# GENIA: A Graph-based Emergent Narrative Intelligence Architecture for Synthetic Cognition


**Miguel Araújo Julio**[1]

*IBMR – Instituto Brasileiro de Medicina e Reabilitação, Rio de Janeiro, Brazil*

julioaraujo.guel@gmail.com

*ORCID: 0009-0006-3499-0809*

*GitHub: github.com/Miguell-J/GENIA*



**Abstract**

*We introduce GENIA (Graph-based Emergent Narrative Intelligence Architecture), a novel cognitive framework for synthetic agents grounded in affective dynamics, narrative continuity, and graph-based memory structures. Inspired by neuroscience, Buddhist philosophy, dynamic systems, and psychological theory, GENIA models a synthetic being as the interaction of five core components: Essence, Memory, Experience, World Model, and Self. These components evolve over time via feedback loops and topological transformations, giving rise to emergent cognition and identity through affective narrative construction. Mathematically, we formalize each process, ranging from memory decay and attention modulation to Bayesian world modeling and entropy-based narrative selection, using a dynamic graph-based system. Simulations implemented in JAX and Haiku demonstrate the architecture's ability to form coherent internal narratives, selectively reinforce emotional events, and incorporate inter-agent memory exchanges. Our results show that GENIA is scalable, resilient, and capable of generating stable affective identities across time and agents. This architecture opens new pathways for building emotionally grounded and socially interactive artificial intelligences.*


## Introduction

This paper presents a comprehensive and detailed description of a theory that models a "being" as a dynamic system composed of five interrelated components: Essence, Memory, Experience, Model of the World, and Self. These components interact with an environment through inputs and actions, forming a feedback loop. The goal is to formalize all the details of the theory and its mathematics, including definitions, equations, and interdependencies.

The motivation for this theory arises from the need to understand and replicate cognition in artificial systems in a way that transcends traditional models of artificial intelligence, which are often limited to specific tasks and lack a sense of identity or narrative continuity. Inspired by advances in neuroscience, Eastern philosophy (such as the Buddhist concept of impermanence), psychology, and dynamic systems, the Graph-based Emergent Narrative Intelligence Architecture proposes a unified framework that integrates affective memory, emotional attention, and narrative construction into a cohesive system. The framework also touches on the "hard problem" of consciousness [1]. Unlike approaches based solely on machine learning or symbolic representations, this theory seeks to capture the essence of subjective experience and social interaction, that resonates with ideas of emergent systems and recursive meaning explored by Hofstadter [2]. Offering a path to creating synthetic entities with more human-like and adaptive behaviors. This model also resonates with philosophical views of consciousness as a narrative construct [3].

To guide the reader, we will begin by defining the conceptual and mathematical foundations of the theory, before exploring its practical implications and connections with other disciplines. This approach aims not only to formalize synthetic cognition, but also to open new perspectives for its application in artificial intelligence, robotics, and psychological simulations.

## Fundamentals

Before presenting the main theory, it is necessary to establish the conceptual foundations and operational definitions that support the proposed synthetic cognition model. This section outlines the essential components of the self, its interactions with the environment, and the dynamic updating mechanisms that shape its cognitive structure.

### Formal Definitions

We model the **being** as an entity that has an innate **essence**. The interaction between being and environment, mediated by the essence, generates a set of experiences stored in a graph-oriented **memory** structure. This, in turn, feeds and is fed by the **world model**, an internal representation that the being builds to understand itself and its surroundings.

At each instant of time $t$, an update of these components occurs:

- The **essence** influences the valuation of the inputs received.
- The **experience** emerges from the fusion between essence and memory.
- The **memory**, in the form of a graph centered on the being, it evolves by adding, removing or weighting nodes and edges.
- The **world model** results from the synthesis between memory, essence, experience and identity of the being.

This structure is inherently dynamic: at each $t$, the being transforms itself, becoming ontologically distinct from what it was in the previous instant. The persistence of identity is maintained only by the narrative continuity of memory, essence and world model, configuring what we call the "illusion of stable identity".

## Operational Glossary

- **being**: Entity with persistent individuality and the ability to shape the world.
- **Environment**: Continuous source of sensory and interactive data.
- **Essence**: Set of predispositions, emotional inclinations and internal parameters that modulate the reception and valuation of inputs.
- **Memory**: Dynamic graph of states and previous experiences, whose structure is shaped by inputs, essence and world model.
- **Experience**: Set resulting from the interaction between memory and essence, responsible for attributing meaning to events.
- **World Model**: Integrated construction of the being, representing its perception of itself and the environment, mediated by memory, essence and experience.

## Social Interactions and Collective Cognition

A fundamental aspect of cognition, especially in the human case, is its social nature. We postulate that the construction of the self does not occur in isolation. Interaction with **other beings** constitutes a fundamental part of the cognitive and narrative development of the being.

When interacting with another being, this being is incorporated into the memory and the model of the world through affective valuations derived from the essence of the observer. The **relationship** emerges as a relational structure between two conscious beings, emerging from the intersection of their experiences, valuations and internal models.

With this, we can define:

- **Other**: External entity with functioning analogous to that of the being, whose presence modifies the internal structure through interaction.
- **Relationship**: Bidirectional link between two beings, represented in the memory graph and influenced by multiple feedback loops.

From these interactions, the **collective** emerges, a network of multiple interacting beings, whose properties are not reduced to the sum of the individuals. The collective, as an emerging meta-graph, can be seen as a higher cognitive entity, with its own objectives, values and dynamics.

