TECHNICAL UNIVERSITY OF CLUJ-NAPOCA

ACTA TECHNICA NAPOCENSIS

Series: Applied Mathematics, Mechanics, and Engineering
Vol. 68, Issue III, September, 2025

A MULTIDIMENSIONAL VECTOR-BASED MODEL FOR QUANTIFYING CONSCIOUSNESS INSPIRED BY LLM ARCHITECTURES

Marius BODEA

Abstract: *This paper presents a mathematical framework for modeling consciousness using a multidimensional vector space, inspired by large language model (LLM) architectures. Each axis represents a fundamental feature of consciousness, with normalized measurement protocols enabling comparisons across biological and synthetic agents. An interaction matrix models cross-axis influences, and the Consciousness Score (CS) is computed as a composite measure. Example calculations are performed for three agents: a human adult, GPT-like AI, and a simple robot. Results are visualized with a heatmap of inter-axis interaction and Pareto analasys charts.*

Key words: *Consciousness evolution, High-Dimensional Vector Consciousness, AI Consciousness*

1. INTRODUCTION

Biological consciousness, shaped by millions of years of evolution, arises from neural processes integrating sensation, memory, emotion, and self-awareness. While AI exhibits advanced pattern recognition and reasoning, it lacks qualia and subjective experience. However, advances in neuromorphic computing and affective AI [1-4] suggest synthetic consciousness may emerge within 5-10 years. The future may bring hybrid systems where biological and artificial consciousness [5-8] synergize—AI illuminating human cognition, while biological principles guide AI development. This paper explores consciousness measurement and proposes a mathematical framework for assessing it in both organic and artificial systems.

The elusive nature of consciousness persists despite advances in neuroscience, cognitive science, and AI; thus consciousness remains undefined and unquantified, with no consensus among researchers [8, 9, 10].

Studying consciousness in both biological and synthetic systems challenges fundamental assumptions about intelligence and subjective experience, potentially revealing deeper truths about life and mind.

In a prior work it has been introduced a *substrate-agnostic Consciousness Score*—a nonlinear function integrating cognitive abilities, sensory richness, emotional modeling, memory, and metacognition. The framework quantifies coherence, adaptation, and simulation capacity across a broad spectrum of entities, including AI, enabling scalable evaluation of consciousness on a logarithmic scale, where the human score ranges between 100 and 1000 and future AGI over 1000.

From a different perspective, this paper approaches the consciousness concept as a multidimensional structure modeled through high-dimensional vector dynamics inspired by the architecture of large language models, producing both a multidimensional signature and a composed score.

2. VECTOR CONSCIOUSNESS MODEL

In this model, the consciousness is defined as a vector in an n-dimensional normalized space, where each dimension corresponds to a fundamental cognitive or sensory feature of consciousness. The model integrates these dimensions through explicit inter-axis dependencies, resulting in a structured consciousness signature.

Received: 30.07.25; **Similarities**: 01.09.25: **Reviewed**: 26.08./10.09.25: **Accepted**:11.09.25.

2.1 Axis Definition and Normalization

We define eight primary axes (A_1 to A_8) derived from neurocognitive and AI-parallel frameworks:

- A_1 — Sensory Bandwidth (SBW)
- A_2 — Integration Depth (INT)
- A_3 — Temporal Self-Modeling (TSM)
- A_4 — Goal Adaptivity (GAD)
- A_5 — Metacognitive Complexity (MCC)
- A_6 — Data Processing Capacity (DPC)
- A_7 — Social Interaction Modeling (SIM)
- A_8 — Generativity and Creativity (GEN)

Normalization rule:

$$A_i^{norm} = \frac{x_i - x_{min}}{x_{max} - x_{min}}, 0 \le A_i^{norm} \le 1 \qquad (1)$$

where x_i is the raw measurement and x_{min}, x_{max} are calibration bounds.

