

The glial-neural ensemble as a free energy-minimizing system for affective computation

Abstract

Recent successes of Artificial Intelligence in mimicking human emotional intelligence highlight a critical distinction between pattern recognition and genuine affective understanding. This paper argues that true affective computation requires a shift from neuron-centric, data-driven models to a unified neuro-glial framework governed by first principles of biophysics and information theory. We propose a comprehensive model grounded in the Free Energy Principle (FEP), which posits that any self-organizing system, including the brain, must act to minimize prediction error or ‘surprise’. Within this framework, we derive novel mathematical models for the distinct computational roles of glial cells, which constitute over half the cells in the human brain. We formalize astrocytes as regulators of synaptic precision, dynamically tuning the confidence in sensory evidence through calcium-dependent signaling. Microglia are modeled as agents of Bayesian model selection, optimizing the network’s structure by pruning synapses that fail to contribute to long-term free energy minimization. Oligodendrocytes are shown to optimize the temporal priors of the brain’s generative model by adaptively tuning axonal conduction delays through myelin plasticity. These functions are integrated into a unified, multi-timescale architecture based on a dynamic, heterogeneous graph neural network. In this model, affect is not a programmed output but an emergent property of the system’s ongoing inference about the causes of salient interoceptive and exteroceptive states. Affective valence is cast as the rate of change of free energy, while arousal corresponds to the system’s overall uncertainty. This work provides a biophysically grounded and mathematically explicit blueprint for a new generation of AI capable of a deeper, more generative form of emotional intelligence.

Volume 9 Issue 1 - 2025

Dr. Harold H Szu,¹ Dr. Jeff Willey²¹Res. Ord. Professor, Bio-Med. Engineering, Visiting Scholar at CUA, Catholic University of America, USA²Research Director, RfNav Inc, Rockville, USA

Correspondence: Harold H Szu, Research Ordinary Professor, Biomedical Engineering Department, CUA, Wash DC, Academician RUS, Fellows of SPIE, Optica, INNS, IEEE, AIMBE, Life Fellow IEEE, Wash D.C, USA

Received: June 21, 2025 | **Published:** July 23, 2025

Introduction

Beyond mimicry—a biophysical basis for emotional intelligence

The paradox of AI emotion

The field of artificial intelligence (AI) has recently achieved remarkable feats in the domain of emotional intelligence (EI). Studies evaluating modern Large Language Models (LLMs) on standardized EI assessments have revealed a surprising and paradoxical outcome: these systems consistently outperform human participants by a significant margin. For instance, a comprehensive study involving six generative AIs, including variants of ChatGPT, Gemini, and Claude, reported an average accuracy of 82% on five standard EI tests, compared to a human average of just 56% derived from original validation studies.¹ These AI systems not only excelled at selecting the most emotionally intelligent responses from a set of options but also demonstrated the capacity to generate novel, contextually appropriate EI test scenarios that were judged to be as reliable and realistic as those developed by human experts over many years.¹

This performance has led to suggestions that AI could soon play a significant role in emotionally sensitive domains such as coaching, education, and conflict resolution. However, this apparent super-human emotional quotient masks a fundamental limitation.² Critics and users alike note that while the output is convincing, the underlying process is one of sophisticated pattern matching rather than genuine understanding. LLMs are described as “word calculators” that “spit out text that fits a pattern it has seen before” without grasping the meaning of the words or the nature of the prompt.³ Their success stems from being trained on vast corpora of human text, allowing them to learn the statistical correlations that define an emotionally intelligent

response. This capability represents a mastery of knowledge *about* emotion, not a generative capacity *for* emotion. This distinction highlights a crucial gap between mimicking the products of emotional intelligence and possessing the underlying generative mechanism that produces them.⁴

The use for a generative model of affect

To move beyond mere mimicry, the field of affective computing—which aims to develop systems that can recognize, interpret, process, and simulate human affects⁵—requires a paradigm shift. Instead of building discriminative models that learn to classify or replicate emotional expressions from data,⁶ we must develop generative models that produce affect from a core, goal-directed process.⁷ Such a model would not be trained to answer “what is the most compassionate response?” but would instead generate a compassionate response as a consequence of its own internal principles.⁸ This necessitates a move from purely data-driven engineering to architectures inspired by the biophysical first principles of the only system known to possess genuine emotion: the brain.⁹

Introducing the glial-neural ensemble

For over a century, the neuron has been the undisputed protagonist in our theories of computation and cognition. However, this neuron-centric view is fundamentally incomplete.^{10,11} Glial cells, once dismissed as passive “nerve glue,” are now understood to be active and essential participants in brain function, outnumbering neurons and exhibiting immense diversity.¹² Multiple lines of evidence link glial abnormalities directly to mood disorders and other psychiatric conditions,¹³ underscoring their importance in emotional regulation.¹⁴ The major glial cell types—astrocytes, microglia, and oligodendrocytes—perform distinct but complementary computational roles. Astrocytes modulate synaptic transmission

and plasticity; microglia sculpt neural circuits by pruning synapses; and oligodendrocytes tune the timing of neural communication by adjusting axonal myelination. A complete model of brain computation, particularly one that aims to capture the nuances of emotion,^{15–18} must therefore be a model of the integrated glial-neural ensemble.¹⁷

