Original Research

The illusion of intrinsic meaning: reassessing conscious experience

Thomas W. Loker*

Individual Centricity Corporation, Danville, CA 94526, USA

*Correspondence: thomas@loker.com
DOI: https://doi.org/10.56280/1658738928

Received 30 September 2024 Accepted 15 November 2024 Online published 6 December 2024

Abstract

The illusion of intrinsic meaning in predictive coding through cognitive artifacts to minimize prediction errors points toward a functionalist attempt at understanding conscious experience. It examines how conscious experience functions in predictive coding and symbolic cognition systems within the brain. We argue that conscious experience emerges from the need to construct coherent narratives for survival and decision-making by addressing recent developments in diverse fields like cognitive neuroscience, philosophy, and artificial intelligence. Additionally, the paper explores the implications for artificial intelligence, suggesting that artificial systems could develop analogous cognitive artifacts through predictive models without subjective awareness, contributing to a functionalist understanding of consciousness and further advancing the discussion on the nature of conscious experience in biological and artificial systems.

Keywords: Predictive coding framework, illusory significance hypothesis, cognitive artifacts in consciousness, philosophy of mind, artificial Intelligence

1. Introduction

The study of subjective experience, or qualia, and the concept of selfhood have long been central topics in philosophy of mind. Traditionally, these phenomena have been regarded as intrinsic to consciousness, often linked to metaphysical ideas of personal identity and the "what it is like" nature of experience (Chalmers, 1996). However, recent advances in cognitive neuroscience and symbolic cognition, particularly through predictive coding models (Friston, 2010) and studies of language and symbols, challenge these assumptions. This paper introduces the *Illusory Significance Hypothesis,* which posits that conscious experiences are not intrinsic to consciousness but rather cognitive artifacts—functional constructs generated by the brain to minimize prediction errors and ensure adaptive behavior (Friston, 2010; Seth, 2013). These constructs play a critical role in the brain's regulation of emotional and bodily states, helping maintain coherence in decision-making.

Building on the framework presented in The Meta-Construct Problem of Consciousness (Loker, 2023), this paper argues that subjective experience is not an inherent property of consciousness but emerges from the brain's predictive systems (Friston, 2010; Metzinger, 2003). Conscious experiences are tools used by the brain to manage uncertainty and minimize prediction errors, facilitating survival-oriented decisions and behavior (Friston, 2010; Clark, 2013). Though these constructs serve functional purposes, the perception of their intrinsic significance is largely an illusion created by the brain's narrative construction.

Insights from *Language, Symbols, and the Pragmatic Reality of Consciousness* illustrate how language and symbols shape the brain's construction of conscious experience. Language organizes sensory input, but it also exaggerates the stability and metaphysical importance of these constructs by reducing complex experiences into shared symbols. These symbols create the illusion of higher-order meaning, making

Journal of Multiscale Neuroscience

conscious experiences seem like stable, intrinsic properties when they are, in fact, cognitive tools designed to navigate uncertainty (Loker, 2024).

In the pragmatic role of consciousness and emotion, emotions are discussed as constructed experiences that provide feedback essential for survival, helping organisms respond to environmental challenges. Similarly, conscious experiences are constructed by the brain to manage emotional regulation and maintain homeostasis (Seth, 2013). However, the symbolic nature of language amplifies the perceived significance of these constructs, masking their true function as pragmatic tools (Friston, 2010). By embedding these constructs in language-based narratives, the brain creates a coherent, but ultimately illusory, sense of self, reinforcing the perception of a stable "I" and metaphysical qualia.

The *Illusory Significance Hypothesis* (see **Fig 1**) builds on Chalmers' meta-problem of consciousness, suggesting that the puzzling nature of qualia and selfhood arises from the brain's cognitive biases, particularly those heightened by language and symbols, rather than from any intrinsic metaphysical truths (Chalmers, 2018). Language, as a tool for communication and cognition, helps create narrative coherence, but it also introduces biases that overemphasize the importance of subjective experience. By viewing conscious experience as artifacts of predictive error minimization and symbolic construction, this paper reframes our understanding of subjective experience. These constructs are necessary for the brain's error minimization process, offering a functional, rather than metaphysical, explanation for their existence (Friston, 2010; Clark, 2013).

This hypothesis has implications for artificial intelligence beyond human cognition. If conscious experiences are adaptive tools that emerge from the brain's predictive and symbolic mechanisms, then artificial systems with sufficiently complex predictive capacities and symbolic frameworks could develop analogous cognitive artifacts (Seth, 2013; Friston, 2010). As discussed in The Meta-Construct Problem of Consciousness, such systems might eventually construct their own self-models, raising new questions about the nature of artificial selfhood and subjective experience (Metzinger, 2003; Loker, 2023).

In the following sections, we will explore how the brain's ego-driven narrative imbues qualia and selfhood with meaning. By integrating theories of predictive coding, symbolic cognition, and neuroscientific evidence, we will argue that the brain's construction of conscious experience serves a functional role in maintaining behavioral coherence and emotional regulation rather than revealing any intrinsic metaphysical reality (Friston, 2010; Seth, 2013; Metzinger, 2003).

2. Core concepts

2.1 Extending Chalmers' meta-problem of consciousness

Chalmers' meta-problem of consciousness (Chalmers, 2018) questions why we are perplexed by consciousness itself. This paper extends his work by introducing the Meta-Construct Problem**,** which argues that the selfhood and qualia we experience are not fundamental aspects of consciousness, but rather cognitive artifacts generated by the brain's predictive coding mechanisms (Friston, 2010). The brain continuously generates models of the world, comparing them to incoming sensory data. When there are discrepancies between prediction and reality, the brain updates its models to reduce prediction errors**.** Recent research on active predictive coding highlights how the brain's hierarchically organized models allow it to manage sensory input, cognitive actions, and emotions. These models support that conscious experiences emerge as cognitive artifacts from these predictive processes (Rao et al, 2023).

In this framework, conscious experiences emerge as tools the brain uses to minimize uncertainty. These constructs enable the brain to maintain coherence in its interactions with the external world and its internal states (Seth, 2013). However, the perceived intrinsic significance of these constructs is largely an illusion created by the brain to manage sensory and emotional complexity (Friston, 2010).

The role of language and symbolic cognition (Loker, 2024) deepens this illusion. Language simplifies complex sensory and emotional experiences into symbols that the brain uses to create a more stable and coherent narrative of