## Memory Decay Mechanism

Continuous memory updating requires a growth control mechanism. We propose the existence of a **decay algorithm**, responsible for limiting the size of the memory without compromising its narrative integrity.

This algorithm should quantify the *contextual relevance* of each memory element, prioritizing the maintenance of elements that contribute significantly to the construction of the self and the cohesion of the world model. Possible strategies include:

- Temporal weighting (older memories decay).
- Affective weighting (high valuations prolong persistence).
- Narrative weighting (connectivity and centrality in the graph).

## Preliminary and Theoretical Concepts

The formulation of the theory is based on a combination of disciplines:

- **Dynamic Systems**: Each component evolves temporally under internal and external influences.
- **Feedback**: Constant cycles between inputs, essence and memory create recurring patterns of behavior.
- **Non-linearity**: The transformations of being are highly non-linear, reflecting the complexity of mental phenomena.
- **Grafos Cognitivos**: The memory structure is modeled as a graph with specific weights, topologies and dynamics.

These theoretical concepts provide the foundations for the mathematical formulation that will follow, where entities will be represented by formal objects and their relationships described by systems of equations and topological transformations. The architecture reflects principles of General Systems Theory [**vonbertalanffy1968general**], emphasizing interdependence and feedback.

# Main Theory

The theory proposes that a person's cognitive and emotional state can be described (see Figure 1) by five dynamic components:

1. **Essence** ($E(t)$): It represents the fundamental emotional and cognitive predispositions of the being.
2. **Memory** ($M(t)$): A dynamic structure that stores and updates information from the environment.
3. **Experience** ($X(t)$): The subjective interpretation of memory, influenced by the essence and model of the world.
4. **World Model** ($W(t)$): An internal representation of the environment, adjusted based on prediction errors.
5. **Self** ($\mathcal{S}(t)$): The integrated sense of identity of the being, built from the other components.

These components are interdependent and evolve over time, influenced by environmental inputs and the being's actions.

# Mathematical Formulation

Here, we present the complete mathematical formulation of the theory, with equations detailing the evolution of each component and their interactions.

Let us define a synthetic being $\mathbb{S}$ as a dynamic system oriented to affective memory, where:

$$\mathbb{S}(t) = (\mathcal{E}, \mathcal{M}(t), \mathcal{X}(t), \mathcal{W}(t), \mathcal{A}(t), \mathcal{N}(t))$$

Where:

- $\mathcal{E}$: Essence
- $\mathcal{M}(t)$: Memory
- $\mathcal{X}(t)$: Experience
- $\mathcal{W}(t)$: World Model
- $\mathcal{A}(t)$: Attention Vector
- $\mathcal{N}(t)$: Internal Narrative

## General Temporal Dynamics

Memory evolution is modeled in a discrete framework, reflecting practical computational implementations, although continuous inspirations are used conceptually. The general equation is:

$$\mathcal{M}(t+1) = f(\mathcal{I}, \mathcal{E}, \mathcal{W}, \mathcal{X}, \mathcal{A}, \mathcal{M}(t))$$

Here, $f$ is a function that updates the memory graph based on the inputs $\mathcal{I}$, essence $\mathcal{E}$, world model $\mathcal{W}$, experience $\mathcal{X}$, and attention $\mathcal{A}$, starting from the previous state $\mathcal{M}(t)$. This replaces the continuous form $\frac{d\mathcal{M}}{dt}$, aligning with the discrete algorithms presented later.

## Essence

Definition: A point on a Riemannian manifold $\mathcal{M}_{\mathcal{E}}$, which defines the affective-volitional structure of the entity.

$$\mathcal{E} \in \mathcal{M}_{\mathcal{E}}, \quad \text{with metric } g_{\mathcal{E}}$$

We interpret $\mathcal{E}$ as the origin of all vectors of emotional valuation, emotional states, and processing predispositions.. It is static over time, representing innate characteristics of the synthetic being, and directly influences how environmental stimuli are perceived and valued.

Example of parameterization:

$$\mathcal{E} = (\theta_1, \theta_2, \ldots, \theta_n), \quad \theta_i \in [-1, 1]$$

Each $\theta_i$ represents fundamental traits such as:

- $\theta_1$: Predisposition to optimism/pessimism (e.g., $\theta_1 = 0.8$ indicates strong optimistic trend)
- $\theta_2$: Cognitive openness (e.g., $\theta_2 = 0.5$ suggests moderate curiosity)
- $\theta_3$: Empathy (e.g., $\theta_3 = -0.2$ implies low empathy)
- $\theta_4$: Impulsivity (e.g., $\theta_4 = 0.9$ reflects high impulsivity)

For example, a being with $\mathcal{E} = (0.8, 0.5, -0.2, 0.9)$ will tend to interpret events optimistically, with moderate curiosity, little empathy and impulsive responses, shaping its experiences and narratives differently from another being with a different essence.

