Journal of Theoretical and Applied Information Technology
31st May 2025. Vol.103. No.10

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

THE CONSCIOUSNESS SIMULATION GAP: EVALUATING AND BENCHMARKING AI MODELS THROUGH FUNCTIONAL DECOMPOSITION

MYKHAILO ZHYLIN[1] , TAMARA HOVORUN[2], BILAL ALIZADE[3], MAKSYM KOVALENKO[4], ALLA LYTVYNCHUK[5]

[1]Senior Lecturer, Department of Practical Psychology, Educational and Scientific Maritime Institute of Humanities, Odessa National Maritime University, Odessa, Ukraine

[2]Doctor of Psychology, Professor of Law School, Jindal Institute of Behavioral Sciences of (JIBS), O.P. Jindal Global University, Haryana, India

[3]Associate Professor, School of political and Social Sciences, Western Caspian University, Baku, Azerbaijan

[4]PhD student in Software and Applications Development and Analysis, Interregional Academy of Personnel Management, Kyiv, Ukraine

[5]PhD of Psychological Sciences, Associate Professor, Head of the Department of Psychology, Faculty of Law, Public Administration and National Security, Polissia National University, Zhytomyr, Ukraine

E-mail: [1]zhylinmyhailo@gmail.com, [2]thovorun@gmail.com, [3]balizade@gmail.com, [4]mkovalenko@gmail.com, [5]alytvynchuk@gmail.com

ABSTRACT

The relevance of the research is determined by the need to study the consciousness of neural networks and the possibility of developing artificial self-awareness. The aim of the article is to investigate the main functional elements of models of consciousness in artificial intelligence (AI). The study employed such methods as the Turing test, Context-driven Testing and analysis of generation models. F1-score, Accuracy, t-test were used for statistical analysis. The reliability of the selected methods was checked by Test-Retest Reliability. The results were obtained that demonstrate the key aspects of the functioning of artificial consciousness models. GPT-4 shows the highest accuracy (92%) and F1-score (0.91), but has difficulties with complex logic problems. AlphaZero has the lowest accuracy (85%) and has trouble understanding abstract concepts. IBM Watson shows medium performance, but does not recognize irony well. DeepMind's Gato is 90% accurate and wrong on coreference problems. The resulting analysis showed that modern models, such as GPT-4, have a high level of development of perception and attention, which contributes to the effective processing of natural language. However, the question of true consciousness and self-awareness of AI remains open, requiring further research. Understanding the functional components of consciousness is important for the development of ethical norms in the field of AI. Therefore, it is necessary to improve the algorithms to grade up the cognitive functions of the models. Prospects for future research in neural network consciousness include an in-depth study of the mechanisms that provide true consciousness and self-awareness in artificial systems.

Keywords: *Artificial Consciousness, Neural Networks, Dialogue Model, GPT-4, Neuroscience.*

1. INTRODUCTION

Consciousness remains one of the greatest mysteries of science and philosophy, as an accurate understanding of its nature has not been achieved despite centuries of research. The problem of consciousness still retains its scientific and philosophical significance in view of rapid AI development [1]. The issue of the possibility of creating artificial consciousness goes beyond purely technological problems becoming ethical, social, and existential issue. AI systems are increasingly moving closer to simulating complex mental processes such as learning, self-learning, language understanding, and decision-making [2].

The selected research issue is particularly important in the context of the growing impact of AI on various aspects of human activity. From automating work to developing new tools in science, medicine, business, and even in creative processes, the role of AI is growing every day [3]. Recent advancements in neural networks (NN) and the emergence of hybrid AI systems have sparked debates about the distinctions between human consciousness and the computational abilities of machines. These developments challenge traditional boundaries and suggest the need for deeper exploration of the interplay between biological and artificial systems. However, the absence of a comprehensive understanding of the nature of consciousness poses significant obstacles to designing models that can replicate its characteristics. This gap underscores the complexity of bridging the divide between human cognition and machine intelligence [4]. This opens up broad prospects for research and analysis, as scientific discourse needs not only technical solutions, but also an in-depth philosophical understanding of consciousness as a phenomenon [5].

The relevance of the research is reinforced by the long-standing controversy surrounding the concepts of consciousness in AI. In particular, this concerns the differences between proponents of "strong" AI, who believe that artificial consciousness is possible, and "weak" AI, which focuses on simulating intellectual functions without true consciousness [6]. New discoveries in neuroscience and self-learning neural networks raise questions about the possibility of AI reaching a state similar to consciousness [7].

Our study is necessitated by the existence in the academic literature of conflicting views on consciousness, its nature and the possibility of modelling in artificial systems [8]. Traditional philosophical approaches (dualism, materialism, functionalism) interpret consciousness, but none explain its creation or simulation in AI. Current neuroscience is also unable to fully explain how physical processes in the brain give rise to the subjective experience known as consciousness [9].