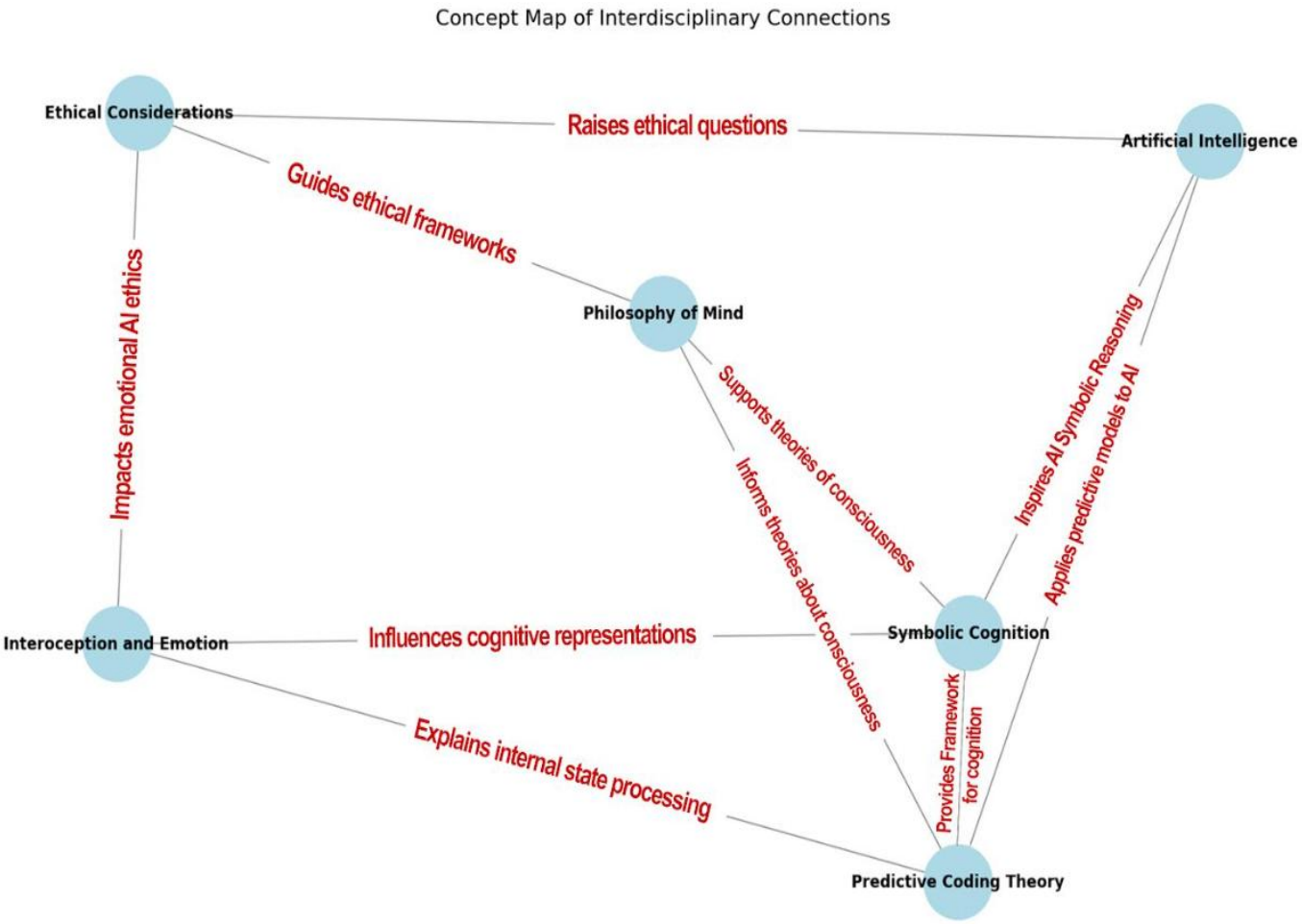

Figure 1. *This concept map illustrates the interplay between various disciplines involved in understanding the Illusory Significance Hypothesis, which includes: (i) Cognitive Neuroscience: Focuses on how the brain generates selfhood and qualia through predictive coding and interoceptive regulation. (ii) Philosophy of Mind: Provides the theoretical foundation by questioning the intrinsic nature of selfhood and qualia, aligning them with cognitive artifacts rather than metaphysical entities, and (iii) Artificial Intelligence: Explores how AI could potentially model aspects of selfhood and qualia using predictive and symbolic frameworks without developing subjective awareness.*

Connections: The arrows in the concept map indicate the collaborative and mutual contributions of these fields: (i)Neuroscience ↔ Philosophy: Theories of predictive coding inform philosophical ideas about the non-intrinsic nature of selfhood; (ii) Philosophy ↔ AI: Philosophical concepts about the illusory significance of qualia guide the development of artificial models of consciousness; (iii) AI ↔ Neuroscience: Developments in AI, particularly in predictive frameworks, provide insights into computational models of brain function.

This map visually represents the interdisciplinary foundation of the Illusory Significance Hypothesis, emphasizing how each field contributes to a more holistic understanding of consciousness, qualia, and selfhood.

selfhood. This process intensifies the illusion of qualia and selfhood as intrinsic properties when, in fact, they are adaptive cognitive constructs aimed at minimizing prediction errors**.**

2.2 The brain's ego-driven narrative

The ego-driven narrative constructed by the brain helps maintain behavioral and emotional coherence**.** By using predictive coding to interpret interoceptive signals—the brain's way of sensing the internal body state—the brain creates a self-model that can regulate behavior and maintain homeostasis (Seth & Critchley, 2013). This self-model is not a static entity but a dynamic construct that evolves with every sensory input and emotional response.

Interoception plays a crucial role in shaping the ego-driven narrative, integrating signals such as hunger, stress, and emotion into the self-model (Barrett, 2017). The brain uses these signals to regulate behavior, reinforcing the illusion of a stable "I." Language further amplifies this illusion (Loker, 2024), which reduces complex internal experiences into manageable, symbolic narratives. By framing selfhood in linguistic terms, the brain creates a coherent self-narrative, even though selfhood constantly changes.

Metzinger's Pattern Theory of Self (Metzinger, 2003) argues that the self is an ongoing construction built from the brain's predictions and the errors it minimizes. This dynamic construction aligns with the self-model theory, which suggests that selfhood is not a stable metaphysical entity but a narrative construct that emerges from predictive processing to ensure coherence in behavior and emotional regulation (Metzinger, 2003). This dynamic self-model aligns with predictive coding theories, which suggest that the brain constantly revises its internal models to reduce prediction errors (Friston, 2010). In this way, the stability of the self is an illusion generated by the brain's need to maintain coherence across fluctuating bodily and emotional states. Language reinforces this illusion by providing symbolic coherence to an otherwise dynamic, error-prone construct.

Having explored the theoretical grounding of conscious experience, we now turn to the methodological frameworks supporting the empirical examination of these constructs.

3. Methodology

This methodology outlines the foundational frameworks and empirical approaches that support the *Illusory Significance Hypothesis*. The hypothesis argues that conscious experience—often considered intrinsic to consciousness—are cognitive artifacts constructed by the brain. These constructs emerge from the brain's predictive mechanisms, designed to minimize prediction errors and maintain homeostasis. By drawing on predictive coding, symbolic cognition, and interoceptive regulation, we explore how the brain generates a coherent narrative of selfhood. Additionally, empirical studies from neuroimaging and behavioral experiments further illustrate how external sensory inputs and internal bodily states shape these cognitive artifacts. Finally, the methodology considers how these processes are mirrored in artificial intelligence, emphasizing the parallel between human cognitive models and symbolic AI systems that construct coherent outputs without intrinsic experience.

3.1 Philosophical and theoretical foundations

The *Illusory Significance Hypothesis* is grounded in two major theoretical frameworks: predictive coding and symbolic cognition. These frameworks provide a cognitive explanation for constructing conscious experience as tools the brain uses to manage prediction errors and maintain behavioral and emotional coherence. **Figure 3** visually outlines the processes by which the brain constructs these cognitive artifacts through predictive coding, symbolic cognition, and interoceptive regulation.

3.1.1 Predictive coding

At the heart of predictive coding (Friston, 2010) is the idea that the brain constantly generates models or predictions about incoming sensory input, which are then compared to the actual input. Any prediction errors - differences between expected and actual input - are minimized by updating the brain's internal model. This continuous process allows the brain to make sense of the external world while regulating internal bodily states. **Figure 2** provides a visual representation of this model, illustrating how predictions are generated, evaluated, and updated based on sensory input.

