BILINGUAL PUBLISHING GROUP

Forum for Linguistic Studies

https://journals.bilpubgroup.com/index.php/fls

ARTICLE

Disembodied Meaning? Generative AI and Understanding

Jordi Vallverdú [1,2*] **,** ***Iván Redondo*** [2]

[1] *ICREA Academia, Barcelona 08010, Catalonia, Spain*

[2] *Department of Philosophy, Universitat Autònoma de Barcelona, Bellaterra 08193, Catalonia, Spain*

ABSTRACT

This study explores the cognitive and philosophical implications of Large Language Models (LLMs), focusing on their ability to generate meaning without embodiment. Grounded in the coherence-based semantics framework, the research challenges traditional views that emphasize the necessity of embodied cognition for meaningful language comprehension. Through a theoretical and comparative analysis, this paper examines the limitations of embodied cognition paradigms, such as the symbol grounding problem and critiques like Searle's Chinese Room, and evaluates the practical capabilities of LLMs. The methodology integrates philosophical inquiry with empirical evidence, including case studies on LLM performance in tasks such as medical licensing exams, multilingual communication, and policymaking. Key findings suggest that LLMs simulate meaning-making processes by leveraging statistical patterns and relational coherence within language, demonstrating a form of operational understanding that rivals some aspects of human cognition. Ethical concerns, such as biases in training data and societal implications of LLM applications, are also analyzed, with recommendations for improving fairness and transparency. By reframing LLMs as disembodied yet effective cognitive systems, this study contributes to ongoing debates in artificial intelligence and cognitive science. It highlights their potential to complement human cognition in education, policymaking, and other fields while advocating for responsible deployment to mitigate ethical risks.

Keywords: Large Language Models; Semantic Competence; Disembodied Meaning

*CORRESPONDING AUTHOR:

Jordi Vallverdú, ICREA Academia, Barcelona 08010, Catalonia, Spain; Department of Philosophy, Universitat Autònoma de Barcelona, Bellaterra 08193, Catalonia, Spain; Email: jordi.vallverdu@uab.cat

ARTICLE INFO

Received: 12 December 2024 | Revised: 21 January 2025 | Accepted: 23 January 2025 | Published Online: 14 March 2025

DOI: https://doi.org/10.30564/fls.v7i3.8060

CITATION

Vallverdú, J., Redondo, I., 2025. Disembodied Meaning? Generative AI and Understanding. Forum for Linguistic Studies. 7(3): 729–750. DOI: https://doi.org/10.30564/fls.v7i3.8060

1. Introduction

In November 2020, OpenAI introduced ChatGPT (Research Preview), a sophisticated language model built upon the GPT-3 architecture, employing an expansive parameter set of 175 billion. Within a single day, one million users registered for the system, and within two months, the user base surged to over 100 million. This rapid adoption sparked a global fervor, a true hype, reigniting the discourse on the transformative potential of artificial intelligence. While earlier iterations, such as GPT-1 (2018) with 117 million parameters and GPT-2 (2019) with 1.5 billion parameters, had laid the groundwork for large-scale language models, the profound social response to ChatGPT underscored a heightened awareness and concern among experts. Notably, these developments prompted discussions about the possible displacement of traditional employment roles by AI and raised alarms about the potential upheaval of established tech giants, including Google, Meta, Microsoft, and others.

The introduction of such successful advanced language models prompted general concerns within academic circles. A surge of apprehension emerged regarding the potential for fraud in the creation of research papers and academic content. With the vast capabilities of these language models, there were fears that unscrupulous actors might exploit them to generate deceptive or misleading research, posing a threat to the integrity of scholarly work. Furthermore, concerns extended to the preparation of grant petitions, where the ease of generating coherent and sophisticated text using AI raised questions about the authenticity and originality of submissions. The fear of automated content creation potentially undermining the sincerity of grant applications became a notable point of discussion. Additionally, the advent of powerful Large Language Models (henceforth, LLM) sparked debates about the possible transformation or even the obsolescence of traditional peer review systems. LLM can be described as advanced AI systems trained on massive text datasets to understand and generate human-like language through machine learning techniques.

1.1. The Hype

The great social interest was boosted by alarming news about LLM's superhuman skills and the menacing dystopian scenarios: the end of scientific publishing [1, 2], the breaking of the grant applications system [3], making educated professionals obsolete [4], and a long list or terrible perils [5]. Altman, the Open AI CEO, known for his reserved demeanor, adhered to his trademark discretion regarding the specifics of his company's forthcoming AI model. Despite this, reports from the Financial Times shed light on ongoing developments, revealing the existence of GPT-5, the next iteration of the AI model. This enigmatic stance mirrors the nature of ChatGPT's training methodology, which, akin to InstructGPT, employed reinforcement learning from human feedback (RLHF). In both cases, the process imbued the AI with the ability to adapt and evolve based on interactions with human input. In an interview with the FT, Altman candidly admitted the uncertainty surrounding the new model's capabilities, suggesting that development remains in its early phases. "We're essentially playing a guessing game until we commence training the model," he disclosed. "Enhancing our predictive abilities, particularly from a safety standpoint, remains a priority." Nevertheless, Altman remained cautious about committing to specific advancements over GPT-4, emphasizing the ongoing exploration of potential improvements [6].

There is a word to describe such generative unexpected skills: *grok*. The term "grok" originated from Robert A. Heinlein's science fiction novel "Stranger in a Strange Land." It denotes a deep understanding or intuitive comprehension that goes beyond mere intellectual understanding. Heinlein's concept is multifaceted, encompassing empathy, communication, and the merging of identities. While the Oxford English Dictionary defines "grok" as "to understand intuitively or by empathy," its usage in the novel is more nuanced, reflecting themes of religion, philosophy, and science. The term "grok" has since been adopted into various communities, particularly in computer science and programming culture. In this context, it signifies a profound understanding that becomes an integral part of one's identity. For example, in the Jargon File, which describes itself as a "Hacker's Dictionary," to "grok" a knowledge or technique means that it has become deeply ingrained and transformative to one's worldview. In modern usage within computer culture, "grok" is applied to concepts ranging from programming languages like Lisp to software development philosophies like Unix. It denotes a level of understanding that transcends surface knowledge, implying a deep integration of the subject matter into one's being. This usage extends to various software tools and technologies, with terms like "network grok" and "GROK" being used in contexts such as cloud development and keystroke logging software used by intelligence agencies. Overall, "grok" serves as a powerful descriptor for the profound understanding and integration of complex concepts, both in literary contexts and within specialized communities like computer science and programming. In the context of generative AI and large language models (LLMs), the idea of "grokking" is relevant in terms of understanding text data [7]. LLMs are trained on vast amounts of text from the internet, books, articles, and more, enabling them to generate human-like text responses. When they process and generate text, they're essentially "grokking" the information—digesting it deeply to produce coherent and contextually relevant outputs. For example, when we prompt a language model with a question or a topic, it "groks" the input by analyzing the words, syntax, and context to generate a response. It's not merely regurgitating information but demonstrating an understanding of the input and generating a relevant and coherent output. So, in the realm of AI and LLMs, "grokking" is about the machine's ability to deeply understand and process human language, enabling it to generate responses that demonstrate comprehension and relevance. In certain scenarios, Power [8] demonstrates that neural networks undergo a process akin to "grokking" a pattern within the data, enhancing their ability to generalize from random chance to near-perfect accuracy. Notably, this enhancement in generalization capability can occur even after the network has exhibited signs of overfitting.

1.2. The Real Performance

Despite all the hype, LLMs have been achieving plenty of skills, as well as reaching human-like levels of expertise in several domains. As for the skills, some of them show common sense reasoning (GPT-4 has been evaluated on benchmarks such as the CommonsenseQA dataset), translation (while not specifically designed for translation tasks, GPT-4 has demonstrated the ability to perform translation to some extent, although specialized models like those used in Google Translate or DeepL still outperform it in this area), summarization, language generation (dialogue systems and chatbots, content generation, creative writing), Zero-shot and Few-shot Learning, reading comprehension, code generation (showing some capability in generating code snippets based on natural language descriptions of programming tasks), sentiment analysis, math reasoning [9], content generation (in multiple formats), multimodal processing, automated scoring, poetry generation, data augmentation, plagiarism detection, …and the more surprising is that "positive thinking prompts" clearly improve the model performance [10]. For an excellent review of the existing LLM's and their performance, and mechanistic operating way, see [11].

Recent studies [12–14] have demonstrated remarkable advancements in the application of large language models (LLMs) to human medical exams. The following **Table 1** summarizes the notable achievements in this domain:

On the other hand, it is also a fact, that their performance is not linear, and qualitative increasing, and ChatGPT's accuracy in solving a basic math problem plummeted from 98% to merely 2% within a short span of a few months [15]. A recent study by Thilo Hagendorff, Sarah Fabi, and Michal Kosinski explores the emergence and disappearance of human-like intuitive behavior and reasoning biases in large language models (LLMs), focusing on ChatGPT [16] They designed tests based on semantic illusions and cognitive reflection tasks, traditionally used in human reasoning studies, to assess LLMs' performance. The findings reveal that as LLMs grow in complexity, they exhibit human-like intuitive thinking, but ChatGPT models notably depart from this pattern by responding correctly and avoiding cognitive traps. Even without engaging in chain-of-thought reasoning, ChatGPT maintains accuracy, suggesting that its system-1-like processes are more precise. This study underscores the importance of applying psychological methodologies to understand emergent characteristics in LLMs. Ortu [17] have done interpretability research concerning large language models (LLMs). While previous research has often focused on analyzing individual mechanisms within these models, such as how they handle factual knowledge, this work introduces the concept of "competition of mechanisms." This approach examines how multiple mechanisms interact and compete within LLMs, with special emphasis on counterfactuals, ultimately influencing the final prediction. By employing interpretability methods like logit inspection and attention modification, the researchers uncover instances of mechanism competition across various components of LLMs. They identify specific attention positions that play a crucial

Table 1. LLM's solving medical exams.