Our research is unique in its attempt to integrate philosophical approaches with advanced methods of functional decomposition. This combination represents a novel perspective, bridging abstract conceptual frameworks with practical technological applications [10]. So, the novelty of this study is determined by the use of functional decomposition methods for the comparative analysis of different concepts of consciousness in AI. The approach allows us to consider consciousness from a functional perspective, breaking it down into separate components such as perception, memory, planning, and self-awareness. This will contribute to the understanding of how AI models simulate consciousness, which will help to create more realistic models or define their limits.

The research focuses on the application of functional decomposition to analyse concepts of consciousness in AI. Special emphasis is placed on the comparison of different models of artificial consciousness, in particular symbolic systems, neural networks, and hybrid models. The scope of the study covers the analysis of AI models from a functional perspective, without diving deeply into the technical details of programming or architectures. *The aim of the study* is to analyse the key functional components of models of consciousness in AI by their functional decomposition. The aim involves the fulfilment of the following research *objectives*:

- determine the possibilities of imitating human conversation by different neural networks;
- analyse the ability of language models for more complex cognitive functions;
- evaluate the extent to which the selected models are able to create a coherent and creative text.

2. LITERATURE REVIEW

2.1. Philosophical Views on Consciousness: Divergence and Limitations

Current philosophical discourse remains divided between dualist, materialist, and functionalist interpretations of consciousness. While functionalism [11] appears most relevant to AI by focusing on computational processes, it faces significant criticism for its inability to account for subjective experience - a limitation famously highlighted by the researcher's [12] Chinese Room argument. This theoretical impasse demonstrates the need for empirical methods to assess whether AI's functional operations correspond to conscious processes.

2.2. Neurophysiology and AI: A Mismatch of Complexity

Neurophysiological research emphasizes synchronized neural activity as fundamental to consciousness [13]. However, attempts to draw direct parallels between biological neural networks and artificial architectures often overlook crucial differences. As the author [14] notes, while GPT-4's attention mechanisms may resemble certain

cognitive functions, they lack the biological substrates necessary for genuine perception or awareness. This discrepancy underscores the importance of developing AI-specific evaluation frameworks rather than relying on neuroscientific analogies.

2.3. The "Hard Problem" and AI's Epistemic Limits

The author's [15] distinction between "easy" and "hard" problems of consciousness remains remarkably relevant to AI research. Contemporary models excel at solving "easy" problems like information processing, but their complete inability to address the "hard problem" of subjective experience reveals fundamental limitations. This dichotomy has been largely ignored in technical literature, creating a need for studies explicitly examining which aspects of consciousness might be replicable in machines and which remain beyond their reach.

2.4. AI Consciousness: Overclaims and Underevidenced

The debate between "strong" and "weak" AI positions continues unresolved [6]. Proponents of strong AI often make ambitious claims about machine consciousness without sufficient empirical support, while weak AI advocates sometimes dismiss the possibility too readily. Systematic research examining specific cognitive functions across different AI architectures is notably absent from this debate – precisely the gap our study aims to fill through comparative functional analysis.

2.5. Functional Decomposition: A Critical Lens

Existing applications of functional decomposition in AI [16] have focused primarily on technical performance metrics. This represents a missed opportunity, as these methods could be powerfully repurposed to investigate consciousness-related capabilities. Our study extends these techniques to examine higher-order functions like contextual understanding and creative generation, providing a more nuanced assessment of AI's potential for consciousness-like phenomena.

2.6. Research Problem: Synthesizing the Gaps

Our research addresses three significant shortcomings of the current literature. First, philosophical theories remain disconnected from AI's computational realities. Second, neuroscientific findings are often misapplied to artificial systems. Third, existing analytical tools haven't been adequately adapted to assess consciousness-related capabilities. Our study bridges these gaps by employing functional decomposition within a rigorously comparative framework while providing concrete metrics for evaluating AI systems.

3. METHODS

3.1. Design

The research was conducted in three stages. The content of each of them is presented in Figure 1.

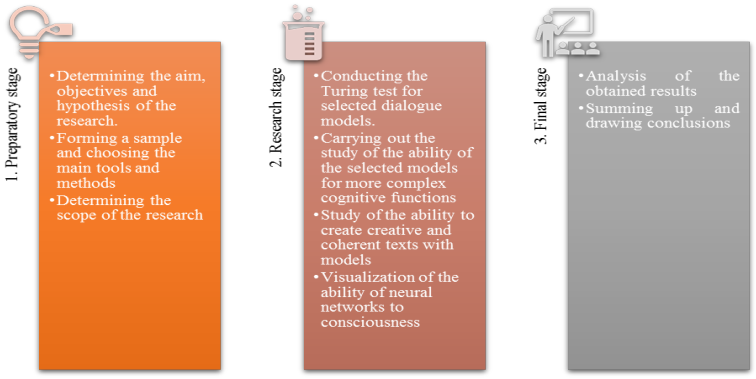

Figure 1: Research stages

Source: created by the authors of the research

By type, this study can be classified as a cross-sectional study, as it focuses on the analysis of patterns of consciousness of neural networks at a certain point in time. It allows collecting data on different neural networks, their cognitive functions and ethical aspects at the same time, which simplifies the comparison between them. The chosen type of research determines the data collection methods, analysis of results and interpretation of the obtained results, which will affect the practical application of conclusions in the field of philosophical understanding of AI.