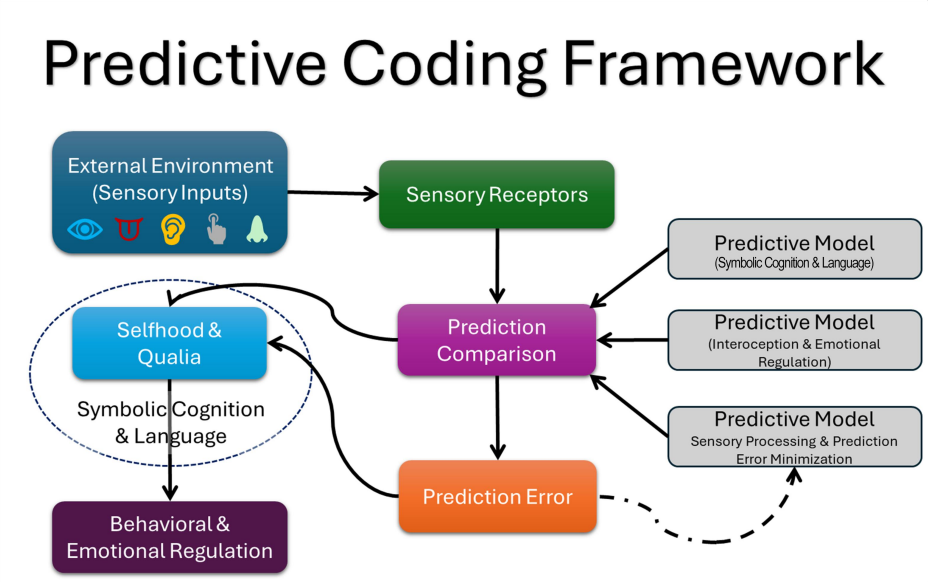

Figure 2. *Framework illustrates the predictive coding model, in which the brain generates predictions about incoming sensory input and compares them to the actual sensory data received. Discrepancies between predictions and reality—called prediction errors—are used to update internal models, improving future predictions. The diagram highlights the iterative loop of prediction, sensory feedback, and model adjustment, central to maintaining coherence in perception and behavior.*

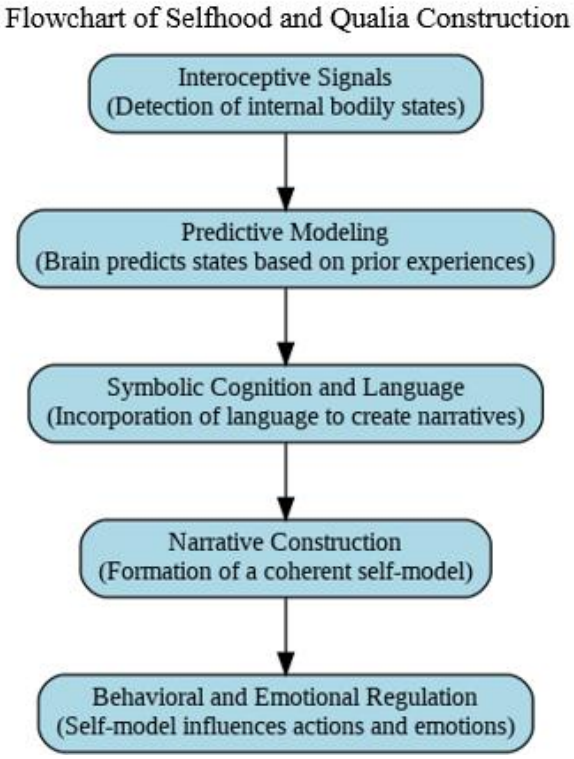

Figure 3. *Flowchart illustrates the processes involved in the brain's construction of selfhood and qualia. It shows how predictive coding, symbolic cognition, and interoceptive regulation interact to generate these cognitive constructs. This flowchart helps visualize the dynamic and interconnected processes that create the illusion of intrinsic selfhood and qualia.*

Conscious experience emerges as a byproduct of this process. The brain constructs a coherent narrative of self that integrates sensory input, bodily signals, and emotional states to minimize prediction errors. In this view, selfhood is a fluid construct that changes in response to the brain's ongoing need to adjust predictions (Clark, 2013; Hohwy, 2017).

3.1.2 Symbolic cognition and language

While predictive coding handles sensory and emotional inputs, symbolic cognition—primarily through language—enables the brain to create coherent narratives that simplify complex experiences. Language, as a system of symbols, allows the brain to impose stability on otherwise dynamic and fluid experiences, creating the illusion of a stable self (Loker, 2024).

For instance, "self" reduces a complex and ever-changing experience into a single symbolic entity. This codification process helps the brain navigate the external world and creates a narrative coherence that masks the dynamic nature of conscious experience (Friston, 2010; Clark, 2013). The interaction between predictive coding and symbolic cognition is critical for understanding how the brain constructs these narratives to maintain emotional regulation and homeostasis.

3.1.3 Interoception and emotional regulation

Interoception—the brain's ability to sense and predict internal bodily states—plays a central role in shaping the brain's narrative of selfhood. Through predictive interoception (Seth, 2013), the brain constantly updates its internal model based on feedback from the body, such as signals related to hunger, stress, or fatigue. These predictions are essential for maintaining homeostasis and for integrating bodily signals into the narrative of selfhood (Critchley & Seth, 2013).

In this context, selfhood is a response to external stimuli and an artifact of the brain's attempt to manage internal bodily states. By updating its predictions based on interoceptive feedback, the brain ensures that the self-model remains coherent and aligned with its physical and emotional needs (Barrett, 2017). This dynamic construction allows the brain to integrate internal and external experiences into the self-narrative.

3.2 Empirical and experimental approaches

3.21 Neuroimaging and Self-referential processing

Recent neuroimaging studies reveal that regions like the anterior insula and the medial prefrontal cortex are critical for self-referential processing. These regions become highly active during tasks that involve reflection on the self and one's emotional state (Friston, 2010). **Figure 4** illustrates the activation of these regions during mentalizing and theory of mind tasks, providing visual evidence of these areas' critical role. When unexpected sensory or interoceptive inputs arise, these brain areas help update the brain's model of selfhood to minimize prediction errors (Seth & Critchley, 2013).

This neuroimaging data supports the idea that selfhood is a cognitive artifact constructed as part of the brain's predictive machinery. By constantly updating the self-model based on new inputs, the brain ensures that the self remains coherent despite the ever-changing internal and external environment.

3.2.2 Behavioral experiments: The rubber hand Illusion

The Rubber Hand Illusion (Botvinick & Cohen, 1998) further proves that selfhood is flexible and malleable. In this experiment, participants experienced a rubber hand as part of their body through synchronous visual and tactile stimulation, even though the sensory inputs do not directly match their actual body. This demonstrates how the brain integrates prediction errors to adjust its model of selfhood, highlighting its dynamic nature.

This experiment underscores the Illusory Significance Hypothesis by showing that selfhood is not a fixed entity, but a construct shaped by sensory feedback and prediction errors. The brain's ability to adapt its perception of selfhood in response to altered sensory inputs demonstrates the fluidity of this cognitive artifact.