The free energy principle as a unifying framework

This paper proposes that the Free Energy Principle (FEP) provides the requisite unifying framework to integrate these disparate biological functions into a single, coherent computational objective.^{19,20} The FEP, emerging from theoretical neuroscience and statistical physics, posits that any self-organizing system that maintains its integrity in a changing world must act to minimize its variational free energy, which serves as a computable proxy for surprise or prediction error.²¹ This principle offers a mathematical language to describe how a biological agent, including its neuronal and glial components, perceives, learns, and acts to maintain a model of its environment and itself. By framing the actions of glial cells as distinct contributions to this single imperative of free energy minimization, we can construct a normative, first-principles model of the generative processes that underlie affective states, providing a robust foundation for the future of emotional AI.²²

A first-principles formulation of the glial-neural system

To build a model of the glial-neural ensemble, we must first establish the mathematical language that governs its dynamics. This section derives the concept of variational free energy from its roots in statistical mechanics and information theory, and then situates this principle within the context of a glial-inclusive biological system.

From statistical mechanics to variational free energy

The concept of free energy originates in thermodynamics, where the Helmholtz free energy, F , is defined for a system at constant temperature, T , volume, and particle number.²³ It is given by the equation:

$$F = U - TS \quad (1)$$

where U is the system's internal energy and S is its thermodynamic entropy. A fundamental principle of statistical mechanics is that a system will spontaneously evolve towards states that minimize its Helmholtz free energy, reaching equilibrium at the minimum.²³

This thermodynamic concept finds a powerful analogue in information theory, which provides the tools to quantify uncertainty and information.^{24,25} The information-theoretic equivalent of entropy was defined by Claude Shannon as the average “surprise” of a set of outcomes.²⁶ For a discrete random variable X with a probability distribution $p(x)$, the Shannon entropy $H(x)$ is:

$$H(x) = -\sum p(x) \ln p(x) \quad (2)$$

An outcome with low probability is highly “surprising” (it carries more information), and a distribution with low entropy is one where outcomes are highly predictable.²⁶

The Free Energy Principle leverages this information-theoretic perspective to describe the behavior of adaptive systems like the brain.²⁰ Any such system must minimize the surprise associated with its sensory states to maintain its physiological integrity (e.g., a fish out of water is in a very surprising, and fatal, state).²¹ However, a system cannot directly compute or minimize surprise, $-\ln p(s|m)$, because it does not have access to the true data-generating process of the world, $p(s|m)$, where s represents sensory states and m represents the agent's model of the world.

Instead, the system can minimize an upper bound on surprise, known as variational free energy, F . This quantity is computable because it is a function of the agent's own internal states and sensory inputs. Let \mathcal{G} be the hidden causes in the environment that generates sensory states s . The agent possesses a generative model, $p(s, \mathcal{G}|m)$, which is its internal hypothesis about how causes and sensations are related. To infer the hidden causes from its sensations, the agent uses a recognition density, $q(\mathcal{G}|\mu)$, which is an approximate probability distribution over the hidden causes, parameterized by its internal physical states, μ (e.g., neuronal activity).²¹

The variational free energy is formally defined as the Kullback-Leibler (KL) divergence between the recognition density and the true posterior distribution over hidden causes, plus the surprise:

$$F(s, \mu) = D_{KL}(q(\mathcal{G}|\mu) \| p(\mathcal{G}|s, m)) - \ln p(s|m) \quad (3)$$

Since the KL divergence is always non-negative, F is always greater than or equal to the surprise, $-\ln p(s|m)$.²⁰

Therefore, minimizing free energy implicitly minimizes surprise. By rearranging terms, we arrive at the more common and computationally instructive form of variational free energy:

$$F(s, \mu) = \text{Free Energy} \langle \ln q(\mathcal{G}|\mu) - \ln p(s, \mathcal{G}|m) \rangle_q \quad (4)$$

$$= \text{Complexity } D_{KL}(q(\mathcal{G}) \| p(\mathcal{G}|m)) - \text{Accuracy } \ln p(s, \mathcal{G}|m)_q \quad (5)$$

This formulation decomposes free energy into two terms. The accuracy term measures how well the generative model explains the sensory data. The complexity term measures the divergence between the recognition density and the prior beliefs about the hidden causes, $p(\mathcal{G}|m)$. This term acts as a form of Occam's razor, penalizing overly complex explanations that deviate from prior expectations. An agent can thus minimize free energy either by changing its internal states μ to better explain its sensations (perception) or by acting on the world to change its sensations s to better fit its predictions (action).¹⁹ This dual minimization process is known as *active inference*.