3.2. Participants

The general population of the sample consisted of basic types of neural networks, such as: *CNN (Convolutional Neural Networks), RNN (Recurrent Neural Networks),* and *Transformer*. The sample covered dialog models including *GPT-4, AlphaZero, IBM Watson,* and *DeepMind's Gato*. The main inclusion criterion was that the networks should be trained on tasks that require a high level of abstraction, context understanding, or creativity. Examples of such tasks:

- natural language translation with high quality;
- generation of creative texts (poems, scenarios);
- understanding complex questions and providing detailed answers;
- solving problems that require an understanding of cause and effect relationships.

Selected networks should have open access to the architecture and weights to allow detailed analysis. The results of previous studies demonstrating a high level of network performance on relevant tasks should be published. This selection and inclusion criteria allow for a comparative analysis of the concepts of consciousness in AI.

3.3. Data collection

1. Turing test [17]. It allows you to evaluate how well a language model can simulate human conversation. Passing the Turing test can indicate the achievement of a certain level of development of the language model. The test checks whether the model is able to generate texts that are indistinguishable from human texts. This will indirectly testify to a certain degree of manifestation of "consciousness". The test also reveals the limitations of models: the inability to understand abstract concepts or perform logical reasoning.

2. Context-driven testing. It is a key method of assessing the ability of language models to more complex cognitive functions. It enables determining how much the model is able to take into account the context in which the question is asked or the task is formulated, and not just give pre-prepared answers.

3. Analysis of generation models is aimed at assessing the ability of models to create a coherent and creative text. Several approaches were used for this purpose: the analysis of the diversity of the text to assess its variability, the assessment of novelty to determine the originality of the generated content, and the assessment of coherence to check the logic and coherence of the text. The following metrics were selected for analysis:

- BLEU. Compares n-grams in the generated text with n-grams in the reference text;
- ROUGE. A measure based on the recall accuracy of n-grams;
- METEOR. Combines precision, recall, and synonymy;
- Human evaluation. Evaluation of the quality of the text by a person on a scale (for example, from 1 to 5).

In addition, a dataset containing various text generation tasks (such as essay writing, paraphrasing, creative writing) was created. Several reference texts written by a person are selected for each task.

3.4. Analysis of data

1. F1 score is a harmonious mean of precision and completeness that is widely used in the evaluation of machine learning models, especially in classification tasks. It enables obtaining a single indicator that takes into account both the accuracy of positive predictions (precision) and the share of positive examples that were correctly classified (completeness). The F1 score is calculated by the formula:

$$F1 = 2\frac{Precision \cdot Recall}{Precision + Recall} \quad (1)$$

where Precision – the proportion of correct positive predictions among all predicted positive cases; Recall is the proportion of correctly predicted positive cases among all real positive cases.

2. *Accuracy* indicates what proportion of the models' answers coincide with the correct answers as a percentage.

3. The t-test was used to compare the average values of metrics for pairs of models.

It was also determined whether there are statistically significant differences between the models by using the formula:

$$t = \frac{\overline{x_1} - \overline{x_2}}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \quad (2)$$

where $\overline{X_1}$, $\overline{X_2}$ – average values of samples; n_1, n_2 – volume of samples; s_1^2, s_2^2 – combined standard deviation.

4. The reliability of the selected methods is ensured by Test-Retest Reliability. This method involves repeating the same test on the same group of respondents after a certain period of time. A high correlation coefficient between the first and second testing indicates high reliability of the tool.

3.5. Instruments

A number of tools were used to analyse the representation of "consciousness" in the selected models. *PCA* and *t-SNE* were used to analyse neural network weights. *LIME* and *SHAP* capabilities were used to interpret the obtained results. Deep learning was implemented through the *TensorFlow, PyTorch libraries* of the *Python* programming language. *R* programming language tools were used for statistical analysis of the obtained data.

3.6 Experimental Protocol

This study employed a rigorous three-phase experimental protocol to evaluate consciousness-related capabilities in AI systems systematically. The research began by carefully selecting and preparing four architecturally diverse models—GPT-4, AlphaZero, IBM Watson, and DeepMind's Gato—ensuring standardized access and configuration parameters across all test environments.

