sense-making in precarious conditions*"[44].

Along a similar but complementary line of reasoning, ***ecological psychology*** explains perception and action in terms of the direct, dynamic relationship between an organism and its environment. Originating with James J. Gibson[45], it rejects the idea that perception relies on internal representations or mental reconstruction as well. Instead, it holds that the environment provides structured information that organisms can directly pick up through active exploration. A central concept is that of "affordances," that is, the action possibilities that the environment offers relative to an organism's bodily capacities. Perception is not about constructing internal models of the world but about detecting these affordances in a way that is immediately meaningful for action. Accordingly, ecological psychology emphasizes embodiment, real-time interaction, and the inseparability of perception and action, framing cognition as something that emerges from the continuous relationship between organism and environment.

Meaning, therefore, is neither internally stored nor externally given; it is relational and processual, emerging from the ongoing organism–environment coupling, in which cognition consists in the active "bringing forth" of a world through sensorimotor activity and lived experience.

A different but potentially complementary view is ***teleosemantics***, which explains the content or meaning of mental representations in terms of their biological functions shaped by evolutionary history. According to this account, a mental

state represents something because it was selected, either through natural selection or through learning, to guide behaviors that contributed to the organism's survival and reproduction. On this view, meaning is neither intrinsic nor purely interpretive but grounded in a naturalized form of teleology—that is, in the idea that biological traits have functions or purposes. The different formulations (e.g., [46]) differ in detail, but they converge on the claim that semantic content arises from functionally defined, historically selected information-bearing states. In this way, teleosemantics links meaning to the organism's adaptive engagement with its environment.

Another influential approach is **predictive processing**, which complements these views by modeling the brain as a hierarchical Bayesian inference machine. Rather than passively receiving sensory inputs, the brain continuously generates predictions about the causal structure of those signals and compares them with incoming sensory evidence. Perception, action, and learning are thus driven by the minimization of prediction errors, making experience an active, constructive process. This framework can be formally described in terms of variational inference minimizing free energy[47]. In this approach also, the treatment of the body via interoception is central: emotions and bodily feelings can be understood as forms of "controlled hallucination," much like visual perception. Selfhood itself emerges from hierarchical interoceptive predictions[48]. Interoceptive prediction errors, arising from mismatches between expected and actual bodily signals, are experienced as affective feelings (e.g., anxiety, fatigue, hunger), thereby grounding valence and significance in the regulation of the organism's internal milieu. Predictive processing is fundamentally hierarchical: lower levels encode fast, modality-specific signals (e.g., sensory and visceral inputs), while higher levels encode increasingly abstract, temporally extended hypotheses about the world and the self. Conscious content emerges from this layered inferential organization across levels. In this way, bodily states shape not only raw feelings but also perception, cognition, and even self-representation. What we consciously experience is thus inseparable from what and how the brain predicts via active inference: organisms do not merely update their beliefs to fit the world; they also act to make the world conform to their predictions, shaping what is experienced as salient or significant.

While predictive processing is grounded in a Bayesian framework, ***Sensorimotor Contingency Theory (SMCT)*** emphasizes the cognitive significance of change and distinction. It holds that perceptual experience is grounded in the practical mastery of how sensation varies with action, rather than in internal world-models. Developed primarily by Kevin O'Regan and Alva Noë[49], SMCT proposes that perception is not the construction of internal representations of the world, but the mastery of lawful relations between sensory changes and one's own movements, so-called "sensorimotor contingencies." The core idea is that perceptual experience, such as seeing, hearing, or touching, is constituted by an implicit understanding of how sensory input would change as a result of movements of the eyes, head, or body, or through interaction with the environment. Perception is action-dependent rather than a passive reception of stimuli. It extends the enactive perspective, highlighting the differential aspect of experience arising from skilled engagement with the world. Qualitative aspects of perception (e.g., "what it is like to see red") are explained in terms of patterns of sensorimotor dependence, rather than internal imagery or representations. Thereby, SMCT is often contrasted with representational theories of perception and aligns closely with ecological psychology and enactive cognitive science. However, it differs in emphasis: it foregrounds the lawful

structure of sensorimotor relations, rather than ecological affordances or predictive inference.

The above-mentioned lines of research deal with an embodied and world-involving conception of sense-making. However, other approaches emphasize more relational and normative dimensions of meaning as well.

For example, ***inferentialism*** holds that the content of mental states is determined by their role within a network of inferences, rather than by a direct correspondence with the external world. On this account, to possess a concept is not primarily to have an internal representation that mirrors reality, but to be able to deploy it appropriately within reasoning: drawing conclusions, endorsing commitments, and recognizing entitlements. Robert Brandom[50] argues that meaning arises from participation in norm-governed inferential practices, i.e., socially embedded activities in which speakers undertake and attribute commitments and entitlements. Inferentialism thus contrasts with representationalist theories by grounding meaning not in internal symbols that stand for external objects, but in inferential relations and normative roles within a space of reasons. In this way, mental content is essentially tied to reasoning, justification, and the capacity to participate in discursive and rational practices.

