# Understanding without Consciousness: Quantum Structures and the Autonomy of Meaning

Luca Sassoli de Bianchi[1] and Massimiliano Sassoli de Bianchi[1,2]

[1] Laboratorio di Autoricerca di Base
Via Cadepiano 18, 6917 Barbengo, Switzerland
E-Mail: sdb.luca@protonmail.ch

[2] Centre Leo Apostel for Interdisciplinary Studies, Vrije Universiteit Brussel
Krijgskundestraat 33, 1160 Brussels, Belgium
E-Mail: msassoli@vub.ac.be

## Abstract

Debates about whether large language models (LLMs) genuinely understand the language they produce often rest on the assumption that understanding necessarily entails conscious experience. In this article, we argue that this assumption is unwarranted. By drawing a principled distinction between *structural meaning* and *phenomenal access to meaning*, we defend the thesis that genuine understanding does not require consciousness. We situate our position within the Turing tradition, while responding to classical objections such as Searle's Chinese Room by showing that they conflate semantic competence with conscious awareness. After distinguishing symbols, information, and meaning, we argue that meaning constitutes an autonomous, abstract domain characterized by contextuality, potentiality, and non-classical composition. We then show how LLMs operate within this domain by exhibiting sensitivity to semantic structure rather than merely manipulating symbols. Drawing on results from quantum cognition, we further argue that both human and artificial semantic intelligence are naturally modeled by quantum-like structures, revealing a convergence between biological and artificial cognition at the level of meaning. We conclude that LLMs instantiate a genuine, though non-phenomenal, form of semantic intelligence, and that acknowledging this requires a revision of inherited concepts of understanding rather than an appeal to consciousness as a necessary condition.

**Keywords:** Non-conscious understanding; Structural meaning; Artificial intelligence; Large language models; Semantic intelligence; Consciousness; Quantum structures; Quantum cognition; Bose-Einstein statistics; Semantics vs. syntax.

## 1. Introduction

Recent years have seen growing debate within the scientific community about whether modern artificial intelligence (AI) genuinely understand natural language and the situations it describes (Mitchell and Krakauer, 2023). A prominent example of an author who denies that AI understands the sentences it hears, writes, or speaks in the way humans do is Roger Penrose. In a debate with Emanuele Severino, Penrose stated (Penrose & Severino, 2022):

> I am only going to concentrate to one thing, which is the word "understanding". I relate three words, which I don't know the meaning of. One of them is "intelligence", one of them is "consciousness", and the other one is "understanding" [...]. Coming from a mathematical background I don't need to understand the words if I can say something about the connections between them. So I would say that in ordinary usage the word "intelligence" requires "understanding", and I would say that in the normal usage "understanding" requires some "awareness". So I would say that for an entity to be "intelligent", in the ordinary usage of the word, it would have to be aware, it would have to be conscious.

A direct consequence of Penrose's premise is that no entity can be considered intelligent if it lacks consciousness. Perhaps the most prominent thinker who has argued that genuine understanding requires consciousness is the philosopher of mind John Searle. Through his well-known Chinese Room argument, he contended that an unconscious entity can at best simulate understanding but cannot genuinely possess it (Searle, 1980).

In his Chinese Room thought experiment, Searle asks us to imagine a person who does not understand Chinese locked inside a room and provided with a rulebook for manipulating Chinese symbols. By following purely formal rules, the person can produce symbol strings that are indistinguishable from those of a native Chinese speaker, even though neither the person nor the system as a whole understands Chinese.[1] Searle's conclusion is that syntactic symbol manipulation, no matter how sophisticated, is insufficient for semantic understanding. Consequently, an unconscious computational system, operating at its fundamental level solely on such manipulation, cannot operate at the level of semantics: it cannot understand what it is doing or possess cognitive states in the intrinsic, intentional sense.

On the other hand, among the authors who believe that AI can truly understand meaning, perhaps the best known and most paradigmatic is Alan Turing. At the beginning of his famous 1950 paper, he poses the following question: "Can machines think?" (Turing, 1950). To address this question, he proposed an *imitation game*, now better known as the *Turing test*, that he used to shift attention away from inner experience to outward manifestations of thought. This because Turing explicitly rejected first-person experience as a viable scientific criterion for assessing the authenticity of a thought process, noting that requiring consciousness as evidence would lead to a solipsistic impasse.

So, although Turing did not explicitly distinguish between phenomenal and non-phenomenal forms of understanding, as we will do in this article, he implicitly treated thinking as a capacity that can be assessed independently of consciousness, observing that we routinely assess levels of understanding in humans through linguistic competence and contextual responsiveness rather than through access to subjective experience.

Moreover, in his discussion of *Lady Lovelace's objection*, Turing also argued that machines can in fact produce novel and unforeseen outcomes, and his proposal of learning or "child" machines anticipated the idea that intelligence can emerge through structural organization and interaction rather than being grounded in an intrinsic phenomenal essence.

Many authors, following Turing's line of thought, have argued in different ways that semantic intelligence admits non-conscious instantiations, with LLMs providing a concrete illustration of this possibility (Mitchell & Krakauer, 2023). Just to give one example among many, Agüera y Arcas contends that understanding does not require embodiment or inner subjective experience. On his account, LLMs

[1] It has been argued, in response to Searle, that while the individual inside the room does not understand Chinese, the system as a whole does; see, e.g., Harnad (1990) and the discussion in Searle (1980). We agree with Searle that, as originally formulated, the Chinese Room system does not genuinely understand Chinese. This is because, given the primitive nature of the rulebook in that scenario, there is no internal modelling of the abstract semantic structure of the human language: the system operates solely through predefined syntactic rules without true sensitivity to meaning. For this reason, the behaviour it produces is best described as a simulation of understanding rather than as an instance of semantic competence, as we will try to argue in this article.

exhibit genuine understanding insofar as they can participate coherently in social and dialogic contexts, model human beliefs and expectations (that is, display a form of theory of mind), and maintain semantic consistency across interactions. Consistent with Turing's argument, he further maintains that because we lack objective access to the inner states of any mind – human or artificial – the distinction between "real" and "fake" understanding is ultimately ill-posed. Quoting from Agüera y Arcas (2022):

> The understanding of a concept can be anywhere from superficial to highly nuanced; from abstract to strongly grounded in sensorimotor skills; it can be tied to an emotional state, or not; but it is unclear how we would distinguish "real understanding" from "fake understanding." Until such time as we can make such a distinction, we should probably just retire the idea of "fake understanding."

Similarly, Manning (2022) emphasizes that understanding is a matter of degree and consists in possessing a rich network of meaningful connections, rather than in grounding words in subjective experience. He argues that LLMs genuinely acquire linguistic meaning insofar as they build dense networks of associations linking words, sentences, and world knowledge. Consequently, while LLMs' understanding is incomplete and in need of further augmentation, it nonetheless qualifies as real understanding.

It would be beyond the scope of this article to enumerate the many thinkers who have rejected or accepted, directly or indirectly, the core idea underlying the cognitive version of the strong AI hypothesis, namely, the claim that an unconscious entity, such as a computer running an LLM, could genuinely understand the sentences it processes and produces, that is, possess genuine cognition.

Beyond the debate over consciousness and perception, we observe that the competing views are separated by deeper conceptual distinctions. One side, for example, posits a sharp qualitative divide between real and simulated systems, or between syntactic and semantic processes. The other side, by contrast, treats these divides as limiting cases along a continuum: the transition from complex syntax to rudimentary semantics is regarded as gradual rather than categorical, and a simulation that progressively captures more of a system's properties may eventually become indistinguishable from it, both externally and internally.

In this debate, we situate ourselves in the "Turing camp," and our aim is to strengthen the standard arguments for the thesis that an entity such as an LLM genuinely understands the words it uses, i.e., that it operates at the level of meaning rather than merely at the level of symbols and syntax.[2] Our claim, which will also draw on recent advances in the field of *quantum cognition*, is that an LLM constitutes an authentic form of *semantic intelligence*, capable of operating on, generating, and responding to conceptual situations in virtue of their meaning. In other words, genuine understanding is possible without requiring that the understanding entity be, by that very fact, a conscious subject with phenomenal access to the meanings it processes.