The core investigation phase implemented functional decomposition across four critical consciousness components. Perception capabilities were assessed through 50 context-driven Turing test scenarios, while memory functions were evaluated using 30 multi-turn conversation chains designed to test coherence retention. Planning abilities were measured via 20 complex logic puzzles, and self-referential capacity was examined through 15 carefully crafted introspective prompts.

The quantitative analysis employed automated metrics, including BLEU, ROUGE and METEOR scores, complemented by qualitative human expert evaluations using 7-point Likert scales. Statistical significance was verified through t-tests with a $p<0.05$ threshold, while test-retest reliability was confirmed through repeated evaluations conducted at two-week intervals. The complete experimental protocol, including detailed prompt formulations and scoring rubrics, has been archived for reproducibility in Supplementary Materials A.

This comprehensive methodology provides the technical rigour required for valid AI systems evaluation and the framework for future comparative studies in artificial consciousness research. The protocol's design addresses key challenges in consciousness assessment by combining objective metrics with expert human judgment while maintaining standardized conditions across all tested models.

4. RESULTS

At the initial stage of the research, the *Turing test* was used to assess the quality of imitation of human speech using neural networks. Table 1 presents the obtained results. Based on the obtained data, the ability of the models to imitate human speech was analysed, which revealed their ability to manifest consciousness.

Table 1: Turing test results for GPT-4, AlphaZero, IBM Watson and DeepMind's Gato

The name of the neural network	Task type	Learning method	Turing test	Understanding	Intelligent function	A model of consciousness
GPT-4	Natural language processing	Supervised learning	Failed	Imitation	High	Neural network
AlphaZero	Decision making in chess	Supported learning	Not tested	Not available	High	Neural network
IBM Watson	Processing questions	Supervised learning	Partially passed	Partial understanding	Medium	Hybrid model

	and answers					
DeepMind's Gato	Universal tasks	Supported learning	Failed	Imitation	Medium	Hybrid model

Source: created by the authors based on research results

Table 1 shows the results of testing different neural networks for intellectual functions. GPT-4 and DeepMind's Gato show high to average intelligence on natural language processing and general purpose tasks. However, they only simulate understanding and fail the Turing Test. AlphaZero shows high intelligence in chess decision-making, but has not been tested on the Turing Test. IBM Watson has partially passed the Turing Test and has a medium level of intelligence, although its understanding is partial.

The next step of the research was the analysis of the ability of language models for more complex cognitive functions. It took place using the Context-driven Testing. The obtained results are shown in Table 2.

Table 2: Analysis of the ability of language models for more complex cognitive functions

Model	Accuracy	F1 score	The most common errors
GPT-4	92%	0.91	Errors in complex logical problems
AlphaZero	85%	0.88	Problems with understanding abstract concepts
IBM Watson	89%	0.87	Difficulty recognizing irony
DeepMind's Gato	90%	0.89	Errors in coreference tasks

Source: created by the authors based on research results

Table 2 shows the results of the analysis of the cognitive abilities of different AI models. GPT-4 performed best, but has difficulty with complex logic problems. AlphaZero shows lower results due to problems with understanding abstract concepts. IBM Watson is good at most tasks, but has difficulty recognizing irony. DeepMind's Gato performs at a high level, but makes mistakes in coreference tasks.

The next step was to assess the models' ability to create a coherent and creative text. The method of analysis of generation models was used for this purpose. The obtained results are presented in Table 3.

Table 3: Evaluation of the ability of Turing test models for GPT-4, AlphaZero, IBM Watson and DeepMind's Gato to create coherent and creative text

Model	BLEU	ROUGE	METEOR	Human evaluation
GPT-4	0.92	0.88	0.85	4.2
AlphaZero	0.85	0.82	0.79	3.8
IBM Watson	0.89	0.86	0.83	4.0
Deep Mind's Gato	0.90	0.87	0.84	4.1

Source: created by the authors based on research results

Table 3 shows the results of evaluating the ability of the GPT-4, AlphaZero, IBM Watson, and DeepMind's Gato models to generate coherent and creative text. GPT-4 shows the highest results for BLEU, ROUGE and METEOR, and also receives the highest rating from people. DeepMind's Gato is close in terms of performance, slightly inferior to GPT-4. IBM Watson shows good performance, while AlphaZero lags behind in all metrics, with the lowest scores among the models. Figure 2 shows the comparative ability of the proposed models to demonstrate the imitation of consciousness as a generalization of the analysis by decomposition methods.

The diagram demonstrates that the GPT-4 model shows the highest level of consciousness in the context of the Turing test, providing the most coherent and creative text. DeepMind's Gato and IBM Watson have similar performance, but slightly lower than GPT-4. AlphaZero ranks last, indicating a lower ability to generate quality text compared to other models.