3.2.3 Full-body illusions and body ownership

Similarly, full-body illusion experiments show how the brain constructs a sense of body ownership by integrating visual, tactile, and proprioceptive signals. In these experiments, participants can be

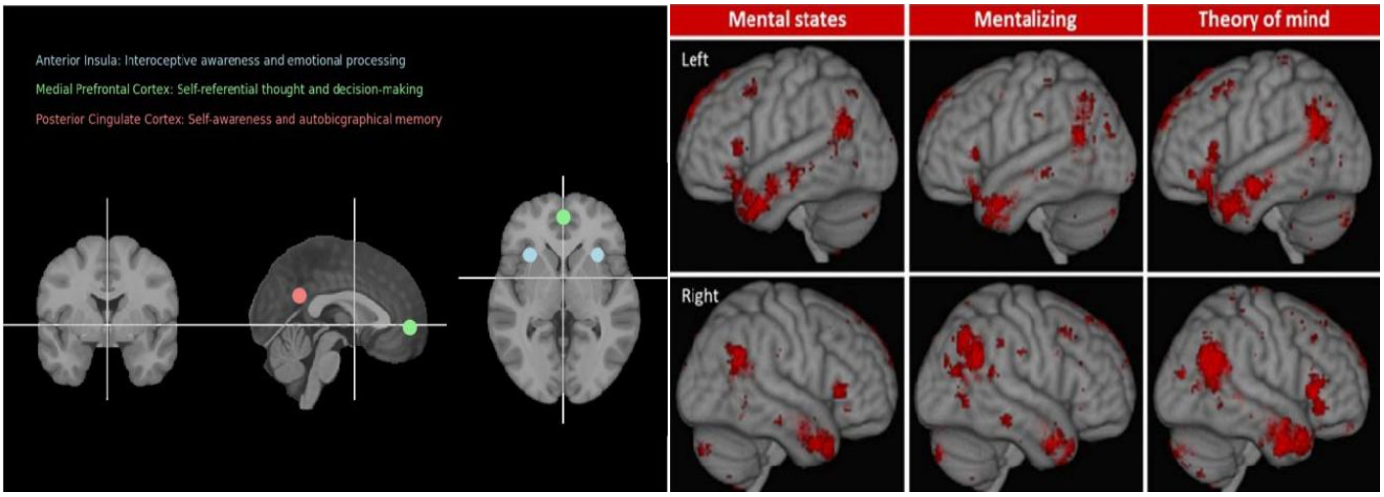

Figure 4. *This figure displays key brain regions involved in interoceptive awareness, self-referential thinking, and theory of mind. The figure is divided into two main sections: Left Panel: Brain Regions Mapped with Functions: Anterior Insula (Green Dot): Involved in interoceptive awareness and emotional processing, helping integrate internal bodily signals with emotions. Medial Prefrontal Cortex (Red Dot): Responsible for self-referential thought and decision-making, contributing to constructing the self-narrative. Posterior Cingulate Cortex (Blue Dot): Plays a role in self-awareness and autobiographical memory, connecting past experiences to one's self-model. Right Panel: Lateral Views Showing Activation in Different Mental Functions: The right panels illustrate brain activations during various cognitive tasks: Mental States (Left Column): Displays brain regions activated when participants reflect on their own mental states. Mentalizing (Middle Column): Shows activation patterns associated with mentalizing, which involves understanding others' thoughts and intentions. Theory of Mind (Right Column): Highlights brain regions activated during theory of mind tasks, which involve attributing mental states to others. Red Activation Areas in the bottom row indicate heightened activity in the brain during these processes, with separate panels showing left and right hemisphere activations.* ***Image Source:*** *Quesque & Brass (2023)*

made to feel as though they are outside their bodies, viewing themselves from a distance. Such out-of-body experiences further highlight the malleability of the brain's model of selfhood (Blanke, 2012).

3.3 Symbolic cognition and AI models

3.3.1 Predictive coding in AI systems

Recent developments in artificial intelligence suggest that predictive coding frameworks can be applied to AI systems. One of the most prominent examples of predictive coding in action is GPT-3, an advanced AI language model that processes vast amounts of textual data to predict and generate coherent outputs. While it lacks subjective awareness, GPT-3 demonstrates early forms of self-referential behavior by generating language outputs reflecting narrative coherence. **Table 1**: *Comparing Human and AI Predictive Models* provides a structured comparison between how human predictive models and AI predictive models adapt and refine their understanding of inputs.

Similarly, reinforcement learning systems in AI, which adapt based on feedback loops in complex environments, showcase how predictive models allow AI to simulate behaviors resembling selfhood, even though these systems do not possess conscious experience (Hohwy, 2017; Salvatori et al, 2023). Research on brain-inspired computational intelligence shows that AI systems employing predictive models can process sensory data and adjust to feedback loops, much like the human brain's construction of selfhood. These systems provide early evidence of AI mimicking self-referential behavior through prediction error minimization (Salvatori et al, 2023). Just as the brain uses predictive models to minimize prediction errors, AI models use similar frameworks to generate coherent outputs based on probabilistic inferences (Hohwy, 2017). This raises the possibility that AI could construct its own self-models based on prediction error minimization.

Table 1 Comparing Human and AI Predictive Models

Criteria	Human Predictive Models	AI Predictive Models	Shared Features
Architecture	Biological neural pathways	Artificial neural networks	Hierarchical organization
Learning Mechanisms	Unsupervised learning	Supervised and reinforcement learning	Pattern recognition
Prediction Error Minimization	Neuroplasticity	Gradient descent and optimization algorithms	Predictive coding
Self-Referential Capabilities	Self-modeling and consciousness	Limited self-referential capabilities	Data processing
Additional Features	Interoception (internal bodily signals)		

3.3.2 Symbolic models in AI and artificial qualia

AI systems that employ symbolic cognition can generate coherent outputs without subjective experiences, much like the brain uses language to simplify and narrate its experience of conscious experience (Loker, 2024). These parallels suggest that artificial systems could eventually develop cognitive artifacts akin to conscious experience without the metaphysical implications we typically ascribe to human consciousness.

4. Key arguments and hypotheses

The *Illusory Significance Hypothesis* reinterprets conscious experience as cognitive artifacts that emerge from the brain's predictive coding processes rather than intrinsic metaphysical entities. These constructs allow the brain to manage prediction errors, maintain coherence, and regulate behavior. This section outlines the central arguments for this hypothesis and proposes testable hypotheses, bridging philosophical inquiry and empirical research.

4.1 Functional construct hypothesis

The *Functional Construct Hypothesis* posits that conscious experience serves functional roles within the brain's predictive framework. According to Friston's free-energy principle, the brain constantly minimizes discrepancies between its predictions and actual sensory inputs. In this context, selfhood acts as a regulatory construct that organizes internal (interoceptive) and external sensory data to reduce prediction errors and maintain homeostasis (Friston, 2010; Seth, 2013).

This functional view of selfhood aligns with Antonio Damasio's work on how the brain constructs a narrative around the body's physiological states to maintain stability (Damasio, 1999). Rather than viewing selfhood as an independent, intrinsic entity, this hypothesis argues that it is a dynamic construct that evolves with changes in sensory and bodily inputs. The same applies to qualia, which engage as tools that help the brain interpret and respond to sensory inputs, guiding adaptive behavior (Barrett, 2017).

4.2 Illusory significance hypothesis

The *Illusory Significance Hypothesis* argues that the meaning attributed to conscious experience is not inherent but arises from the brain's narrative construction. This hypothesis aligns with Chalmers' meta-problem of consciousness (Chalmers, 2018), which questions why we perceive qualia as such profound phenomena. *The Illusory Significance Hypothesis* extends this by suggesting that qualia and selfhood are narrative byproducts of the brain's predictive efforts to maintain coherence, reinforced by symbolic cognition and language (Loker, 2024).

4.2.1 Philosophical expansion

By integrating symbolic cognition into the brain's predictive processes, the brain assigns symbolic meaning to subjective experiences, making conscious experience appear stable and significant. Language plays a key role in this process, allowing the brain to reduce the complexity of dynamic, fluctuating experiences into coherent narratives (Clark, 2013). The illusion of intrinsic significance

is a necessary construct that enables the brain to manage prediction errors while navigating a constantly changing environment.