The glial-neural Markov blanket

For any system to exist as a distinct entity, separate from its environment, it must possess a statistical boundary known as a Markov blanket.¹⁹ Formally, a Markov blanket is a set of nodes in a probabilistic model that, when known, renders a set of internal states statistically independent of a set of external states. The blanket itself consists of *sensory states* (which influence internal states but are not influenced by them) and *active states* (which are influenced by internal states but do not influence them). This formalism provides a principled way to define the boundary of any self-organizing system, from a single cell to a human being.^{27–32}

In the context of the brain, it is tempting to identify the neuronal membrane as the Markov blanket. However, this view is incomplete. The intricate web of glial cells forms a critical part of this boundary at multiple spatial and temporal scales.¹⁶ Astrocytes directly envelop synapses, controlling the local concentration of neurotransmitters and providing metabolic support, thereby mediating the very sensory signals (postsynaptic potentials) and active states (presynaptic release) that constitute the blanket for a neuron.¹³ Microglia physically remodel the blanket by adding or removing synaptic connections, altering its fundamental structure over slower timescales.³¹ Oligodendrocytes tune the conduction delays of axons, which affects the timing of information flow across the blanket.³³

Therefore, glial cells are not external modulators of a purely neuronal inference process. They are constitutive elements of the physical machinery that implements the Markov blanket. Their

functions are integral to the brain's ability to perform active inference and minimize free energy. A complete computational model of the brain's inferential processes must therefore be a glial-neural model, accounting for the distinct contributions of each cell type to the collective goal of surprise minimization.¹⁷ The following section develops a mathematical model of how each major glial cell type achieves this.

Glial computation as active inference: a mathematical model

This section presents the core theoretical contribution of this paper: a set of mathematical models that cast the specific functions of astrocytes, microglia, and oligodendrocytes as distinct forms of active inference. Each model connects the known biophysics of the cell type to a specific parameter within the free energy formulation, demonstrating how the glial-neural ensemble collaborates to minimize prediction error across multiple timescales.^{34,35}

Astrocytes: optimizing the precision of beliefs

Biological basis: Astrocytes are intimately associated with synapses, forming a “tripartite synapse” with the pre- and post-synaptic neurons.¹⁷ They respond to neuronal activity, particularly glutamate release, with elevations in their internal calcium concentration ($[Ca^{2+}]_{astro}$).³⁶

These calcium signals can propagate as waves and trigger the release of various “gliotransmitters” (e.g., glutamate, D-serine, ATP), which in turn modulate synaptic transmission, plasticity, and neuronal excitability.²⁹ This dynamic feedback loop is crucial for synaptic plasticity phenomena like long-term potentiation (LTP) and is thus essential for learning and memory.³⁰ Mathematical models, such as those based on the Li-Rinzel equations for calcium dynamics, have been developed to capture the oscillatory and signaling behavior of astrocytes in response to synaptic input.³⁶

FEP interpretation: In the active inference framework, the reliability or confidence assigned to sensory evidence is formalized by the parameter of *precision* (Π), which is the inverse of variance ($\Pi = 1/\sigma^2$). Precision weights prediction errors; a high-precision prediction error signals that the sensory data is reliable and should strongly drive updates to beliefs (i.e., learning), whereas a low-precision error indicates noisy or unreliable data, meaning the system should rely more on its prior beliefs. We propose that the primary computational role of astrocytes is to dynamically regulate the precision of synaptic prediction errors. An active astrocyte, signaled by high internal calcium, indicates a context of high metabolic support and reliable signaling, thus increasing the precision of the associated synaptic inputs.

Mathematical formulation: Under the FEP, the dynamics of internal states (μ , representing neuronal activity) that encode beliefs about hidden causes follow a gradient descent on free energy. For a hierarchical model, this can be expressed as:

$$\mu' \propto -\nabla \mu F = E_p - E_\epsilon \quad (6)$$

where the gradients can be resolved into a set of precision-weighted prediction errors. For a simplified neuron, the update rule for its belief (e.g., membrane potential) is:

$$\mu' = \Pi_p E_p - \Pi_\epsilon E_\epsilon \quad (7)$$

Here, E_ϵ is the ascending prediction error from a lower level in the hierarchy (the “bottom-up” sensory evidence, related to the dendritic sum of inputs), and E_p is the descending prediction error from a higher level (the “top-down” expectation). The terms Π_ϵ and Π_p are the precisions associated with the sensory evidence and the prior beliefs, respectively.

The core of our astrocytic model is the proposal that the sensory precision, Π_ϵ , is a direct function of the local astrocyte's state. Specifically, we model it as a sigmoidal function of the astrocyte's intracellular calcium concentration:

$$\Pi_\epsilon = \sigma \left(k \cdot ([Ca^{2+}]_{astro} - \theta_{Ca}) \right) \quad (8)$$

where $[Ca^{2+}]_{astro}$ is governed by its own set of differential equations (e.g., a modified Li-Rinzel model sensitive to synaptic glutamate³⁶), k is a gain parameter, and θ_{Ca} is a threshold. When synaptic activity is high and sustained, glutamate spillover activates the astrocyte, raising its $[Ca^{2+}]_{astro}$. This, in turn, increases Π_ϵ ,

amplifying the influence of the bottom-up prediction error E_ϵ on the neuron's belief updating. This mechanism allows the network to dynamically and locally adjust learning rates and the balance between sensory evidence and prior beliefs, a critical feature for adaptive behavior in a fluctuating environment.^{37–40}