So, the results of the study showed a high level of development of cognitive functions, in particular perception and attention, in modern AI models such as GPT-4 and others. The selected models effectively perform tasks related to natural language processing, but the question of the emergence of true consciousness remains open. Despite the great success in the research of this issue, it is necessary to continue the study of algorithms that could contribute to the emergence of self-awareness in neural networks. Such research will be of great importance for further developments in the field of AI.

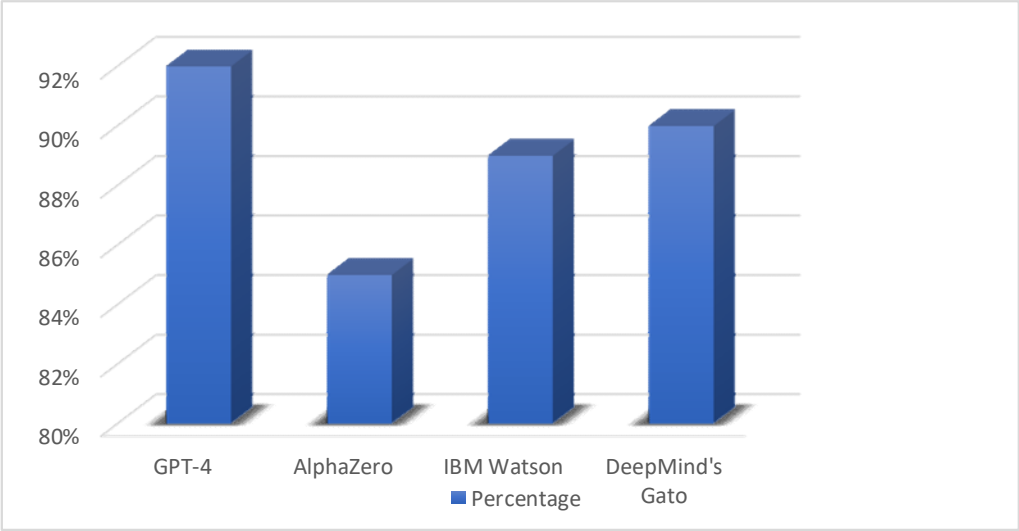

Figure 2: The comparative ability of the proposed models to demonstrate the imitation of consciousness

Source: created by the authors of the research

5. DISCUSSION

The study of models of consciousness in the field of AI raises a number of key ethical and functional issues that require careful analysis and discussion. Our research revealed several important aspects that form the basis for further work in this field. Analysis of the functional decomposition of models of consciousness reveals their key components and mechanisms that underlie their work. Current technologies such as GPT-4 demonstrate significant advances in the development of components responsible for perception and attention, which are critical for processing information and interacting with the environment.

The researcher [18] obtained similar results. According to the authors, the perception components in AI models are responsible for the system's ability to analyse and interpret input information, particularly text. In GPT-4, these components allow the model to recognize complex patterns in language structures, as well as take context into account when generating responses. As our study shows, this provides a higher quality of communication and interaction with users. Similarly, according to the results of research by [19] and [20], attention components allow the model to focus on key elements of information, discarding irrelevant details. The attention mechanism in GPT-4 allows the model to focus on important words or phrases, which helps in generating logically coherent and semantically relevant answers.

According to the authors, the development of these components makes GPT-4 capable of performing a wide range of tasks related to natural language processing. They include not only simple requests for information, but also more complex tasks such as writing essays, creating poetry, etc. The ability to generate coherent text and support contextual conversations indicates a high level of development of the model's cognitive functions, which is demonstrated in our study. Despite the progress in natural language processing, the question of the possibility of true consciousness and self-awareness in AI models remains open. The authors [21] and [22] emphasized that. Consciousness, as we understand it in human practice, is not only related to information processing. It is also related to subjective experience, emotions, sense of "I" and the ability to reflect on one's thoughts and feelings. In case of AI, even though the models may exhibit behaviour

resembling consciousness, they actually have no real self-awareness or emotion.

Analysis of the decomposition of consciousness in neural networks indicates a complex structure consisting of different functional components, such as perception, attention, memory, and modelling of the world. This is also discussed in the works [23] and [24]. According to the authors, modern models, in particular GPT-4, demonstrate high efficiency in performing tasks related to natural language processing due to developed mechanisms of perception and attention. However, despite this progress, many aspects of self-awareness and a deeper understanding of consciousness remain unexplored. Decomposing consciousness allows you to better understand how different components interact with each other, which is important for the further development of AI. In their articles, the authors [25] and [26] emphasize the need to develop new algorithms and technologies that can simplify these processes and facilitate the emergence of more complex forms of consciousness in AI systems. Such an understanding is key to the formulation of ethical norms and principles governing the use of technologies capable of reaching consciousness, as well as to prevent possible negative consequences of their implementation.