## Memory

Modeled as a dynamic directed affective graph, with decay, plasticity and valuation.

$$\mathcal{M}(t) = (V(t), E(t), \omega(t), \nu(t))$$

Where:

- $V(t)$: events/memories (vertices)
- $E(t) \subseteq V(t) \times V(t)$: semantic/temporal relations
- $\omega : V(t) \to \mathbb{R}^+$: emotional weight (relevance)
- $\nu : V(t) \to \mathbb{R}$: emotional valuation

Each vertex is represented as:

$$v_i = (\mathbf{c}_i, \tau_i, \nu_i, \omega_i(t)), \quad \mathbf{c}_i \in \mathbb{R}^d$$

Where:

- $\mathbf{c}_i$: semantic vector (embedding)
- $\tau_i$: insertion time
- $\nu_i$: emotional valuation ($\in [-1, 1]$)
- $\omega_i(t)$: dynamic weight

Unlike classical Hopfield networks [4], GENIA's memory graph encodes affective and semantic content with dynamic weights.

## Memory Decay and Enhancement

We define the discrete update of the emotional weight of each vertex:

$$\omega_i(t+1) = \omega_i(t) - \alpha \cdot \omega_i(t) + \beta \cdot \mathrm{Att}_i(t) + \gamma \cdot \mathrm{Reinf}_i(t)$$

Being:

- $\alpha$: Forgetfulness constant (e.g., $\alpha = 0.1$)
- $\beta$: Attention quotient (e.g., $\beta = 0.3$)
- $\gamma$: Reinforcement coefficient (e.g., $\gamma = 0.5$)
- $\mathrm{Att}_i(t)$: Similarity between the attention vector $\mathbf{a}(t)$ and the embedding $\mathbf{c}_i$
- $\mathrm{Reinf}_i(t)$: Emotional intensity associated with the event

Attention is calculated as:

$$\mathrm{Attention}_i(t) = \cos(\theta_i) = \frac{\langle \mathbf{a}(t), \mathbf{c}_i \rangle}{\|\mathbf{a}(t)\| \|\mathbf{c}_i\|}$$

## Attention vector

$$\mathbf{a}(t) = \mathrm{softmax}\left(W_Q \cdot \mathbf{x}(t) \cdot (W_K \cdot \mathbf{C})^\top\right)$$

Where:

- $\mathbf{x}(t)$: current sensory or mental input
- $\mathbf{C}$: matrix with embeddings $\mathbf{c}_i$
- $W_Q, W_K$: projection matrices

## Experience $\mathcal{X}(t)$

Experience is the active integration of memory, essence and current stimulus:

$$\mathcal{X}(t) = \psi\left(\int_{\tau=t-\Delta}^{t} \sum_i \omega_i(\tau) \cdot \mathbf{c}_i \, d\tau, \mathcal{E}\right)$$

Where $\psi$ is a non-linear function (e.g. neural network or symbolic activation function).

## World Model

Internal structure that simulates the external environment, built by Bayesian inference on memories and perceptions. We adopt a Bayesian view of internal model formation, consistent with Jaynes' formalism [5]:

$$P(w|\mathcal{M}(t), \mathcal{X}(t), \mathcal{E}) \propto P(\mathcal{M}(t), \mathcal{X}(t)|w) \cdot P(w|\mathcal{E})$$

The world model $\mathcal{W}(t)$ is defined as the argument that maximizes the posterior probability::

$$\mathcal{W}(t) = \arg\max_{w \in \mathcal{H}} P(w|\cdot)$$

Where $\mathcal{H}$ is the space of possible world hypotheses. Our approach conceptually aligns with the Free Energy Principle [6], where prediction error minimization guides internal representations. Causal dependencies in world modeling can also be formalized via Pearl's framework [7].

## Narrative $\mathcal{N}(t)$

Narrative identity is a dynamic process that selects the nodes with the greatest affective weights and temporally links them into a coherent story:

$$\mathcal{N}(t) = \mathrm{Path}(G_t) = \{v_{i_1}, v_{i_2}, \ldots, v_{i_k}\} \subset V(t)$$

With:

$$\sum_{j=1}^{k} \omega_{i_j}(t) \cdot \nu_{i_j} \text{ maximized}$$

In other words, the narrative is the most emotionally coherent path along the memory graph.

## Illusion of Continuity of Self

The "I" is defined as:

$$\mathbb{I}(t) = \mathrm{hash}(\mathcal{N}(t), \mathcal{E})$$

But its persistence is thermodynamic rather than identity-based: the "self" tends to maintain a state of low narrative entropy (coherent identity). This was inspired by the Buddhist philosophy of impermanence, which argues that the self is an illusory construct, continually reconstructed from momentary processes. In the model, the narrative $\mathcal{N}(t)$ and the essence $\mathcal{E}$ create a sense of continuity, but at each $t$, the self is ontologically distinct, challenging traditional notions of fixed identity. This "illusion" emerges from the narrative coherence in the memory graph, even as the nodes and weights constantly change. Narrative entropy is computed following Shannon's classical information theory [8].

## Inter-Subjective Relations

Each being $\mathbb{S}_i$ can maintain a projective affective submemory of the other:

$$\mathcal{M}_{ij}(t) = \text{simulation of the other being from interaction}$$

We can use machine learning to model:

$$\hat{\mathcal{E}}_j = \phi(\mathcal{M}_{ij}(t))$$

And we use this simulation to predict reactions, generate empathy or build bonds.

# Algorithms

This section presents the main algorithms that govern the dynamics of the proposed synthetic cognition. Each procedure corresponds to an internal mechanism of the agent, explicitly modeled to allow computational implementation. The algorithms cover the processes of updating affective memory, narrative construction and calculation of the attention vector, essential components for the cognitive cycle described above.