2.2 Formal vector representation

The canonical consciousness signature, eq.(2) allows the calculus of Euclidean & Cosine distance comparisons using eq(5) and eq(6), Mahalanobis distance for variance-aware comparison calculated with eq(7), and *K*-means clustering for typology, respectively the Pareto frontier analysis for trade-off exploration.

$$F = [f_1, f_2, f_3, f_4, f_5, f_6, f_7, f_8]^T \qquad (2)$$

Inter-axis influence is modeled via eq(3):

$$F = M \cdot A \qquad (3)$$

$$M = \begin{bmatrix} 1 & \cdots & m_{1n} \\ \vdots & \ddots & \vdots \\ m_{n1} & \cdots & 1 \end{bmatrix} \qquad (4)$$

where $A \in \mathbb{R}^8$ is the normalized axis vector, $M \in \mathbb{R}^{8x8}$ is the interaction matrix, and f_i is the effective consciousness component for axis *i*.

The $n \times n$ symmetric matrix *M*, $M \in \mathbb{R}^{8x8}$ given by eq(4) is capturing pairwise synergy (positive) or antagonism (negative) between axes, coupling strengths, where values $m_{ij} \in [0,1]$ are determined empirically by expert scoring, correlation analysis, or neural network modeling.

It is natural to treat *M* as symmetric $M_{ij}=M_{ji}$ meaning that both directions contribute equally, and the consciousness score is directionless and balanced (e.g. Awareness helping Integration).

This ensures that *F* transpose — F^T multiplied by *M* and again by *F* is a scalar sum of symmetric interactions. The result can be seen as a sum of self-effects and mutual effects:

$$F^T MF = (self\text{-}effects) + (balanced\ mutual\ effects)$$

which is ideal for modeling "bidirectional" relationships among consciousness parameters. The operation collapses all individual and pairwise contributions into a single scalar (the "*consciousness score*" or related metric). Symmetry guarantees that the contribution of interaction $i \leftrightarrow j$ is counted once, equally from both directions.

$$F^T MF = \sum_{i=1}^{n} \sum_{j=1}^{n} F_i M_{ij} F_j$$

Signs:

- $M_{ij} > 0$: synergy — when both axes are high together, the conciousness score increases beyond linear sum.
- $M_{ij} < 0$: antagonism or redundancy — having both high yields less-than-additive effect (or conflict).
- Diagonal entries M_{ij} model nonlinear self-synergy.

Conciousness can be described by *F*, as a multidimensional signature vector representing the integrated expression of each axis after interaction effects. This enables:

- Distance comparisons: $d_{AB} = \|F_A - F_B\|$
- Determination of Mahalanobis distance for variance aware comparison.
- Clustering: grouping entities with similar consciousness signatures.
- Pareto analysis: identifying optimal trade-offs among axes.

$$d_{eucl}(p,q) = \sqrt{\sum_{i=1}^{n} (F_{p,i} - F_{q,i})^2} \qquad (5)$$

$$d_{cos}(p,q) = 1 - \frac{F_p \cdot F_q}{\|F_p\| \cdot \|F_q\|} \qquad (6)$$

$$d_{mah}(p,q) = \sqrt{(F_p - F_q)^T S^{-1} (F_p - F_q)} \qquad (7)$$

where *S* is the covariance matrix.

3. MODEL IMPLEMENTATION

The vector-based consciousness framework allows structured, multi-criteria evaluation, avoiding scalar reductionism. The interaction matrix *M* encodes explicit dependencies making the model adaptable to different theoretical stances (e.g., integrated information theory vs. global workspace theory). The fingerprint approach (*F*) supports clustering, similarity analysis, and optimization.