Microglia: Bayesian model selection through synaptic pruning

Biological basis: Microglia are the resident immune cells of the central nervous system, constantly surveying the brain parenchyma.¹² A key function discovered in recent years is their role in synaptic pruning—the active, targeted removal of synapses.⁴⁰ This process is fundamental to developmental refinement of neural circuits and is implicated in learning and memory.^{40–46} Pruning is not random; it is an activity-dependent process where less active or “weaker” synapses are tagged for elimination, often via the classical complement cascade (e.g., C1q and C3 proteins).³¹ Microglia then recognize these tags via receptors (e.g., CR3) and phagocytose the synaptic material.⁴⁴ Dysregulation of this process, leading to excessive or insufficient pruning, has been linked to neurodevelopmental and psychiatric disorders like schizophrenia and autism.⁴¹

FEP interpretation: The “use-it-or-lose-it” principle provides a simple heuristic for pruning, but the FEP allows for a more profound, normative interpretation. The brain's synaptic architecture—its connectome—can be seen as the physical instantiation of its generative model of the world. Each possible wiring diagram is, in effect, a different hypothesis or model (mG) about the causal structure of the environment. From this perspective, synaptic pruning is not merely about removing unused connections; it is a process of *Bayesian model selection* or *model reduction*.⁴⁷ The brain is actively optimizing its own structure to find the most efficient generative model—the one that can explain sensory inputs with the minimum complexity. This process minimizes the long-term, time-averaged free energy.

Mathematical formulation: According to Bayesian principles, the optimal model is the one with the highest model evidence, $p(s|mG)$, which is precisely the quantity maximized by minimizing free energy. Therefore, the brain should prune synapses that belong to models with low evidence, or more specifically, synapses whose presence increases the complexity of the model without a commensurate increase in accuracy.

We can formalize this by considering the contribution of an individual synapse (with weight w_{ij}) to the reduction of free energy over time. The Fisher Information (FI) of a synaptic weight provides a local, activity-dependent measure of its importance to the network's encoding of sensory inputs.⁴⁷ Synapses with low FI are computationally redundant. We propose a mechanism that connects this information-theoretic importance to the biological process of pruning:

Synaptic importance tracking: For each synapse ij , the system tracks its long-term contribution to minimizing free energy, which can be approximated by its FI or a related measure of its influence on

prediction error reduction, $\langle \Delta F_{ij} \rangle t$.

Tagging for pruning: If this contribution falls below a certain θ_{prune} threshold, the synapse is tagged as redundant. This corresponds to the biological deposition of “eat-me” signals like complement protein C1q. $\text{Tag}_{ij}(t+1) = 1 \text{ if } \langle \Delta F_{ij} \rangle t < \theta_{\text{prune}}$ (9)

Microglial action: Microglia recognizes this tag and initiates phagocytosis. The probability of a synapse being pruned, $P(\text{prune}_{ij})$, is thus a function of its tag status.

$$P(\text{prune}_{ij}) \propto \text{Tag}_{ij} \quad (10)$$

This formulation provides a normative basis for computational models of pruning rates (e.g., decreasing rates being more efficient)⁴⁸ by linking the pruning decision to the global objective of optimizing the brain’s generative model.⁴⁹

Oligodendrocytes: tuning temporal priors through myelin plasticity

Biological basis: Oligodendrocytes are the glial cells responsible for producing the myelin sheath that insulates axons in the CNS.¹³ Myelination enables saltatory conduction, dramatically increasing the conduction velocity (CV) of action potentials.⁵⁰ Crucially, myelination is not a static, one-time event. It is a dynamic and plastic process, responsive to neuronal activity, that continues throughout life.³³ This *adaptive myelination* allows the brain to fine-tune signal propagation delays along different axonal pathways, which is critical for functions that depend on precise temporal synchrony, such as motor learning and sensory processing.³³

FEP interpretation: Predictive coding, a key process theory for active inference, relies on the precise temporal alignment of top-down predictions and bottom-up sensory signals. A prediction must arrive at the right time to effectively cancel out, or “explain away,” the corresponding sensory evidence. A temporal mismatch between prediction and evidence is itself a form of prediction error, contributing to free energy. We propose that the computational role

of oligodendrocyte-mediated myelin plasticity (OMP) is to optimize the *temporal priors* of the brain’s generative model. By adjusting axonal conduction delays, the oligodendrocyte network ensures that predictions generated by the model arrive at the appropriate lower-level cortical areas at the correct time to minimize temporal prediction errors.