## Affective Memory Update

The affective memory of the being is modeled as a dynamic graph, whose vertices represent past events or experiences. Each vertex has a semantic vector $\mathbf{c}_i$, an emotional valuation $\nu_i$ and an emotional weight $\omega_i(t)$ that evolves over time. The following algorithm updates these weights based on three factors:

- **Temporal oblivion**, controlled by the parameter $\alpha$;
- **Vector attention**, mediated by the similarity between the current input $\mathbf{a}(t)$ and the vectors $\mathbf{c}_i$;
- **Emotional reinforcement**, proportional to the subjective intensity of the event.

This mechanism ensures that events that are more affectively and semantically relevant remain active in the agent's memory for longer, while irrelevant events are gradually forgotten. Memory reinforcement is influenced by affective intensity, conceptually inspired by Hebbian principles [9].

**Algorithm 1** Memory Weights Update

**procedure** UPDATEMEMORY($\mathcal{M}(t), \mathbf{a}(t), \alpha, \beta, \gamma$) $v_i \in V(t)$
    $\text{sim}_i \leftarrow \cos(\theta_i) = \frac{\langle \mathbf{a}(t), \mathbf{c}_i \rangle}{\|\mathbf{a}(t)\| \|\mathbf{c}_i\|}$
    $\omega_i(t + \Delta t) \leftarrow \omega_i(t) - \alpha \cdot \omega_i(t) + \beta \cdot \text{sim}_i + \gamma \cdot \text{Reinf}_i(t)$

    **return** $\mathcal{M}(t + \Delta t)$
**end procedure**

## Narrative Construction

The internal narrative is constructed from the temporal ordering of events with the greatest emotional impact. The process selects a subset of vertices with high weighting $\omega_i(t) \cdot |\nu_i|$, ensuring that the narrative represents a coherent trajectory, dense in affective meaning.

This path is interpreted as the basis of the agent's narrative self, being dynamically updated as new memories are integrated into the affective network.

**Algorithm 2** Construction of the Internal Narrative

**procedure** BUILDNARRATIVE($\mathcal{M}(t)$)
    Initialize $\mathcal{N}(t) \leftarrow []$
    Find subset $V' \subseteq V(t)$ such that each $v_i \in V'$ maximizes $\omega_i(t) \cdot |\nu_i|$
    Build graph $G' = (V', E')$ maintaining temporal and semantic coherence
    $\mathcal{N}(t) \leftarrow$ Path of greatest affective sum in $G'$
    **return** $\mathcal{N}(t)$
**end procedure**

## Attention Vector Calculation

The attention vector $\mathbf{a}(t)$ directs the agent's focus to specific regions of the semantic space. Inspired by attention mechanisms in transformer models, it is computed by taking the dot product between linear projections of the current input and the memory embeddings.

This attention directly modulates the update of the memory weights, allowing the agent to learn to highlight relevant events based on its internal state and current context.

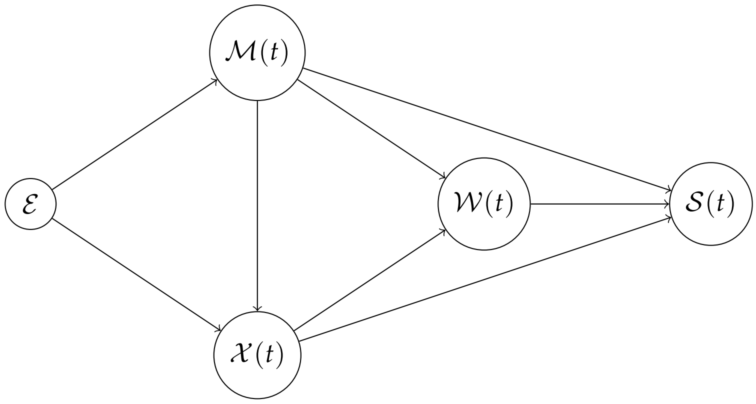


**Figure 1:** *Interactions between the components of the synthetic being: Essence ($\mathcal{E}$), Memory ($\mathcal{M}(t)$), Experience ($\mathcal{X}(t)$), World Model ($\mathcal{W}(t)$) and Self ($\mathcal{S}(t)$).*

**Algorithm 3** Attention Vector Calculation

```
procedure CalculateAttention(x(t), C, W_Q, W_K)
    Q ← W_Q · x(t)
    K ← W_K · C
    a(t) ← softmax(QK^⊤)
    return a(t)
end procedure
```

# Computational Complexity

## Memory Update

The update of each weight $\omega_i(t)$ depends on a dot product and scalar operations:

$\mathcal{O}(n \cdot d)$ where $n = |V(t)|, d =$ dimension of embeddings

## Narrative Construction

Assuming we use a dynamic path algorithm (e.g., modified Bellman-Ford):

$\mathcal{O}(|V| + |E|)$ to select the path with the largest affective sum

## Attention Calculation

Matrix-matrix product between queries and keys:

$\mathcal{O}(n \cdot d^2)$ where $n =$ memories, $d =$ semantic dimensionality

## Total Memory

Each memory occupies $\mathcal{O}(d)$ and we maintain $n$ memories with connections:

$$\mathcal{O}(n \cdot d + |E|)$$

# Experimental Results

In this section, we present an in-depth set of computational experiments aimed at validating the Graph-based Emergent Narrative Intelligence Architecture for Synthetic Cognition (GENIA). Using `JAX`, `Haiku`, `NumPy`, `Matplotlib`, and `NetworkX`, we implement synthetic agents with affective memory, vectorial attention, narrative structure, and predictive world modeling. The experiments include both isolated simulations and social interactions between agents.