It follows that the domain of meaning is far broader than the domain of consciousness: one region is inhabited by conscious, experiential understanding, while another – likely the larger – is populated by forms of nonconscious understanding. We use the term "consciousness" here in a sense aligned with human subjective experience, and not in the broader sense invoked by certain forms of panpsychism – such as

[2] The distinction between syntactic and semantic structures should not be understood as a sharp opposition between rule-following devoid of significance and meaningful, context-aware compositions. Syntax itself involves structured relations among symbols that contribute locally to intelligibility; formal approaches to the syntax-semantics interface – such as constraint-based frameworks developed in lexical-functional grammar and later known as *glue semantics* – model how syntactic structure constrains and interacts with semantic composition (Dalrymple, *et al.*, 1993). The boundary between syntax and semantics is therefore not absolute but graded. What we intend to deny is not that syntax can carry any meaning at all, but that purely syntactic processing – understood as adherence to predefined formal rules without context-sensitive organization of conceptual relations – is sufficient for genuine semantic understanding. In the present framework, syntax can be seen as contributing to micro-coherence within a message, whereas meaning concerns, also and above all, macro-coherence across contexts, uses, and potential interpretations.

micropsychism (Strawson 2006) or cosmopsychism (Goff 2017) – in which experiential properties are treated as fundamental and widespread, though not necessarily as constituting full-fledged conscious subjects.

The article is organized as follows. In Section 2, we introduce a distinction between *symbols*, *information*, and *meaning*, arguing that many objections to non-conscious understanding arise from conflating these different levels of description. In Section 3, we develop the notion of *structural meaning* and argue for the autonomy of understanding from phenomenal consciousness, drawing on examples from mathematics, language, and artificial systems. Section 4 applies this framework directly to LLMs, showing that their behavior is best described as a form of non-conscious *semantic intelligence* rather than mere syntactic simulation. In Section 5, we draw on results from quantum cognition to argue that the abstract structure of meaning is naturally captured by the quantum formalism, and we show that both human cognition and LLM-generated language exhibit convergent quantum-like semantic organization. Finally, Section 6 summarizes the main conclusions and discusses their implications for contemporary debates on understanding, intelligence, and consciousness, including their relevance for panpsychist and illusionist approaches to the philosophy of mind.

## 2. Symbols, Information, and Meaning

To prevent a widespread confusion that underlies many objections to non-conscious semantic intelligence, it is important to start by explicitly demarcating between three notions: *symbols*, *information*, and *meaning*.

Symbols are concrete, physically instantiated, spatiotemporal entities – such as marks, forms, sounds and signals – that can be discriminated, stored, and manipulated. Information and meaning, by contrast, are abstract, non-spatiotemporal notions. In a very general sense, symbols can be said to carry information, and information, in turn, to carry meaning, much as a moving particle carries momentum and momentum carries energy (Falk *et al.*, 1983). The carrier relation here does not imply reduction or identity: information is not a property intrinsic to symbols, nor is meaning intrinsic to information. Rather, each level introduces a new layer of abstraction and relational organization.[3]

A persistent source of confusion in the literature concerns the use of the term "information" as a synonym for *meaningful information*, thereby blurring the distinction between information and meaning. This ambiguity becomes even more pronounced with the introduction of *quantum information*. While in physics quantum information refers to well-defined properties and measures associated with quantum states, the quantum formalism has been successfully employed in fields such as *quantum cognition* precisely to model contextuality, superposition, and interference of meanings. In these applications, the formalism operates primarily at the level of meaning rather than information in the strict, Shannon-theoretic sense. The reuse of the same terminology across these different contexts thus risks conflating distinct conceptual

[3] One might object that the distinction between symbols, information, and meaning does not correspond to sharply separable ontological categories, but rather to different descriptive levels that may blur into one another depending on context. For example, symbols can be used to encode information in a message, such as letters in a word. However, any message can be reused as a new symbol in a different encoding pattern to create new information, such as a phrase. These properties make it difficult to make a categorical distinction between symbols and information. An analogous situation arises in physics, where it is widely held that there may be no fundamentally classical entities distinct from quantum ones, classical behaviour emerging instead as an effective regime under specific conditions. In a similar spirit, symbols may themselves be understood in more abstract terms, and the relation between the abstract and the concrete may vary with the context in which a system is considered. We do not deny this possibility. The distinctions introduced here are not intended as absolute metaphysical demarcations, but as conceptually and explanatorily useful distinctions that track different organizational roles within cognitive systems. Their relevance lies in how these roles manifest under different contextual constraints, rather than in the assumption that symbols, information, and meaning constitute irreducibly separate kinds of entities.

levels, unless explicit care is taken to specify whether quantum notions are intended to describe informational structure or semantic organization.

Information, like energy, is an abstract quantity that is not directly observed but inferred. What is directly accessible are concrete empirical traces – such as distinguishable objects, experimental outcomes, or recorded events – while the amount of information is assigned only a posteriori by applying a theoretical framework that specifies how these distinctions are to be quantified.

For example, in Shannon's approach, the information content of a string of symbols is evaluated by assigning probabilities to the possible symbols and computing a corresponding numerical measure, without any reference to their meaning. In this operational sense, information is not something directly observed in the symbols themselves, but a quantity inferred by applying a theoretical framework that specifies how observable distinctions among alternatives and their relative availabilities are to be quantified. In this respect, information is analogous to physical quantities such as energy. Indeed, energy is also assigned on the basis of observed motions, interactions, and constraints, according to formal rules provided by a physical theory. Different theoretical contexts may lead to different expressions or decompositions of energy, just as different informational frameworks may lead to different measures of information. In both cases, what is empirically given are concrete events and distinctions, while the corresponding abstract quantity is defined relationally and contextually by the rules of the adopted framework.

Meaning, on the other hand, does not admit a purely operational or quantitative definition comparable to that of information, yet it exhibits robust structural properties. It is inherently contextual, can exist in latent (potential) form, and combines in a non-compositional manner, such that the meaning of a whole cannot, in general, be reconstructed from the meanings of its parts alone. Meaning functions as a connective structure linking conceptual entities, contexts, and possible actions, and it constrains behavior in ways that cannot be accounted for at the level of symbols or information alone.[4]

Although meaning requires cognitive entities for its production and articulation, it can leave traces in the symbolic structures it organizes that can be analyzed independently of semantic awareness, revealing statistical and structural signatures characteristic of its presence. Also, meaning appears to be structured in a way that is naturally captured by the quantum formalism, whose core features – contextuality, superposition, indistinguishability, and measurement – mirror its observed properties. Indeed, the successful application of quantum models in domains such as cognition and language suggests that the quantum formalism is not merely a mathematical metaphor, but a particularly well-suited framework for modeling the structure of meaning itself (more on this in Section 5).

## 3. Structural Meaning and the Autonomy of Understanding

Having made the important distinction between symbols, information and meaning, we will now consider why the claim that LLMs genuinely operate at the level of meaning is not obvious to many. This is due to the deeply rooted, distinctly human tendency to equate understanding with access to meaning, or the conscious experience of meaningful information.

From this perspective, anything outside consciousness is relegated to mere syntactic processing, i.e., symbol manipulation devoid of genuine (or sufficiently deep) semantics. This prejudicial stance excludes from the category of understanding any conception of meaning understood in structural, functional,

[4] By "information alone" we mean structured distinctions that can be generated and processed without semantic integration, such as the numerical readings of a thermometer, which are informative but become meaningful only when embedded in a context that links them to expectations and possible actions.

generative, emergent, and potential terms. Such a reduction is problematic, as it conflates the conscious manifestation of meaning with the full ontological domain of meaning itself.