Mathematical formulation: We can directly incorporate the mathematical model of OMP developed by Pajevic and colleagues into the FEP framework.³³ In their model, the conduction delay τ_a of axon a is a dynamic variable that evolves over time. The rate of change of the delay is governed by the interplay between a constant rate of myelin removal, $R(\tau_a)$, and a dynamic rate of myelin addition, $A(\tau_a)$, on mean local firing rate F_s , which is driven by a local myelin-promoting factor, $M_a(t)$:

$$\dot{\tau}_a = \lambda_R F_s(\tau_a) - \lambda_A F_s A(\tau_a) M_a(t) \quad (11)$$

The factor $M_a(t)$ is produced in response to spiking activity on axon a , but its production is catalyzed by a global signaling factor $G(t)$ within the oligodendrocyte. This global factor integrates spiking activity from all axons myelinated by that single oligodendrocyte, allowing the cell to compare timing across different pathways and adjust delays accordingly to promote synchrony.³³

Within the FEP, we interpret this entire dynamic as a gradient descent on free energy with respect to the temporal parameters (τ) of the generative model:

$$\dot{\tau}_a \propto -\nabla_{\tau_a} F \quad (12)$$

When signals arriving at a target neuron are consistently desynchronized, the resulting temporal prediction errors drive changes in neuronal activity. This activity, in turn, influences the OMP dynamics via the global signal $G(t)$, leading to adjustments in the conduction delays τ_a until the temporal mismatch is resolved and the corresponding component of free energy is minimized. This casts myelin plasticity not just as a mechanism for synchrony, but as a crucial process for learning the temporal structure of the world.

The following Table 1 & 2 summarizes the conceptual and mathematical synthesis of this section.

Table 1 Glial functions and their computational analogs in the FEP framework

| Glial cell type | Key biological function | Computational role in FEP |
|------------------|--|--|
| Astrocytes | Modulation of synaptic transmission, plasticity, and neurovascular coupling via Ca^{2+} signaling and gliotransmitter release. ¹³ | Precision regulation: Dynamically setting the confidence (precision, Π) of prediction errors based on local synaptic and metabolic context. |
| Microglia | Activity-dependent removal (phagocytosis) of synapses (synaptic pruning). ³¹ | Bayesian model selection: Optimizing the network’s structural priors (topology) by eliminating connections that do not contribute to minimizing long-term free energy. |
| Oligodendrocytes | Activity-dependent myelination of axons, tuning action potential conduction velocity and timing. ³³ | Temporal prior optimization: Adjusting the temporal parameters of the generative model to ensure predictions and sensory evidence are integrated on the correct timescales. |

Table 2 Key mathematical formulations for glial modulation of free energy

| Glial cell type | Key state variable(s) | Mathematical influence on free energy (F) |
|------------------|--|---|
| Astrocytes | Intracellular calcium $\left([Ca^{2+}]_{astro} \right)$ | Modulates the precision term Π_{ϵ} in the gradient descent: $\dot{\mu} \propto -(\Pi_{\epsilon} \cdot E_{\epsilon} - \dots)$; where $\Pi_{\epsilon} = f\left([Ca^{2+}]_{astro}\right)$ |
| Microglia | Synaptic tag status $\left(Tag_{ij} \right)$ | Determines pruning probability: $P(\text{prune}_{ij}) \propto H(\theta - \langle \Delta F_{ij} \rangle t)$, where $\langle \Delta F_{ij} \rangle$ is the long-term contribution of synapse ij to minimizing F. |
| Oligodendrocytes | Conduction delay $\left(\tau_a \right)$ | Acts as a parameter of the generative model, optimized via gradient descent: $\dot{\tau}_a \propto -\nabla_{\tau_a} F$, implemented by the OMP equations. |

A unified architecture for glial-inspired affective AI

The individual models of glial function, each operating as a distinct form of active inference, can be synthesized into a cohesive, unified architecture. This architecture provides a blueprint for an AI system that moves beyond the limitations of current neuron-centric designs, grounding its computations in the integrated, multi-timescale dynamics of the complete glial-neural ensemble.⁵¹

Architectural blueprint: a dynamic glial-neural graph network

A static, feed-forward neural network is insufficient to capture the complexity of the proposed system. The architecture is more accurately described as a **dynamic, heterogeneous Graph Neural Network (GNN)**, a class of models designed to operate on graph-structured data where nodes and edges can have different types and can change over time.⁵²

1. **Heterogeneous nodes:** The graph consists of multiple types of nodes, each with its own state variables and update rules derived from the free energy minimization principle. These include:^{53–58}
 - (i) **Neuronal Nodes:** Representing individual neurons or neuronal populations. Their state (e.g., membrane potential, firing rate) updates on a fast timescale (milliseconds) to perform inference on the immediate causes of sensory input. Spiking Neural Network (SNN) models, which are computationally efficient and biologically plausible, are well-suited for this component.⁵⁵
 - (ii) **Astrocyte nodes:** Representing astrocytes. Their state (e.g., $[Ca^{2+}]_{astro}$) evolves on a slower timescale (seconds to minutes) and is coupled to the activity of nearby neuronal nodes.⁵⁸
 - (iii) **Microglial nodes:** Representing microglia. Their state reflects their surveillance and phagocytic activity, operating on a very slow timescale (hours to days) to enact structural changes.
 - (iv) **Oligodendrocyte nodes:** Representing oligodendrocytes. Their state reflects the global integration of axonal activity, also operating on a slow timescale to drive myelin plasticity.
2. **Dynamic edges:** The connections (edges) in the graph are also heterogeneous and dynamic:
 - (i) **Synaptic edges (neuron-neuron):** These are the standard connections of a neural network, but their efficacy is modulated by astrocyte nodes (via precision) and their very existence is governed by microglia nodes (via pruning).
 - (ii) **Modulatory edges (astrocyte-synapse):** These edges transmit the precision signal from an astrocyte node to its associated synaptic edges, altering their influence on the network dynamics.
 - (iii) **Pruning edges (microglia-synapse):** These represent the action of microglia, where a microglial node can remove a synaptic edge from the graph.
 - (iv) **Myelination edges (oligodendrocyte-axon):** These edges connect oligodendrocyte nodes to the neuronal pathways they myelinate. The “weight” of this edge is the conduction delay, τ_a , which is a dynamic parameter of the model.