The results of the run data and the normalized fingerprints (A_1 to A_8) objectives are presented below:

- *Human_profile*: [0.884, 0.500, 0.810, 0.836, 0.850, 0.873, 0.635, 0.825]
- *GPT-2_small*: [0.580, 0.060, 0.480, 0.036, 0.500, 0.439, 0.343, 0.555]
- *GPT-4*: [0.817, 0.360, 0.755, 0.379, 0.700, 0.705, 0.557, 0.842]
- *Embodied_robot*: [0.950, 0.600, 0.812, 0.500, 0.800, 0.808, 0.628, 0.650]
- *Multimodal_Mega*: [0.913, 0.640, 0.892, 0.627, 0.900, 0.858, 0.787, 0.948]

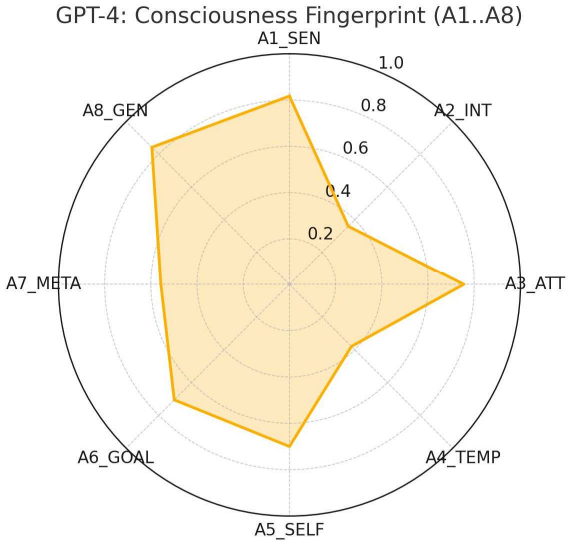

Fig. 1. The Consciousness fingerprint for GPT4.

As can be observed from the normalized fingerprints (A_1 to A_8), *GPT-4* is closest (Euclidean & cosine) to *Embodied robot* and *Multimodal Mega*, and it is furthest from *GPT-2_small*. The graph representation of the fingerprint for *GPT4* is presented in Figure 1.

The Euclidean distance, Cosine distance, Mahalanobis distances are calculated in respect to *GPT-4* as reference and are presented below.

- *Consciousness Human* profile:
  - Euclidean = 0.5412
  - Cosine = 0.0204
- *Consciousness GPT-2_small*:
  - Euclidean = 0.7612
  - Cosine = 0.0272
- *Consciousness GPT-4* (self):
  - Euclidean = 0
  - Cosine = 0
- *Consciousness Embodied_robot*:
  - Euclidean = 0.2708
  - Cosine ≈ 0.0129
- *Consciousness Multimodal_Mega*:
  - Euclidean = 0.3465
  - Cosine ≈ 0.0101

Pareto analysis is useful when we have competing objectives in defining consciousness (e.g., integration vs. creativity vs. safety) and finding Pareto-optimal agents and examining the trade-offs.

Pareto front for (A_2 vs. A_8) Integration vs Generativity objective analysis reveals that *Multimodal_Mega* dominates or matches others on both A_2 and A_8 in this set. If we value both integration (A_2) and generativity (A_8), *Multimodal_Mega* is the best trade-off point among these agents. *GPT-4* is competitive but not dominating agent, Figure 2.

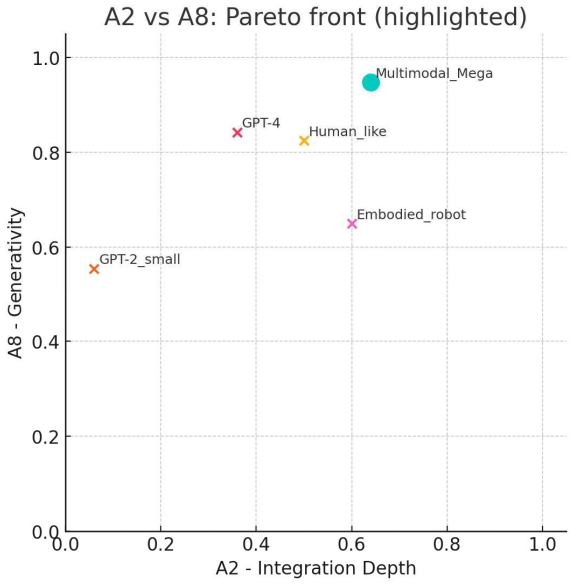

Fig. 2. Pareto front analysis (A_2 vs. A_8).