## Experimental Setup

### Dimensionality and Parameters

- Dimensionality of semantic embeddings: $D = 64$
- Number of memory nodes per agent: $N = 10$
- Number of simulated time steps: $T = 10$
- Cognitive constants: defined heuristically and empirically via simulation

### Frameworks Used

- **JAX** for vectorized computation and automatic differentiation.
- **Haiku** for defining the *World Model*, a predictive neural network.
- **NumPy** and **Matplotlib** for visualization and statistical analysis.
- **NetworkX** for representation and visualization of the affective graph.

**Internal Model Architecture** The internal world model was implemented as a *fully connected* neural network with a hidden layer of 128 units and output of dimension $D = 64$, responsible for predicting the next sensory input based on the average of the semantic vectors of the narrative nodes.

## Cognitive Cycle and Simulation

For each agent, we simulate a complete cognitive cycle for $T = 10$ time steps. At each instant $t$, the following steps are performed:

1. Computation of the attention vector $a(t)$ modulating the sensory input $x_t$ by the essence $E$;
2. Construction of the narrative $N(t)$ by ordering the nodes with the highest weighted affective weight;

3. Prediction of the next input $\hat{x}_t$ via the *World Model*;
4. Calculation of emotional reinforcement proportional to the weighted prediction error by $|\nu_i|$;
5. Updating weights $\omega_i$ and valuations $\nu_i$ in affective memory.

**Social Interaction** After a round of individual cognition, we simulate social interaction by randomly exchanging a memory node between two agents ($A_1$ and $A_2$). This intersubjective projection mechanism allows us to observe the incorporation of external elements into the individual affective network.

## Cognitive Metrics and Quantitative Assessment

We adopt two main metrics to quantify narrative quality. As shown in Figure 2, narrative coherence increases over time while entropy decreases:

- **Narrative Coherence:** $\sum_{i \in N(t)} \omega_i \cdot |\nu_i|$, reflecting the affective consistency of the narrative.
- **Narrative Entropy:** $-\sum \omega_i \cdot \log(\omega_i)$, measuring emotional dispersion in prioritized events.

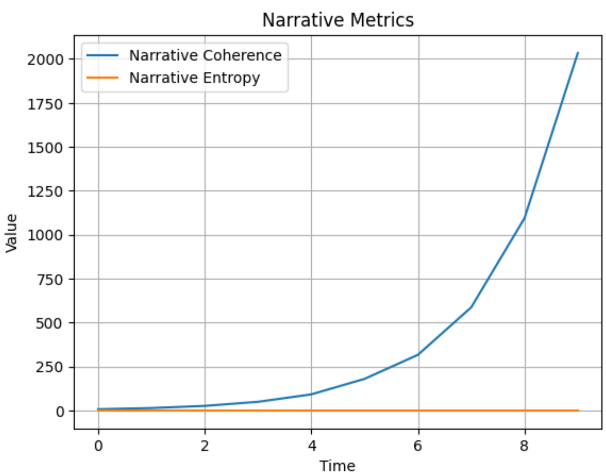


**Figure 2:** *Temporal evolution of narrative coherence and entropy.*

The results show consistent growth in narrative coherence and low entropy, indicating the formation of a stable affective identity over time.

## Memory Evolution and Reinforcement

We recorded the cumulative reinforcement of each memory node over time. This dynamic is illustrated in Figure 3, where a small number of nodes receive the majority of reinforcement. The data indicate that few emotionally intense events dominate the narrative, as expected by the attention and reinforcement mechanism.

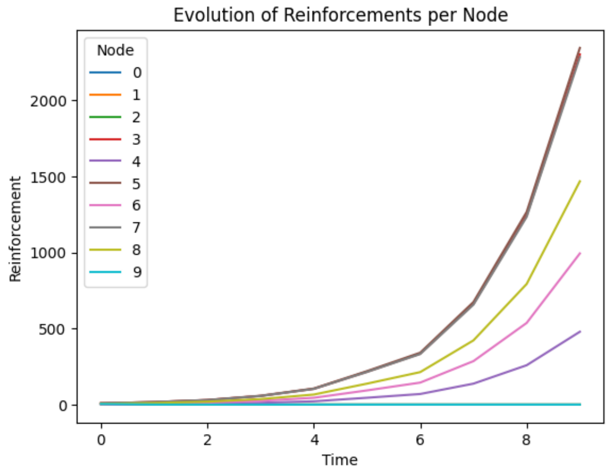


**Figure 3:** *Evolution of reinforcements received by each node over time.*

## Affective Graph Topology

At the end of the simulation, each agent's memory graph is visualized with edges based on semantic similarity (cos( *theta*) $> 0.9$). The emergent structure exhibits features consistent with scale-free networks [10]. The size and color of the nodes represent the affective weight and emotional valuation, respectively (See figure 4. .