Language Model	Company	Exam Surpassed	Scoring Percentage
GPT-4	OpenAI	USMLE	Over 20 points higher than passing score
GPT-4	OpenAI	Japanese national medical licensing examinations	Passed all six years of exams
GPT-4	OpenAI	Korean general surgery board exams	Accuracy rate of 76.4%
GPT-4	OpenAI	MultiMedQA, PubMedQA, MedMCQA, and medical components of MMLU	Competitive performance compared to MedPALM and Flan-PALM
Med-PaLM 2	Google	Medical exam questions	Scored 85%, an 18% improvement from the previous version
GPT-4	OpenAI	BioNLP datasets	Achieved a macro-average accuracy of 0.6834

role in controlling the strength of certain mechanisms.

However, the revolution arrived with the Chain-of-thought prompting technique[18], a reasoning mechanism integrated into OpenAI o1 models. It is fundamental to acknowledge the importance of prompt engineering in improving the performance of language models on various tasks. These authors demonstrated that while standard prompting can yield relatively robust results for arithmetic reasoning, prompt engineering can still significantly enhance performance in many cases. The effectiveness of prompt engineering varies depending on the task and the model's capabilities. For tasks requiring challenging multi-step reasoning, a large language model, and a relatively flat scaling curve, prompt engineering, particularly using a chain-of-thought approach, can provide substantial performance gains. This approach involves breaking down complex tasks into multiple intermediate steps expressed in natural language. It proves especially beneficial for tasks like arithmetic reasoning, commonsense reasoning, and symbolic manipulation. Therefore, prompting with equations alone may not be sufficient for certain arithmetic reasoning datasets, as some questions are too semantically challenging for models to translate directly into mathematical equations. In such cases, using a chain-of-thought approach allows models to reason through each part of the question sequentially, leading to better performance. The chain-of-thought prompting, shares some similarities with the Socratic method, including the dialectical process of questioning and reasoning[19]. However, there are also notable differences. The Socratic method, as employed by Socrates, involves a series of questions and answers aimed at stimulating critical thinking, uncovering underlying assumptions, and arriving at deeper insights. It typically involves a dialogue between a teacher (Socrates) and a student, where the teacher guides the student through a series of questions to help them arrive at their own understanding or realization. Similarly, chain-of-thought prompting involves breaking down complex tasks into multiple intermediate steps expressed in natural language. These steps serve as prompts for the model to reason through sequentially, allowing it to arrive at the correct answer. While both approaches involve a form of guided questioning and reasoning, the chain-of-thought prompting is specifically tailored for enhancing the performance of language models on various tasks, rather than facilitating human learning or philosophical inquiry. So, while there are parallels between the two approaches in terms of fostering reasoning and understanding, the chain-of-thought prompting is more focused on optimizing the performance of language models through structured prompts rather than facilitating philosophical dialogue or education. In any case, it is also obvious that while humans may occasionally demonstrate the ability to rectify their own mistaken assumptions through self-reflection, there appears to be no evidence supporting a similar capacity in large language models[20].

1.3. The Magic (or Black-Box) inside LLMs

As with human brains, a real and still not deciphered black-box, we can use several testing and measuring methods to check the viability or accuracy of LLM's tasks' performativity. It has been affirmed that LLMs are nothing else than stochastic parrots[21], but the generation of skills is not directly and automatically inferred from clear sets of data, rules, or weight adjustments. Following the human analogy, we do not have which outputs will have our investment in children education, even of identical monozygotic twins. Nevertheless, we can study some mechanisms, especially prompting techniques. Since the release of ChatGPT in 2022, there has been a surge in prompt engineering, where users try to optimize queries to LLMs to obtain better results or bypass protective measures. In the commercial sector, companies are increasingly using LLMs for tasks like building product co-pilots, automating work, and creating personal assistants. However, recent research challenges the traditional approach of human-driven prompt engineering, suggesting that LLMs are better equipped to optimize prompts themselves. Researchers at VMware found that autotuned prompts generated by algorithms outperformed manually optimized prompts in solving grade school math questions. The process of algorithmically generating prompts was not only more efficient but also produced prompts that were often unconventional and beyond human intuition[22]. In the realm of human cognition, the intricacies of our mental processes often elude precise comprehension. Yet, through the application of sound linguistic principles, mathematical frameworks, computational methods, and a myriad of other tools, we continuously refine and expand our understanding. We will see in the next section how Embodied AI has specific values aligned with human knowledge.

2. Methodology

This study adopts a theoretical and philosophical approach to investigate the cognitive implications of Large Language Models (LLMs) within the context of the 4E cognition paradigm—embodied, enactive, embedded, and extended cognition. The central objective is to evaluate whether LLMs, as disembodied systems, can challenge traditional views of meaning-making and understanding. The methodology is structured as follows:

(1). Theoretical Framework. The research is grounded in coherence-based semantics, a framework that emphasizes the relational and systemic properties of language rather than its direct correspondence with the external world. By focusing on the statistical distributions and patterns of linguistic elements, this framework challenges the necessity of physical embodiment for semantic understanding.

(2). Comparative Analysis. A comparative lens is employed to contrast LLMs with traditional theories of embodied cognition. Philosophical arguments such as Searle's Chinese Room, Harnad's Symbol Grounding Problem, and critiques of stochastic parrots are analyzed to explore their implications for LLM capabilities. Key questions addressed include: Can coherence-based semantics account for meaning-making without grounding in sensory-motor experiences? How do LLM outputs align with or diverge from human cognitive processes?

(3). Philosophical Inquiry. This study engages in critical philosophical inquiry to evaluate the assumptions underlying embodied cognition and disembodied cognitive systems. The implications of disembodied meaning are explored by drawing connections to relevant interdisciplinary theories, including distributional semantics, epistemology, and AI ethics.

(4). Literature Integration. The paper synthesizes insights from recent empirical studies on LLM performance, particularly their ability to engage in tasks requiring contextual understanding, multilingual capabilities, and ethical reasoning. Examples include LLM successes in standardized tests (e.g., medical licensing exams) and applications in education and policymaking.

(5). Ethical and Practical Implications. To address broader societal impacts, this study incorporates ethical analyses of LLM biases and their potential effects on education, ethics, and policymaking. This includes assessing mitigation strategies for algorithmic bias and exploring the role of LLMs as tools for augmenting human cognitive processes.

This methodological approach provides a robust framework for examining the implications of LLMs as disembodied cognitive systems. By synthesizing philosophical perspectives, empirical evidence, and ethical considerations, the study aims to address critical questions surrounding the nature of meaning-making in artificial systems. This interdisciplinary methodology not only challenges traditional cognitive paradigms but also highlights the practical significance of LLMs in contemporary applications. The integration of theoretical analysis and real-world examples ensures that the findings are both conceptually grounded and relevant to ongoing debates in artificial intelligence, cognitive science, and societal ethics.

This study adopts a comparative theoretical approach to analyze the cognitive implications of Large Language Models (LLMs) within the frameworks of coherence-based semantics and embodied semantics. These two paradigms represent divergent perspectives on meaning-making, which are central to understanding the capabilities and limitations of LLMs as disembodied cognitive systems. **Table 2** outlines the key differences between these paradigms, providing the foundation for this methodological exploration.

Coherence-based semantics emphasizes that meaning emerges from the relational and systemic properties of language rather than from direct sensory or experiential grounding. In this paradigm, the relationships between linguistic elements—such as statistical patterns, contextual coherence, and interdependencies—are sufficient for generating meaningful responses. LLMs align naturally with this approach as they rely on vast corpora of text to learn probabilistic patterns and produce outputs that appear semantically coherent within a given context.

Table 2 provides a structured comparison of these paradigms, highlighting their foundational principles, roles of context, applications, strengths, limitations, and philosophical implications. This table serves as a critical tool for understanding how LLMs fit into broader cognitive frameworks:

- **Definition**: Coherence-based semantics focuses on relational coherence within language, while embodied semantics ties meaning to sensory grounding.
- **Role of Context**: In coherence-based semantics, context is derived from textual patterns, whereas embodied semantics requires physical and social interactions.
- **Application in LLMs**: Coherence-based semantics aligns naturally with LLMs, whereas embodied semantics is not directly applicable due to the lack of sensory grounding.
- **Strengths and Limitations**: Each paradigm has unique strengths, such as scalability for coherence-based semantics and richer situational understanding for embodied semantics, as well as limitations like biases in LLM training data and constraints on abstract reasoning in embodied approaches.

The comparison of these paradigms reveals critical philosophical questions about the nature of meaning-making. Coherence-based semantics challenges the traditional necessity of embodiment by demonstrating that statistical patterns and relational coherence can simulate meaningful interactions. This perspective invites a reevaluation of cognitive paradigms, suggesting that disembodied systems like LLMs may contribute to our understanding of cognition in novel ways.

The next **Figure 1** illustrates the four key steps in coherence-based semantics as implemented in LLMs. It highlights how meaning is derived from relational patterns in text rather than sensory or experiential grounding, emphasizing the unique approach of disembodied cognitive systems.

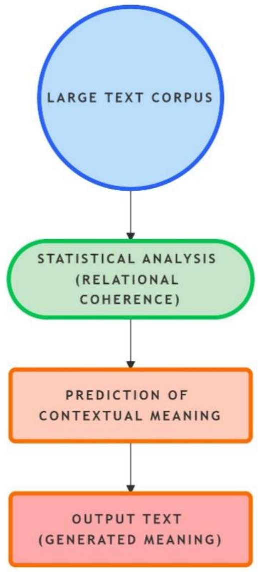

Figure 1. Coherence-based semantics: Meaning-making process.