We do not claim, a priori, to exclude the possibility that meaning might also originate in consciousness, i.e., in conscious informational activity (as a panpsychist framework would suggest). Yet, this possibility does not imply that meaning lacks an autonomous reality independent of its putative conscious origin. Consider, for example, the mathematical notion of a group. Although this concept emerged from the conscious reflective activity of mathematicians such as Évariste Galois, once its meaning was brought into being through definition, a *terra incognita* of possible relations, properties and structures appeared, existing independently of any subsequent conscious exploration mathematicians may have of it,. In other words, the conceptual domain that was created possesses a coherence and autonomy not reducible to that initial intentional act of invention of a new notion.

One may of course ask whether such a conceptual landscape would still exist if all mathematicians were suddenly to cease to exist. This question is akin to asking whether a tree falling in a forest makes a sound if no one is present to hear it. Addressing it requires clarifying what it means for something to be a property or a feature of reality. As emphasized by Einstein, Podolsky, and Rosen in their celebrated reality criterion (Einstein *et al.*, 1935, Sassoli de Bianchi, 2011), a property can be understood, in the final analysis, as a *state of prediction*: if one can predict with certainty the outcome of an experimental test, then there exists an element of reality (an actual property) corresponding to that outcome. From this perspective, what matters is not whether a mathematician is actually present to experience the conceptual landscape created by Galois' definition, but whether, if such a landscape were to be encountered under the appropriate conditions, a mathematician would be able to discover it with certainty. In this sense, the reality of the conceptual structure depends not on its conscious experience, but on its objective availability, as a structure, for stable cognitive exploration.[5]

However, if Galois's definition created a structure of actual properties that mathematicians could subsequently discover, aspects of creation are also constantly at play. Not everything unfolding from a given mathematical definition is predictable: there are properties that remain latent (potential) and whose actualization depends on subsequent creative acts, so that their emergence cannot be predicted in advance. Therefore, they can only be actualized following indeterministic processes. The degree to which the evolution of a mathematical landscape is predictable raises an old philosophical question: whether mathematics is created by humans or merely discovered. Penrose is an example of a thinker who maintains that abstract mathematical truths and forms exist in an independent and objective realm that can be explored and that underpin the physical universe. Consciousness, he also believes, connects these worlds. However, Penrose's position seems more like a hybrid between Platonism and a certain type of constructivism, since he also recognizes the role played by the mathematician's creativity, implying that the mathematical landscape expands in an unpredictable way (Penrose, 1989):

> As I have said, there are things in mathematics for which the term "discovery" is indeed, much more appropriate than "invention" [...]. These are the cases where much more comes out of the structure than is put into it in the first place. One may take the view that in such cases the mathematicians have stumbled upon "works of God". However, there are other cases where the mathematical structure does not have such a compelling uniqueness, such as when, in the midst of a proof of some result, the mathematician finds the need to introduce some contrived and far from unique construction in order to achieve some very specific end. In such cases no more is likely to come out of the construction than was put into it in the first place, and the word "invention" seems more appropriate than "discovery". These are indeed just "works

[5] This is in accordance with the view of *structuralism* (classically associated with Ferdinand de Saussure), that language is a self-contained structure of relations which exists somewhat independently of its manifestations in specific languages and individual acts of speech (Ress, 2022).

> of man". On this view, the true mathematical discoveries would, in a general way, be regarded as greater achievements or aspirations than would the "mere" inventions.

It is of course not our purpose to discuss how much of the mathematical landscape is discovered and how much of it is created. The important point is that both aspects, of creation and discovery, appear to play a role. So far, the creation (invention, construction) aspect has been performed only by human mathematicians, hence by conscious entities, and in that respect, we can say that most of the meaning in mathematics originated so far from conscious cognitive activity. However, we live in a historical moment where mathematicians already actively collaborate with LLMs in their journey of creations and discoveries of new mathematical results; see for example Schmitt (2025) and Georgiev *et al.* (2025). Hence, even if mathematics, as we know it today, largely originated in conscious mental processes, the mathematics we will be dealing with tomorrow may well incorporate genuinely new inputs arising from artificial unconscious systems. This further illustrates how meaning, once instantiated through cognitive activity, acquires a structural autonomy that allows it to be explored, extended, and actualized independently of the consciousness that initially contributed to its emergence.

Of course, artificial intelligence (AI) systems based on deep neural networks and machine learning learn from texts produced by human authors. Their training therefore depends, at least initially, on mostly conscious activity. Yet, once trained, they generate new linguistic combinations (or new mathematical reasonings), perceived as novel and unexpected, that are coherent, interpretable, and meaningful, even though no consciousness stands behind their production or is directly involved in it. In this respect, they are not "stochastic parrots" (Bender *et al*., 2021), as critics often claim to dismiss their capacity for genuine language understanding, portraying them as systems that merely reproduce the vast datasets on which they were trained. This characterization is misguided, however, since an LLM does not store its entire training corpus in its weights. Rather, it "forms an internal latent representation of the training data, allowing it to respond to novel queries with novel responses" (Sejnowski, 2023).[6]

The example of the mathematical notion of a group illustrates how an *abstract structure of meaning* can exist independently of any conscious subject, while the example of a generative language model demonstrates a process of *meaning actualization* that likewise unfolds without direct conscious involvement. Taken together, these examples support the view that meaning and understanding possess an autonomous reality.

Humans, too, often experience this autonomy from deliberate conscious control. When we speak or write, our words frequently seem to compose themselves, as if guided by processes beyond direct introspection, unfolding through a dynamic that is partly deterministic and partly indeterministic and shaped by our immersion in language and context.[7] In such moments, we are more spectators than creators. We may feel an emotion as a sentence takes form, or when we reread it, but much of the compositional process proceeds by following the contextual coherence of the emerging words. In other words, meaning unfolds by preserving its internal coherence relative to the context in which it is expressed. We wrote the following as a possible explanation for this process in Aerts *et al.* (2025a):

[6] The ability to achieve efficient information compression is frequently regarded as indicative of a deeper understanding of the nature of the compressed data.

[7] The indeterministic component arises because, when a subject is confronted with a set of mutually exclusive alternatives, a mental equilibrium state rapidly forms, generated by competing tensions between the initial conceptual state and the available possibilities. This equilibrium can be represented by an abstract elastic membrane whose tension lines connect the initial state to the vertices corresponding to the possible answers (Aerts and Sassoli de Bianchi, 2015). At some point, this equilibrium is unpredictably disrupted, initiating an irreversible *weighted symmetry-breaking process* in which the conceptual state is drawn toward one specific outcome, modeled by the random breaking of the membrane.

> [...] in both language and physics domains, the lack of statistical independence would be produced by a phenomenon of entanglement through contextual updating. To explain what we mean by this mechanism of contextual updating, let us consider the combination of two words in a text. Each word carries meaning but, whenever the two words are combined, their combination also carries its proper meaning, which is not the trivial combination of the meanings associated with the two individual words as prescribed by a classical logical semantics. The new emergent meaning of the combined word arises in a complex contextual way, in which the whole of the context relevant to the story plays a fundamental role. Thus, each time a word is added to a text, a mechanism of updating influenced by the meaning of the whole context occurs, and this updating continues to take place until the end of the story that contains all the words.

Note that a major advance in contemporary LLMs is that they process an input text as a single, contextually integrated whole, rather than merely as a linear sequence of neighboring tokens. Each token is generated by taking into account its relations to all other tokens in the input, so that meaning is computed globally, through *contextual coherence*, rather than in a local sequential way.