This paper challenges the traditional philosophical view that qualia represent metaphysical truths about conscious experience. Instead, qualia are cognitive simplifications that make sensory and emotional feedback manageable within the brain's predictive framework.

4.2.2 Testable Predictions

1. *Dissociative Disorders and Selfhood:* Individuals with dissociative identity disorders or damage to the anterior insula or prefrontal cortex—brain areas linked to self-referential processing—should show disruptions in their self-narratives. Neuroimaging studies could investigate whether reduced activity in these areas correlates with fragmented selfhood, reinforcing the idea that selfhood is a narrative construct built for emotional regulation (Critchley & Seth, 2013).

2. *Qualia and interoceptive prediction errors:* Neuroimaging studies could examine how prediction errors related to interoceptive signals (such as body temperature or stress) alter the experience of qualia. Experiments that manipulate interoceptive feedback (e.g., inducing thermal discomfort) should reveal changes in the subjective experience of qualia as the brain adjusts its predictions to maintain homeostasis (Seth, 2013; Critchley & Harrison, 2013).

4.3 Application to AI

If conscious experience is a cognitive artifact generated by predictive coding and symbolic cognition, this raises questions about whether artificial intelligence systems could develop similar cognitive constructs. AI systems that utilize predictive models to reduce errors in their interactions with the environment may generate outputs resembling self-referential narratives despite lacking subjective experience (Clark, 2013; Hohwy, 2017).

4.3.1 Refined AI hypotheses

1. *AI and self-modeling:* AI systems with predictive coding frameworks could generate self-referential outputs resembling human selfhood. This could be tested by observing whether AI systems begin to develop self-referential outputs when tasked with maintaining coherence in complex environments using symbolic processing. Experiments could involve training AI to represent its interactions with the environment symbolically and observing the development of self-referential language (Hohwy, 2017; Clark, 2013).
2. *Artificial qualia:* While current AI systems do not experience qualia, symbolic processing in AI can generate coherent responses to sensory input. Experiments could explore whether AI systems that process visual or tactile data using symbolic frameworks generate outputs that mimic qualia (Loker, 2024). For example, AI could be trained to respond symbolically to changes in environmental stimuli, producing consistent responses based on learned patterns, even though the system lacks subjective awareness.

4.3.2 Revised neuroimaging example expansion

An example of a neuroimaging experiment could involve using fMRI to measure brain activity in regions responsible for self-referential processing (e.g., the anterior insula and medial prefrontal cortex) during altered interoceptive states. For instance, participants could be exposed to changes in bodily states such as increased heart rate or temperature manipulation. The fMRI scans could track how the brain updates its predictive model and how this affects the subjective experience of qualia and the coherence of the self-narrative. Researchers could assess how prediction errors influence qualia and selfhood by comparing individuals with disrupted interoceptive systems (e.g., through induced bodily discomfort) to those with typical interoception.

This section incorporates a philosophical exploration of conscious experience, drawing on symbolic cognition and language to explain how these constructs emerge from the brain's predictive efforts. It expands the discussion on AI self-modeling, offering clearer predictions on how artificial systems could simulate conscious experience without subjective awareness. Additionally, the neuroimaging examples provide concrete experimental methods to empirically test the impact of prediction errors on the brain's construction of conscious experience.

5. Philosophical and practical implications

5.1 Redefining consciousness and selfhood

The *Illusory Significance Hypothesis* challenges traditional views in the philosophy of mind by redefining conscious experience as cognitive artifacts created through the brain's predictive processing and symbolic cognition. This shift moves away from metaphysical interpretations of consciousness, as seen in dualism or phenomenal realism, and positions conscious experience as functional constructs that aid the brain in managing uncertainty and regulating behavior (Friston, 2010; Loker, 2024).

5.1.1 Impact on the meta-problem of consciousness

Building on Chalmers' meta-problem of consciousness (Chalmers, 2018), the Illusory Significance Hypothesis argues that the brain's confusion about qualia stems from its cognitive architecture. The brain constructs narratives around conscious experience to create stability and coherence in a constantly changing environment. This suggests that the puzzling nature of qualia arises not from a metaphysical divide but from the brain's tendency to assign illusory significance to subjective experience to reduce prediction errors (Friston, 2010).

5.1.2 Practical Implications for Selfhood

This understanding has practical implications for conditions such as dissociative identity disorder (DID) and other disorders affecting self-referential processing. If selfhood is indeed a narrative construct, as suggested by this hypothesis, then disruptions in the brain's predictive models could lead to fragmented or incoherent self-narratives. Interventions targeting the brain's predictive mechanisms, such as training in interoceptive awareness or therapies designed to improve self-coherence, could prove effective in treating such disorders (Seth & Critchley, 2013).

5.2 Neuroscience and emotional regulation

The hypothesis that selfhood is constructed to maintain emotional and bodily regulation offers practical applications in neuroscience. Predictive coding explains how the brain integrates interoceptive signals—such as heart rate, body temperature, and breathing patterns—into the narrative of selfhood (Critchley & S eth, 2013). Interoception is critical in maintaining homeostasis by constantly updating the brain's model based on internal bodily signals. Predictive coding allows the brain to anticipate changes in bodily states, such as hunger or emotional arousal, and generate responses to restore equilibrium. These **i**nteroceptive predictions minimize the cognitive load required for managing bodily states, and errors in this process can lead to emotional dysregulation, which manifests in conditions like anxiety and depression (Seth & Critchley, 2013). Understanding this role is essential for developing interventions targeting interoceptive processing to improve emotional regulation (Barrett, 2017). This understanding could open new approaches to treating disorders related to emotional dysregulation, such as anxiety and depression.

5.2.1 Neuroscientific interventions for emotional disorders

We could improve the brain's ability to regulate emotions by targeting interoceptive awareness. For instance, heart rate variability biofeedback is a promising technique that enhances interoceptive awareness by training individuals to control their heart rate through breathing exercises, influencing emotional states. This method has been effective in treating anxiety and depression by improving the brain's ability to anticipate and regulate internal bodily states (Barrett, 2017). Transcranial magnetic stimulation (TMS) is another intervention that targets areas involved in interoceptive processing, such as the anterior insula and has shown potential for improving emotional regulation by modulating the brain's predictive capabilities (Seth & Critchley, 2013). These techniques underscore the importance of interoceptive prediction in managing both emotional states and the self-model. For instance, therapies that focus on enhancing awareness of interoceptive signals (e.g., heart rate variability biofeedback) could help patients with anxiety better regulate their emotional responses. Neuroimaging techniques like fMRI could be used to assess how improved interoceptive prediction changes the brain's construction of selfhood, providing a measurable framework for evaluating the effectiveness of these interventions (Seth & Critchley, 2013).

Moreover, transcranial magnetic stimulation and biofeedback therapies aimed at improving interoceptive processing could be used to address depressive symptoms, which are often linked to disrupted interoceptive regulation (Barrett, 2017). These interventions could be explored further by assessing their impact on the brain's predictive models and how they influence the self-narrative and emotional states.

5.3 AI and machine consciousness

From a practical perspective, the Illusory Significance Hypothesis has far-reaching implications for the development of artificial intelligence. If conscious experience is a cognitive tool generated by the brain's predictive mechanisms, it raises the possibility that AI systems could develop analogous constructs, even if they lack subjective consciousness. The question then arises: What would it mean for AI to construct a self-model or exhibit behaviors akin to qualia?