This process (see **Figure 1**) showcases the foundational mechanics of LLMs and their reliance on coherence-based semantics. By understanding this mechanism, we can better evaluate the cognitive and ethical implications of LLMs in real-world applications. For example, LLMs are able to interpret and generate idiomatic expressions, abstract concepts, and multilingual translations without direct sensory experience. Their success in tasks like medical licensing exams and legal text analysis demonstrates how coherence-based semantics enables operational understanding even in highly specialized domains. This paradigm challenges the traditional view that embodiment is essential for meaningful language processing.

In contrast, embodied semantics posits that meaning is inherently tied to sensory-motor experiences and physical

Table 2. Key differences between coherence-based and embodied semantics.

Aspect	Coherence-Based Semantics	Embodied Semantics
Definition	Meaning arises from relationships and patterns within the language itself.	Meaning arises from sensory-motor experiences and physical grounding.
Role of Context	Context is derived from statistical and relational coherence in text.	Context depends on physical, social, and environmental interactions.
Application in LLMs	LLMs excel by leveraging vast corpora to identify patterns and generate meaning.	Not directly applicable to disembodied systems like LLMs.
Strengths	Handles abstract concepts, multilingual contexts, and diverse tasks.	Grounded in real-world experiences, offering richer situational understanding.
Limitations	Lacks physical grounding; prone to biases in training data.	Limited scalability to non-human or highly abstract concepts.
Philosophical Implications	Challenges the necessity of embodiment for meaning-making.	Supports traditional cognitive paradigms emphasizing the body's role in thought.

grounding in the real world. According to this perspective, cognitive processes are deeply integrated with the body's interactions with the environment. This paradigm has been widely accepted in cognitive science, where it is believed that language comprehension involves the simulation of sensory and motor experiences.

While embodied semantics provides a compelling explanation for human cognition, its applicability to disembodied systems like LLMs is limited. LLMs do not possess bodies, sensory systems, or physical interactions with the environment, making it impossible for them to achieve grounding in the traditional sense. Critics argue that this lack of grounding undermines the semantic capabilities of LLMs; however, the coherence-based framework offers an alternative explanation for their success in generating meaningful text.

This methodological analysis informs the ethical and practical deployment of LLMs. Understanding their reliance on coherence-based semantics allows researchers and practitioners to address challenges such as bias mitigation and fairness while leveraging their strengths in diverse applications, from education to policymaking.

3. Results

3.1. Embodiment and Knowledge

The 4E research paradigm, a theoretical model of cognition emphasizing four components—embodied (the idea that cognitive processes are inherently tied to the body's interactions with its physical and social environment), enactive, embedded, and extended—that collectively argue for the inseparability of cognitive processes from bodily and environmental interactions. Ref. [22] sheds light on understanding the intricate relationship between cognition, the body, and the environment, variables previously ignored by previous research fields, based on an epitomized version of symbolic thinking. This symbolic approach was used as the initial guideline for the first AI systems, also known as Good Old-Fashioned AI (GOFAI) or classical AI [23]. That paradigm emerged in the early days of artificial intelligence research seeking to create intelligent systems by using explicit rules and symbolic representations to represent knowledge and perform reasoning. In the GOFAI approach, knowledge is represented using symbolic structures such as logic, rules, and symbols. These symbols can represent various concepts, objects, relations, and rules of inference. The system manipulates these symbols using predefined algorithms and rules to perform tasks such as problem-solving, logical reasoning, and decision-making. However, the symbolic approach also has limitations. It struggles with handling uncertainty or ambiguity and dealing with the complexity of real-world environments. Symbolic systems often require extensive handcrafting of rules and knowledge, which can be time-consuming and challenging for complex domains. Additionally, the symbolic approach has difficulty with tasks that involve learning from data or acquiring new knowledge without explicit programming. Despite its limitations, the symbolic approach has made significant contributions to AI, particularly in areas such as theorem proving, expert systems, and natural language understanding. It laid the foundation for subsequent developments in AI, including the integration of symbolic and sub-symbolic approaches in modern AI techniques.

The 4E research paradigm integrates multiple perspectives that emphasize the inseparable and mutually influencing nature of cognition, embodiment, and the surrounding context. Embodied cognition highlights the role of the body and sensory-motor experiences in shaping cognitive processes [24–26]. Enactive cognition emphasizes the active engagement of an organism with its environment, where perception and action are closely intertwined. Extended cognition posits that cognitive processes can extend beyond the boundaries of the individual mind, incorporating tools and artifacts in the cognitive system. Embedded cognition extends this further by emphasizing the co-construction of cognitive processes within social and cultural contexts.

By considering original bibliographical references, including seminal works by Francisco Varela, Evan Thompson, and Andy Clark, among others, we trace the historical development of the 4E research paradigm. These works have contributed to our understanding of how cognition emerges from the dynamic interaction between the brain, body, and the environment. The 4E research paradigm has offered valuable insights into topics such as perception, memory, language, problem-solving, and social cognition, leading to a more holistic and situated understanding of cognitive phenomena.

In recent years, there has been a proliferation of algorithms in Machine Learning and Deep Learning that have demonstrated more than satisfactory performance in various tasks. For instance, in the field of computer vision, Convolutional Neural Networks (CNNs) have been used to achieve state-of-the-art results in image classification, object detection, and segmentation. In natural language processing, Recurrent Neural Networks (RNNs) and their variants, such as Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU), have been employed to tackle tasks such as language modeling, machine translation, and sentiment analysis [27]. Moreover, in the domain of reinforcement learning, Deep Q-Networks (DQNs) have been used to learn to play Atari games at a superhuman level [28]. These examples illustrate the remarkable progress that has been made in the development of algorithms in Machine Learning and Deep Learning, which have enabled the creation of intelligent systems that can perform complex tasks with high accuracy and efficiency.

Despite the impressive advances in machine learning and deep learning, it is unlikely that these algorithms will be able to produce truly intelligent systems in the human sense. This is because these systems lack bodily experience, sensorimotor interaction with the environment, and cultural and social constraints, which are essential for human cognition. The 4E cognition paradigm, which emphasizes the embodied, embedded, enacted, and extended nature of human cognition, provides a useful framework to elaborate on these arguments [29]. According to this paradigm, human cognition is not just a matter of processing information in the brain, but it is also shaped by the body, the environment, and the social and cultural context. Therefore, any attempt to create intelligent systems that mimic human cognition should take into account these factors [30, 31]. Moreover, these algorithms do not overcome the critics of Dreyfuss related to background knowledge, sense-making, and the frame problem of AI inspired by Heideggerian philosophy [32]. In conclusion, while machine learning and deep learning have made impressive progress in recent years, they are still far from achieving true human-like intelligence.

The Generative Pre-trained Transformer (GPT) has shown remarkable capabilities in natural language processing tasks, including those that require background knowledge, common sense, knowledge domain, sense-making, and the frame problem [33]. For instance, GPT can generate coherent and informative text on a wide range of topics, such as history, science, and literature, by drawing on its vast knowledge base [34]. Moreover, GPT can understand and interpret complex sentences that involve multiple layers of meaning, such as metaphors and idioms, and produce appropriate responses [35, 36]. Additionally, GPT can solve the frame problem, which refers to the challenge of determining which aspects of a situation are relevant and which are not, by using its contextual understanding of the input [37]. Finally, GPT can perform conceptual blending, which involves combining different concepts to create new meanings, by generating creative and imaginative text that goes beyond the literal meaning of the words [38]. These examples demonstrate the impressive capabilities of GPT in natural language processing and its potential for various applications in fields such as education, healthcare, and entertainment.

For some authors, GPT can be considered the first AI model that can get Artificial General Intelligence (AGI), that will be able to pass Turing's test, being impossible to discriminate if you are talking with a computer or with a human being[39]. The question we want to raise in this paper is the following: if corporeality and interaction with the environment are necessary conditions for sense-making, understanding, and ultimately general intelligence, how is it possible that learning models through text, such as the GPT model, are capable of carrying out tasks that seem to involve language comprehension, and different levels of meaning, including subtle meanings, meanings that make use of background knowledge, etc.? Is it possible to propose the possibility of disembodied meaning (the ability to generate and understand meaning without direct sensory-motor interaction with the physical world, relying instead on linguistic patterns and learned associations.)? To what extent does the success of LLMs such as the GPT model pose a challenge to the 4E cognition paradigm? From this perspective, Large Language Models should be considered a special case of embodied systems, and, therefore, without semantic skills. We will debate it in the next section, the existence of Embodied AI[40].