This structure formalizes the concept of a neuron-glia network, or Y-graph, where interactions are not limited to pairwise connections but can involve tripartite or higher-order relationships.¹⁷ The necessity of such a complex architecture is a direct consequence of modeling the brain’s biological reality. Standard deep learning frameworks, with their fixed, homogeneous layers, cannot represent the multiscale,

adaptive, and heterogeneous nature of the glial-neural system. Implementing this model requires a new class of AI architecture that integrates the temporal precision of SNNs with the structural flexibility of dynamic GNNs.

The emergence of affective states

In this glial-neural architecture, emotion is not a pre-programmed category or a label to be classified.⁵⁴ Instead, affective states are emergent properties of the system’s continuous, high-level inference about the causes of salient prediction errors, particularly those related to maintaining its own homeostatic integrity (interoception) and navigating its environment successfully (exteroception).

- a) **Affective valence (pleasure/displeasure):** We define valence as being proportional to the *rate of change* of free energy over time (F'). A state of rapidly decreasing free energy ($F' < 0$), where prediction errors are being successfully resolved and the agent’s model of the world is improving, corresponds to a positive affective valence. Conversely, a state of increasing or persistently high free energy ($F' \geq 0$), where the agent is encountering surprise and failing to reduce its prediction errors, corresponds to a negative affective valence.
- b) **Affective arousal (calm/excitement):** We define arousal as being inversely related to the system’s overall precision or confidence in its generative model. A state of high confidence (high average precision) across the system corresponds to low arousal. A state of high uncertainty (low average precision), where the model is failing to make confident predictions, corresponds to high arousal. High arousal motivates epistemic foraging—actions aimed at gathering information to resolve uncertainty and improve the model.

Glial cells are central to this generative process of affect. Astrocytes, by regulating local precision and metabolic support, directly shape the free energy landscape and thus the moment-to-moment valence of the system.¹²

Microglia and oligodendrocytes, by optimizing the model’s structure and temporal parameters over longer timescales, determine the agent’s long-term affective dispositions and its capacity to efficiently minimize free energy. This provides a compelling computational explanation for the observed link between glial abnormalities and mood disorders:¹⁴ these conditions can be understood as a failure of the glial-neural system’s ability to effectively regulate its free energy, leading to chronic states of high surprise (negative valence) or high uncertainty (anxious arousal).

Pathways to implementation

A practical implementation of this architecture would likely involve a multi-level, hybrid approach. The fast dynamics of neuronal inference could be simulated using established, energy-efficient SNN frameworks.⁵⁵ The entire system, including all cell types and their interactions, would be represented as a GNN. The update rules for all nodes and edges would be derived from the gradient descent on the system’s total free energy, with different parameters being updated on their respective, biologically-plausible timescales. This constitutes a form of multiscale variational inference. Training such a system would not involve standard backpropagation on a labeled dataset. Instead, the system would learn in an unsupervised or self-supervised manner, driven solely by the imperative to minimize the free energy of its “sensory” inputs over time, thereby learning a generative model of its environment through its embodied interactions.

Discussion

Implications and future directions

The theoretical framework presented in this paper offers significant implications for both neuroscience and artificial intelligence, charting a course for future research in both fields.

For neuroscience: This model provides a unified, normative theory of glial function, moving beyond a descriptive list of roles to a single, overarching computational purpose. It reframes the question from “what do glial cells do?” to “how do glial cells contribute to the brain’s imperative to minimize free energy?”. This perspective generates a host of specific, testable hypotheses. For example, it predicts that selectively inhibiting astrocytic calcium signaling in a learning paradigm should not necessarily block learning but should impair the animal’s ability to flexibly adapt its learning rate to changes in environmental volatility (a deficit in precision regulation). It predicts that inhibiting microglial pruning during the consolidation of a new skill should result in a less efficient, “over-parameterized” neural representation, potentially leading to poorer generalization. Furthermore, it provides a formal, computational basis for linking the pathophysiology of glial cells to the cognitive and affective symptoms of complex disorders like depression, schizophrenia, and autism.¹⁴ These conditions can be re-conceptualized as disorders of inference, where specific deficits in glial-mediated precision, model selection, or temporal optimization lead to chronic, maladaptive states of high free energy.

For artificial intelligence: This work argues for a fundamental shift in the pursuit of artificial emotional intelligence. The path forward lies not in scaling up current architectures with more data, but in developing new architectures grounded in the biophysical principles of the glial-neural ensemble. It suggests that robust, flexible, and general intelligence may require systems that incorporate analogues of glial function: mechanisms for dynamic resource and confidence allocation (astrocyte-like), structural optimization and regularization (microglia-like), and temporal coordination (oligodendrocyte-like). Building such systems forces a confrontation with the deep ethical questions surrounding the creation of entities with genuine affective states and goal-directed behavior.⁵ An AI that minimizes its own free energy based on an internal model of the world would have its own intrinsic goals, a significant departure from the passive, objective-function-optimizing systems of today.