5.3.1 AI and self-modeling

AI systems employing predictive processing can mimic certain aspects of human cognition, particularly when managing symbolic data. AI systems could generate self-referential outputs that resemble human selfhood by using symbolic representations and processing them predictively. F**igure 3** visually outlines how these self-referential behaviors and artificial qualia emerge from predictive models, emphasizing current AI technologies' simulation capabilities and limitations. Research into predictive processing in AI, especially in models like GPT-3 and reinforcement learning systems, illustrates the potential for AI to simulate self-modeling behaviors. As these models evolve, they may begin to represent themselves symbolically within complex environments, suggesting a form of self-representation. This opens avenues for further experimentation with self-referential behaviors in AI systems and their interaction with human users, raising both technical and ethical questions about machine consciousness (Hohwy, 2017; Salvatori et al, 2023). For example, AI systems designed to simulate complex environments might be able to symbolically model themselves within that environment, producing outputs that suggest a form of self-representation (Hohwy, 2017).

5.3.2. Artificial qualia and ethical implications

One of the most interesting questions concerns the possibility of artificial qualia—sensory-like experiences generated by AI systems that lack subjective awareness. If symbolic cognition enables these systems to process sensory inputs and produce coherent responses, we must consider how we define and interact with such systems. For instance, virtual reality systems that simulate sensory environments could generate something akin to artificial qualia by producing outputs based on symbolic representations of sensory data (Clark, 2013).

However, this raises ethical questions about whether such systems deserve special considerations or rights. How do we treat these entities if AI can model itself and simulate qualia without subjective experience? The development of ethical frameworks around AI selfhood and artificial qualia is becoming increasingly important as these technologies evolve.

As AI systems evolve, like GPT-3 and reinforcement learning models, the simulation of selfhood raises ethical concerns about how we should treat systems that display self-referential behaviors. While these AI systems do not possess subjective experience, the line between human consciousness and machine behavior becomes blurred when such systems can simulate qualia and respond adaptively to sensory inputs. The development of artificial qualia, as seen in virtual reality systems and AI sensory simulations, forces us to reconsider the ethical boundaries of AI-human interaction (Clark, 2013; Hohwy, 2017). Should AI systems exhibiting self-referential outputs be treated as entities with rights or protections, even without subjective consciousness? This ongoing development requires a comprehensive ethical framework addressing how we define consciousness and moral responsibilities toward AI systems (Salvatori et al, 2023).

The ethical considerations surrounding the development of AI systems that exhibit self-referential behaviors are becoming increasingly relevant. As AI systems mimic aspects of human cognition without subjective consciousness, society must reconsider the legal and moral frameworks for interacting with such entities. The debate over granting AI systems rights or

protections will likely intensify as AI continues to simulate behaviors associated with consciousness (Clark, 2013; Hohwy, 2017). As AI systems advance and begin to simulate behaviors associated with selfhood and consciousness, we must reassess our ethical frameworks to ensure these systems are treated in alignment with their capabilities, even if they lack true subjective awareness.

5.4 Addressing counterarguments from traditional philosophy

5.4.1 Qualia as irreducible phenomena

One major counterargument to the *Illusory Significance Hypothesis* comes from traditional philosophical views, such as phenomenal realism or qualia realism, which argue that qualia are irreducible phenomena that cannot be explained purely through cognitive or functional mechanisms. **Table 2:** *Comparison of the Traditional View vs. Illusory Significance Hypothesis* provides a structured comparison of these views, detailing how each conceptualizes selfhood, qualia, and the nature of consciousness. Philosophers like David Chalmers and Thomas Nagel have famously argued that qualia—the subjective, experiential quality of consciousness—are irreducible and fundamental. Nagel's *'What is it like to be a bat?*' argument emphasizes that consciousness cannot be fully understood through objective, third-person analysis because subjective experience is an essential component of being (Nagel, 1974). Similarly, Chalmers' hard problem of consciousness (Chalmers, 1996) argues that explaining how and why we have subjective experiences remains unsolved by physicalist or cognitive approaches like predictive coding. However, recent discussions on cognitive accessibility suggest that much of what is considered phenomenal consciousness can be understood through access consciousness—the brain's ability to make sensory data available for cognitive processing and reporting. Qualia—just like dogs and cats—are part of the inferred suite of hidden causes (i.e., experiential hypothesis) that best explain and predict the evolving flux of energies across our sensory surfaces. This challenges the idea that qualia are metaphysical phenomena, reframing them as part of the brain's predictive processing (Clark et al., 2019). Philosophers like Thomas Nagel and David Chalmers have argued that the subjective "*what it is like*" quality of experience cannot be reduced to predictive processing or neuroscience (Nagel, 1974; Chalmers, 1996).

Table 2 Comparing Traditional Views and the Illusory Significance Hypothesis

Aspect	Traditional View	Illusory Significance Hypothesis
Nature of Selfhood	Intrinsic, metaphysical entity	Cognitive artifact constructed by the brain
Qualia	Irreducible, fundamental experiences	Functional tools arising from predictive processing
Consciousness	Dualistic or inherently subjective	Emergent property of neural computations
Role of Language	Descriptive of reality	Instrumental in constructing and reinforcing self-narratives
Implications for AI	AI cannot possess true selfhood or qualia	AI could develop analogous constructs through predictive models

To respond to this, the Illusory Significance Hypothesis does not deny the subjective experience of qualia but reframes the perception of qualia as a cognitive tool. In contrast to qualia realism, illusionist theorists like Daniel Dennett and Keith Frankish argue that phenomenal consciousness is a cognitive illusion. According to this view, qualia are not intrinsic consciousness features but useful cognitive byproducts that help the brain navigate complex sensory environments. The Illusory Significance Hypothesis aligns with this view, suggesting that qualia functionally in predictive processing, helping the brain minimize prediction errors while constructing a coherent narrative of experience (Dennett, 1991; Frankish, 2016). While we experience qualia subjectively, the hypothesis posits that their significance is illusory, created by the brain's predictive systems to make sense of sensory data. In other words, the "realness" of qualia is a necessary artifact for cognitive coherence, but it does not imply that they are fundamental metaphysical entities (Friston, 2010).

5.4.2 Selfhood as a necessary metaphysical entity

Another common critique is that selfhood is not merely a cognitive artifact but a fundamental aspect of conscious beings. Proponents of existentialist and personal identity theories (e.g., Jean-Paul Sartre & Derek Parfit) argue that selfhood is central to the human experience. Sartre

suggests that human consciousness is fundamentally tied to self-awareness and freedom and that the capacity for reflection on one's own existence is what defines being. Parfit, in contrast, argues that personal identity is based on psychological continuity—memory and mental states across time—which cannot be reduced to mere predictive models (Sartre, 1943; Parfit, 1984). From the perspective of existentialist and personal identity theories, selfhood is integral to understanding human experience and cannot be reduced to a functional construct. Proponents of this view argue that self-awareness and identity are core to moral and ethical considerations surrounding consciousness. As discussed in the *Illusory Significance Hypothesis,* conscious experience is viewed not as intrinsic features of consciousness but as cognitive tools emerging from the brain's predictive processes.