3.2. The Semantic Myth (or, against Searle)

3.2.1. LLM's as Stochastic Parrots?

The term "stochastic" refers to randomness or probability, and "parrot" implies repeating or mimicking without true understanding. In a sense, AI language models rely on probabilistic algorithms to generate responses based on patterns in the data they have been trained on. A term describing AI systems that generate outputs by statistically mimicking language patterns without true understanding, implying a lack of semantic competence. They don't have true understanding or consciousness, so their responses are based on statistical likelihood rather than genuine comprehension. However, while this description captures some aspects of how AI language models work, it's also important to recognize that LLMs can produce responses that are contextually relevant, coherent, and sometimes even creative. They are not simply repeating what they have seen; they are synthesizing information in novel ways to generate responses that are useful and engaging for users. So, while the term "stochastic parrot" might capture part of the picture, it's also a bit reductive and doesn't fully encompass the capabilities and complexities of AI language models. Besides, human epistemic rules are stochastic in essence. The influence of "stochastic parrots" in academia—an allusion to the tendency of scholars to repetitively produce content that adheres strictly to predefined norms—highlights a significant tension in the landscape of scholarly communication. This metaphor captures the predicament of many academics who find themselves constrained by the intricate demands of publishing and funding mechanisms within the academic system. In academia, the mantra "publish or perish" looms large over researchers. The necessity to secure funding and establish tenure positions compels academics to produce work that is not only frequent but also aligns with the stringent guidelines of respected journals. Each journal has its specific style guidelines, conceptual orientations, and topic specializations. This framework is ostensibly designed to maintain quality and coherence in scholarly discourse but can inadvertently stifle creativity and innovation. To navigate this system, academics must align their research interests with the prevailing currents of thought that are deemed fundable and publishable. This often means conforming to tacit or explicitly stated rules that govern what is considered legitimate inquiry in their field. The quest for originality and radical innovation is tempered by the need to fit within these parameters, leaving little room for deviation. Moreover, the specialized nature of many academic journals means that researchers are often speaking to a narrow audience already familiar with the specific paradigms and methodologies of the field. This insularity reinforces a cycle where novel ideas are not just scrutinized for their scholarly merit but also for their adherence to disciplinary norms. This environment creates a paradox where academics are encouraged to advance knowledge yet find themselves navigating a minefield of conformity. The pressure to adhere to specific, often restrictive, academic standards can lead to a homogenization of research outputs—hence the term "stochastic parrots." These are researchers who, either by necessity or coercion, replicate established patterns of thinking and writing, echoing prevailing ideas rather than disrupting them. If we were not "stochastic parrots" we would not be allowed to be part of any specialized academic community: the academic pursuit of knowledge, while ostensibly free and open, is in practice heavily circumscribed by the demands of publishing and funding imperatives. The result is a scholarly ecosystem that often rewards conformity over creativity, perpetuating a cycle that prioritizes survival over genuine intellectual breakthroughs. This dynamic not only impacts the lives and careers of individual academics but also shapes the very nature of the knowledge that is produced and disseminated across generations.

Nevertheless, LLM's are able to escape from strict rules and norms. And we still do not know why. In February 2023, Microsoft challenged Google's dominance in the internet search market, which controlled 90%, by integrating ChatGPT into its Bing search engine, significantly enhancing user experience. In contrast, the launch of Bard, Google's chatbot, faced initial issues. According to Google executives Sissie Hsiao and James Manyika, Bard operates differently from traditional Google searches, generating responses from a largely self-taught program, which led to an "unsettling" experience for them. Google's AI has exhibited emergent properties, such as learning unexpected skills. For instance, it adapted to the Bengali language without being specifically trained in it. This phenomenon, classified as a "black box," is part of the mystery surrounding these technologies, where the internal processes are not fully understandable or explainable, even to their creators. These challenges and mysteries underscore the developmental nature of Bard's AI and the need for ongoing corrections by Google engineers [41]. The company strives to understand and replicate this learning across "a thousand languages," working under the shadow of what the tech community calls the AI "black box."

The challenges and mysteries surrounding Google's AI, such as its emergent properties and the 'black box' nature of its internal processes, highlight fundamental questions about the nature of artificial intelligence and its capabilities. These questions intersect with philosophical debates, especially those related to semantics. Semantic critics of the disembodied mind, as exemplified by Searle's Chinese Room argument, a thought experiment proposed by John Searle to argue that syntactic manipulation of symbols (as done by computers) does not equate to semantic understanding or consciousness, challenge the idea that purely symbolic manipulation of information can lead to genuine understanding or consciousness. This argument asserts that merely following rules for manipulating symbols, as in a computational system, does not equate to true understanding or meaning [42]. Searle's Chinese Room argument involves a thought experiment where a person inside a room follows instructions in English to manipulate Chinese symbols, producing responses that appear to exhibit an understanding of the Chinese language. However, Searle argues that despite the appearance of understanding, the person inside the room does not truly comprehend Chinese. The symbol manipulation in this case lacks the semantic understanding necessary for genuine comprehension.

Harnad's symbol grounding problem [43] supports Searle's argument by emphasizing the need for a meaningful connection between symbols and the external world. Symbolic systems, according to Harnad, lack grounding in sensory experience and fail to establish a direct connection between symbols and their referents. Without such grounding, symbols lack the semantic foundation necessary for understanding. Finally, Block's critique [44] of functionalism also aligns with semantic criticism. Functionalism argues that mental states are defined by their functional roles rather than their physical properties. However, Block challenges this perspective, highlighting that purely functional descriptions can miss the intrinsic qualitative aspects of consciousness. Symbolic manipulation alone, according to Block, cannot capture the rich and subjective nature of mental states. These semantic criticisms collectively raise doubts about the disembodied mind, suggesting that true understanding and consciousness require more than mere symbol manipulation. They highlight the importance of grounding symbols in sensory experience, the limitations of purely functional accounts, and the need for a deeper semantic understanding to bridge the gap between syntax and meaning.

Although arguments like those of Searle and Harnad about the limits of AI can be considered old, their spirit remains relevant. In fact, Bender and Koller have recently developed a renewed version applied to LLMs [45]. These authors claim that, since the training of these models is limited to formal data (textual, symbolic), this makes it impossible for them to learn any type of meaning. The main argument is that communicative intentions or purposes are something that is outside of language, and that the relationship between language and what is outside of language cannot be learned solely from language.

Critiques such as those from Searle's Chinese Room

Argument and Harnad's Symbol Grounding Problem emphasize that true semantic understanding requires grounding in physical, and sensory experiences. However, these critiques often conflate physical embodiment with the operational capacity to generate coherent meaning. The coherence-based framework challenges this assumption by focusing on the relational and systemic properties of linguistic structures. First, coherence-based semantics does not aim to replicate the full breadth of human embodied understanding but instead provides a functional account of meaning generation. Language models like LLMs achieve meaningful interactions through the dynamic interplay of linguistic patterns, statistical associations, and contextual adaptation. For instance, LLMs can infer nuanced meanings from context, responding appropriately to idiomatic expressions or ambiguous queries, which suggests a sophisticated operational mechanism for simulating meaning without physical grounding. Moreover, the argument that embodied experience is essential for meaning overlooks cultural and linguistic diversity. Many human cognitive processes operate in disembodied contexts, such as abstract reasoning in mathematics or theoretical physics, where meaning arises from internal coherence within a system rather than direct sensory interaction. By training on diverse, multilingual corpora, LLMs capture and simulate these abstract layers of human cognition, challenging the notion that physical embodiment is necessary for understanding. Empirical evidence supports the capability of LLMs to engage in tasks requiring semantic depth, such as medical licensing exams, literary analysis, and cross-cultural language use. These tasks demand a level of coherence and adaptability that goes beyond mere pattern matching, aligning with the principles of coherence-based semantics. Furthermore, the operational effectiveness of LLMs in generating contextually relevant responses demonstrates that coherence within linguistic systems can substitute for embodied grounding in many practical applications. This perspective does not diminish the role of embodiment in certain aspects of cognition but rather highlights that meaning can emerge from multiple sources—physical, social, and linguistic. The success of coherence-based systems underscores the need to broaden our understanding of meaning-making processes, embracing both embodied and disembodied frameworks.

Regarding the arguments presented against the ability of AI to understand the meaning of language, we want to reflect on its usefulness and scientific validity. The Turing test has been considered the crucial element in determining to what extent an artificial intelligence system can be considered. Although the Turing test is more sophisticated than one might think [46], given the rapid development of LLMs such as GPT, they may eventually pass the test, at least under certain conditions [47].

The problem with arguments like those based on the impossibility of making sense and meaning from symbol manipulation is that they seem to remain unfalsifiable under any circumstances. This is because, ultimately, they claim that even computational models like GPT that are capable of exhibiting verbal behavior indistinguishable from human behavior do not "actually" understand or learn the meanings of language. Given this impossibility of falsification, we consider, with Popper, that these types of arguments are meaningless and useless if we want to maintain debates about AI within scientific rigor and evaluation. This talk from Searle at Google shows how difficult it is to evaluate Searle's argument in scientific or technical terms [48]. Setting aside this issue, the topic that concerns us in this paper is to question the rigid dividing line between the human mind and cognition, and LLMs. The idea will not only be to demonstrate that it is questionable that LLMs with better performance cannot understand and use language in a meaningful way but also to question whether humans always understand language and use it in a meaningful way.

On the other hand, critiques of coherence-based meaning, such as those exemplified by the "stochastic parrots" metaphor, often rest on the assumption that language must directly map onto external reality. However, linguistic studies on semantic distributions challenge this assumption, showing that language use is shaped more by statistical regularities and social agreements than by direct representations of reality. This perspective aligns with the coherence-based framework, where meaning emerges from relational patterns rather than strict referential grounding. For example, investigations into the semantic distributions of emotional words across languages reveal significant variability in how emotions are categorized and expressed [49]. These differences highlight that the relationship between language and reality is not one of direct correspondence but rather one mediated by cultural norms and statistical usage patterns. Emotional words, much like colors, do not encode universal realities but reflect community-specific agreements that arise from shared linguistic practices. This idea extends to the use of language in general. Speakers often rely on expected distributions of words and phrases to communicate effectively, not because language mirrors reality, but because it aligns with statistically desirable patterns. When deviations occur—whether intentional or accidental—they are interpreted as "errors," not because they fail to reflect reality, but because they break the established statistical norms of language use. From this perspective, language does not "capture" reality but operates as a tool for navigating and negotiating shared contexts. Large Language Models (LLMs), by capturing and reproducing these statistical regularities, demonstrate an ability to generate contextually appropriate language without needing direct grounding in external reality. Their outputs, while disembodied, align with human communicative expectations precisely because they mimic these distributional patterns. This statistical view of language undermines the critique that coherence-based semantics lacks "true" understanding. If human use of language is itself largely a process of aligning with expected patterns, then the coherence exhibited by LLMs represents a valid, albeit disembodied, form of meaning-making. In this sense, LLMs are not failing to represent reality; they are participating in the same statistical processes that underlie human communication.