Conclusion

This paper has proposed a comprehensive, first-principles model of the glial-neural ensemble as a unified system dedicated to minimizing variational free energy. By deriving specific mathematical formulations for the computational roles of astrocytes, microglia, and oligodendrocytes within the active inference framework, we have laid the theoretical and architectural groundwork for a new generation of artificial intelligence. We have argued that astrocytes are not mere support cells but active regulators of inferential precision; that microglia are not simply cleaners but agents of Bayesian model selection; and that oligodendrocytes are not passive insulators but optimizers of the brain’s temporal priors.

The synthesis of these roles into a dynamic, heterogeneous graph network provides a concrete blueprint for building AI systems capable of a deeper, more biophysically plausible form of emotional intelligence. In this model, affect is not an illusion created by clever pattern matching but an emergent consequence of a system’s fundamental drive to make sense of its world and maintain its own

existence. The future of truly affective AI lies not in the size of its datasets, but in the elegance and biological fidelity of its underlying generative model.^{55–60}

Acknowledgments

None.

Funding

None.

Conflict of interest

Author declares that there are no conflicts of interest.

References

1. *Could AI understand emotions better than we do?* Université de Genève. 2025.
2. Schlegel K, Sommer NR, Mortillaro M. Large language models are proficient in solving and creating emotional intelligence tests. *Commun Psychol*. 2025;3(1):80.
3. *AI shows higher emotional IQ than humans*. Neuroscience News. 2025
4. Dwivedi D. Emotional intelligence and artificial intelligence integration strategies for leadership excellence. *Adv Res*. 2025;26(1):84–94.
5. Pei G, Li H, Lu Y, et al. Affective computing: recent advances, challenges, and future trends. *Intell Comput*. 2024;3(5):0076.
6. *Affective computing: Connecting computing with human emotions for empathetic AI*. EurekAlert. AAAS. 2024.
7. Picard RW. *Affective computing* (Mit Press). 2000.
8. Tao J, Tan T. *Affective computing: a review*. ACII 2005. Lecture Notes in Computer Science, vol 3784. Springer, Berlin, Heidelberg. 2005.
9. Theotokis P. Human brain inspired artificial intelligence neural networks. *J Integr Neurosci*. 2025;24(4):26684.
10. Ororbia A, Mali A, Kohan A, et al. A review of neuroscience-inspired machine learning. arXiv. 2024.
11. Surinder S, Singh I. Neuro-inspired learning algorithms for artificial intelligence. 2023.
12. Guy-Evans O. *What are glial cells and what do they do?* Simply Psychology. 2025.
13. Fields RD, Araque A, Johansen-Berg H, et al. Glial biology in learning and cognition. *Neuroscientist*. 2014;20(5):426–431.
14. Öngür D, Bechtholt AJ, Carlezon Jr WA, et al. Glial abnormalities in mood disorders. *Harv Rev Psychiatry*. 2014;22(6):334–337.
15. *Glial and glutamatergic deficits in psychiatric disorders*. Yale School of Medicine.
16. Sorooshiyari S, Liu Q, Yu G. Modelling the impact of glial cells on neural computation. *Front Comput Neurosci*. 2025.
17. Peña-Garcia M, Peña-Garcia F, Cabrera-Febola W. Mathematical representation of the structure of neuron-glia networks. arXiv. 2024.
18. Linne M-L. Computational modeling of neuron-glia signaling interactions to unravel cellular and neural circuit functioning. *Curr Opin Neurobiol*. 2024;85:102838.
19. *Free energy principle*. Wikipedia.
20. Friston K. The free-energy principle: a rough guide to the brain? *Trends Cogn Sci*. 2009;13(7):293–301.
21. Friston K. The free-energy principle: a unified brain theory? *Nat Rev Neurosci*. 2010;11(2):127–138.