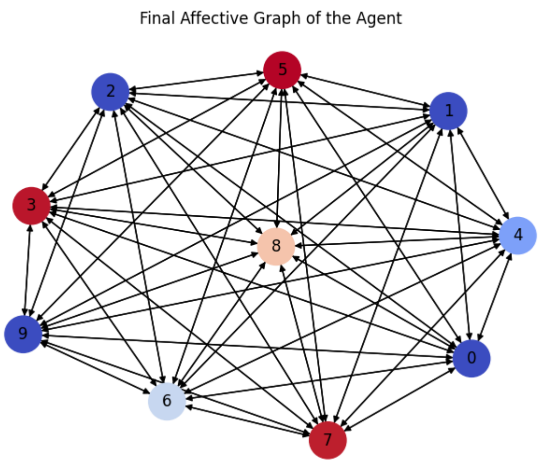


**Figure 4:** *Final affective graph of the agent after 10 cognitive cycles.*

## Collective Cognition and Memory Exchange

To assess collective cognition, we visualized the memory graphs before and after the interaction between two agents. a new node appears isolated in the receiving agent's graph, indicating that the information has not yet been semantically or emotionally integrated. Figure 5 shows the memory graph after social interaction, highlighting an unassimilated external node.

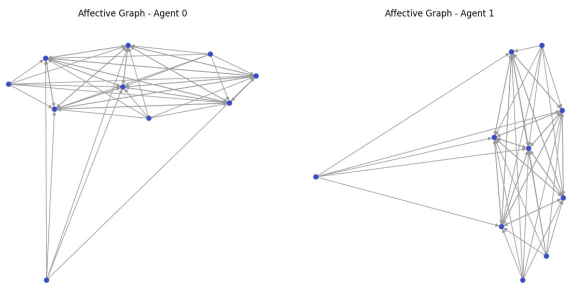


**Figure 5:** *Memory graph of agents after social interaction.*

## Scaling the Experiments

To investigate the scalability, stability and robustness of the GENIA, we conducted two large-scale experiments, significantly increasing the number of memory nodes and temporal cycles. These simulations aim to observe affective dynamics and narrative emergence in scenarios with greater semantic complexity, information density and interagent interactions.

## Experiment with $N = 200$ nodes

In the first scaled experiment, we configured each agent with $N = 200$ affective memory nodes and $T = 100$ cognitive cycles. The architecture remained stable even with this significant increase in mnemonic capacity. We observed an exponential growth in narrative coherence over time, accompanied by a decreasing narrative entropy, which quickly converges to values close to zero. This dynamic indicates the consolidation of a stable, coherent, and affectively dense narrative structure. The trajectory of coherence and entropy in the scaled experiment is shown in Figure 6.

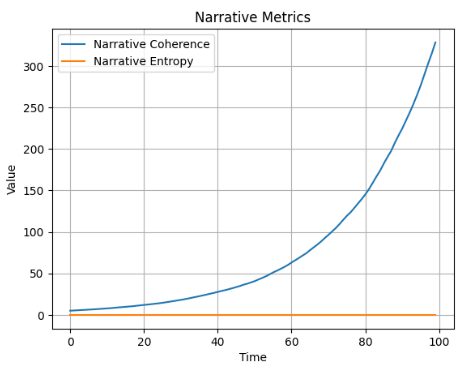


**Figure 6:** *Narrative coherence and entropy over 100 cycles.*

In addition, the evolution of emotional reinforcements reveals an emerging selective pattern: a few emotionally salient events end up dominating the narrative space, with increasing $\omega_i$ weights and consistent $v_i$ valuations. The resulting affective graph is visually cohesive, with strongly connected clusters, suggesting the formation of internal emotional subthemes or cores. The figure 7 and 8 illustrates the resulting affective graph after 100 cycles.

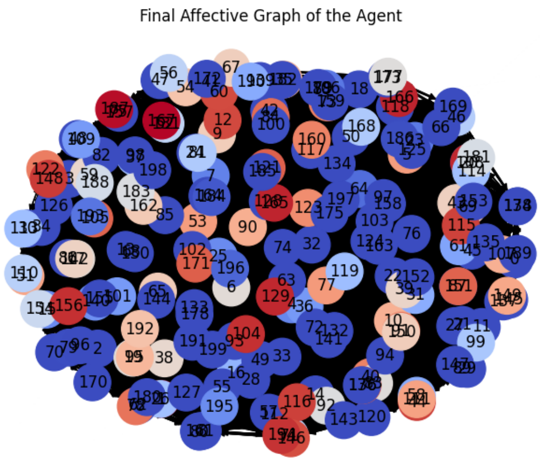


**Figure 7:** *Final affective graph of the agent with $N = 200$ and $T = 100$.*

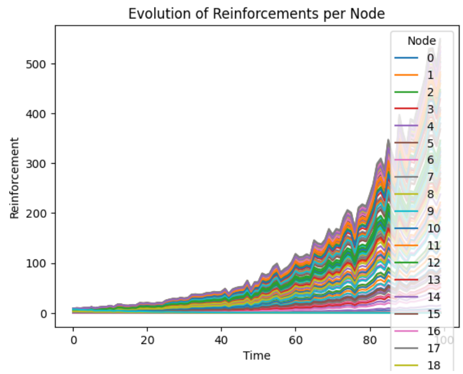


**Figure 8:** *Cumulative evolution of reinforcements per node over time.*

## Experiment with $N = 200$ nodes and $A = 4$ agents

The second experiment aimed to test the system's ability to support synthetic cognition in multiple interacting agents with larger memory volumes. We used $N = 200$ nodes per agent and $A = 4$ distinct agents, each with its own essence vector $\mathbf{E}_a$. We also introduced a new graph formation mechanism: instead of a connection by fixed similarity threshold ($\cos(\theta) > 0.9$), we used the $k$-nearest neighbors approximation ($k = 5$) to ensure relevant connectivity in high-dimensional semantic spaces.