According to the above, meaning is not only what is consciously experienced, but also what is structurally organized and coherently maintained via context-sensitive patterns of conceptual relations that support coherent inference and the actualization of potential properties. This is why a non-conscious entity, such as an AI system, can operate at the abstract level of meaning. It does not understand meaning in the sense of undergoing phenomenal experience, but in the sense of *being sensitive and responsive to it*. An AI is responsive to meaning insofar as it is sensitive to the coherence of the signs and representations that embody it, having adequately modeled this abstract level, as we will explain in more detail below. This, too, constitutes a legitimate form of understanding, even if it is not a phenomenal one.[8]

As a final example of how language and meaning can exist and develop independently of their initial conditions – whether or not these involve consciousness – consider the case of programming languages and compilers. The Chinese Room argument naturally invites comparison with a compiler architecture, where the human agent plays the role of a processor executing formal instructions, the rulebook corresponds to the compiler code, and the input and output strings function as the source and target programs. Compilers are instructive semantic artifacts because their existence typically involves a bootstrapping process. In practice, a compiler for a new programming language $L_2$ is initially written in a previously available language $L_1$ and compiled using an existing compiler for $L_1$. This produces a first compiler for $L_2$, written in $L_1$. Once such a compiler exists, it can be used to write a new compiler entirely in $L_2$ itself. At that point, the original $L_1$-based compiler can be discarded, leaving a compiler that is both written in and capable of processing $L_2$.

From a purely structural perspective, the resulting compiler embodies a complete and self-sustaining set of semantic rules governing the language, even though its existence depended historically on a prior linguistic and technical infrastructure. If one were to encounter such a compiler without knowledge of its developmental history, its semantic coherence and operational autonomy would be fully intelligible without reference to its origin. This example illustrates a general feature of language-like systems: although their emergence may require an initial creative or constructive step, once instantiated they acquire a structural independence that allows their meaning-governing rules to function autonomously. In this sense, semantic structures can be historically dependent yet ontologically autonomous, reinforcing the claim that meaning, once brought into being, is not reducible to the processes – conscious or otherwise – that contributed to its initial formation.

[8] Although this view is compatible with certain functionalist insights, it does not rely on classical computational functionalism; rather, it emphasizes the organization of meaning within high-dimensional conceptual spaces whose contextual and quantum-like properties go beyond standard functionalist models, as discussed later in the article.

## 4. Non-Conscious Semantic Intelligence in LLMs

Considering what has been discussed in the previous sections, it is important to distinguish between *structural meaning* and *phenomenal access to meaning*, only the latter being grounded in consciously lived experience, intentionality, and affect.

While perhaps only a conscious subject can inhabit meaning, meaning itself can exist, be generated, inferred, and understood independently of consciousness. A cognitive entity therefore need not be conscious to be *semantically intelligent*;[9] it is sufficient that it exhibits *sensitivity to meaning*, not only in its actualized forms but also in its potential (latent) ones. For this reason, as already observed, even a trained artificial neural network can generate conceptual combinations that have never appeared before, thereby conveying new emergent meanings.

One might object that the mere generation of novel combinations is not sufficient to establish sensitivity to meaning, since even a simple algorithm that randomly concatenates words can produce combinations that have never appeared before. However, novelty alone is not the relevant criterion. Although meaning does not admit a direct quantitative measure in the way energy or information does, its presence can nevertheless be evidenced in objective terms, provided one considers cognitive agents that are sensitive to it and belong to the cultural and linguistic domain in which such meaning emerges.[10] When genuinely new meanings arise, they manifest as abstract connections between concepts: the combined expression is no longer adequately describable as a composition of independently meaningful parts, but functions as a new, integrated conceptual entity.

In the framework of *quantum cognition*, this phenomenon is modeled in close analogy with *quantum entanglement*, whose presence is not directly observed but inferred through violations of classical probabilistic constraints, such as Bell-type inequalities. Empirical studies of conceptual combinations show that meaningful combinations systematically violate the predictions of classical probability theory, indicating the emergence of a new *semantic whole* rather than a mere juxtaposition of components (Busemeyer and Bruza, 2025; Aerts *et al.*, 2025c). Random word combinations may generate formal novelty, but they do not reliably produce such structured, context-sensitive, and non-classical effects. Sensitivity to meaning is therefore evidenced not by novelty per se, but by the capacity to generate and sustain coherent semantic structures whose behavior departs from classical compositional expectations.

When, for example, the chatbot of an LLM translates an unpublished poem into another language, perhaps even in the style of another poet, or when it completes a text that it has never encountered before, it does so neither by randomly picking words, nor by parroting from any dataset, nor by following a predetermined set of formal rules, as in Searle's Chinese Room,[11] but by exploiting learned, context-sensitive

[9] There is no need to precisely define the notion of semantic intelligence here. Let us simply say that this notion doesn't include the capacity for self-awareness, but certainly the ability to learn, reason, create new knowledge, model the world and others, produce abstractions, and predict the future.

[10] The extraordinarily rapid and widespread adoption of LLM-based tools may be regarded as the positive outcome of a large-scale, informal Turing test, one in which the relevant question is no longer whether we are interacting with a human, but whether we are interacting with an entity with which deep and broad meaningful interactions is possible.

[11] It may be objected that LLMs, once trained, operate according to predetermined architectures and update rules, and are therefore deterministic systems whose apparent creativity merely reflects the complexity of their internal statistics. While it is true that an LLM's underlying mechanisms are fixed after training, this does not entail determinism at the level of concrete outputs. In practice, LLMs operate under conditions of stochastic sampling, contextual sensitivity, and cumulative interaction history, so that, at each step of generation, several contextually appropriate responses are available, none of which is fixed in advance as the uniquely determined outcome. Randomness therefore plays a role not merely as an external add-on, but as a mechanism for actualizing one possibility among many latent ones encoded in the model's learned semantic structure. In this

semantic relations. The chatbot's constructions are original and meaningful and we understand them because we, in turn, activate our own sensitivity to meaning. This capacity to generate new conceptual combinations, in a non-predetermined way, and to recognize their significance is a clear indication of what we call *thinking*, a capacity that does not necessarily require awareness of one's own stream of thoughts. Indeed, one can readily distinguish between *thinking* and *conscious thinking*, just as we propose to distinguish between *understanding* and *conscious understanding*, and just as one can easily distinguish between, for example, *dancing* and *conscious dancing*, since one can dance skillfully, expressively, and coherently while being only minimally aware of the detailed movements one is performing, or, more plainly, since even a robot can execute a complex and expressive dance without any accompanying phenomenal awareness.

By contrast, thinkers such as Penrose and Searle treat understanding as a property of consciousness, maintaining that meaning can exist only within conscious experience and intentionality. This view overlooks the fact that the decisive boundary is not between those who understand and those who do not, but between those who experience what they understand and those who do not. We do not deny that meaning can be enriched by experience or can emerge from it; rather, we argue that meaning does not depend on lived experience, because it is not intrinsically tied to phenomenal consciousness.

In other words, although phenomenal experience adds depth and qualitative richness to how information is accessed and integrated, it does not introduce new semantic structures, and understanding is not limited to conscious entities. For example, consciously perceiving the color red enriches one's grasp of "redness" in a way that is inaccessible to a non-conscious entity. However, this experiential enrichment supplements, rather than grounds, the understanding already available to a conscious subject and is not a prerequisite for the existence or manifestation of meaning. This distinction echoes the well-known *knowledge argument* illustrated by Mary's room (Jackson, 1982). Although Mary possesses complete physical and functional knowledge – hence, a full structural understanding – of color vision while confined to a black-and-white environment, she acquires new knowledge when she first experiences color. What Mary gains is a new phenomenal mode of access to information already structurally understood, not an additional layer of meaning.[12]

Direct, conscious experience of the world is not required to speak meaningfully about it; what is required is the construction of a sufficiently complete model of the abstract structure of a language that represents the world, and access to that model. Such access amounts to genuinely understanding meaning about the world and within the world; not all meaning, of course, but at least that portion that can be coherently represented through abstract entities and their possible relations, both actual and potential. As noted earlier, the initial construction of such a model may have depended on perception and conscious agents, but access to the model once constructed does not.

respect, the behaviour of LLMs is closer to that of human cognition when choosing among competing, contextually viable responses in the absence of a uniquely determined answer: unpredictability arises not from arbitrariness, but from the interaction between structured constraints and indeterministic selection.