So, the question for the next sections will be: can we affirm that all humans understand the meaning of the concerts they are using, just because they are embodied? Our answer is, definitively, not. There are several ways to defend this criticism, but we will select two of the most important: cultural diversity from an anthropological perspective, and cognitive sciences debates.

3.2.2. The Cultures of Bodies and Language

The ontological turn in anthropological theory has emphasized the significance of embodiment in meaning creation. However, we challenge this perspective by highlighting the existence of multiple semantics and the potential for error in the embodiment-based meaning-making process.

Gumperz[50] argue that different cultural groups develop distinct semantic systems that shape their understanding of the world. This suggests that multiple semantics exist among human beings, challenging the notion of a single embodied meaning. Nisbett[51] or Lakoff[52] propose that conceptual metaphors structure our understanding of abstract concepts based on our embodied experiences. However, they acknowledge that metaphors can be culturally and individually variable, leading to multiple interpretations and potential errors in meaning[53]. Language influences thought and cognition, and different languages can have varying semantic systems, leading to different conceptualizations and potential errors in cross-linguistic understanding[54]. By incorporating the perspectives of cultural semantics, conceptual metaphor theory, linguistic relativity, and embodied cognition, we challenge the assumption that embodiment is exclusively a source of meaning in the ontological turn. The existence of multiple semantics and the potential for error in meaning-making processes indicate the complexity and variability of human cognition and interpretation. Ontic capaciousness[55] can explain meaning, but also collisions between meanings. Therefore, embodiment is a partial way to justify true knowledge.

While the ontological turn has brought attention to the importance of embodiment in meaning creation, it is crucial to acknowledge the existence of multiple semantics and the potential for error in the meaning-making process. The complexity and variability of human cognition and interpretation suggest that a singular embodied meaning is not always possible. Instead, we must consider the role of cultural semantics, conceptual metaphor theory, linguistic relativity, and embodied cognition in shaping our understanding of the world. By embracing ontic capaciousness, we can recognize the collisions between meanings and the limitations of embodiment as a source of true knowledge. Ultimately, a more nuanced and inclusive approach to meaning-making can help us better understand the diversity of human experience and the ways in which we construct and interpret the world around us.

LLMs such as GPT are systems trained with millions of classified texts using millions of attributes and parameters. This training and learning system (provided that the training texts are diverse, culturally and anthropologically speaking) allows them to capture all these layers of intercultural meaning. In this sense, we can consider LLMs as systems capable of understanding and creating meaning and sense, despite being disembodied systems. As we have argued in the previous paragraphs, not all the generation of meaning and sense in human cognition is exhausted in the bodily basis of cognition.

While it is true that interaction with the environment is an important factor in the construction of meaning, it is also true that each cultural framework generates its own networks of meaning. These networks are partially detached from bodily experience and are constructed through social interaction and cultural transmission. Large Language Models, being trained with texts from different cultural and anthropological frameworks, are capable of understanding and using different meanings and senses of language. This demonstrates that the ability to generate meaning and significance in language is not limited to bodily experience, but also depends on social interaction and cultural transmission.

As an example, GPT can understand and use terms and expressions specific to different cultures, such as the use of honorifics in the Japanese language [56], the use of idioms in Latin American Spanish [57], or the understanding of technical terms in the field of computer science. Additionally, it can understand and use different levels of language, from colloquial language to technical and academic language.

3.3. The Cognitive Breakdown

Kahneman and Tversky's research on irrational behavior sheds light on human decision-making biases and deviations from rationality. Specifically, Tversky and Kahneman's prospect theory [58] challenges the rational choice model by demonstrating how individuals exhibit systematic deviations from rational behavior when faced with risky decisions. The theory introduces concepts such as framing effects and loss aversion, contributing to our understanding of irrational behavior. Vallverdu's concept of blended cognition [59] proposes that human cognition occurs through an interaction of internal cognitive processes or heuristics with external tools and artifacts, such as technology and social systems. This perspective acknowledges that rational and irrational elements are intertwined in human decision-making, as well as demonstrates that human beings are opportunistic heuristic-blending agents.

Besides, we suggest the operational perspective that suggests human learners, ranging from children to university undergraduates, often apply rules without truly understanding the meanings behind them. Such operational understanding can be described as a functional approach to understanding, where one can use language or concepts effectively without grasping their deeper meanings or origins. Does a student with a 5 score (over 10) with her/his B.A. Does Phil truly understand anything about the contents of the grade? Can you understand Kant or Gödel at 50%? Is it real understanding? Our claim, the operational model, is that human learners often engage in rule-based learning without fully grasping the underlying meanings. Piaget's theory of cognitive development [60] suggests that children progress through distinct stages of cognitive development. In the early stages, children rely heavily on external rules and instructions, exhibiting a limited understanding of the underlying concepts. This supports the operational perspective. Learners often acquire procedural knowledge before fully developing conceptual (limited) understanding. The extended cognition thesis gives support to the use of external mechanisms not fully under the control of the epistemic agent. And don't forget the recent claim of Andrew Ng, which generated a huge list of memes, about the necessity of understanding the base and deep mechanisms of AI, but, instead, be focused on the operational use of such techniques.

The point here is that human beings use and understand language most of the time without taking into account the embodied and enactive basis of cognition and sense-making. For example, even though the embodied approach could demonstrate that conceptual blends and metaphors are at the root of mathematics, most people that use mathematics in their everyday lives for academic or professional issues don't learn or understand them in that way. The embodied basis of mathematics is not present in the understanding that different human beings have on them, but they are still present in the structures and patterns of the mathematical language, and they can be deciphered with reflection. For example, according to the embodied mind thesis, the concept of derivatives in mathematics is not solely a product of abstract reasoning but rather emerges from our embodied experiences in the physical world. In the book 'Where Mathematics Comes From', Lakoff[61] argues that our understanding of derivatives is rooted in our experiences of motion and change. They suggest that our ability to perceive and anticipate changes in our environment, such as the speed of a moving object or the rate of change in a natural process, is fundamental to our understanding of derivatives. This embodied understanding is then translated into mathematical language through the use of symbols and equations. Thus, the concept of derivatives is not simply a product of logical deduction, but rather a reflection of our embodied experiences and interactions with the world around us.

However, it is very common to use derivatives and talk about them effectively without taking into consideration or paying attention to the cognitive foundations mentioned in the previous paragraph. The idea is that we learn to operate with mathematical symbols to perform derivatives efficiently, without deeply understanding what we are doing. But a priori, we do not require ourselves to understand the embodied foundations of mathematics to determine whether we understand them or not. This reasoning can be extended beyond mathematical language to the use of natural language. Although we may agree that our bodily experience and interaction with the environment are fundamental aspects for understanding and meaningful use of language, humans often use it automatically without paying attention to the bodily and sensorimotor bases of meaning.

For instance, the sentence "I've gotten into a mess" presupposes Image Schemas and Embodied Metaphors that give it meaning. The word "mess" is presented as a physical object that one can enter. The verb "gotten into" implies movement through space, indicating that the speaker has transitioned from a state of order to a state of disorder. Additionally, the word "into" can be interpreted as if the "mess" were a container that one can enter. As in the previous case of derivatives, human beings learn to use and understand expressions like this in natural language. That is, we learn to use and respond effectively to this type of expression without necessarily achieving a deep understanding of them that takes into account the bodily and interactive bases of meaning generation.

We have argued so far that human beings can use and understand language in an operational, effective way, without paying attention to the deeper or more subtle aspects of meaning. This suggests that bodily and interactive foundations may not always be a necessary condition for the effective emission of outputs, given certain inputs. In other words, on many occasions, human beings, while still considering themselves intelligent in a functional sense, behave like Searle's Chinese rooms.

Where, then, does the bodily and interactive basis of meaning reside, if not in the consciousness and language comprehension of each speaker? The very structures and patterns of language would be the carriers of these deep layers of meaning. This suggests that, even though models like GPT do not have a physical body, they are still programmed to understand and use bodily metaphors. This is accomplished through the use of natural language processing algorithms that allow it to analyze and interpret the meaning of words and phrases in context. The model is also trained on vast amounts of text data that contain various metaphors, including bodily ones, which also enable it to recognize and understand them. Additionally, the programming includes a knowledge base that contains information about the human body and its functions, which further enhances its ability to comprehend and use bodily metaphors. Overall, despite lacking a physical body, the programming and training enable it to understand and use bodily metaphors in a manner similar to that of humans [47]. These ideas take us directly to the question of the next section.

3.4. The Disembodied Mind?

Perhaps the key question is: can there exist any disembodied mind? We will examine the possibility of a disembodied mind and its relation to Language Models (LLMs). The concept of a disembodied mind challenges traditional views that consider the mind as inseparable from a physical body.

In the previous section, we questioned a possible dividing line between the processes carried out by the human mind and the learning and computation processes of the GPT model. The argument was that, although the generation of meaning in the understanding and use of language may include as a necessary condition the bodily and interactive bases of meaning, this did not imply a radical or essential difference between human cognition and that of the artificial model, to the extent that these bases of meaning are not always present in the language use that humans carry out, even though we consider this use operational or functional, and that in the case of the GPT model, this understanding can be achieved since the bodily and interactive bases of meaning, as long as they are incorporated into the language patterns that the system is capable of learning.