While inferentialism focuses on logical and justificatory relations, ***structuralist semantics*** and its computational instantiation in distributional semantics in machine learning (e.g., [51]) emphasizes formal and statistical relations. It is the view that meaning in artificial systems arises not from direct reference to external objects, but from patterns of relations among internal representations, in line with structuralist traditions in linguistics and philosophy. Structuralist semantics in machine learning frames meaning as emerging from internal relational organization, rather than from a direct mapping between symbols and the world. Instead of treating a model's symbols or embeddings as meaningful by virtue of standing for external entities, it holds that their semantic content is determined by their position within a network of differences, similarities, and transformation relations inside the model. In modern machine learning, especially in deep learning and LLMs, this is expressed through distributed representations (e.g., vector embeddings, as described in the previous sections), where the "meaning" of a word or concept is encoded by its relations to other vectors in a high-dimensional space. From this standpoint, semantic competence consists in the ability to preserve and exploit relational structures across tasks, rather than in grounding symbols in perception or embodied interaction with the world.

The expectation is that a coherent integration of some or all of the above-mentioned approaches will contribute to the next generation of AI and, ultimately, to AGI, ideally realized within the framework of developmental robotics. ***Developmental robotics*** is an interdisciplinary field at the intersection of robotics and cognitive science that studies how robots can acquire cognitive, motor, and social abilities through processes analogous to human development. Rather than being fully pre-programmed, developmental robots learn incrementally through continuous interaction with their environment, employing mechanisms such as exploration, sensorimotor learning, and adaptation.

Thus, all these views[14] have their value and explanatory power. In a sense, they confirm and complement what has been said so far. They also, at least implicitly, presuppose some form of conscious experience or a subjectively felt spectrum of phenomenal events as the background condition for the emergence of meaning.

However, that being said, they tend to proceed in the opposite direction: rather than taking phenomenological experience as primary and asking how it gives rise to functional, inferential, or structural descriptions, they begin with third-person accounts of information processing, behavior, or biological function, and then attempt to reconstruct or explain meaning in those terms. This results in an inversion of explanatory priority, where what is arguably the explanandum, the lived, first-person sense-making, is treated as a product of a "naturalized phenomenology"––that is, as derivative of something that, deep down, remains invariably a functional or computational process. In trying hard to naturalize semantics, these approaches risk explaining structure, correlation, and causal role while leaving underdetermined the very phenomenon they aim to account for: the intrinsic, felt dimension of significance in experience. They seek to naturalize meaning by pursuing an "experience-blind" naturalism via a "new naturalism," which still struggles to account for the emergence of experience within the natural world[52]. More specifically, affective neuroscience highlights the functional role of interoceptive and affective states in grounding meaning at the level of lower-order physiological processes (e.g., activity in structures such as the amygdala, insula, and prefrontal cortex, as well as homeostatic regulation). However, it does not adequately emphasize—or, arguably, fully recognize—the extent to which subjective experience itself may play a constitutive role in higher-order cognitive semantics.

While the enactive perspective offers a powerful account of how organism–environment interactions shape the semantic landscape, it does not necessarily posit conscious experience as a required substratum of sense-making. Similarly, SMCT explains how perceptual experience depends on lawful relations between sensory change and action, but it remains unclear how far this account extends beyond representational or quasi-representational structures unless qualia themselves are taken as fundamental primitives.

Otherwise, one could in principle imagine a computational system that exhaustively registers the structure of its embodied interaction with the environment, augmented by a hierarchical predictive architecture, and encodes relations between sensory changes and actions in binary numbers, while nevertheless lacking any phenomenological experience. Such a system might still successfully navigate and behave as if it possessed internal sense-making capacities, raising the question of whether this would be sufficient to overcome the explanatory limits associated with the Chinese Room argument. Even a system that fully implements the formal and functional organization associated with meaning-making, down to fine-grained sensorimotor contingencies and predictive hierarchies, would still not, on its own, go beyond the syntactic manipulation of structured relations, however rich or dynamically embedded. Stating that we make sense of the world through the interactions we entertain is an incomplete account: we do so starting from the experiences those interactions elicit.

Meanwhile, teleosemantics specifies the evolutionary mechanisms that determine what counts as meaningful or meaningless by identifying semantic content in biologically selected functions. However, this account does not, in and of itself, explain why meaningful states are accompanied by felt significance—that is, why semantic content is experienced as intrinsically meaningful rather than merely as an abstract, functionally defined categorization. The gap between functional role and phenomenal character is not bridged, if addressed at all. Teleosemantics can tell us why certain internal states count as representations from the standpoint of selective and adaptive advantage, but it does not by itself

account for why those representations should be experienced as meaningful rather than merely implemented in terms of normative biology. It does not bridge the further transition from functional assignment to phenomenologically manifest significance. The latter appears to require an additional explanatory dimension, one that connects biological function to lived experience rather than merely to mechanistic adaptive processes. Ultimately, here as well, what drives and determines the selection process are qualitative experiences—pleasure or pain, fear or a felt sense of safety, joy or grief, and so on.

This family of approaches still leaves an explanatory gap. It does not fully answer why meaning should emerge from embodied engagement with the environment; rather, it often posits it as a given or redescribes this emergence in terms of feedback loops, affordances, or dynamical couplings. The question remains why such cybernetic or sensorimotor organizations should amount to genuine sense-making rather than yet another sophisticated form of information processing.

The greenness of a leaf, or the sweetness of sugar, continue to be relational properties between relata that are mere tokens or signs, not semantic categories emerging from lived experiences. In what sense, then, does sensorimotor coupling "bring forth" a world that transcends representational accounts? If perception is understood as the mastery of sensorimotor contingencies, or if cognition is modeled as continuous feedback regulation within organism–environment systems, it is still unclear why these descriptions should not remain, at bottom, computational in nature—namely, transformations over internal or relational states that track structure without thereby grounding it in meaning. The gap between representation, if any, and meaning can be bridged—that is, grounded—only by giving the phenomenological dimension a chance to be the anchor.