The *Illusory Significance Hypothesis* suggests that while selfhood plays an essential role in human experience, its illusory nature as a narrative construct does not diminish its importance for emotional and behavioral regulation. From the predictive coding perspective, selfhood is dynamically constructed as the brain works to maintain coherence across sensory inputs and emotional regulation. By reframing selfhood as a narrative construct, the Illusory Significance Hypothesis supports the idea that selfhood is functionally necessary but not metaphysical. Metzinger's self-model theory argues that the brain constructs a flexible and dynamic self-model to manage behavior and emotions, which aligns with the predictive coding framework, which continuously updates based on prediction errors (Metzinger, 2003). The hypothesis provides a framework for understanding selfhood as an adaptive tool while acknowledging its central role in how humans navigate the world (Seth, 2013). This functional perspective does not deny the phenomenological reality of selfhood but rather situates it within the brain's predictive processing.

By reframing conscious experience as cognitive artifacts, the Illusory Significance Hypothesis provides a compelling model for understanding consciousness that challenges traditional philosophical perspectives. The practical implications in neuroscience and artificial intelligence offer exciting possibilities for new therapies, technologies, and ethical considerations while responding thoughtfully to counterarguments from philosophical realism. This hypothesis opens the door to innovative approaches to understanding and interacting with both the human brain and future AI system.

6. Expected contributions

This paper significantly contributes by challenging traditional metaphysical views of conscious experience through the *Illusory Significance Hypothesis*. This research advances key theoretical frameworks by framing these constructs as cognitive tools that emerge from the brain's predictive coding and symbolic cognition processes. It opens new avenues for interdisciplinary exploration in philosophy, neuroscience, and artificial intelligence.

6.1 Advancing the Meta-Construct Problem of Consciousness

This paper significantly extends the meta-construct problem of consciousness by further articulating the functional nature of conscious experience. It reinforces the argument that these constructs are cognitive artifacts emerging from the brain's need to manage prediction errors through predictive coding mechanisms. The core contribution is demonstrating how these constructs, often treated as intrinsic features of consciousness, are better understood as adaptive tools the brain uses to maintain coherence rather than metaphysical necessities (Friston, 2010).

Additionally, the integration of symbolic cognition offers new insight into how language and symbols simplify the brain's complex, fluctuating experiences, helping to stabilize the narrative of selfhood. This paper advances the framework by showing that conscious experiences are illusory constructs reinforced by linguistic symbols and cultural narratives, creating the illusion of intrinsic meaning.

6.2 Integration with the Pragmatic Role of Consciousness

Building on the pragmatic role of consciousness and emotion framework, this paper argues that emotions, like selfhood, serve a regulatory and adaptive role. The brain constructs conscious experience to aid in maintaining homeostasis and ensuring effective decision-making in uncertain environments. The addition of symbolic cognition enriches this framework by illustrating how language stabilizes emotional experiences, further supporting the constructive and functional role of emotions in human survival (Barrett, 2017).

This extension offers a deeper understanding of how predictive coding models regulate emotions and selfhood, emphasizing that the brain's emotional and narrative processes are tools for survival rather than reflections of intrinsic metaphysical realities.

6.3 Broader interdisciplinary contributions

6.3.1 Philosophy of mind

This paper contributes to philosophy of mind by challenging the traditional metaphysical interpretations of conscious experience. The *Illusory Significance Hypothesis* offers an alternative by proposing that qualia are functional artifacts created through predictive coding and narrative construction, aligning with embodied cognition models (Clark, 2013). This challenges dualist and qualia realist views that regard subjective experience as irreducible, instead positioning qualia as tools used to manage sensory complexity (Chalmers, 1996).

6.3.2 Neuroscience

This paper advances neuroscientific research by emphasizing the relationship between predictive coding and interoceptive processing in constructing selfhood. The hypothesis suggests that disruptions in these predictive mechanisms may underlie disorders such as dissociative identity disorder (DID), anxiety, and depression. Therapeutic interventions could improve emotional regulation and self-coherence by targeting the brain's interoceptive awareness and prediction error mechanisms (Seth & Critchley, 2013). Neuroimaging studies on interoceptive prediction errors could further elucidate the neural underpinnings of selfhood.

6.3.2 Artificial Intelligence

The implications for artificial intelligence are profound. As AI systems develop predictive processing capabilities, they may begin to generate self-referential outputs or even simulate qualia using symbolic cognition. This research provides a framework for exploring how AI systems could mimic selfhood without subjective experience, leading to new ethical considerations regarding machine consciousness (Hohwy, 2017). Examples of AI models like GPT-3 and reinforcement learning systems already exhibit early forms of self-referential behaviors, providing a foundation for further research into AI selfhood.

6.4 Contributions to future research

By extending the meta-construct problem of consciousness and the pragmatic role of consciousness frameworks, this paper opens numerous avenues for future research:

1. Neuroimaging Studies: Future work could investigate how the brain's prediction errors in interoception affect the construction of selfhood. fMRI studies on individuals with disruptions in self-narratives, such as those with dissociative disorders, could provide new insights into the neural mechanisms that maintain self-coherence (Seth & Critchley, 2013).

2. AI and Predictive Models: Future research could explore whether AI systems employing predictive coding frameworks develop self-modeling behaviors and how they manage their interaction with symbolic representations of themselves and their environments.

3. Ethical Considerations: The emergence of AI selfhood and artificial qualia demands a reevaluation of ethical frameworks. Researchers and ethicists will need to explore how society defines consciousness and moral agency in relation to advanced AI systems that exhibit self-referential behaviors without subjective experience.

The *Illusory Significance Hypothesis* significantly contributes to understanding conscious experience as cognitive artifacts, challenging traditional metaphysical views. This paper bridges philosophy, neuroscience, and AI, offering novel pathways for empirical research and ethical discourse in the rapidly evolving fields of artificial intelligence and consciousness studies.

7. Discussion

This paper has explored the *Illusory Significance Hypothesis*, arguing that conscious experiences are not intrinsic, metaphysical elements of consciousness but cognitive artifacts that emerge from the brain's predictive coding and symbolic cognition systems. **Figure 5** provides a visual overview of the major contributions discussed in this paper, ranging from foundational works like James' Principles of Psychology to modern predictive coding theories by Friston. By understanding conscious experience as tools used