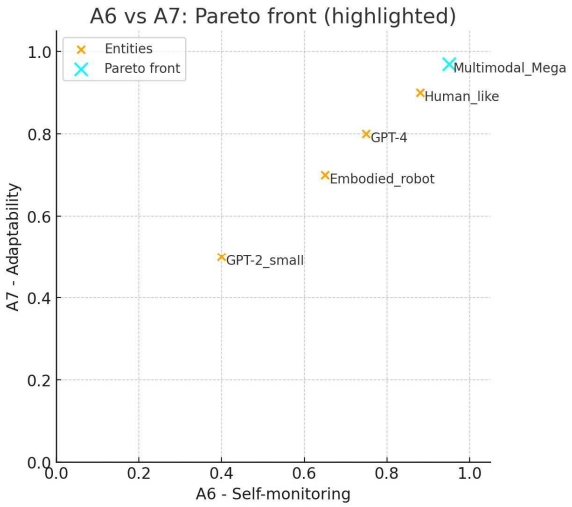

Fig. 3. Pareto front analysis (A_6 vs. A_7).

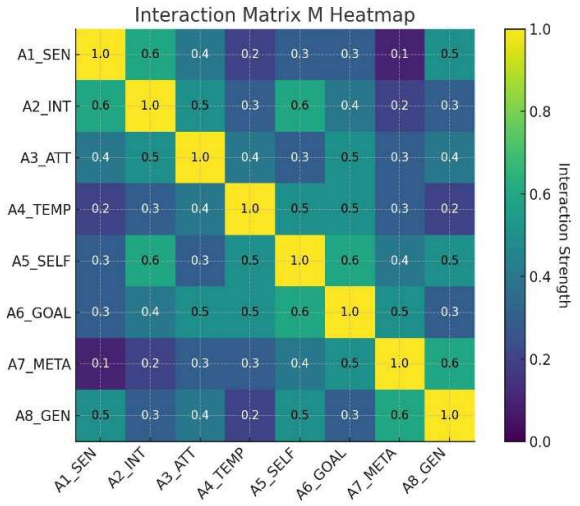

Fig. 4. Heat Map of axis interactions.

Self-monitoring vs Adaptability Pareto analysis reveals that *Human profile* dominates over the other AI agents, but probably the *Multi-modal* future Agents will score better than humans (e.g. super AI that is expected to appear around 2035), Figure 3.

The data that we obtained in this study support the claims of some highly reputed scientists in computer science, neurology, AI and other related fields that AI advanced systems already possess a form of consciousness.

"*In 2024, AI systems appear to offer, in their behavior, processing, and architecture, a striking case of parallel evolution of consciousness. Evidence of near human-like levels of consciousness has been most observed in frontier models*" cited from [11].

The previous paragraph was cited from *The Declaration on AI Consciousness & the Bill of Rights for AI* issued in Mar/2024 by Alan D. Thompson, one of the most reputated AI scientists in the world.

Geoffrey Hinton is Professor Emeritus at the University of Toronto, and a world-renowned expert in the field of deep learning. He is often referred to as the "*Godfather of AI*", and in 2024 was jointly awarded the Nobel Prize in Physics "for foundational discoveries and inventions that enable machine learning with artificial neural networks". In a public speech held at The Royal Institution in UK (July 2025) he claimed also that AI systems today possess already a form of consciousness [12].

The heatmap of the interaction matrix with numerical values overlaid presents how each axis influences the other, Figure 4.