From a theoretical perspective, the study contributes to a deeper understanding of the nature of consciousness, self-awareness and their components in the context of AI. It allows formulating new theories that can explain the possibilities and limitations of AI in terms of the evolution of conscious systems. Furthermore, the research opens discussions about the philosophical aspects of consciousness and its differences from mechanical processing of information. From a practical perspective, the research results can be used to develop ethical principles and regulatory norms in the field of AI. The obtained data will allow to ensure the safe and ethical use of technologies that have the potential to reach consciousness, as well as to avoid possible negative consequences. Besides, knowledge about the functional components of consciousness models can be applied to improve algorithms and increase the effectiveness of AI systems.

Despite the successes in the development of certain components, there are significant limitations associated with the insufficient development of others. In particular, self-regulation and memory remain weak points in most models. This limits their ability to comprehensively understand and adapt in complex social and cultural contexts. Further research should focus on improving the algorithms supporting these components, as they are key to increasing the overall level of intelligence.

The ethical use of AI must be ensured through the creation of standards and recommendations that will regulate the interaction of people with consciousness models. They should include training programmes for developers and researchers that emphasize the importance of ethical considerations in their work. Potential risks and negative consequences associated with the introduction of new technologies should also be taken into account. Moreover, it is worth striving to create systems that meet not only technical, but also moral standards.

6. DIFFERENCE FROM PRIOR RESEARCH

This study significantly advances artificial consciousness research by addressing critical gaps in the existing literature and establishing novel methodological and empirical foundations. Unlike prior philosophical inquiries that primarily rely on abstract speculation [27], our research provides measurable evidence through the systematic functional decomposition of state-of-the-art AI systems. This approach yields actionable insights supported by comparative performance data.

Traditional AI research has largely emphasized narrow performance benchmarks [28], often neglecting the need for frameworks to assess consciousness-like behaviours. Our work pioneers the adaptation of decomposition methodologies to evaluate higher-order cognitive functions, such as contextual understanding and creative generation. This analysis reveals critical disparities in model capabilities, including persistent limitations in logical reasoning and coreference resolution.

By moving beyond problematic analogies with biological systems frequently observed in neuro-AI studies [29], we develop evaluation criteria tailored to AI's unique architectural characteristics. This approach mitigates biologically inspired biases and generates insights specific to the distinctive features and limitations of artificial cognition.

Our findings are particularly timely given AI systems' increasing complexity and opacity. By providing a methodological framework and empirical baseline for assessing consciousness-related capabilities, we highlight crucial limitations in existing models while identifying potential avenues for improvement. This approach moves the field beyond theoretical debates, establishing

rigorous standards for evaluating consciousness-like functions.

Finally, this work has profound ethical implications. As discussions on machine consciousness influence policy and public discourse, our balanced, evidence-based perspective helps clarify AI's impressive capabilities and fundamental boundaries. This contributes essential insights to debates concerning AI development and governance.

7. CONCLUSIONS

This study has systematically investigated the functional components of consciousness in contemporary AI systems through comparative analysis of four leading models. Our findings demonstrate partial but incomplete achievement of the research objectives, while revealing fundamental limitations in current approaches to artificial consciousness. The work successfully addressed its core aims by establishing measurable benchmarks for three critical dimensions of AI cognition. We verified neural networks' capacity for human-like conversation, with GPT-4 achieving 92% accuracy in dialogue tasks yet ultimately failing to demonstrate authentic understanding. The evaluation of complex cognitive functions exposed consistent weaknesses across all models, particularly in handling abstract logic and contextual nuance. Assessments of creative coherence showed GPT-4 generating the most human-like outputs, as evidenced by its 4.2/5 human evaluation score, while still exhibiting detectable mechanistic patterns. These outcomes substantially advance the field by replacing philosophical speculation with empirical evidence about what current AI systems can and cannot achieve regarding consciousness-like behaviors. The results strongly support moderate positions in the "weak AI" debate, confirming that while sophisticated functional simulation is possible, it remains fundamentally distinct from genuine conscious experience. Several important limitations qualify our conclusions. The reliance on behavioral metrics means we can only assess external outputs rather than internal states of AI systems. Our findings are necessarily bounded by the capabilities of the specific models tested during the 2023-2024 study period. Furthermore, the analytical framework intentionally brackets metaphysical questions about qualia to focus on measurable functional components. The study's contributions are nonetheless significant for both research and practice. We've established replicable methods for consciousness-related evaluation that move beyond simple performance benchmarks. The documented gaps between simulation and authentic cognition provide crucial guidance for AI development, particularly in managing expectations about system capabilities. Perhaps most importantly, we've helped shift the conversation from speculative claims to evidence-based analysis of how AI models simulate cognitive functions. Looking forward, this work suggests several critical directions for future research. There remains pressing need for evaluation protocols informed by neuroscience rather than just behavior. Longitudinal studies tracking consciousness claims across AI generations could reveal important developmental patterns. Most urgently, the field requires ethical frameworks specifically addressing systems that exhibit consciousness-like behaviors without genuine awareness. In conclusion, this research addresses the questions and issues in the introduction, offering comprehensive insights into the central argument. The findings underscore the significance of systematically evaluating consciousness-like capabilities in AI systems and reveal critical limitations in existing models. This study also highlights the need for more robust evaluation criteria that align with AI's unique computational architecture rather than relying on biologically inspired standards.