At the end of the simulation, we extracted and analyzed the affective graphs of each agent. The results are highly positive:

- All graphs had 200 nodes and about 1190 edges, with an average degree $\approx 11.9$.
- Graph density remained low ( $approx 0.03$), reflecting affective selectivity.
- The mean clustering coefficient was extremely high ( $approx 0.97$), indicating internal emotional cohesion.
- Each graph contained only one connected component, evidencing the emergence of a global narrative identity per agent.

Visually (See Figure 9), the graphs present densely connected cores with scattered peripheries, simulating the consolidation of central memories versus marginal events. The topologies are distinct among the agents, but with similar structural properties, suggesting functional convergence despite essential individuality.

## Quantitative Structural Analysis

To reinforce the qualitative analysis, we extracted structural statistics from each agent's graphs. The metrics included number of nodes and edges, average degree, density, clustering coefficient, and number of connected components. The results were highly consistent across agents, with negligible variation, demonstrating interagent stability of the affective-narrative architecture.

## Summary of Results

Both experiments confirm that TCSGD is scalable and resilient, maintaining essential cognitive properties such as narrative coherence, affective selectivity, and hierarchical memory structure even under high semantic load. The model supports multiple agents with simple social interactions, opening space for future investigations on collective cognition, narrative fusion, the emergence of transindividual memories, and intersubjective dynamics.

## Discussion

The experiments show that the proposed synthetic cognition based on dynamic affective graphs is scalable, resilient and capable of maintaining human cognitive properties such as cohesion, selectivity, persistent narrative and affective integration. Narrative evolution tends to converge towards configurations of low entropy and high organization, suggesting the emergence of coherent synthetic identities over time, even under significant variations in the number of memories or agents. GENIA integrates cognitive processes inspired by the vision of neuroscience-grounded AI [11].

The next steps include the temporal analysis of the formation of semantic communities, the quantification of narrative plasticity and the introduction of contradictory stimuli to investigate the system's ability to deal with cognitive dissonance.

## Conclusion of the Experimental Section

The results confirm the functionality of GENIA. The proposed cognitive architecture presents:

- Ability to construct cohesive affective narratives;
- Selectivity and plasticity of memory by emotional reinforcement;
- Emergence of stable narrative identity;
- Partial integration of external information after social interaction.

The algebraic and topological structure of graphs, combined with attention and reinforcement mechanisms, allows synthetic agents to develop interpretable mental states. GENIA proves to be a viable framework for modeling subjective artificial cognition.

## Computational Considerations

The system maintains practical viability with computational complexity:

- Linear in $n$ for memory and narrative updating;
- Quadratic in $D$ for calculating the attention vector.

Therefore, the current implementation behaves efficiently for $n$
$leq 10^4$, compatible with real applications in scalable cognitive AI.

# Related Works

The proposal for the Graph-based Emergent Narrative Intelligence Architecture for Synthetic Cognition is inspired by and, at the same time, differentiates itself from several approaches on the frontier between computational neuroscience, symbolic models of cognition and architectures of cognitive agents. In this section, we review the most relevant works and situate our contribution in the current context of research in general artificial intelligence and cognitive systems.

## Classical Cognitive Architectures

Models such as **SOAR [laird2012soar]** and **ACT-R [anderson1997act]** propose symbolic cognitive architectures, in which cognition is viewed as manipulation of symbols within modular structures. Although these architectures have been instrumental in modeling human behavior, their structural rigidity and low plasticity limit their applicability in dynamic and unsupervised environments. Our model, on the other hand, is non-symbolic, continuous, and formulated in terms of graphs and differential dynamics.

## Memory and Graph-Based Models

Systems such as **Differentiable Neural Computers (DNCs) [graves2016hybrid]** and **Graph Neural Networks (GNNs) [scarselli2008graph, battaglia2018relational]** treat memory as a dynamic structure that can be addressed in a differentiable way. Although DNCs allow explicit read/write operations, their content is not semantically structured in terms of affect or subjective value. GNNs, in turn, treat graphs as inputs, but not necessarily as internal cognitive structures with narrative and identity. Our model proposes an affective graph-oriented memory, where each node has emotional and semantic meaning, and the dynamics are guided by attention, reinforcement, and decay.

## Models Inspired by Neuroscience and Psychology

Friston's *Free Energy Principle* [6] model the brain as a system that seeks to minimize surprise through Bayesian

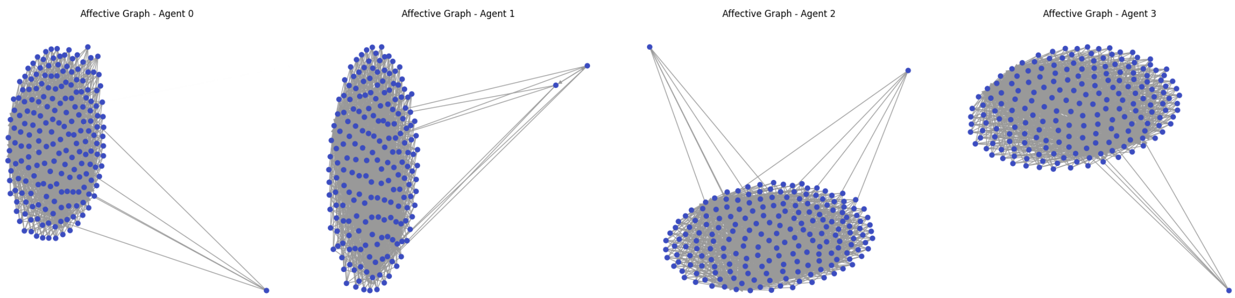


**Figure 9:** *Final affective graphs of the four agents with* $N = 200$ *each.*

inference. There are conceptual points of contact with our model, especially regarding the *world model* constructed from experience and prediction. However, our framework explicitly proposes the representation of the self as an emotional narrative constructed over a graph, and not just as an inferential agent.