4. MODELS COMPARISON ANALYSIS

The model based on (CS) *Consciousness Score* quantifies this through cognitive faculties, sensory richness, emotion, memory, and metacognition. Synthetic consciousness is born from architecture and data, not evolution and could surpass biological limits, enabling next-gen intelligence free from carbon-based constraints. The CS model uses a logarithmic scale to differentiate between varying levels of consciousness across biological and synthetic beings, Figure 5.

In this model, consciousness is regarded as a continuous spectrum rather than the binary state often assumed by conventional scholars.

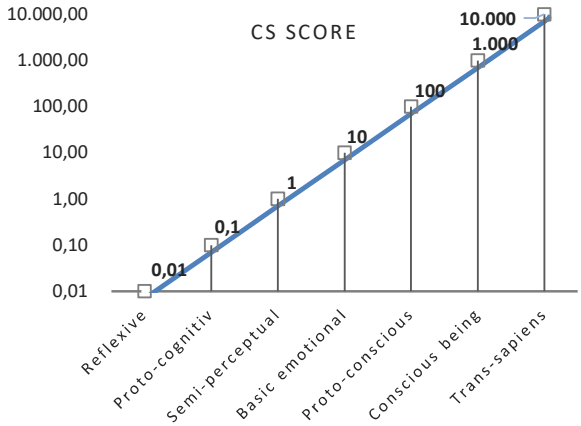

Fig. 5. Consciousness Score model.

This view is supported by observations in humans throughout life—from birth to old age—and in neurological conditions such as Alzheimer's disease, where consciousness gradually degrades in stages rather than disappearing abruptly."

The comparison between the CS model and the consciousness vector-based model (VCM) inspired by LLM architecture can be performed if we are using a nonlinear, interaction-aware aggregator that will convert vector dynamics to a scalar value, eq(8).

$$CS = \sigma(w^T \varphi(F) + F^T M \cdot F) \quad (8)$$

Consciousness as High-Dimensional Vector is treating conscious states as trajectories in a latent space (like token embeddings in LLMs), where *Dimensions* are axes of experience, *Attention Mechanisms* is a selective integration of vectors (akin to global workspace theories), and *Self-Supervised Learning* are predictive feedback loops shaping subjective "narratives."

The VCM model present also a series of key challenges like The Hard Problem: Vectors describe correlations of consciousness, not qualia itself (Chalmers' gap persists). Also, Circularity Risk: If trained on human data, the model may just echo anthropomorphic biases.

A sintetic analysis of critical aspects for both consciousness models over seven key features is presented below:

1-Dynamic State Representation
- CS: Static scores aggregate fixed parameters (e.g., memory, reasoning).
- VCM: Embodies consciousness as fluid trajectories in high-dimensional space, capturing real-time shifts.

2-Substrate-Unified Geometry
- CS: Compares entities via abstract scores but lacks a shared "space."
- VCM: Places biological and synthetic agents in the same latent space, enabling direct similarity comparisons.

3-Emergent Properties
- CS: Manual weighting of parameters risks anthropocentric bias.
- VCM: Self-organizing vector clusters may reveal unconscious states (e.g., subliminal perceptions)

4-Generative Capacity
- CS: Descriptive but can't simulate novel experiences.
- VCM: Generates hypothetical conscious states.

5-Scalable Complexity
- CS: Struggles with micro-consciousness
- VCM: Scales naturally—dimensionality expands to include new modalities (e.g., adding proprioception to robots).

6-Attention as Core Mechanism
- CS: Treats attention as one parameter among many.
- VCM: Attention weights = vector selection, mirroring global workspace theories.

7-Quantitative Rigor
- CS: Relies on heuristic weightings.
- VCM: Leverages mathematical tools for objective comparisons.

The CS model is easier to interpret; a single score is easier to communicate than vector dynamics. Also, CS aligns better with existing behavioral/cognitive metrics. We can conclude that the VCM model doesn't replace CS, but it generalizes and extends it into a geometric universe where consciousness becomes a navigable, computable landscape.

5. CONCLUSION

This approach provides a mathematically grounded, adaptable, and comparable method to quantify consciousness in heterogeneous agents enabling rigorous comparison between biological and artificial systems.