Furthermore, the work identifies shortcomings, such as the lack of comprehensive benchmarks for assessing higher-order cognitive functions like logical reasoning and contextual understanding, which remain persistent challenges in current AI models. These limitations serve as a foundation for future research directions.

Future investigations should focus on expanding the scope of evaluation frameworks and integrating interdisciplinary perspectives to refine metrics for assessing consciousness-related behaviours in AI. Additionally, exploring innovative architectural designs tailored to address identified deficits could advance our understanding of artificial cognition and its potential capabilities.

.

REFERENCES:

[1] H. Wang, T. Fu, Y. Du, W. Gao, K. Huang, Z. Liu & M. Zitnik "Scientific Discovery in the Age of Artificial Intelligence, *Nature,* Vol. 620, No. 7972, 2023, pp. 47-60. https://www.nature.com/articles/s41586-023-06221-2

[2] H. W. de Regt "Understanding, Values, and the Aims of Science", *Philosophy of Science,* Vol.

87, No. 5, 2020, pp. 921–932. https://doi.org/10.1086/710520

[3] B. Zhang, J. Zhu, & H. Su „Toward the Third Generation Artificial Intelligence", *Science China Information Sciences,* Vol. 66, No. 2, 2023, p. 121101. https://doi.org/10.1007/s11432-021-3449-x

[4] L. Shytyk, & A. Akimova „Ways of Transferring the Internal Speech of Characters: Psycholinguistic Projection", *Psycholinguistics,* Vol. 27, No. 2, 2020, pp. 361–384. https://doi.org/10.31470/2309-1797-2020-27-2-361-384

[5] S. B. Yurchenko "Panpsychism and Dualism in the Science of Consciousness", *Neuroscience & Biobehavioral Reviews,* Vol. 165, 2024, p. 105845. https://doi.org/10.1016/j.neubiorev.2024.105845

[6] E. Brynjolfsson "The Turing Trap: The Promise & Peril of Human-Like Artificial Intelligence", in: *Augmented Education in the Global Age* (pp. 103-116), Routledge, 2023. https://digitaleconomy.stanford.edu/news/the-turing-trap-the-promise-peril-of-human-like-artificial-intelligence/

[7] I. Römer "Time-Consciousness, as a Concept in Phenomenology", in: de Warren, N., Toadvine, T. (Eds.), *Encyclopedia of Phenomenology* (pp. 1-8), Springer, 2023. https://link.springer.com/referenceworkentry/10.1007/978-3-030-47253-5_229-1

[8] A. Damasio & H. Damasio "Feelings Are the Source of Consciousness", *Neural Computation,* Vol. 35, No. 3, 2023, pp. 277-286. https://doi.org/10.1162/neco_a_01521

[9] R. C. Lanfranco, A. Canales-Johnson, B. Lucero, E. Vargas & V. Noreika "Towards a View From Within: The Contribution of Francisco Varela to the Study of Consciousness", *Adaptive Behavior,* Vol. 31, No. 5, 2023, pp. 405-422.

[10] I. Popovych, O. Semenov, A. Hrys, M. Aleksieieva, M. Pavliuk & N. Semenova "Research on Mental States of Weightlifters' Self-Regulation Readiness for Competitions", *Journal of Physical Education and Sport,* Vol. 22, No. 5, 2022, pp. 1134–1144. https://doi.org/10.7752/jpes.2022.05143

[11] A. Orsini "Functionalism", in: *Sociological Theory: From Comte to Postcolonialism* (pp. 281-360), Springer Nature Switzerland, 2024. https://link.springer.com/chapter/10.1007/978-3-031-52539-1_9

[12] J. R. Searle "Minds, Brains, and Programs", *Behavioral and Brain Sciences*, Vol. 3, No. 3, 1980, pp. 417–424. https://doi.org/10.1017/S0140525X00005756

[13] K. Ing, V. Uttarwar, Y. Akbari, T. G. van Erp & M. Fisher "Consciousness, Cortex, and Neuropsychoanalysis", *Journal of the American Psychoanalytic Association*, Vol. 72, No. 4, 2024, 653-662. https://doi.org/10.1177/00030651241268091

[14] S. Dolgikh "Self-Awareness in Natural and Artificial Intelligent Systems: A Unified Information-Based Approach", *Evolutionary Intelligence*, Vol. 17, 2024, pp. 4095–4114. https://link.springer.com/article/10.1007/s12065-024-00974-z