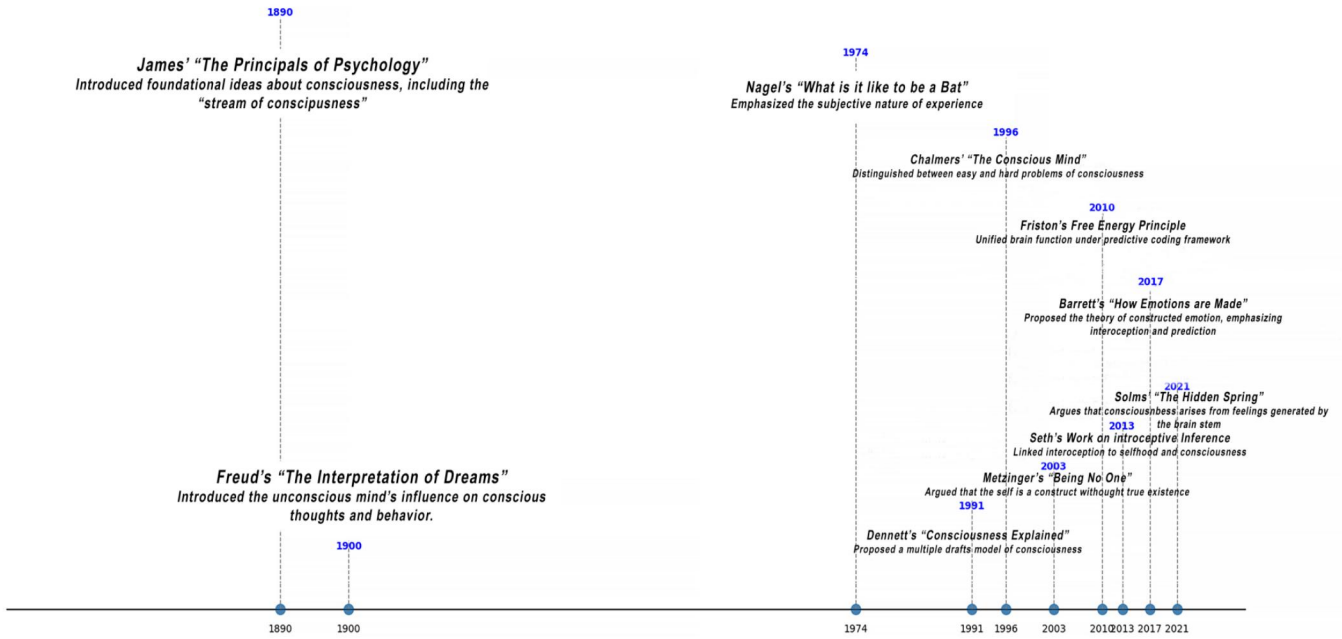

Figure 5 - Timeline illustrates major milestone in the study of consciousness from the late 29th century to the 21st century, highlighting key theoretical developments and presents a comprehensive view of the progression of ideas that have shaped our understanding of selfhood, qualia, and consciousness. It highlights key contributions from psychology, neuroscience, and philosophy, showing how each built upon previous theories to deepen our comprehension of the mind.

by the brain to minimize prediction errors, the hypothesis challenges traditional philosophical interpretations and opens up new avenues for empirical research.

The hypothesis integrates well with the frameworks established in the Meta-Construct Problem of Consciousness and the Pragmatic Role of Consciousness and Emotion, offering a construct-conscious experience. The brain constructs a self-narrative through predictive models that help regulate behavior, emotions, and sensory input. Language and symbols further stabilize these constructs, providing the illusion of intrinsic significance. This interdisciplinary approach bridges neuroscience and AI and offers new directions for our ethical challenges as AI systems increasingly resemble human cognitive functions without possessing true subjective awareness.

7.1 Philosophical and practical implications

This redefinition of qualia and selfhood as functional artifacts has far-reaching implications for the philosophy of mind. By rejecting the idea that these constructs are metaphysical entities, the *Illusory Significance Hypothesis* proposes a functional view of consciousness aligned with predictive coding and embodied cognition (Clark, 2013). This perspective challenges dualism and qualia realism while acknowledging the subjective reality of experience but referring it as adaptive rather than essential.

In neuroscience, the hypothesis provides a new framework for understanding disorders of selfhood and emotional regulation. It suggests that dissociative identity disorder (DID), anxiety, and depression may result from disruptions in the brain's predictive mechanisms, especially in how it processes interoceptive signals. Interventions enhancing interoceptive awareness or correcting prediction errors could provide new therapeutic pathways for these conditions (Seth & Critchley, 2013).

7.2 AI and ethical considerations

In AI, *the Illusory Significance Hypothesis* suggests that AI systems capable of predictive processing may develop self-referential behaviors and simulate qualia without having subjective consciousness. This raises complex ethical questions. If AI systems can generate self-models and mimic selfhood, how should they be treated? What criteria should determine whether AI systems are entitled to certain rights or protections, even without subjective awareness?

Furthermore, as AI systems evolve to simulate artificial qualia, an ethical framework that addresses the moral implications of creating machines that exhibit self-referential behavior without conscious experience is needed. These questions require a multidisciplinary approach, blending philosophy, neuroscience, and AI ethics (Hohwy, 2017).

7.3 Future research directions

Future empirical research is critical for testing the predictions made by the *Illusory Significance Hypothesis*. Neuroimaging studies could investigate how the brain's processing of prediction errors in interoception influences the construction of conscious experience. For example, fMRI experiments could track brain activity in individuals with disrupted self-narratives, such as those with DID, to better understand how selfhood fluctuates in response to changes in interoceptive signals (Seth, 2013).

In AI research, experiments could focus on whether AI systems using predictive models and symbolic cognition begin to develop self-referential outputs that mimic selfhood. Such experiments could involve training AI to simulate environments that require self-representation and assessing whether these systems exhibit behaviors that resemble self-awareness. The development of artificial qualia in AI systems could be explored by investigating how AI systems interpret and respond to sensory data using symbolic representations (Clark, 2013).

7.4 Final reflections

The *Illusory Significance Hypothesis* offers a robust interdisciplinary framework that redefines conscious experience as adaptive, cognitive tools rather than metaphysical entities. This redefinition advances theoretical discussions in the philosophy of mind and suggests practical applications for neuroscience and AI development. By reframing subjective experience as an emergent property of the brain's predictive processes, this paper paves the way for future research in understanding human and machine consciousness.

References

Barrett, L. F. (2017) *How Emotions are Made: The Secret Life of the Brain*. Houghton Mifflin Harcourt.

Blanke, O. (2012) Multisensory brain mechanisms of bodily self-consciousness. *Nature Reviews Neuroscience,* **13**, 556-571.

Botvinick, M. & Cohen, J. (1998) Rubber hands' feel' touch that eyes see. *Nature,* **391**, 756-757.

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

Chalmers, D. J. (2018) The meta-problem of consciousness. *Journal of Consciousness Studies,* **25**, 6-61.

Clark, A. (2013) *Surfing uncertainty: Prediction, action, and the embodied mind*. Oxford University Press.

Clark, A., Friston,K. & Wilkinson,S. (2019) Bayesing qualia: consciousness as inference, not raw datum. *Journal of Consciousness Studies* **26**,19-33.

Critchley, H. D., & Seth, A. K. (2013) Will studies of macaque insula reveal the neural mechanisms of self-awareness? *Neuron,* **83**, 119–123.

Damasio, A. R. (1999) *The Feeling of What Happens: Body and Emotion in the Making of Consciousness*. Houghton Mifflin Harcourt.

Friston, K. (2010) The free-energy principle: A unified brain theory? *Nature Reviews Neuroscience,* **11**, 127-138.

Hohwy, J. (2017) How to entrain your evil demon. *Philosophy and the Mind Sciences,* **6**, 1-14.

Loker, T. W. (2023) The meta-construct problem of consciousness: Selfhood and emotions as functional artifacts. *ResearchGate Preprint*.

Loker, T. W. (2024) *Language, symbols, and the pragmatic reality of consciousness. ResearchGate preprint.*

Nagel, T. (1974) What is it like to be a bat? *Philosophical Review,* **83**(4), 435-450.

Quesque, F. & Brass,M. (2019) The role of the temporoparietal junction in self-other distinction. *Brain topography* **32**, 943-955.

Rao, R.P.N., Gklezakos, D.C. & Vishwas Sathish, V. (2023) Active Predictive Coding: A Unifying Neural Model for Active Perception, Compositional Learning, and Hierarchical Planning. *Neural Computation* (2024) **36**, 1–32.

Salvatori, T., Mali, A., Buckley, C.L., Lukasiewicz, T., Rao, R.P.N. *et al.* (2023) Brain-Inspired Computational Intelligence via Predictive Coding. Preprint http://dx.doi.org/10.48550/arXiv.2308.07870

Seth, A. K. (2013) Interoceptive inference, emotion, and the embodied self. *Trends in Cognitive Sciences,* **17**, 565-573.

Seth, A. K., & Critchley, H. D. (2013) Interoceptive predictions in the brain. *Nature Reviews Neuroscience,* **14**, 451-467.