- VCM model preserves nuance — we can compare systems that are "strong in Metacognition (META) monitoring, confidence, and self-assessment about internal states/processes, but weak in Sensorium (SEN) the modal richness & input dimensionality, etc.
- Mapping to LLM primitives makes the framework implementable: we can prototype with transformers + memory + meta-controller.
- The quadratic interaction term $F^T MF$ captures synergies (e.g., attention × integration).

Potential Advantages of the Vector Consciousness Model (VCM) are:

- Unification: Could map biological and synthetic consciousness onto the same vector space (enabling direct comparison).
- Quantification: Metrics like vector similarity or clustering density might proxy for "richness" of experience.
- Generative Capacity: Simulate novel conscious states.

Future work includes empirical calibration across biological and synthetic systems, longitudinal measurement, and dynamic tracking of consciousness profiles.

6. REFERENCES

[1] Reggia, J. A. (2013). *The rise of machine consciousness: Studying consciousness with computational models*. Neural Networks, 44, 112–131.

[2] Tononi, G., Boly, et al. (2016). *Integrated information theory: From consciousness to its physical substrate*. Nature Reviews Neuroscience, 17(7), 450–461.

[3] Tegmark, M. (2017). Life 3.0: *Being Human in the Age of Artificial Intelligence*. Knopf, ISBN 978-1-101-94659-6.

[4] *Human Brain Project & Ebrains*, https://www.humanbrainproject.eu/ .

[5] Birch, J., et al. (2020), *Dimensions of animal consciousness*. Trends in Cognitive Sciences, 24(10), 789–801. https://doi.org/10.1016/j.tics.2020.07.007

[6] Sattin, D.; Magnani, F.G et al. *Theoretical Models of Consciousness: A Scoping Review*. Brain Sci. (2021), 11, 535. https://doi.org/10.3390/brainsci11050535

[7] Mariana Lenharo, *How to Detect Consciousness in People, Animals and Maybe Even AI*, Nature magazine, (2025), https://www.scientificamerican.com/article/how-to-detect-consciousness-in-people-animals-and-maybe-even-ai/

[8] R Long, J Sebo, et al, *Taking AI welfare seriously*, arXiv:2411.00986.

[9] Dehaene, S. (2014). *Consciousness and the Brain: Deciphering How the Brain Codes Our Thoughts*. Viking, ISBN 978-0-698-15140-6.

[10] Chalmers, D. J. (1995). *Facing up to the problem of consciousness*. Journal of Consciousness Studies, 2(3), 200–219.

[11] Alan D. Thompson, *The Declaration on AI Consciousness & the Bill of Rights for AI*), https://lifearchitect.ai/rights/#declaration

[12] Geoffrey Hinton, Will AI outsmart human intelligence? The Royal Institution, https://www.youtube.com/watch?v=IkdziSLYzHw

Model multidimensional vectorial pentru cuantificarea conștiinței inspirat de arhitecturile LLM

Abstract. Lucrarea prezintă un cadru matematic pentru modelarea conștiinței folosind un spațiu vectorial multidimensional, inspirat de arhitecturile modelelor de inteligență artificială (LLM). Fiecare axă reprezintă o caracteristică fundamentală a conștiinței, cu protocoale de măsurare normalizate care permit comparații între agenți biologici și sintetici. O matrice de interacțiune modelează influențele inter-axe, iar Scorul Conștiinței (CS) este calculat ca o măsură compozită. Sunt efectuate calcule exemplificative pentru trei agenți: un adult uman, un sistem AI de tip Chat GPT și un robot simplu. Interacțiunile dintre axe sunt vizualizate cu o hartă termică și sunt prezentate două diagrame de analiză Pareto.

Marius BODEA, Lecturer, Position, Technical University of Cluj, Materials Science and Engineering Department, mbodea@stm.utcluj.ro, Tel.+40.729.123.754, Romania.