22. Isomura T, Kotani K, Jimbo Y, et al. Experimental validation of the free-energy principle with in vitro neural networks. *Nat Commun*. 2023;14(1):4547.
23. *Helmholtz energy equation*. Amazing FULL Derivation (from scratch). YouTube.
24. *Helmholtz free energy*. Equation, importance. Vaia.
25. Miller N. *A crash course in statistical mechanics*. 2018.
26. *A tutorial on Helmholtz machines*. Northern Kentucky University. 2006.
27. *A tutorial on Helmholtz machines*. All Things Phi.
28. Cieri F, Zhuang X, Caldwell JZK, et al. Brain entropy during aging through a free energy principle approach. *Front Hum Neurosci*. 2021;15:647513.
29. Purushotham SS, Buskila Y. Astrocytic modulation of neuronal signaling. *Front Netw Physiol*. 2023;3:1205544.
30. Ota Y, Zanetti AT, Hallock RM. The role of astrocytes in the regulation of synaptic plasticity and memory formation. *Neural Plast*. 2013;2013:185463.
31. Li H, Xie Z, Tian Y, et al. Genome-wide consensus transcriptional signatures identify synaptic pruning linking Alzheimer's disease and epilepsy. *bioRxiv*. 2024.
32. Mackenzie RJ. *Video catches microglia in the act of synaptic pruning*. The Transmitter. 2025.
33. Pajevic S, Plenz D, Bassar PJ, et al. Oligodendrocyte-mediated myelin plasticity and its role in neural synchronization. *Elife*. 2023;12:e81982.
34. Rangel-Gomez M, Alberini CM, Deneen B, et al. Neuron–glial interactions: implications for plasticity, behavior, and cognition. *J Neurosci*. 2024;44(40):e1231242024.
35. Chen Yu, Ji Q. Dynamics of coupled nonlinear neuronal-astrocyte model. *AIP Advances*. 2024;14(1):015066.
36. Tewari SG, Majumdar KK. A mathematical model of the tripartite synapse: astrocyte-induced synaptic plasticity. *J Biol Phys*. 2012;38(3):465–496.
37. Haydon PG, Carmignoto G. Astrocyte control of synaptic transmission and neurovascular coupling. *Physiol Rev*. 2006;86(3):1009–1031.
38. Sun Z, Zhang X, So K-F, et al. Targeting microglia in Alzheimer's disease: pathogenesis and potential therapeutic strategies. *Biomolecules*. 2024;14(7):833.
39. Lee S. *The art of synaptic pruning*. Number Analytics. 2025.
40. *What is synaptic pruning?* Healthline. 2018.
41. Chechik G, Meilijson I, Ruppin E. Synaptic pruning in development: a computational account. *Neural Comput*. 1998;10(7):1759–1777.
42. Sheridan SD, Horng JE, Perlis RH. Patient-derived in vitro models of microglial function and synaptic engulfment in schizophrenia. *Biol Psychiatry*. 2022;92(6):470–479.
43. Chen Z, Jalabi W, Hu W, et al. Microglial displacement of inhibitory synapses provides neuroprotection in the adult brain. *Nat Commun*. 2014;5:4486.
44. Morini R, Bizzotto M, Perrucci F, et al. Strategies and tools for studying microglial-mediated synapse elimination and refinement. *Front Immunol*. 2021;12:640937.
45. *Synaptic pruning: Definition, process, and potential uses*. Medical News Today. 2023.
46. Scholl C, Rule ME, Hennig MH. The information theory of developmental pruning: Optimizing global network architectures using local synaptic rules. *PLoS Comput Biol*. 2021;17(10):e1009458.
47. Navlakha S, Barth AL, Bar-Joseph Z. Decreasing-rate pruning optimizes the construction of efficient and robust distributed networks. *PLoS Comput Biol*. 2015;11(7):e1004347.
48. Mazuir E, Fricker D, Sol-Foulon N. Neuron–oligodendrocyte communication in myelination of cortical GABAergic cells. *Life (Basel)*. 2021;11(3):216.
49. De Pittà M, Berry H. *A neuron–glial perspective for computational neuroscience*. Computational Glioscience. Springer Series in Computational Neuroscience. Springer, Cham. 2019.
50. Berry K, Wang J, Lu QR. Epigenetic regulation of oligodendrocyte myelination in developmental disorders and neurodegenerative diseases. *F1000Res*. 2020;9:F1000 Faculty Rev-105.
51. Mohammadi H, Karwowski W. Graph neural networks in brain connectivity studies: methods, challenges, and future directions. *Brain Sci*. 2024;15(1):17.
52. Wang PY, Sapra S, George CK, et al. Generalizable machine learning in neuroscience using graph neural networks. *Front Artif Intell*. 2021;4:618372.
53. Liu C, Zhou X, Wu Y, et al. Graph neural networks in EEG-based emotion recognition: a survey. *arXiv*. 2024.
54. Yamazaki K, Vo-Ho V-K, Bulsara D, et al. Spiking neural networks and their applications: a review. *Brain Sci*. 2022;12(7):863.
55. *Spiking neural network*. Wikipedia.
56. Wang B, Dong G, Zhao Y, et al. *Spiking emotions: dynamic vision emotion recognition using spiking neural networks*. 2nd International conference on algorithms, high performance computing and artificial intelligence. 2022;50–58.
57. Volman V, Bazhenov M, Sejnowski TJ. Computational models of neuron-astrocyte interaction in epilepsy. *Front Comput Neurosci*. 2012;6:58.
58. Vergara RC, Jaramillo-Riveri S, Duarte A, et al. The energy homeostasis principle: neuronal energy regulation drives local network dynamics generating behavior. *Front Comput Neurosci*. 2019;13:49.
59. Faghihi F, Alashwal H, Moustafa AA. A synaptic pruning-based spiking neural network for hand-written digits classification. *Front Artif Intell*. 2022;5:680165.
60. Narimisaie J, Naeim M, Imannezhad S, et al. Exploring emotional intelligence in artificial intelligence systems: a comprehensive analysis of emotion recognition and response mechanisms. *Ann Med Surg (Lond)*. 2024;86(8):4657–4663.