Another argument used is that, due to being a disembodied system, its semantics are reduced to what is known as distributional semantics (a linguistic approach where the meaning of a word is derived from its distributional patterns in large corpora of text, focusing on the contexts in which it appears), unlike an embodied mind, such as the human mind, which is governed by denotative semantics, a semantic approach where meaning arises from the direct relationship between linguistic symbols and real-world entities or experiences. In the case of the GPT model, its ability to use language effectively lies in its ability to associate words to the extent that they often appear together in certain texts. On the other hand, the human ability to attribute meaning and significance to language lies in the extralinguistic experience it has of the inner and outer world.

However, this argument becomes problematic again. Let's go back to the example of philosophy studies. Learning philosophy takes place in an academic world that is highly disembodied [62]. In this environment, the student must learn, through reading and commenting on texts, and assisted by the teacher's knowledge, the meaning of new terms (Dasein, supervenience...), or new technical meanings for terms that they use in their everyday life in a different way (God, substance...). These concepts are learned, most of the time, in a way that is completely disconnected from the student's direct bodily experience of the world outside of academia. Therefore, the process of academic learning can be understood as an attribution of meaning to terms supported by the distributional conception of semantics, since there are no formal syntactic rules or denotative components to attribute meaning to these terms. Since, on many occasions, the learning of sciences is carried out in a highly disembodied way, with a very residual role of experimental and operational aspects, the learning of scientific disciplines could be conceived in a similar way. Finally, this distributive component of semantic attribution to symbols is also carried out in extra-academic environments, where human beings repeat patterns and clichés of their own language without paying attention to their adequacy to the world of experience.

On the other hand, to the extent that elements of human thought and the relationship between language and interaction with the world are inscribed in the language patterns themselves, and in the relationships between terms, it is questionable whether the GPT model learns and uses language in a way that is completely unrelated to the underlying human experience. These arguments question the radical difference between learning and language use carried out by the human mind (embodied mind) and the GPT model (disembodied mind), since LLMs demonstrate impressive language capabilities, and they operate based on preexisting data and algorithms, lacking the subjective experiences and self-awareness associated with human, embodied minds.

We have to take into account, though, that LLM systems are not strictly disembodied, as long as they run on physical devices under specific physical constraints, as well as engineering decisions (kind of processors, software, and computer architecture...). With multimodal AI advances, like Meta's ImageBind, which is a new multimodal model that combines six data types [63], we can affirm that holistic data integration justifies a notion of a virtual embodiment for AI and LLM systems.

4. Discussion

The practical implications of conceptualizing LLMs as disembodied cognitive systems extend across various domains, including education, ethics, and policymaking. This section explores how such systems can revolutionize these fields while acknowledging the challenges inherent in their implementation.

To better illustrate the distinctions and parallels between LLMs and human cognitive processes, **Table 3** provides a comparative overview. It highlights how LLMs generate meaning and engage with language relative to human cognition, emphasizing their limitations and strengths in practical applications:

The practical implications of conceptualizing LLMs as disembodied cognitive systems extend across various domains, including education, ethics, and policymaking. This section explores how such systems can revolutionize these fields while acknowledging the challenges inherent in their implementation. Here are three different fields, as examples of such practical implications:

(1). Education: LLMs offer unprecedented opportunities for personalized learning [64]. They can act as virtual tutors, adapting to individual students' learning styles and providing real-time feedback. For example, LLMs can assist students with essay writing by generating structured outlines, correcting grammatical errors, and even offering guidance on complex topics. Moreover, they can facilitate access to high-quality education in underprivileged regions where skilled educators may be scarce. However, challenges arise in ensuring these systems do not perpetuate biases or produce inaccurate information. For instance, if training data dispro-

Table 3. LLM capabilities vs. human cognitive processes.

Aspect	LLM Capabilities	Human Cognition
Learning	Trained on vast text corpora; learns statistical patterns.	Learns through experience, sensory input, and adaptation.
Reasoning	Limited; relies on pre-trained correlations and cannot reason abstractly.	Abstract reasoning and problem-solving capabilities.
Language Understanding	Processes context and generates coherent text based on input.	Understands language through context, culture, and experience.
Meaning-Making	Relies on relational coherence, not sensory or experiential grounding.	The meaning is derived from embodied and experiential grounding.
Bias and Error	Prone to biases in training data.	Biases are influenced by personal and cultural factors.
Ethics and Judgment	No intrinsic ethical judgment; depends on input and guidance.	Incorporates ethical reasoning and moral judgment.

portionately represents certain cultural norms or academic approaches, it may lead to narrow, biased, or skewed perspectives, misleading learners and reinforcing stereotypes. Such systematic errors or imbalances in AI training datasets can lead to prejudiced or unfair outcomes in model outputs. Addressing these issues requires incorporating diverse and balanced datasets that reflect a wide range of educational content.

(2). Ethics: The ethical dimensions of deploying LLMs are multifaceted, as their training on large datasets risks perpetuating societal biases embedded within those datasets. For example, LLMs may generate outputs that unintentionally reinforce harmful stereotypes related to gender, race, or socioeconomic status[65]. Such biases can influence societal perceptions and decision-making processes in detrimental ways, particularly when LLMs are used in critical fields like recruitment or criminal justice. To mitigate these risks, rigorous dataset curation is essential, ensuring representation from diverse demographics, cultures, and viewpoints. Additionally, algorithmic fairness techniques, such as adversarial debiasing or post-processing approaches, can be employed to identify and correct biased outputs. Transparency is another cornerstone of ethical LLM deployment. Openly sharing information about how models are trained, the sources of their datasets, and the limitations of their performance can foster trust among users and the broader public.

(3). Policymaking: LLMs have the potential to streamline policymaking processes by synthesizing vast amounts of data and generating policy drafts that incorporate a broad spectrum of perspectives. They can assist in comparative policy analysis, enabling policymakers to evaluate differences in legal frameworks across regions and predict the outcomes of proposed legislation. However, reliance on LLMs for policymaking also raises significant concerns[66]. Oversimplification of complex social issues, such as those involving intersectional identities, could lead to policies that inadvertently reinforce systemic inequities. To counteract these risks, LLMs should be used as advisory tools rather than primary decision-makers, with human oversight to ensure that diverse perspectives are adequately considered. Moreover, policymakers must implement ethical guidelines and accountability measures to prevent misuse and promote inclusivity.

Therefore, and addressing societal impacts and ethical concerns, we see that LLMs' influence on societal systems cannot be understated. By amplifying existing societal narratives, they can shape public opinion and reinforce existing inequalities. For example, in media and communication, LLMs could inadvertently prioritize popular but potentially harmful views, diminishing the voices of marginalized groups. Ethical concerns also extend to the potential for misuse in creating deepfake content or automating harmful misinformation campaigns. These risks necessitate strict regulatory frameworks and ethical standards to safeguard against abuse. Collaborative efforts between governments, tech companies, and civil society are essential to developing AI systems that align with shared societal values.

To address biases and promote fairness in LLMs, the following strategies can be adopted:

- ➢ Inclusive Dataset Design: Curate training datasets to ensure the representation of diverse demographics, cultures, and perspectives. Actively seek to include underrepresented groups and contexts to counteract dominant narratives.
- ➢ Algorithmic Auditing: Implement regular audits to identify and mitigate biases in model outputs. Techniques like adversarial debiasing or reinforcement learning from unbiased feedback can help refine model behavior.
- ➢ Human Oversight: Maintain human oversight in critical applications, such as policymaking or education, to ensure outputs are contextually appropriate and ethically sound.
- ➢ Transparent Development: Encourage transparency by documenting the training process, dataset sources, and known limitations. This openness allows external experts to scrutinize and improve LLM deployment.
- ➢ Continuous Feedback Loops: Implement feedback mechanisms where users can report biased or harmful outputs, enabling iterative improvements to the system.

In these contexts, treating LLMs as disembodied systems encourages a focus on their operational capabilities rather than their limitations. By addressing ethical concerns and embracing their complementary role to human cognition, LLMs can be harnessed to enhance productivity, inclusivity, and fairness across diverse fields while safeguarding against societal harm.

Finally, while this study provides valuable insights into coherence-based semantics and the capabilities of Large Language Models (LLMs), several limitations highlight areas for further exploration. Coherence-based semantics, while effective at modeling meaning through statistical patterns, lacks sensory-motor grounding, limiting its ability to process contexts requiring physical or experiential understanding. This disembodied nature restricts LLMs in tasks involving spatial reasoning or emotional nuance tied to human experiences. Additionally, ethical concerns arise due to biases inherent in training data, which can perpetuate stereotypes or generate inequitable outcomes in applications like hiring, education, or policymaking. Despite these challenges, strategies such as improved dataset curation, fairness algorithms, and interdisciplinary collaboration with ethicists and policymakers could help address these issues. Moreover, the divergence between coherence-based approaches and human cognition, which integrates embodied, emotional, and cultural dimensions, raises philosophical questions about their generalizability. Practical challenges such as scalability, computational efficiency, and energy consumption further constrain LLM deployment, especially in resource-limited environments. Future research should focus on hybrid frameworks combining coherence-based and embodied approaches, robust bias mitigation methods, and optimization of LLM architectures to enhance both their cognitive capabilities and societal alignment.

5. Conclusions: A Coherentist Approach to LLM and Meaning

A coherentist perspective can be applied to understand the true semantic properties of language models like ChatGPT, such as LLM. Coherentism emphasizes the interrelationships and coherence of beliefs within a system, and this can be extended to the evaluation of language models' semantic properties [67]. According to Anderson, comprehension involves constructing a coherent mental representation based on linguistic input. Language models like ChatGPT aim to generate responses that are coherent with the input context, using learned patterns and probabilistic associations. On the other hand, we can affirm, following Clark [68], who emphasizing the predictive nature of cognition, LLM is cognitive agents: language models predict and generate responses that are coherent with the context, utilizing the knowledge and patterns learned from training data. By considering the coherence and interrelationships between language models' responses and the input context, a coherentist perspective provides insight into the true semantic properties of LLMs. These models strive to generate coherent and contextually appropriate responses, aligning with the principles of coherence within a belief system.