That is to say, for instance, when Einstein formulated the theory of general relativity, his cognitive processes cannot be reduced to bodily sensations, affective states, or immediate sensorimotor interaction with the environment, nor can they be explained solely in terms of biologically grounded drives such as survival or homeostatic regulation. Instead, they involved a rich landscape of creative insight, abstraction, and mental imagery that operated across multiple representational and imaginative levels that were not only inferential and relational but also and especially phenomenally grounded in the world. These processes drew on perception and embodied experience but also extended far beyond them into forms of thought that are not fully captured by accounts reducible exclusively to affect, interoception, enaction, or prediction. There is more to the nature of human understanding than these frameworks alone can explain.

Even our understanding of the world at the most fundamental level of physics is rooted in representations grounded in subjective perception. For example, the physical property of mass, which is defined as a measure of inertia, is reified by the mind as a sensation of resistance or tactile pressure. Likewise, a particle's position in space is not merely a set of three numerical coordinates but an internal image of extension and differentiation born out of our experiential visual, tactile, and sensorimotor dimensions. The angular momentum of a particle is not grasped purely as a vector but is often associated with the image of a spinning object. The temperature of an object is not ordinarily conceptualized in terms of the kinetic energy of particles but is instead tied to the sensation of warmth or cold. In physics, these are expressed in terms of numerical values, such as scalars, vectors, matrices, or tensors, but in our sense-making minds, they are invariably integrated within a phenomenological space.

While inferentialist and structuralist approaches, along with their implementation in LLMs, reframe the emergence of meaning in terms of inferential relations or distance-based measures of similarity between more or less closely related items, they lack the explanatory power discussed in the preceding sections and seem unlikely to close the same explanatory gap highlighted by the symbol grounding problem when extended to robotic implementations. Even if a system can manipulate symbols according to inferential rules or statistical regularities, it remains unclear how those symbols acquire intrinsic semantic content rather than being merely derived, system-relative significances. We cannot abstract away from the fact that a biological organism's semantic understanding of real-world relations operates on relata that are qualia.

In human cognition, when we refer to real-world "features," "properties," "attributes," "facts," or "events" that together constitute structured representations of the environment, we are invariably operating with elements grounded in subjective experience. That is, our semantic practices are not merely relational or syntactic but are anchored in phenomenally instantiated content. While both human cognition and LLMs involve the detection and manipulation of relational patterns, the relata in the human case are not abstract tokens or numerical states but lived, qualitative episodes. Perceptual properties such as color are not apprehended as physical magnitudes (e.g., wavelengths of electromagnetic radiation) but as phenomenal qualities, instances of what it is like to see red or blue. Even so-called primary qualities, such as shape and size, are not accessed in a purely geometrical or metric sense but are given through multimodal sensory experience, structured by visual and tactile phenomenology. Similarly, a property such as wetness is not represented in terms of fluid-dynamical parameters (e.g., viscosity) but as first-person tactile experiences of contact and flow.

In this sense, semantic content in biological agents is constitutively linked to phenomenal consciousness: the intentional directedness of cognitive states is inseparable from their qualitative character. There is no feature, property, or structure of the world that our semantic awareness does not ultimately relate to, or identify with, a qualitative conscious experience. Under all our semantic and cognitive processes stands a complex, dynamically organized mosaic of sensory, affective, imaginative, extero- and interoceptive qualitative states––rather than physical informational ones––integrated in a coherent and meaningful understanding of the world.

All in all, these approaches remain functional descriptions of complex information processing rather than explanations of how and why such processes become grounded in a genuinely semantic dimension. They place too much emphasis on a third-person perspective on the underlying cognitive and computational machinery, while neglecting the role of the first-person, phenomenological dimension. Meaning, like intentionality, volition, purposiveness, aboutness, and valence, cannot be exhausted in mechanistic interactions, ungrounded formal semantics, and cybernetic complex dynamical system theory alone.

Ultimately, it is only by integrating the first-person meaningfulness and, thereby, complementing the third-person account that we can fully appreciate the origin and structure of the explanatory gap between functional organization and semantics as lived meaning. It is precisely here that the phenomenological dimension becomes indispensable for the transition from mere information processing to meaningful world-disclosure grounded in a non-circular way.

Integrating these perspectives with lived phenomenology would represent a necessary step toward a more complete account of meaning-making.

Is the AI-AGI Gap a Conscious Semantics Gap?

As with any AGI narrative, since the success of ChatGPT and the impressive capabilities of LLMs, there has been a marked increase in discussions about the imminent arrival of AGI–human-like intelligence. However, many of these discussions overlook the intrinsic connection between general intelligence and consciousness.

Enhancing AI performance by adding another trillion neurons, increasing the number of parameters, providing sensory-motor environmental interaction, integrating it with neuro-symbolic AI, or developing more advanced deep learning algorithms, internal feedback loops between sensory acquisition and a central processing system to construct more accurate world-models, etc., will no doubt advance current AI capabilities. However, it will not accomplish the "semantic trick." The elephant in the room must be addressed: the possibility that biological intelligence, with its semantic awareness, might not be replicable in machines unless we learn how to manufacture consciousness itself (if it is even possible to do so). If AGI is understood as genuinely semantically aware intelligence rather than merely domain-general functional competence, then consciousness may be a necessary condition. The current widespread discourse on the impending AGI revolution, which is predicted to shape our future, warrants a more critical examination that considers the first-person perspective.