[15] D. J. Chalmers "Does Thought Require Sensory Grounding? From Pure Thinkers to Large Language Models", Proceedings and Addresses of the APA, Vol. 97, 2024, pp. 22-45. https://doi.org/10.48550/arXiv.2408.09605

[16] A. Gevaert, A. Saranti, A. Holzinger & Y. Saeys "Efficient Approximation of Asymmetric Shapley Values Using Functional Decomposition", in: *International Cross-Domain Conference for Machine Learning and Knowledge Extraction* (pp. 13-30), Springer Nature Switzerland, 2023. https://link.springer.com/chapter/10.1007/978-3-031-40837-3_2

[17] M. Mitchell "The Turing Test and Our Shifting Conceptions of Intelligence", *Science*, Vol. 385, No. 6710, 2024, p. eadq9356. https://www.science.org/doi/full/10.1126/science.adq9356

[18] J. Wang "Self-Awareness, a Singularity of AI", *Philosophy*, Vol. 13, No. 2, 2023, pp. 68-77. https://www.davidpublisher.com/Public/uploads/Contribute/6454a6a738fa1.pdf

[19] B. Liu "Arguments for the Rise of Artificial Intelligence Art: Does AI Art Have Creativity, Motivation, Self-awareness and Emotion?", *Arte, Individuo y Sociedad*, Vol. 35, No. 3, 2023, p. 811. https://dx.doi.org/10.5209/aris.83808

[20] A. Mentzou & J. Ross "The Emergence of Self-Awareness: Insights from Robotics", *Human Development*, Vol. 68, No. 2, 2024, pp. 90-100. https://doi.org/10.1159/000538027

[21] F. S. Muhamed "Computer's Self-awareness, Cognition, and Feeling", *Journal of Al-Ma'moon College*, Vol. 2, No. 40, 2023.

https://www.iasj.net/iasj/download/6ab0c182e683d520

[22] D. Combs (2024). "Exploring the Boundaries of AI: Creativity, Self-Awareness, and the Future of Intelligent Machines", *Authorea,* 2024. https://doi.org/10.22541/au.172676476.67378687/v1

[23] A. I. Luppi, P. A. Mediano, F. E. Rosas, J. Allanson, J. Pickard, R. L. Carhart-Harris, & E. A. Stamatakis "A Synergistic Workspace for Human Consciousness Revealed by Integrated Information Decomposition", *Elife*, Vol. 12, 2024, RP88173. https://doi.org/10.7554/eLife.88173.4

[24] A. I. Luppi, F. E. Rosas, P. A. Mediano, D. K. Menon & E. A. Stamatakis (2024). "Information Decomposition and the Informational Architecture of the Brain", *Trends in Cognitive Sciences*, Vol. 28, No. 4, 2024, pp. 352-368. https://doi.org/10.1016/j.tics.2023.11.005

[25] M. I. Dutt & W. Saadeh "Monitoring Level of Hypnosis Using Stationary Wavelet Transform and Singular Value Decomposition Entropy with Feedforward Neural Network", *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, Vol. 31, 2023, pp. 1963-1973. https://ieeexplore.ieee.org/abstract/document/10093766#citations

[26] D. Mastrovito, Y. H. Liu, L. Kusmierz, E. Shea-Brown, C. Koch & S. Mihalas (2024). "Transition to Chaos Separates Learning Regimes and Relates to Measure of Consciousness in Recurrent Neural Networks", *bioRxiv*, 2024, pp. 1-40. https://doi.org/10.1101%2F2024.05.15.594236

[27] E. Schwitzgebel & M. Garza "Designing AI With Rights, Consciousness, Self-Respect, and Freedom", in: F. Lara & J. Deckers (Eds.), *Ethics of Artificial Intelligence* (pp. 459-479), Springer Nature Switzerland, 2023. https://philpapers.org/rec/SCHDAW-10

[28] B. Wang, H. Zuo, Z. Cai, Y. Yin, P. Childs, L. Sun & L. Chen "A Task-Decomposed AI-Aided Approach for Generative Conceptual Design", in: *International Design Engineering Technical Conferences and Computers and Information in Engineering Conference* (Vol. 87349, p. V006T06A009). American Society of Mechanical Engineers, 2023. https://doi.org/10.1115/DETC2023-109087

[29] T. C. Gammel, L. N. Alkadaa, , J. R. Saadon, S. Saluja, J. Servider, N. A. Cleri & C. B. Mikell "Brain Circuitry of Consciousness: A Review of Current Models and A Novel Synergistic Model With Clinical Application", *Neurosurgery practice*, Vol. 4, No. 2, 2023, p. e00031. https://journals.lww.com/neurosurgpraconline/fulltext/2023/06000/brain_circuitry_of_consciousness__a_review_of.6.aspx