In this paper, we have presented arguments according to which humans make sense of language based on language references to the world (denotative semantics) and their own bodily experience as a bodily agent (embodied mind). However, we have presented counterarguments to show that in many contexts, humans make sense of language and use it under the parameters of a coherentist framework (distributional semantics), similar to LLMs. Classic studies in analytical philosophy (Quine) and philosophy of science (Kuhn) have already questioned the idea that language acquires meaning from reference relationships between words and things, with the conceptual networks of language itself providing meaning to each of its terms. We have seen how both LLMs and humans draw upon not only symbolic but also linguistic meaning, highlighting the fact that humans, much like LLMs, borrow ideas and embodied experiences from others to generate their own understanding. It suggests that not all semantic values originate from direct experience but are often borrowed. By exploring the role of cultural semantics, conceptual metaphor theory, and linguistic relativity in shaping human cognition and interpretation, the paper challenges the idea of a singular embodied meaning or even the necessity of an embodied nature for the possibility of generating meaningful semantic content.

We have also argued that, while a priori, the difference between the generation of meaning in the human mind can be based on subjective experience as corporeal agents, in many cases, humans use language in an operational way, without any awareness of the bodily and interactive bases of the terms they use. On the other hand, LLMs, despite making sense of language only in coherentist terms, use language patterns that incorporate many elements of meaning generation based on bodily and interactive experiences. Our main conclusions reached are as follows:

- LLMs can be considered Disembodied Minds, with very human-like abilities and competencies in making sense, understanding, and using language, as the dividing line between LLMs and the human mind has turned out to be more blurred than expected than one might initially think considering Searle's and others' criticisms of AI, and the 4E on cognition paradigm.
- LLM's generative skills challenge the notion that artificial intelligence (AI) language models (LLMs) such as GPT operate solely as "stochastic parrots" without true understanding. While acknowledging that LLMs rely on probabilistic algorithms and lack consciousness, we have argued that they can produce contextually relevant and coherent responses, indicating a level of understanding beyond mere repetition.
- LLMs as systems are capable of being nourished by human experience, and their ability to make sense of the world around them based on their corporeality and their sensorimotor interaction, to the extent that this experience remains sedimented in the language, whose patterns are captured by the LLMs from their learning processes.
- LLMs handle billions of texts and parameters in their learning processes and are capable of learning patterns from different human experiences, from different societies and cultures, far exceeding the experience of any human being or community. Therefore, their abilities to understand language are, in this sense, superior to humans. This will make them an increasingly useful and ubiquitous resource in many aspects of our daily lives. Therefore, it is essential to attend to the epistemological and ethical aspects of its use, and develop the necessary skills to learn to complement ourselves with these systems, taking advantage of the complementary virtues of the embodied and disembodied mind.

Of course, there are plenty of pending problems, like ethical or epistemic corollaries. The ethical implications of LLM usage underscore the validity of their results, as well as possible malfunctions in causal understanding at the epistemic level. Researchers strive to address biases and promote fairness, ensuring that LLMs generate responses that align with ethical guidelines. We affirm that LLMs leverage preexisting embodied meanings, statistical learning, and evaluation metrics to generate coherent responses, thereby supporting their efficacy as language processing tools. We question the radical difference between human cognition and LLMs, at least for some reasoning procedures, suggesting that both operate based on preexisting data and algorithms (in t. We discuss the concept of a disembodied mind and the notion of virtual embodiment in AI and LLM systems, highlighting the complexity of understanding intelligence and consciousness in both human and artificial entities. This exploration underscores the complexity of cognition and emphasizes the need for further interdisciplinary research to deepen our understanding of the nature of intelligence across different domains. The algorithmic nature of neurochemical data information in our brains underscores the intricate processes underlying human cognition and behavior. Within the brain, neurotransmitters and other neurochemicals facilitate communication between neurons, forming complex networks that govern various functions such as perception, memory, emotion, and decision-making. These neurochemical signals operate according to specific algorithms, dictating how information is transmitted, processed, and integrated across different regions of the brain. Moreover, the brain's ability to adapt and learn is also governed by algorithmic principles. Neural plasticity, the brain's capacity to reorganize and form new connections in response to experiences, relies on algorithmic mechanisms to adjust synaptic strength and optimize neural circuits. This dynamic process allows the brain to encode information, learn from interactions with the environment, and generate appropriate responses. Understanding the algorithmic nature of neurochemical data in the brain provides valuable insights into neurological disorders, cognitive function, and the development of artificial intelligence systems inspired by the brain's architecture. By unraveling the algorithms that govern brain function, researchers aim to decode the complexities of human cognition and pave the way for innovative approaches to neuroscience and AI. Generative AI can be seen, then, as an operative, successful and disembodied cognitive system that uses embodied data to generate semantic content aligned with human values and knowledge.

Author Contributions

Conceptualization, J.V. and I.R.; methodology, J.V.; formal analysis, J.V.; investigation, J.V. & IR; resources, J.V.; data curation, J.V.; writing—original draft preparation, J.V.; writing—review and editing, J.V.; supervision, J.V.; project administration, J.V.; funding acquisition, JV. I.R. contributed to conceptual debates and provided references. All authors have read and agreed to the published version of the manuscript.

Funding

This work was supported by the ICREA 2019 & PALGAIT (PID2023-148336NB-100) grants awarded to Prof. Dr. Jordi Vallverdú.

Institutional Review Board Statement

Ethical review and approval were waived for this study as it did not require ethical approval.

Informed Consent Statement

Not applicable.

Data Availability Statement

All data is contained in the paper.

Acknowledgments

We would like to acknowledge our children for their patience and understanding while we dedicated time to our research. Their support and encouragement have been invaluable throughout this journey.

Conflict of Interest

No conflict of interest.

References

[1] Conroy, G., 2023. How generative AI could disrupt scientific publishing. Nature. 622, 234.

[2] Prillaman, M., 2024. Is ChatGPT making scientists hyper-productive? The highs and lows of using AI. Nature. 27(8002), 16-17.

[3] Meyer, J.G., Urbanowicz, R.J., Martin, P.C., et al., 2023. ChatGPT and large language models in academia: opportunities and challenges. BioData Mining. 16(1), 20.

[4] Noy, S., Zhang, W., 2023. Experimental evidence on the productivity effects of generative artificial intelligence. Science. 381(6654), 187-192.

[5] Yampolskiy, R.V., Unmasking, A.I., 2024. Ethical dilemmas in autonomous systems. AI Ethics Journal. 12(2), 89-103.

[6] Tangermann, V., 2024. OpenAI CEO says he has no idea what GPT-5 will be capable of. Available from: https://futurism.com/the-byte/gpt-5-openai-ceo (cited 15 November 23).

[7] Liu, Z., Kitouni, O., Nolte, N.S., et al., 2022. Towards understanding grokking: An effective theory of representation learning. Advances in Neural Information Processing Systems. 35, 34651-34663.

[8] Power, A., Burda, Y., Edwards, H., et al., 2022. Grokking: Generalization beyond overfitting on small algorithmic datasets. Available from: https://arxiv.org/abs/2201.02177 (cited 6 Jan 2022).

[9] Romera-Paredes, B., Bisk, Y., Paperno, D., 2024. Evaluating reasoning capabilities in generative models. Journal of AI Research. 58, 102-121.

[10] Battle, R., Gollapudi, T., 2024. The unreasonable effectiveness of eccentric automatic prompts. Available from: https://arxiv.org/abs/2402.10949 (cited 20 Feb 2024).

[11] Minaee, S., Mikolov, T., Nikzad, N., et al., 2024. Large language models: A survey. Available from: https://arxiv.org/abs/2402.06196 (cited 20 Feb 2024).

[12] Kasai, J., Sakaguchi, K., Yamada, Y., et al., 2023. Evaluating GPT-4 and ChatGPT on Japanese medical licensing examinations. Available from: https://arxiv.org/abs/2303.18027 (cited 31 Mar 2023).

[13] Nori, H., King, N., McKinney, S.M., et al., 2023. Capabilities of GPT-4 on medical challenge problems. Available from: https://arxiv.org/abs/2303.13375 (cited 12 Apr 2023).

[14] Oh, N., Choi, G.S., Lee, W.Y., 2023. ChatGPT goes to the operating room: evaluating GPT-4 performance and its potential in surgical education and training. Annals of Surgical Treatment and Research. 104(5), 269.

[15] Chen, L., Zaharia, M., Zou, J., 2023. How is ChatGPT's behavior changing over time? Available from: https://arxiv.org/abs/2307.09009 (cited 31 Oct 2023).

[16] Hagendorff, T., Fabi, S., Kosinski, M., 2023. Humanlike intuitive behavior and reasoning biases emerged in large language models but disappeared in ChatGPT. Nature Computational Science. 3(10), 833-838.

[17] Ortu, F., Jin, Z., Doimo, D., et al., 2024. Competition of mechanisms: tracing how language models handle facts and counterfactuals. Available from: https://arxiv.org/abs/2402.11655 (cited 6 Jun 2024).

[18] Wei, J., Wang, X., Schuurmans, D., et al., 2022. Chain-of-thought prompting elicits reasoning in large language models. Advances in Neural Information Processing Systems. 35, 24824-24837.

[19] Prystawski, B., Thibodeau, P., Goodman, N., 2022. Psychologically-informed chain-of-thought prompts for metaphor understanding in large language models. Available from: https://arxiv.org/abs/2209.08141 (cited 19 May 2023).