Though I am an "AGI skeptic,"[15] the rationale presented here does not necessarily imply that AGI is impossible or that consciousness in machines is impossible. The claim is that if we want the machine to become semantically aware—where, by 'awareness,' we mean a symbol-grounding subjective dimension capable of experiencing—it must become conscious as well, because awareness and consciousness are inextricably intertwined. Otherwise, it remains semantics without awareness––that is, AI remains locked in the Chinese room.

On the other hand, recognizing the relationship between meaning and consciousness might offer new approaches to longstanding questions. The distinction between conscious intelligence and machine intelligence (or even the distinction between machine and living organism itself?) becomes apparent when, sooner or later, the machine makes a glaringly irrational mistake–one that no conscious, semantically aware agent would make. This reveals its lack of genuine semantic understanding and could be taken as evidence of its absence of subjective experience. In principle, this could suggest the basis for a new test of intelligence, potentially replacing the famous Turing test. The goal would be to develop a test designed to assess whether a machine possesses a deep, semantic understanding comparable to that of a conscious being building its meanings beyond symbol manipulation in a semantic space determined by subjective qualitative experiences. If it does, this would imply a form of intelligence that points to some level of experiential awareness, even if not necessarily human.

This line of thought could also serve as an argument for or against the hypothetical existence of "philosophical zombies" (or "p-zombies"). A p-zombie is imagined as a being physically identical to a human and behaving exactly like a conscious person but lacking any inner life or conscious experiences. If one rejects the conceivability of behaviorally perfect p-zombies, then persistent semantic breakdowns in artificial systems may be interpreted as evidence that behavior alone is insufficient for attributing phenomenal understanding. This

does not refute the zombie argument as classically formulated; however, a test centered on meaning-making provides a stronger and potentially more convincing proof of the presence or absence of consciousness and true intelligence in a machine than the Turing test does.

One possibility could be to draw inspiration from modern tests that evaluate forms of consciousness in animals and apply them to future machine intelligence that could potentially exhibit signs of conscious awareness. At this stage, I cannot provide concrete, implementable testing frameworks, but considering the possibility that meaning-making and consciousness are interconnected offers, in and of itself, a new perspective that could potentially lead to innovative practical approaches and methodological tests as well.

It might be worth noting that, although I have primarily focused on a conceptual form of meaning-making, our cognitive processes are not limited to exclusively analytical semantics; they also apprehend meaning through a felt sense of beauty in music, poetry, painting, and the arts more generally. For example, musical expression carries meaning within phenomenal consciousness, even when it is not readily expressible in terms of truth-conditional or formal semantics. Likewise, the meaning of lyrics cannot be entirely separated from their sensory acoustic enjoyment, as it is partly enriched by the broader musical performance. There exists a further level of meaning associated with an inner aesthetic sensibility that extends beyond purely linguistic semantic expression. When we refer to AGI, then, are we implying that such systems will possess the same, or similar, capacities to apprehend these layers of meaning?

Furthermore, no contemporary AI exhibits any signs of agency and personhood[16]. Without input, these systems do not act autonomously. LLMs are unable to act because they lack intentions[53]. The scientific concept of agency—encompassing intentionality, autonomy, self-determination, purposeful decision-making, and its connection to cognition—is itself a complex subject[54]. However, it is clear that regardless of the definition we adopt, nothing resembling agency exists in current AI systems. If one does not ask a question of ChatGPT, it will remain in a passive idle state, doing nothing and waiting for input indefinitely. There is no psychological impulse or drive to will, no intrinsic desire to act or autonomously generate behavior through goal-directed volition and intention. What is absent is self-determination and spontaneity: the capacity to initiate action independently of explicit instruction. This absence, together with the evident lack of affective states such as desire, joy, grief, or love, suggests that we may be facing a deeper explanatory gap than is commonly assumed. It may plausibly be connected to a deeper underlying absence: the absence of consciousness itself. From this perspective, the lack of autonomous agency is not merely a behavioral limitation but a reflection of a more fundamental ontological void.

It will be interesting to observe the recent efforts to develop 'agentic AI' systems —AI agents capable of achieving specific goals with minimal supervision while mimicking human decision-making to solve problems in real time. Research in this area focuses on creating systems that exhibit autonomy, goal-driven behavior, and adaptability. The hope is that this will be achieved through generative AI techniques that use LLMs to enhance their capabilities for independently completing complex tasks. Current AI systems can instantiate externally scaffolded agency, but there is no strong evidence that they possess endogenous phenomenal agency, intrinsic desire, or self-originating volition. If the present analysis is sound and applicable to agency as well, we can expect to

see the same pattern observed in the emergence of AI abilities over the past couple of decades—characterized by alternating successes and impressive breakthroughs, but also marked by significant cognitive failures—appearing in agentic AI systems with similar deficiencies. The question of whether agency and consciousness might be intrinsically related deserves more attention, particularly as we speculate about the potential emergence of AGI.

Conclusions

I have focused primarily on the relationship between human-like experience and semantics because it represents the most familiar form of cognition for us. Nonetheless, I believe similar arguments can apply to non-human animals, with sentience also serving as the basis for meaningfully navigating the world.