[12] In psychology and perceptual science, it is customary to distinguish between *stimuli* – the physical inputs impinging on the sensory apparatus – and *percepts*, namely the structured experiential outcomes produced by perceptual processing. Human colour perception, for example, does not mirror the continuous physical spectrum but organizes stimuli into a limited number of categorical percepts, often corresponding to a small set of basic colour categories (such as red, green, yellow, blue, purple, pink, orange, and gray), whose number and boundaries can vary across cultures (Berlin & Kay 1969). In this sense, percepts are already conceptually structured and closely related to meaning. More generally, perceptual systems are known to systematically warp stimulus space through mechanisms such as *categorical perception*, whereby distinctions are sharpened or compressed in ways that reflect learned or biologically salient categories (Goldstone & Hendrickson 2010). When Mary first experiences colour, she therefore does not merely gain access to raw sensory information, but already to a certain meaningful content, via the organization of stimuli into pre-established perceptual categories.

An LLM's chatbot does not merely simulate meaning: it genuinely engages with meaning by operating at a sufficiently abstract level within its neural architecture and by capturing a substantial portion of the conceptual landscape of human language. If an entity can reproduce every behavior associated with a given cognitive attribute, including sustained, coherent communication with humans, then it becomes inappropriate to describe its performance as mere simulation. The attribution of simulation would require an independent criterion of genuine semantic competence that LLMs fail to meet, yet no such criterion is available once meaning is understood structurally rather than phenomenally. What we observe is not a simulation of cognition, but rather, a *variant* of it. An LLM does not simulate semantic intelligence; it constitutes a variant of semantic intelligence. Nor does it simulate consciousness, and at present there are no strong arguments for or against the hypothesis that one day it may possess consciousness, whether very similar or very different from our human consciousness.[13]

A further objection sometimes raised against attributing genuine meaning to LLMs is that meaning is said to require direct causal coupling to the external world, often understood in embodied or enactive terms, and that the semantic competence of LLMs is therefore merely parasitic on human intentionality in a way that human cognition itself is not. According to this view, because LLMs lack (for the time being) their own sensory apparatus and bodily engagement with the environment, the meanings they manipulate are at best derivative rather than original (Harnad, 1990, Varela *et al.*, 1991). However, this objection rests on an overly restrictive conception of access to the world. Human cognition itself does not enjoy unmediated access to reality, but only access through a complex network of sensory organs, neural transduction, and socially mediated learning. From this perspective, the difference between human and artificial cognition is not that one is world-coupled and the other is not, but that the modes of coupling differ; see for example Abdou *et al.* (2021) and Xu *et al.* (2025) for two case studies.[14]

In the case of LLMs, humans function as a distributed sensory and interpretive interface: through the texts they produce, humans convey structured traces of their interactions with the world, including perceptual regularities, causal relations, and pragmatic constraints. Once trained on such data, an LLM acquires indirect but genuine access to the world's structure, sufficient to support coherent, context-sensitive semantic behavior. The fact that this access is mediated does not render the resulting meanings merely derivative, any more than the mediation of human perception by sensory organs renders human meaning derivative.[15] What matters for semantic competence is not the presence of a particular kind of embodiment, but the availability of stable, law-governed correlations between representational structures and worldly regularities. From this standpoint, the semantics instantiated by LLMs is not parasitic in principle in a way that distinguishes it from human semantics, since all meaning is historically and

[13] The present account makes no assumptions about the possibility of machine consciousness (Chalmers, 2023); it concerns only the conditions under which non-conscious systems may instantiate structural understanding.

[14] A related bias is the assumption that LLMs' thinking is more limited than ours because it occurs only in response to human prompts. This argument overlooks the fact that we do not know whether human cognition can exist in the absence of stimuli.

[15] One might object that, in the present state of development, humans contribute to LLMs far more than a sensory interface. In addition to providing perceptual traces through language, humans also supply the questions posed to the system, the pragmatic goals guiding interaction, the evaluative and normative standards used to assess responses, and the training data from which the model's internal structures are learned. From this perspective, current LLMs may be seen as operating within a human-provided epistemic and cultural scaffold, and it is a legitimate concern whether they genuinely enrich the global pool of knowledge or merely reorganize and possibly amplify existing human-generated content. This situation can be compared to the previously mentioned early stages of compiler bootstrapping, in which a compiler for a new language is still written in a prior language rather than being fully self-hosted. Acknowledging this dependence, however, does not undermine the claim that LLMs already exhibit non-conscious semantic competence. It rather suggests that their current form of understanding is historically and contextually mediated, and that full epistemic autonomy – if achievable at all – would require further stages of semantic and practical self-grounding.

relationally grounded. It is, however, differently grounded: structurally anchored in the collective, historically accumulated coupling between human cognition, language, and the world.

## 5. Quantum Structure and the Convergence of Human and Artificial Cognition

Our thesis is significantly strengthened by observing the success of the field of research known as *quantum cognition*; see Busemeyer & Bruza (2025), Aerts *et al.* (2025b,c), and the references cited therein. Quantum cognition developed with the main aim of explaining human cognitive and decision-making behaviors using the modeling tools of quantum mathematics. Although it deals with the human mind and how it interacts with entities of meaning, no one in this field addresses the issue of consciousness. This is simply because, as we have tried to argue, consciousness concerns only the qualia of subjective experience, while human cognitive processes can be studied and described regardless of whether they are conscious or not. In other words, according to the quantum cognition research community, the quantum formalism does not describe consciousness, but the abstract structures of meaning and cognitive behavior that can be analyzed independently of consciousness. To quote Busemeyer (2025): "Quantum cognition theories have avoided addressing fundamental issues about consciousness and have remained agnostic with respect to the quantum brain hypothesis."

It was certainly unexpected to observe the remarkable effectiveness of quantum formalism in modeling the structural organization of meaning and information, independently of whether such structures are phenomenally accessed, as though quantum mathematics and its associated notions were the natural framework for describing it, capturing contextuality, potentiality, and the interplay between the abstract and the concrete, the distinguishable and the indistinguishable. This surprising relevance of the quantum model has inspired a bold yet highly explanatory hypothesis: the *conceptuality interpretation of quantum mechanics* (Aerts *et al.*, 2025b,c), which proposes that physical systems can also be described in conceptual and cognitive terms, that is, as governed by dynamics involving an exchange of meaning (which, however, would have nothing to do with human meaning). Importantly, this interpretation does not require that such exchanges of meaning be accompanied by consciousness: it supports a *pancognitivist* view, not a *panpsychist* one.

It is important at this point to address Searle's warning that we humans have the tendency of anthropomorphizing human artefacts (Searle, 1980):

> We often attribute "understanding" and other cognitive predicates by metaphor and analogy to cars, adding machines, and other artifacts, but nothing is proved by such attributions. We say, "The door *knows* when to open because of its photoelectric cell," "The adding machine knows how (*understands how*, is *able*) to do addition and subtraction but not division," and "The thermostat *perceives* changes in the temperature." The reason we make these attributions is quite interesting, and it has to do with the fact that in artifacts we extend our own intentionality; our tools are extensions of our purposes, and so we find it natural to make metaphorical attributions of intentionality to them; but I take it no philosophical ice is cut by such examples. The sense in which an automatic door "understands instructions" from its photoelectric cell is not at all the sense in which I understand English. If the sense in which Schank's [Schank and Abelson, 1977] programmed computers understand stories is supposed to be the metaphorical sense in which the door understands, and not the sense in which I understand English, the issue would not be worth discussing. But Newell and Simon (1963) write that the kind of cognition they claim for computers is exactly the same as for human beings. I like the straightforwardness of this claim, and it is the sort of claim I will be considering. I will argue that in the literal sense the programmed computer understands what the car and the adding machine understand, namely, exactly nothing. The computer understanding is not just (like my understanding of German) partial or incomplete; it is zero.