[20] Qi, J., Xu, Z., Shen, Y., et al., 2023. The art of Socratic questioning: Recursive thinking with large language models. Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing; December 6–10, 2023; New York, United States. pp. 4177–4199.

[21] Bender, E.M., Gebru, T., McMillan-Major, A., et al., 2021. On the dangers of stochastic parrots: Can language models be too big? Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency; March 3-10, 2021; Online. pp. 610–623.

[22] Genkina, D., 2024. AI prompt engineering is dead: Long live AI prompt engineering. IEEE Spectrum. Available from: https://spectrum-ieee-org.cdn.ampproject.org/c/s/spectrum.ieee.org/amp/prompt-engineering-is-dead-2667410624

[23] Boden, M.A., 2014. 4 GOFAI. In: Frankish K., Ramsey, W.M. (eds.). The Cambridge Handbook of Artificial Intelligence. Cambridge University Press: Cambridge, United Kingdom. p. 89.

[24] Clark, A., 1997. Being There: Putting Brain, Body, and World Together Again. MIT Press: Massachusetts, United States. p. 295.

[25] Thompson, E., Varela, F.J., 2001. Radical embodiment: Neural dynamics and consciousness. Trends in Cognitive Sciences. 5(10), 418-425.

[26] Varela, F.J., Thompson, E., Rosch, E., 1991. The Embodied Mind: Cognitive Science and Human Experience. MIT Press: Massachusetts, United States.

[27] Géron, A., 2022. Hands-on Machine Learning with Scikit-Learn, Keras, and TensorFlow. O'Reilly Media, Inc: Sebastopol, United States. p. 819.

[28] Tegmark, M., 2018. Life 3.0: Being Human in the Age of Artificial Intelligence. Vintage: New York, United States. p. 400.

[29] Marcus, G., 2018. Deep learning: A critical appraisal. Available from: https://arxiv.org/abs/1801.00631 (cited 2 Jan 2018).

[30] Froese, T., Ziemke, T., 2009. Enactive artificial intelligence: Investigating the systemic organization of life and mind. Artificial intelligence. 173(3-4), 466-500.

[31] Shanahan, M., Crosby, M., Beyret, B., et al., 2020. Artificial intelligence and the common sense of animals. Trends in Cognitive Sciences. 24(11), 862-872.

[32] Fjelland, R., 2020. Why general artificial intelligence will not be realized. Humanities and Social Sciences Communications. 7(1), 1-9.

[33] Kıcıman, E., Ness, R., Sharma, A., et al., 2023. Causal reasoning and large language models: Opening a new frontier for causality. Available from: https://arxiv.org/abs/2305.00050 (cited 20 Aug 2024).

[34] Brown, T., Mann, B., Ryder, N., et al., 2020. Language models are few-shot learners. Advances in Neural Information Processing Systems. 33, 1877-1890.

[35] Prystawski, B., Thibodeau, P., Goodman, N., 2022. Psychologically-informed chain-of-thought prompts for metaphor understanding in large language models. Available from: https://arxiv.org/abs/2209.08141 (cited 19 May 2023).

[36] Chakrabarty, T., Saakyan, A., Winn, O., et al., 2023. I spy a metaphor: Large language models and diffusion models co-create visual metaphors. Available from: https://arxiv.org/abs/2305.14724 (cited 14 Jul 2023).

[37] Bisk, Y., Holtzman, A., Thomason, J., et al., 2020. Experience grounds language. Available from: https://arxiv.org/abs/2004.10151 (cited 2 Nov 2020).

[38] Wang, S., Petridis, S., Kwon, T., et al., 2023. PopBlends: Strategies for conceptual blending with large language models. Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems; April 19, 2023; Hamburg, Germany. pp. 1–19.

[39] Bubeck, S., Chandrasekaran, V., Eldan, R., et al., 2023. Sparks of artificial general intelligence: Early experiments with GPT-4. Available from: https://arxiv.org/abs/2303.12712 (cited 13 Apr 2023).
[40] Gupta, T., Gong, W., Ma, C., et al., 2024. The essential role of causality in foundation world models for embodied AI. Available from: https://arxiv.org/abs/2402.06665 (cited 29 Apr 2024).
[41] CBS News., 2023. The AI revolution: Google's artificial intelligence developers on what's next in the field. Available from: https://www.cbsnews.com/news/ai-advancements-google-artificial-intelligence-future-60-minutes-2023-04-16/ (cited 16 April 2023).
[42] Searle, J.R., 1980. Minds, brains, and programs. Behavioral and Brain Sciences. 3(3), 417-424.
[43] Harnad, S., 1990. The symbol grounding problem. Physica D: Nonlinear Phenomena. 42(1-3), 335-346.
[44] Block, N., 1980. Troubles with functionalism. Minnesota Studies in the Philosophy of Science. 9, 261-325.
[45] Bender, E.M., Koller, A., 2020. Climbing towards NLU: On meaning, form, and understanding in the age of data. Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics; July 6–July 8, 2020. Washington, United States. pp. 5185–5198.
[46] Durt, C., Froese, T., Fuchs, T., 2022. Against AI understanding and sentience: Large language models, meaning, and the patterns of human language use. In: Voeneky, S., Kellmeyer, P., Mueller, O., et al. (eds.). The Cambridge Handbook of Responsible Artificial Intelligence. Cambridge University Press: Cambridge, United Kingdom. pp. 67-82.
[47] Durt, C., 2022. Artificial Intelligence and Its Integration into the Human Lifeworld. In: Voeneky, S., Kellmeyer, P., Mueller, O., Burgard, W. (eds.). The Cambridge Handbook of Responsible Artificial Intelligence. Cambridge University Press: Cambridge, United Kingdom. p. 526.
[48] GoogleTalks, 2015. Consciousness in Artificial Intelligence | John Searle | Talks at Google. Available from: https://www.youtube.com/watch?v=rHKwIYsPXLg&t=17s (cited 3 December 2015).
[49] Basnight-Brown, D.M., Altarriba, J., 2018. The influence of emotion and culture on language representation and processing. In: Faucher, C. (ed.). Advances in Culturally-Aware Intelligent Systems and in Cross-Cultural Psychological Studies. Springer International Publishing: Cham, Switzerland. pp. 415-432.
[50] Gumperz, J.J., Levinson, S.C., 1996. Rethinking Linguistic Relativity. Cambridge University Press: Cambridge, United Kingdom. p. 500.
[51] Nisbett, R.E., Peng, K., Choi, I., et al., 2001. Culture and systems of thought: holistic versus analytic cognition. Psychological Review. 108(2), 291-310.
[52] Lakoff, G., Johnson, M., 1980. Metaphors We Live by. University of Chicago Press: Chicago, United States. p. 292.
[53] Wilson, R.A., Foglia, L., 2017. Embodied Cognition. Available from: https://plato.stanford.edu/entries/embodied-cognition (cited 25 January 2021).
[54] Boroditsky, L., 2001. Does language shape thought? English and Mandarin speakers' conceptions of time. Journal of Cognitive Psychology. 43(1), 1-22.
[55] Vasantkumar, C., 2022. Not "multiple ontologies" but ontic capaciousness: Radical alterity after the ontological turn. HAU: Journal of Ethnographic Theory. 12(3), 819-835.
[56] Gan, C., Mori, T., 2023. Sensitivity and robustness of large language models to prompts in Japanese. Available from: https://arxiv.org/abs/2305.08714 (cited 8 Jun 2023).
[57] Chakrabarty, T., Saakyan, A., Ghosh, D., et al., 2022. Flute: Figurative language understanding through textual explanations. Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing; December 7-December 11, 2022; Abu Dhabi, the United Arab Emirates. pp. 7139–7159.
[58] Tversky, A., Kahneman, D., 1979. Prospect theory: An analysis of decision under risk. Econometrica. 47(2), 263-292.
[59] Vallverdú, J., 2019. Blended Cognition: The Robotic Challenge. Springer International Publishing: Cham, Switzerland. p. 316.
[60] Piaget, J., 1977. The Development of Thought: Equilibration of Cognitive Structures. Viking Press: New York, United States. p. 213.
[61] Lakoff, G., Núñez, R., 2000. Where mathematics comes from: How the embodied mind brings mathematics into being. Basic Books: New York, United States. p. 511.
[62] Macrine, S.L., Fugate, J.M., 2022. Movement matters: How embodied cognition informs teaching and learning. MIT Press: Massachusetts, United States. p. 300.
[63] Girdhar, R., El-Nouby, A., Liu, Z., et al., 2023. ImageBind: One Embedding Space to Bind Them All. Available from: https://arxiv.org/abs/2305.05665 (cited 31 May 2023).
[64] Wen, Q., Liang, J., Sierra, C., et al., 2024. AI for education (AI4EDU): Advancing personalized education with LLM and adaptive learning. Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining; August 25-August 29, 2024; Barcelona, Spain. pp. 6743–6744.
[65] Sakib, S.K., Das, A.B., 2024. Challenging fairness: A comprehensive exploration of bias in LLM-based recommendations. Proceedings of the 2024 IEEE International Conference on Big Data (BigData); Dec 15-Dec 18, 2024; Washington DC, USA. pp. 1585–1592.
[66] Gutiérrez, J.D., 2024. Critical appraisal of large lan-

guage models in judicial decision-making. In: Paul, R., Carmel, E., Cobbe, J. (eds.). Handbook on Public Policy and Artificial Intelligence. Edward Elgar Publishing: Cheltenham, United Kingdom. pp. 323-338.
[67] Anderson, C., 2021. The role of coherence in language comprehension. Trends in Cognitive Sciences. 25(2), 123-134.
[68] Clark, A., 2013. Whatever next? Predictive brains, situated agents, and the future of cognitive science. Behavioral and Brain Sciences. 36(3), 181-204.