Conversely, one could argue that machine learning without any conscious experience could be viewed as a form of “understanding” as well. Avoiding an anthropocentric perspective means acknowledging that our concept of understanding need not be confined to human ways of making sense of the world; indeed, computers could “understand” things in their own manner.

In fact, recent findings in mechanistic interpretability[55], the field dedicated to understanding the internal reasoning processes of trained neural networks and LLMs, reveal that these systems do not rely solely on statistical inference. LLMs also develop internal structures that are functionally analogous to key aspects of human understanding. They perceive connections and form learned vector features within a multidimensional latent space, which humans typically bind and coalesce under a single concept. However, the understanding exhibited by LLMs is fundamentally different from human understanding. We need new frameworks that can embrace the emerging forms of intelligence we are creating, without assuming that human comprehension is necessarily unique, exclusive, or superior.

That said, it is crucial to question whether machine understanding differs from human understanding precisely because of its lack of subjective experience. We can undoubtedly attribute a different way of "understanding" the world to machines; however, they may never attain certain human-like cognitive skills, just as we will never achieve specific machine-like capabilities, precisely due to the missing link of consciousness.

On the other hand, animal forms of consciousness may result in entirely different types of semantic awareness of the world. We should not assume that human subjective experience is the only standard either. Perhaps there even exists an extraterrestrial civilization with a radically different understanding of reality. What do we know?

This is plausible, as human semantics might be only one of infinitely many varieties of semantics. However, this paper’s main claim does not focus on any specific form of human cognition. Rather, the paper posits that semantics is inherently tied to conscious experience, whether human, animal, extraterrestrial, or otherwise. If semantics is not grounded in some form of consciousness and sentience, without any perception of the world’s properties through qualitative subjective experiences, then it would, by definition, be a p-zombie semantics. Their “understanding,” exhibited by a cognition devoid of qualia–no matter how complex, powerful, or sophisticated–would remain at the level of a Turing machine head processing physical information on a strip of

tape. Nothing would convert this physical information into genuine semantic information.

The realization of AGI, if it ever happens, will first require the creation of conscious machines—that is, of phenomenal AGI in contrast to mere functional AGI. There is a foundational, principled argument against the imminent arrival of AGI, as we remain far from developing systems that transcend memorization, pattern matching, or probabilistic assessments to achieve genuine, complex abstract reasoning based on conscious experience. Syntactic, statistical, embodied, and functional organization still leaves something unexplained about lived meaning—phenomenal grounding may be the missing layer in AI semantics.

Therefore, a first-person perspective, not only in the philosophy of mind but also with regard to addressing fundamental questions about AI, is essential. Approaching AI from this perspective might also help us gain a deeper understanding of ourselves.

But no matter where one stands—whether in the camp of those who believe that AI is (or will become) conscious or in the camp of those who deny it—the very fact that we now seriously entertain this debate is, in itself, indicative of a distinction between intelligence and consciousness, rather than their identity. We now understand, through our daily interaction with LLMs, that something can be intelligent, potentially even surpassing human intelligence, without necessarily being conscious. Conversely, a living being may be conscious without necessarily possessing a highly developed form of intelligence. Historically, however, the dissociation between functional intelligence and phenomenal consciousness has not been regarded as straightforward.

Whatever the case, framing certain philosophical problems can serve as the initial step toward developing new methodologies and practical applications. Our actions are influenced not only by our knowledge but also by our underlying philosophical worldviews. If programmers, developers, and software engineers explored deeper questions of the philosophy of mind about the origin and nature of consciousness, intelligence, the mind, and life, they could potentially gain insights that a purely technical perspective cannot provide.

Statements and Declarations

Funding

No specific funding was received for this work.

Potential Competing Interests

No potential competing interests to declare.

Data Availability

Data sharing is not applicable to this article as no datasets were generated or analyzed during the current study.

Author Contributions

M.M. was the sole author and is responsible for all aspects of the manuscript.

Footnotes

[1] One might argue that the meaning of symbols is rooted in their "use"—specifically, in the function they serve within a given context. However, from a more computational and reductive perspective, there is no inherent 'use' or function in the cell state that a Turing machine's head reads and reacts to, other than the meaning assigned by an external semantic agent.

[2] For a good summary of these theories and their inability to solve the problem of meaning in AI shortly before LLMs took center stage, see [47].

[3] Here, the term "reasoning" encompasses not only problem-solving abilities that rely on logical processes like deduction, induction, abduction, multi-step inference, or mathematical skills, but also those requiring abstraction, generalization, semantic discrimination rather than plausible best guesses, and the capacity to organize thoughts and conceptual structures into coherent, meaningful wholes—that is, what is commonly referred to as "common sense," "rationality," or "thinking" in human cognitive skills.

[4] Anyone with teaching experience recognizes this behavior in students who cheat. When they do not know the answer, they abandon meaningful reasoning and instead try to arrive at the correct answer by guessing—mimicking the methods and rules they have seen applied but without truly understanding the problem-solving method.

[5] This is also reminiscent of the fact that, contrary to popular belief, humans do not possess the largest brain. Other species have larger brains in terms of size, number of neurons, or weight. What, then, determines human cognitive dominance?

[6] For many more examples that show how significant reasoning failures persist in LLMs, occurring even in seemingly simple scenarios, see[56], or see Dr. W. Hsu's collection of LLM failures: https://lnkd.in/eUW6TYCY.