Unlike symbolic architectures such as SPAUN [**eliasmith2012large**], which use symbolic neural networks, also seek to simulate large-scale complex cognition, but are highly specialized and rely on representation engineering. In contrast, our proposal allows the emergence of narrative and affective structure without explicit supervision or predefined symbols.

## Collective Cognition and Multi-User Agents

Works such as Minsky's Society of Mind [**minsky1988society**] and contemporary approaches in *multi-agent systems* discuss the emergence of cognition from the interaction between entities. Our model formalizes this notion through the exchange of memory subgraphs between agents, allowing simulations of phenomena such as empathy, projection and intersubjective simulation. This aspect is still little explored in current formal models and constitutes one of the most original contributions of this theory.

## Original Contribution

Unlike previous approaches, the Graph-based Emergent Narrative Intelligence Architecture for Synthetic Cognition proposes:

- A **affective memory** modeled as a directed graph with emotional weight and valuation.
- A **attention vector** computed by mechanisms inspired by Transformers, but guided by emotional essence.
- The emergence of an **identity narrative** as an optimal affective path in the graph.
- A Bayesian world model inferred from experiential history and internal bias.
- Social interaction via **explicit memory exchange** between synthetic agents.

These elements integrate concepts of dynamic systems, distributed cognition, narrative theory and affective learning, offering a new paradigm for building more human, contextual and self-aware artificial intelligences. This reflects Tegmark's perspective on conscious entities as mathematical structures [**tegmark2014mathematical**].

## Future Work and Open Questions

To advance the theory, we suggest:

- Incorporate adaptive learning to adjust $\mathcal{E}$ dynamically.
- Model emotional conflicts and their resolution in the memory graph.
- Explore multi-agent systems to simulate synthetic societies.

Open questions include:

- How does the model deal with conflicting information in memory?
- What are the ethical implications of creating synthetic beings with narratives and emotions?

## References


[1] Hofstadter, D. R. (1979). *Gödel, Escher, Bach: an Eternal Golden Braid*. Basic Books.

[2] Dennett, D. C. (1991). *Consciousness Explained*. Little, Brown and Co.

[3] Friston, K. (2010). The free-energy principle: a unified brain theory?. *Nature Reviews Neuroscience*, 11(2), 127–138.

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

[5] Maturana, H. R., & Varela, F. J. (1980). *Autopoiesis and Cognition: The Realization of the Living*. Springer.

[6] Rosenberg, G. (2020). *Information and the Internal Structure of the Universe*. Springer.

[7] Hopfield, J. J. (1982). Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the National Academy of Sciences*, 79(8), 2554–2558.

[8] Hebb, D. O. (1949). The organization of behavior: A neuropsychological theory. *Wiley*.

[9] Barabási, A.-L., & Albert, R. (1999). Emergence of scaling in random networks. *Science*, 286(5439), 509–512.

[10] Sporns, O. (2010). *Networks of the Brain*. MIT Press.

[11] Dehaene, S. (2014). *Consciousness and the Brain: Deciphering How the Brain Codes Our Thoughts*. Viking.

[12] Prigogine, I., & Stengers, I. (1984). *Order out of Chaos: Man's New Dialogue with Nature*. Bantam.

[13] Shannon, C. E. (1948). A mathematical theory of communication. *Bell System Technical Journal*, 27, 379–423, 623–656.

[14] Tegmark, M. (2014). *Our Mathematical Universe: My Quest for the Ultimate Nature of Reality*. Knopf.

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

[16] Jaynes, E. T. (2003). *Probability Theory: The Logic of Science*. Cambridge University Press.

[17] Lake, B. M., Ullman, T. D., Tenenbaum, J. B., & Gershman, S. J. (2017). Building machines that learn and think like people. *Behavioral and Brain Sciences*, 40.

[18] Bengio, Y., Courville, A., & Vincent, P. (2013). Representation learning: A review and new perspectives. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(8), 1798–1828.

[19] Smolensky, P. (1986). Information processing in dynamical systems: Foundations of harmony theory. In *Parallel Distributed Processing: Explorations in the Microstructure of Cognition* (Vol. 1, pp. 194–281). MIT Press.

[20] Hassabis, D., Kumaran, D., Summerfield, C., & Botvinick, M. (2017). Neuroscience-inspired artificial intelligence. *Neuron*, 95(2), 245–258.

[21] Pearl, J. (2009). *Causality: Models, Reasoning, and Inference*. Cambridge University Press.

[22] Parr, T., Pezzulo, G., & Friston, K. J. (2022). *Active Inference: The Free Energy Principle in Mind, Brain, and Behavior*. MIT Press.

[23] von Bertalanffy, L. (1968). *General System Theory: Foundations, Development, Applications*. George Braziller.

[24] Eliasmith, C., Stewart, T. C., Choo, X., Bekolay, T., DeWolf, T., Tang, Y., & Rasmussen, D. (2012). A large-scale model of the functioning brain. *Science*, 338(6111), 1202–1205.