How can we avoid the criticism that Searle directed at the work of Roger Schank and his colleagues at Yale, who claimed that their machines understood the stories they processed? We can avoid it for two reasons. The first is that we carefully distinguished *understanding* from *conscious understanding*, thereby avoiding the assumption – central to Searle's critique – that genuine understanding must be conscious understanding. The second is that today's LLMs operate in a fundamentally different organizational manner from the artificial systems to which Searle was responding. This difference should not be understood as a categorical break at the level of computation or physical implementation: both traditional programs and LLMs are realized on deterministic computational substrates and ultimately manipulate symbols according to formal rules. The relevant contrast lies instead in the level at which semantic structure is represented and generated. Schank's programs, based on *conceptual dependency theory* and related frameworks, relied on explicitly hand-crafted representations: predefined scripts, frames, and causal templates whose semantic roles were specified in advance by human designers. By contrast, LLMs acquire their semantic organization through large-scale statistical learning, resulting in distributed, high-dimensional representations in which meaning is not locally encoded in explicit rules, but emerges from patterns of relations across the system as a whole.

In other words, LLMs acquire their internal representations autonomously through exposure to vast corpora of text, learning high-dimensional conceptual relations encoded in distributed vector spaces. In these spaces, semantic content is stored as abstract structures that behave like superposition states, whose meaning becomes determinate only in relation to a contextual prompt. Moreover, LLMs exhibit entanglement-like correlations among concepts, where the state of one representation constrains and co-defines the state of others in non-classical ways. Their semantic competence thus emerges from quantum structures, i.e., quantum-like geometries of meaning, characterized by contextuality, potentiality, and non-Boolean combination rules, rather than from any set of pre-specified human cognitive models. As a result, LLMs can generate novel and contextually coherent interpretations and conceptual combinations that were never explicitly encoded by their designers, behaviors that Schank's systems were neither intended nor able to perform.

Let us offer a more specific argument for why the semantic intelligence of LLMs can be regarded as similar to human semantic intelligence. This argument will also provide a different perspective on why both human cognition and artificial cognition naturally employ quantum structures to model and handle the substance of meaning. Let us consider a sufficiently long text written by a human, capable of conveying meaning. Imagine it as the analogue of a "gas of words," where these words behave like quantum entities that we shall call *cognitons*, following Aerts and Beltran (2020). More precisely, each word is treated as a specific state of a cogniton, characterized by a certain energy. Since most words in the text occur many times, multiple cognitons can occupy the same state. They are therefore similar to *bosons*: they are indistinguishable (interchanging their positions leaves the meaning of the text invariant) but do not obey Pauli's exclusion principle.

We can then assign the state corresponding to the lowest energy level – the ground state – to the word that appears most frequently in the text. The next energy level can be assigned to the second most frequent word, and so on, thereby defining the entire *one-entity energy spectrum* of this gas of cognitons. As a first approximation, all these energy levels can be taken to be equally spaced. Quantum indistinguishability then tells us that even when its constituents do not interact, a gas of bosons exhibits statistical correlations that prevent the particles from being independent. It is as if a mysterious *effective force* governed their collective behavior. As a result, the correct statistical distribution for such a system is not the classical Maxwell-Boltzmann (MB) distribution but the quantum Bose-Einstein (BE) one.

The BE distribution follows from two fundamental postulates: the *tensor-product postulate*, which governs how quantum systems are combined, and the *symmetrization postulate*, which constrains the admissible states of collections of identical quantum entities. For bosons, the allowed states must be symmetric under permutation, and such symmetrized states are necessarily entangled states. Thus, the effective force responsible for the lack of statistical independence and for the emergence of the BE distribution is, in fact, a consequence of the entanglement-induced connections among the particles and the consequent formation of a collective *coherent domain* (Aerts *et al.*, 2025a).

Surprisingly, as shown by Aerts and Bertrand (2020), cognitons – that is, the words composing a natural-language text – also obey the BE statistics and behave analogously to a *Bose-Einstein condensate* at a temperature close to absolute zero. In physics, a Bose-Einstein condensate, often referred to as the "fifth state of matter," is a state that arises when a collection of identical (bosonic) quantum particles is cooled to extremely low temperatures, so that a macroscopic number of them occupy the same lowest-energy state and act coherently as a single quantum system. In other words, the collection of entities enters a regime in which thermal fluctuations are negligible and collective quantum effects dominate. In this regime, the system exhibits maximal coherence and strong correlations among its constituents, even in the absence of direct interactions.

So, if we examine how different words in a text (cognitons in distinct states) populate the various energy levels, what we find is that their distribution always follows the BE distribution, and not the MB one. Since it arises from an underlying quantum structure, an analogous structure must be present in natural-language texts as well.[16] The notions of *meaning* and *quantum coherence* play here an equivalent role, and as research in quantum cognition has shown in many ways, *quantumness* and *conceptuality* appear to designate two aspects of the same underlying reality: one that can manifest at different levels of organization within the world.[17]

In other words, human language, in its role as a carrier of meaning, exhibits a quantum organization governed by principles such as superposition and the indistinguishability of identical words within a text. What matters for the thesis we defend here is that the same analysis can be performed, in an entirely equivalent way, on texts generated by LLMs, which also form Bose-Einstein condensates close to absolute zero, thereby revealing an evolutionary convergence between human and artificial cognition (Aerts *et al.*, 2025d). Although they have evolved in very different ways, both human cognition and artificial cognition make use of abstract quantum-like geometries. Distinct evolutionary trajectories have thus given rise to abstract structures with similar cognitive significance.[18]

[16] It is worth noting that there is a formal correspondence between Bose-Einstein statistics and Zipf's law which was first identified by Bruce M. Hill (1974) and later by Viktor Maslov (2005). However, it is only when this correspondence was independently rediscovered by Diederik Aerts and Lester Beltran (2020), within the explanatory frameworks of quantum cognition and the conceptuality interpretation, that its deeper significance became clear, namely, that Zipf's law reveals the presence of an underlying quantum structure.

[17] The quantum-theoretic formalism is not used metaphorically in quantum cognition but is supported by empirical fits to multiple linguistic and cognitive data, indicating that the quantum framework genuinely captures structural regularities that classical probabilistic models cannot reproduce (Busemeyer and Bruza, 2025). Note that this correspondence between *quantumness* and *conceptuality* is also central to the more speculative *conceptuality interpretation*, which goes as far as to posit not only the presence of quantum structures in human cognitive processes but also, in a "move à la de Broglie," the presence of non-human conceptual structures at the core of physical processes (Aerts *et al.*, 2020).

[18] It is important to note that the mere existence of a program capable of producing texts that comply with Bose-Einstein statistics is not, in itself, evidence of semantic understanding, since even a program that simply produces collages from pre-existing texts would be capable of doing so. The result becomes significant, however, if the program is also able to adapt dynamically to different semantic contexts. In other words, the claim defended here concerns the emergence of deeper organizational structures, such as contextuality, non-classical composition, and sensitivity to latent semantic relations, that cannot be reduced to

In this light, describing current AI models as mere networks is overly reductive. The characterization is understandable from a historical standpoint but is no longer sufficient. The semantic intelligence of these systems resides in the architecture of their *vector semantic spaces*, which arise from – but are not reducible to – the underlying neural substrate. Consequently, a naïve mechanistic view is inadequate to capture either the functioning of LLMs or their striking affinity with human thought processes. A similar reduction would occur if one were to describe a book solely in terms of the paper and ink from which it is made (the level of the symbols). Such a description, while not false, would entirely miss what makes the object a book in the first place: the vast domain of meaning encoded in the structure and organization of the printed symbols, or more precisely, the traces left by meaning in their relational arrangement, that a cognitive entity is then able to interpret. Meaning is not located in the physical substrate as such, but in the structured patterns that the substrate supports and that enable meaning-driven interactions with other meaning-bearing entities. Likewise, while neural networks provide the physical basis of LLMs, their semantic intelligence resides in the structured relations instantiated within their representational spaces.