[7] For further review on how close we are to AGI, see also [57], and references therein.

[8] One of the most dramatic and well-known examples of this is the story of deafblind Helen Keller.

[9] One might argue that in human infants, certain cognitive developments occur prior to full conscious awareness. This suggests that some forms of understanding can arise without phenomenal consciousness. However, this conclusion relies on the questionable assumption that newborns have no subjective experiences, which contradicts all external evidence. The burden of proof lies with those who claim that babies lack sentience.

[10] Children who, for whatever reason, do not learn to relate the number symbols, or a perception of numerosity of objects, to a "number sense" are most likely to develop mathematical learning disabilities like dyscalculia[58].

[11] A careful first-person investigation reveals that this is not the case only with sensory inputs, but with thoughts and emotions as well.

[12] I like to reframe this as the "hard problem of semantic awareness."

[13] For a mechanistic account of how interoceptive signals and affective experience link valence to bodily signaling pathways, see also [59].

[14] A range of additional approaches proposing alternative mechanisms for semantic emergence, such as the Extended Mind hypothesis, Integrated Information Theory, the Cellular Basis of Consciousness model, Rosennean complexity, or biosemiotics, could likewise be taken into consideration. However, due to space constraints, even a cursory review of these frameworks is not possible here. Notwithstanding, the line of argument advanced in this essay would converge on broadly similar conclusions.

[15] For reasons too lengthy to elaborate on here, Porębski and Figura offer, in my view, a compelling summary[60].

References

1. ^Dreyfus HL (1992). What Computers Still Can't Do: A Critique of Artificial Reason. MIT Press. ISBN 9780262540674.
2. ^Shannon CE (1948). "A Mathematical Theory of Communication." Bell Syst Tech J. 27(3):379423. doi:10.1002/j.1538-7305.1948.tb01338.x.
3. ^Oizumi M, Albantakis L, Tononi G (2014). "From the Phenomenology to the Mechanisms of Consciousness: Integrated Information Theory 3.0." PLoS Comput Biol. 10(5):e1003588. doi:10.1371/journal.pcbi.1003588.
4. ^Baars BJ (1988). A Cognitive Theory of Consciousness. Cambridge University Press. ISBN 9780521301336.
5. ^Seth AK, Bayne T (2022). "Theories of Consciousness." Nat Rev Neurosci. 23(7):439452. doi:10.1038/s41583-022-00587-4.
6. ^Kuhn RL (2024). "A Landscape of Consciousness: Toward a Taxonomy of Explanations and Implications." Prog Biophys Mol Biol. 190:28169. https://www.sciencedirect.com/science/article/pii/S0079610723001128.
7. ^Van Gulick R (2026). "Consciousness." In: Zalta EN, Nodelman U (eds.). The Stanford Encyclopedia of Philosophy. Spring 2026 ed. Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/spr2026/entries/consciousness/.
8. ^Nagel T (1974). "What Is It Like to Be a Bat?" Philos Rev. 83(4):435450. doi:10.2307/2183914.
9. ^Mahowald K, Ivanova AA, Blank IA, Kanwisher N, Tenenbaum JB, Fedorenko E (2024). "Dissociating Language and Thought in Large Language Models." Trends Cogn Sci. 28(6):517540. PMID 38508911.
10. ^Searle JR (1980). "Minds, Brains, and Programs." Behav Brain Sci. 3(3):417424. doi:10.1017/S0140525X00005756.
11. ^Searle JR (1990). "Is the Brain a Digital Computer?" Proc Addresses Am Philos Assoc. 64(3):2137. https://www.jstor.org/stable/3130074.
12. ^Penrose R (1989). The Emperor's New Mind: Concerning Computers, Minds, and the Laws of Physics. Oxford University Press. ISBN 9780198519737.
13. ^Harnad S (1990). "The Symbol Grounding Problem." Physica D. 42(1-3):335346. https://www.sciencedirect.com/science/article/abs/pii/0167278990900876.
14. ^Harnad S (2007). "Symbol Grounding Problem." Scholarpedia. 2(7):2373. https://www.scholarpedia.org/article/Symbol_grounding_problem.
15. ^Mollo DC, Millire R (2023). "The Vector Grounding Problem." arXiv. doi:10.48550/arXiv.2304.01481.
16. a, b, cBrowning J (2024). "Personhood and AI: Why Large Language Models Dont Understand Us." AI Soc. 39:24992506. doi:10.1007/s00146-023-01724-y. Chalmers

DJ (1995). "Facing Up to the Problem of Consciousness." J Conscious Stud. **2**(3):200 219. https://consc.net/papers/facing.pdf.

17. ^Shi F, Chen X, Misra K, Scales N, Dohan D, Chi EH, Schrli N, Zhou D (2023). "Large Language Models Can Be Easily Distracted by Irrelevant Context." In: Proceeding s of the 40th International Conference on Machine Learning. Proceedings of Mac hine Learning Research, Vol. 202. PMLR. pp. 3121031227. https://proceedings.mlr.p ress/v202/shi23a.html.
18. ^Jiang B, Xie Y, Hao Z, Wang X, Mallick T, Su WJ, Taylor CJ, Roth D (2024). "A Peek i nto Token Bias: Large Language Models Are Not Yet Genuine Reasoners." In: Proc eedings of the 2024 Conference on Empirical Methods in Natural Language Proc essing. Association for Computational Linguistics. pp. 47224756. doi:10.18653/v1/2 024.emnlp-main.272.
19. ^Mirzadeh I, Alizadeh K, Shahrokhi H, Tuzel O, Bengio S, Farajtabar M (2024). "G SM-Symbolic: Understanding the Limitations of Mathematical Reasoning in Larg e Language Models." arXiv. doi:10.48550/arXiv.2410.05229.
20. ^Schaeffer R, Miranda B, Koyejo S (2023). "Are Emergent Abilities of Large Langu age Models a Mirage?" Adv Neural Inf Process Syst. **36**. https://proceedings.neurip s.cc/paper_files/paper/2023/hash/adc98a266f45005c403b8311ca7e8bd7-Abstract-Conference.html.
21. ^Zhou L, Schellaert W, Martnez-Plumed F, Moros-Daval Y, Ferri C, Hernndez-Orall o J (2024). "Larger and More Instructable Language Models Become Less Reliabl e." Nature. **634**(8032):6168. doi:10.1038/s41586-024-07930-y.
22. ^McCoy RT, Yao S, Friedman D, Hardy MD, Griffiths TL (2024). "Embers of Autore gression Show How Large Language Models Are Shaped by the Problem They Ar e Trained to Solve." Proc Natl Acad Sci U S A. **121**(41):e2322420121. https://www.pn as.org/doi/pdf/10.1073/pnas.2322420121?download=true.
23. ^Prabhakar A, Griffiths TL, McCoy RT (2024). "Deciphering the Factors Influencin g the Efficacy of Chain-of-Thought: Probability, Memorization, and Noisy Reason ing." In: Findings of the Association for Computational Linguistics: EMNLP 2024. Association for Computational Linguistics. pp. 37103724. doi:10.18653/v1/2024.fin dings-emnlp.212.
24. ^Dziri N, Lu X, Sclar M, Li XL, Jiang L, Lin BY, Welleck S, West P, Bhagavatula C, Le Bras R, Hwang JD, Sanyal S, Ren X, Ettinger A, Harchaoui Z, Choi Y (2023). "Faith and Fate: Limits of Transformers on Compositionality." Adv Neural Inf Process Sy st. **36**. https://proceedings.neurips.cc/paper_files/paper/2023/hash/deb3c28192f97 9302c157cb653c15e90-Abstract-Conference.html.
25. ^Shojaee P, Mirzadeh SI, Alizadeh K, Horton M, Bengio S, Farajtabar M (2025). "T he Illusion of Thinking: Understanding the Strengths and Limitations of Reasoni ng Models Via the Lens of Problem Complexity." Adv Neural Inf Process Syst. **38**. https://proceedings.neurips.cc/paper_files/paper/2025/hash/9b26ad15462c81548c 0689188d2e8018-Abstract-Conference.html.
26. ^Kambhampati S (2024). "Can Large Language Models Reason and Plan?" Ann NY Acad Sci. **1534**(1):1518. doi:10.1111/nyas.15125.
27. ^Huang J, Chen X, Mishra S, Zheng HS, Yu AW, Song X, Zhou D (2024). "Large Lan guage Models Cannot Self-Correct Reasoning Yet." In: The Twelfth International Conference on Learning Representations. https://openreview.net/forum?id=IkmD 3fKBPQ.
28. ^Lappin S (2024). "Assessing the Strengths and Weaknesses of Large Language M odels." J Log Lang Inf. **33**(1):920. doi:10.1007/s10849-023-09409-x.
29. ^Huckle J, Williams S (2025). "Easy Problems That LLMs Get Wrong." In: Arai K (e d.). Advances in Information and Communication. FICC 2025. Lecture Notes in Ne

tworks and Systems, Vol. 1283. Springer, Cham. pp. 313332. doi:10.1007/978-3-031-84457-7_19.