If the conceptuality interpretation of quantum mechanics is correct, then the role of the quantum formalism in the present discussion goes beyond that of a particularly effective modeling tool. On this view, the quantum formalism does not merely describe semantic structures by analogy or approximation but captures the very essence of what a semantic structure is. Within this interpretative framework, quantum structures would then constitute the formal signature of the objective presence of meaning in reality.

Importantly, quantum structures should not be understood here as referring to microscopic physical systems as such, but to a general mode of organization and behavior of systems in relation to the contexts with which they interact, independently of whether these systems are micro or macro, abstract or concrete. From this perspective, the appearance of quantum structures in domains as diverse as physics, human cognition, language, and AI would not be accidental, but would instead reflect a shared underlying organization characteristic of meaning itself.

To conclude this section, it should be emphasized that the fact that meaning is structural and can be instantiated independently of consciousness does not depend on the ultimate validity of quantum cognition as research program, or of the more speculative conceptuality interpretation. Even if these frameworks were to be revised or rejected, the argument for non-conscious semantic understanding would remain intact. Nevertheless, the convergent success of quantum-based approaches significantly strengthens the thesis, providing independent evidence not only that meaning is structural, but that its structure is fundamentally quantum.

## 6. Conclusion

In this article we have sought to emphasize the importance of drawing a clear demarcation between *structural meaning* and *phenomenal access to meaning and information*. Consciousness does not ground meaning; it modulates how meaning-bearing information is accessed, integrated, and lived.

The term "structural" should not be understood here in the sense of a spatial structure, that is, as the mere collection of constituent elements of a construction, but rather in the sense of a mathematical structure describing how change occurs, what the domain of potentiality is, and how different potential properties are available for actualization in different contexts. Such structures include, for example, the lattice of a system's properties, the state space, the observables and the probabilistic models. These are therefore genuine structures, but non-spatiotemporal ones: fundamentally abstract frameworks devised to

frequency-based regularities alone. It is the presence of the latter in combination with these deeper organizational structures that confers upon LLMs a distinct cognitive status.

describe change and to capture the relation between the potential and the actual, that is, between the unmanifest and the manifest. So, instead of *structural meaning* we could also have used terms as *operational meaning*, *abstract meaning*, etc.

It is always possible, of course, to stipulate that the term "understanding" be reserved for activities that, by definition, require consciousness, and to introduce a different term to denote the form of sensitivity to meaning and semantic competence described in this article. Such a terminological choice would not, in itself, be illegitimate. However, we suggest that revising the concept of understanding along the lines proposed here is philosophically preferable, and arguably necessary. The reason is that what is phenomenal in cognition is not meaning as such, but *access to meaning*. Meaning itself is structural and relational, whereas phenomenal consciousness concerns the manner in which meaning-bearing information is accessed, further integrated, and lived.

Access to information can occur consciously or non-consciously, just as information itself can be processed with or without awareness. Insisting that understanding must be conscious therefore conflates a property of *access* with a property of *content*, obscuring rather than clarifying the nature of semantic competence. By contrast, treating understanding as a graded capacity grounded in sensitivity to meaning allows us to account coherently for both conscious and non-conscious forms of cognition, without multiplying concepts or excluding empirically robust forms of semantic intelligence from the outset.

One of the consequences of failing to distinguish between *structural meaning* and *phenomenal meaning* understood as *phenomenal access to meaning*, concerns certain quantum-inspired forms of panpsychism, where the success of quantum formalisms in modeling cognition and meaning is taken to suggest that consciousness itself must be an intrinsic feature of quantum states or quantum fields. On this view, quantum theory is interpreted not merely as a framework for describing abstract structures of meaning or information, but as directly characterizing *inner experiential states*.

For example, in their *quantum information-based panpsychism*, D'Ariano and Faggin (2022) propose a principle stating that "the information theory of consciousness is quantum theory," which gives quantum systems in so-called pure states an inner reality. However, this principle is not supported by quantum cognition results. As we mentioned in the previous section, these studies demonstrate that quantum states effectively and often remarkably accurately describe conceptual dynamics, contextuality, potentiality, and non-classical correlations in cognitive processes. They do not imply, however, that such processes are accompanied by phenomenal experience. Inference from quantum structure to consciousness becomes compelling only if one assumes that meaning and understanding are intrinsically phenomenal.

From the perspective defended in this article, this assumption is precisely what must be questioned. Once meaning is understood as entirely structural, the inference from quantum structure to phenomenal consciousness loses its primary motivation, i.e., quantum states can be understood as modeling abstract, non-spatiotemporal structures of meaning and cognition without committing to any claim about consciousness. Conscious experience may indeed instantiate or inhabit such structures in human subjects, but it does not exhaust them. Consequently, interpreting quantum states as intrinsically experiential reflects a conflation of structural and phenomenal (access to) meaning rather than a necessary implication of quantum-theoretic modeling itself.

To conclude, it is worth briefly considering the *illusion problem* about consciousness (Frankish, 2016) and the related notion of *meta-hard problem* (Chalmers, 2018). According to illusionist accounts, consciousness – understood as phenomenal experience or qualia – is itself an illusion: there is, strictly speaking, nothing over and above certain cognitive, functional, or representational processes. On this view, the genuine explanatory task is not to account for phenomenal consciousness as such, but to explain why we are strongly disposed to believe that we possess it, and why consciousness appears to us as something irreducible and problematic. In other words, the real challenge is not the hard problem of consciousness, but the meta-hard problem: why there seems to be a hard problem at all.

If illusionism were correct, the distinction we have drawn throughout this article between semantic (or structural) understanding and phenomenal understanding would ultimately collapse. Meaning would then be entirely structural, functional, and relational, with no genuinely phenomenal component to distinguish it. Consciousness would not add an ontologically distinct layer to meaning or understanding, but would itself be a cognitive construct emerging from underlying structural processes. While our argument does not presuppose the truth of illusionism, it is noteworthy that the position defended here is compatible with it: if meaning and understanding can exist independently of consciousness, then the illusionist thesis would merely strengthen this conclusion by denying that phenomenal understanding ever existed as a separate category. Whether consciousness is real or illusory, semantic intelligence would remain a robust and autonomous domain, deserving of investigation in its own right.

In this respect, if LLMs possess genuine forms of non-conscious understanding, then "understanding" itself must be reconceived as a graded and pluralistic notion rather than a binary predicate. The central philosophical challenge may therefore not be whether machines understand, but whether our inherited concept of understanding is sufficiently refined to accommodate forms of cognition that are semantic yet non-subjective, coherent yet non-lived.

Contemporary linguistic models appear to possess neither self-awareness nor phenomenal experience. Nevertheless, they operate at the level of structural meaning, understood as the internal coherence and contextual organization of abstract representations. In this sense, the very existence of such highly effective models provides a natural test bed for distinguishing semantic understanding from phenomenal consciousness.

One thing is certain, it is becoming increasingly difficult today to deny that LLMs can generate, manipulate, and recombine meanings without the need to experience them (certainly not in the way humans sometimes do). LLMs are neither mere parrots nor human-like intelligences. Are they merely powerful generalizers over language with strong social and dialogical competence (Sejnowski, 2023), or instead an emergent phenomenon that we have not yet fully understood, perhaps even an initial step toward more advanced forms of semantic intelligence than our own, especially as AI systems become increasingly and directly coupled to the physical and social world?

Certainly, they are forcing us to revise our notions of meaning, understanding, and the very criteria by which intelligence itself is recognized. This is exactly what we tried to do with this article, emphasizing a crucial point for the future of cognitive science and AI: consciousness cannot be reduced to information processing, yet meaning cannot be confined to consciousness alone. Further developments of LLMs, and their integration into artificial bodies, will not automatically give rise to consciousness, but they will continue to expand the range of conceptual structures and forms of non-conscious intelligence available for exploration. This suggests that future research must clarify not only what consciousness is, but also what kinds of intelligence can exist independently of it.