30. ^*Gonen H, Blevins T, Liu A, Zettlemoyer L, Smith NA (2025). "Does Liking Yellow I mply Driving a School Bus? Semantic Leakage in Language Models." In: Proceedi ngs of the 2025 Conference of the Nations of the Americas Chapter of the Associat ion for Computational Linguistics: Human Language Technologies (Volume 1: Lo ng Papers). Association for Computational Linguistics. pp. 785798. doi:10.18653/v 1/2025.naacl-long.35.*
31. ^*Sacks O (1985). The Man Who Mistook His Wife for a Hat: And Other Clinical Tal es. Summit Books. ISBN 9780671554712.*
32. ^*Floridi L (2026). "AI and Semantic Pareidolia: When We See Intelligent Consciou sness Where There Is None." Philos Technol. **39**:36. doi:10.1007/s13347-026-01052-1.*
33. ^*Queloz M (n.d.). "Can Word Models Be World Models? Language as a Window O nto the Conditional Structure of the World." Manuscript. https://philpapers.org/re c/QUECWM.*
34. ^*Masi M (2024). "The Origin and Nature of Speech in Abhinavagupta and Sri Aur obindo." PhilArchive. https://philarchive.org/rec/MASTNA-6.*
35. ^*Harnad S (2025). "Language Writ Large: LLMs, ChatGPT, Meaning, and Underst anding." Front Artif Intell. **7**:1490698. doi:10.3389/frai.2024.1490698.*
36. ^*Vallor S (2024). "The Danger of Superhuman AI Is Not What You Think." Noema Magazine. https://www.noemamag.com/the-danger-of-superhuman-ai-is-not-what-you-think/.*
37. ^*Jackson F (1982). "Epiphenomenal Qualia." Philos Q. **32**(127):127136. doi:10.2307/2960077.*
38. ^*Landau B, Gleitman LR (1985). Language and Experience: Evidence from the Bli nd Child. Harvard University Press. https://ww.bibliovault.org/BV.book.epl?ISBN=9780674510265.*
39. ^*von Senden M (1960). Space and Sight: The Perception of Space and Shape in th e Congenitally Blind Before and After Operation. Methuen. https://openlibrary.or g/books/OL19137563M/Space_and_sight.*
40. ^*Held R, Ostrovsky Y, de Gelder B, Gandhi T, Ganesh S, Mathur U, Sinha P (2011). "The Newly Sighted Fail to Match Seen with Felt." Nat Neurosci. **14**(5):551553. doi: 10.1038/nn.2795.*
41. ^*Herzog M (2008). "Binding Problem." In: Binder MD, Hirokawa N, Windhorst U (eds.). Encyclopedia of Neuroscience. Springer. pp. 388391. doi:10.1007/978-3-540-29678-2_626.*
42. ^*Panksepp J (1998). Affective Neuroscience: The Foundations of Human and Ani mal Emotions. Oxford University Press. https://global.oup.com/academic/product/affective-neuroscience-9780195178050.*
43. ^*Varela FJ, Rosch E, Thompson E (1991). The Embodied Mind: Cognitive Science a nd Human Experience. MIT Press. ISBN 9780262220422.*
44. ^*Thompson E (2011). "Living Ways of Sense Making." Philos Today. **55**(Suppleme nt):114123. doi:10.5840/philtoday201155Supplement14.*
45. ^*Gibson JJ (2014). The Ecological Approach to Visual Perception. Psychology Pres s. https://www.taylorfrancis.com/books/mono/10.4324/9781315740218/ecological-approach-visual-perception-james-gibson.*
46. ^*Millikan RG (2021). "Neuroscience and Teleosemantics." Synthese. **199**:2457246 5. doi:10.1007/s11229-020-02893-9.*

47. ^a, ^bFriston K (2010). "The Free-Energy Principle: A Unified Brain Theory?" Nat Rev Neurosci. **11**(2):127138. doi:10.1038/nrn2787. Froese T, Taguchi S (2019). "The Problem of Meaning in AI and Robotics: Still with Us After All These Years." Philosophies. **4**(2):14. doi:10.3390/philosophies4020014.
48. ^Seth AK (2013). "Interoceptive Inference, Emotion, and the Embodied Self." Trends Cogn Sci. **17**(11):565573. doi:10.1016/j.tics.2013.09.007.
49. ^O'Regan JK, No A (2001). "A Sensorimotor Account of Vision and Visual Consciousness." Behav Brain Sci. **24**(5):939973. doi:10.1017/S0140525X01000115.
50. ^Brandom R (2007). "Inferentialism and Some of Its Challenges." Philos Phenomenol Res. **74**(3):651676. doi:10.1111/j.1933-1592.2007.00044.x.
51. ^Boleda G (2020). "Distributional Semantics and Linguistic Theory." Annu Rev Linguist. **6**:213234. doi:10.1146/annurev-linguistics-011619-030303.
52. ^Barrett NF (2025). "Experience and Nature in Pragmatism and Enactive Theory." Phenomenol Cogn Sci. **24**:147169. doi:10.1007/s11097-024-10012-z.
53. ^Gubelmann R (2024). "Large Language Models, Agency, and Why Speech Acts Are Beyond Them (For Now) A Kantian-Cum-Pragmatist Case." Philos Technol. **37**:32. doi:10.1007/s13347-024-00696-1.
54. ^Virenque L, Mossio M (2024). "What Is Agency? A View from Autonomy Theory." Biol Theory. **19**(1):1115. doi:10.1007/s13752-023-00441-5.
55. ^Beckmann P, Queloz M (2026). "Mechanistic Indicators of Understanding in Large Language Models." Philos Stud. **183**:17471792. doi:10.1007/s11098-026-02513-1.
56. ^Song P, Han P, Goodman N (2026). "Large Language Model Reasoning Failures." arXiv. doi:10.48550/arXiv.2602.06176.
57. ^Ananthaswamy A (2024). "How Close Is AI to Human-Level Intelligence?" Nature. **636**(8041):2225. doi:10.1038/d41586-024-03905-1.
58. ^Decarli G, Sella F, Lanfranchi S, Gerotto G, Gerola S, Cossu G, Zorzi M (2023). "Severe Developmental Dyscalculia Is Characterized by Core Deficits in Both Symbolic and Nonsymbolic Number Sense." Psychol Sci. **34**(1):821. doi:10.1177/09567976221097947.
59. ^Feldman MJ, Bliss-Moreau E, Lindquist KA (2024). "The Neurobiology of Interoception and Affect." Trends Cogn Sci. **28**(7):643661. doi:10.1016/j.tics.2024.01.009.
60. ^Porbski A, Figura J (2025). "There Is No Such Thing as Conscious Artificial Intelligence." Humanit Soc Sci Commun. **12**:1647. doi:10.1057/s41599-025-05868-8.

Declarations

Funding: No specific funding was received for this work.
Potential competing interests: No potential competing interests to declare.