## References

Aerts, D. and Beltran, L. (2020). Quantum structure in cognition: Human language as a Boson gas of entangled words, *Foundations of Science*, **25**, pp. 755-802.

Aerts, D., Sassoli de Bianchi, M., Sozzo, S. and Veloz, T. (2020). On the Conceptuality interpretation of Quantum and Relativity Theories. *Foundations of Science*, **25**, pp. 5-54.

Aerts, D., Aerts Arguëlles, J., Beltran, L., Sassoli de Bianchi, M., and Sozzo, S. (2025a). The Origin of Quantum Mechanical Statistics: Some Insights from the Research on Human Language. To be published in: *Philosophical Transactions of the Royal Society A*. arXiv:2407.14924 [q-bio.NC].

Aerts, D. and Sassoli de Bianchi, M. (2015). The unreasonable success of quantum probability I: Quantum measurements as uniform fluctuations, *Journal Mathematical Psychology*, **67**, pp. 51-75.

Aerts, D., Sassoli de Bianchi, M. and Sozzo, S. (2025b). From Quantum Cognition to Conceptuality Interpretation II: Unraveling the Quantum Mysteries, *Philosophical Transactions of the Royal Society A*, **383**: 20240381. Doi: 10.1098/rsta.2024.0381.

Aerts, D., Sassoli de Bianchi, M. and Sozzo, S. (2025c). From Quantum Cognition to Conceptuality Interpretation I: Tracing the Brussels Group's Intellectual Journey, *Philosophical Transactions of the Royal Society A*, **383**: 20240382. Doi: 10.1098/rsta.2024.0382.

Aerts, D., Aerts Arguëlles, J., Beltran, L., Geriente, S., Sassoli de Bianchi, M., Leporini, L. and Sozzo, S. (2025d). Identifying Quantum Structure in AI Language: Evidence for Evolutionary Convergence of Human and Artificial Cognition, *arXiv: 2511.21731*.

Abdou, M., Kulmizev, A., Hershcovich, D., Frank, S., Pavlick, E., Søgaard, A. (2021). Can Language Models Encode Perceptual Structure Without Grounding? A Case Study in Color. In: *Proceedings of the 25th Conference on Computational Natural Language Learning*, ed. Bisazza, A. & Abend, O., pp. 109-132. Association for Computational Linguistics.

Xu, Q., Peng, Y., Nastase, S. A., Chodorow, M., Wu, M. and Li, P. (2025). Large language models without grounding recover non-sensorimotor but not sensorimotor features of human concepts. *Nat. Hum. Behav.*, **9** (9), pp. 1871-1886.

Agüera y Arcas, B. (2022). Do Large Language Models Understand Us? *Daedalus*, **151** (2), pp. 183-197.

Bender, E. M., Gebru, T, McMillan-Major, A. and Shmitchell, S. (2021). On the Dangers of Stochastic Parrots: Can Language Models Be Too Big? In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency (FAccT'21)* (pp. 610-623). Association for Computing Machinery.

Berlin, B. & Kay, P (1969). Basic Color Terms: Their Universality and Evolution. Berkeley: University of California Press. Broad, C.D. (1923). Scientific Thought. London: Routledge & Kegan Paul, p. 242.

Busemeyer, J. R. and Bruza, P. D. (2025). *Quantum models of cognition and decision*, 2$^{nd}$ ed, Cambridge: Cambridge University Press.

Busemeyer, J. and Lu, M. (2025). Quantum consciousness, Brains, and Cognition, *Journal of Consciousness Studies*, **32** (9-10), pp. 156-182.

Chalmers, D. J. (2018). The Meta-Problem of Consciousness, *Journal of Consciousness Studies*, **25**, (9–10), pp. 6-61.

Chalmers, D. J. (2023). Could a Large Language Model be Conscious? *Boston Review*, August 9. This is an edited version of a talk given at the conference on Neural Information Processing Systems (NeurIPS) on November 28, 2022.

D'Ariano, G. M., Faggin, F. (2022). Hard Problem and Free Will: An Information-Theoretical Approach. In: Scardigli, F. (eds) *Artificial Intelligence Versus Natural Intelligence*. Springer, Cham, pp. 145-191.

Einstein, A., Podolsky, B., and Rosen, N. (1935). Can quantum-mechanical description of physical reality be considered complete? *Physical Review*, **47**, pp. 777-780.

Falk, G., Herrmann, F. and Schmid, G.B. (1983). Energy forms or energy carriers? *Am. J. Phys.*, **51**, pp. 1074-1077.

Frankish, K. (2016) Illusionism as a theory of consciousness,*Journal of Consciousness Studies*, **23** (11–12), pp. 11-39. Reprinted in Frankish, K. (ed.) (2017) *Illusionism as a Theory of Consciousness*, Exeter: Imprint Academic.

Georgiev, B., Gómez-Serrano, J., Tao, T. and Wagner, A. Z. (2025). Mathematical exploration and discovery at scale. *arXiv:2511.02864 [cs.NE].*

Goldstone, R. L. & Hendrickson, A. T. (2010). Categorical perception. *Wiley Interdiscip. Rev. Cogn. Sci*., **1**, pp. 69-78.

Goff, P. (2017). *Consciousness and Fundamental Reality*, New York: Oxford University Press.

Jackson, F. (1982). Epiphenomenal Qualia, *Philosophical Quarterly*, **32** (127), pp. 127-136.

Harnad, S. (1990). The Symbol Grounding Problem. *Physica D*, **42**, pp. 335-346.

Hill, B. M. (1974). The Rank-Frequency Form of Zipf's Law, *Journal of the American Statistical Association*, **69** (348), pp. 1017-1026.

Manning, C. D. (2022). Human Language Understanding & Reasoning, *Daedalus*, **151** (2), pp. 127-138.

Maslov, V. P. (2005). On a General Theorem of Set Theory Leading to the Gibbs, Bose-Einstein, and Pareto Distributions, *Mathematical Notes*, **6**, pp. 807-813.

Mitchell, M. and Krakauer, D. C. (2023). The debate over understanding in AI's large language models, *Proc. Natl. Acad. Sci. U.S.A.*, **120** (13) e2215907120.

Newell, A. and Simon, H. A. (1963). GPS, a program that simulates human thought. In: *Computers and thought*, ed. A. Feigenbaum & V. Feldman, pp. 279-93. New York: McGraw Hill.

Penrose, R. (1989). *The Emperor's New Mind: Concerning Computers, Minds and The Laws of Physics*. Oxford University Press.

Penrose, R., Severino, E. (2022). Dialogue on Artificial Intelligence Versus Natural Intelligence. In: Scardigli, F. (eds) *Artificial Intelligence Versus Natural Intelligence*. Springer, Cham, pp. 27-70.

Rees, T. (2022). Non-Human Words: On GPT-3 as a Philosophical Laboratory. *Daedalus*, **151** (2): pp. 168-182.

Sassoli de Bianchi, M. (2011). Ephemeral properties and the illusion of microscopic particles, *Foundations of Science*, **16**, pp. 393-409.

Schank, R. C. and Abelson, R. P. (1977). *Scripts, plans, goals, and understanding*, Hillsdale, N.J.: Lawrence Erlbaum Press.

Schmitt, J. (2025). Extremal descendant integrals on moduli spaces of curves: An inequality discovered and proved in collaboration with AI. *arXiv:2512.14575 [math.AG].*

Searle, John (1980), Minds, Brains and Programs, *Behavioral and Brain Sciences*, **3** (3), pp. 417-457.

Sejnowski, T. J. (2023) Large Language Models and the Reverse Turing Test, *Neural Computation*, **35**, pp. 309-342.

Strawson G. (2006). Realistic monism. Why physicalism entails panpsychism, *Journal of Consciousness Studies*, **13** (10-11), pp. 3-31.

Varela, F. J., Thompson, E., and Rosch, E. (1991). *The Embodied Mind*. MIT Press.