

Operationalizing Minsky: A Modular Cognitive System Convergent with the Society-of-Mind Model

Edervaldo José de Souza Melo
edersouzamelo@gmail.com
ORCID: 0009-0003-6835-135X

December 2025

Abstract

Marvin Minsky’s *Society of Mind* (1986) proposed that intelligence emerges from the interaction of many simple agents, each with partial and context-dependent functions. Although widely cited, the model has lacked operational formulations capable of bridging its conceptual insights with implementable cognitive architectures. This paper proposes a contemporary modular cognitive system that is *convergent* with Minsky’s hypothesis: a framework developed independently yet arriving at structurally compatible premises regarding agency, symbol manipulation, and layered control. The system treats cognition as a dynamic interplay among specialized modules whose coordination yields emergent higher-level functions, offering a practical pathway for revisiting and formalizing Society-of-Mind principles in modern contexts such as human–AI assisted cognition and symbolic reasoning. We outline the architecture, justify its modular principles, compare it to classical and contemporary cognitive frameworks, and argue that convergent formulations of Minsky’s ideas remain fertile ground for computational and hybrid cognitive systems.

1 Introduction

Marvin Minsky’s *Society of Mind* [8] remains one of the most influential conceptual models of cognition. Rather than positing a unified or centralized men-

tal mechanism, Minsky argued that intelligent behaviour arises from the cooperation of numerous simple agents, each performing narrow functions under local constraints. Intelligence, in this view, is not a property of a single process but an emergent consequence of structured interactions among heterogeneous components. Although originally formulated as a theoretical and metaphorical account of mind, the model anticipated later developments in modular cognitive architectures, hierarchical control systems, and distributed representations.

Despite its historical and conceptual importance, the Society-of-Mind hypothesis has rarely been *operationalized* into a coherent computational framework. Existing cognitive architectures such as SOAR [7], ACT-R [1], LIDA [5], and GWT [3, 4] incorporate modular principles, yet none directly map or formalize Minsky’s idea of interacting agents with symbolic interfaces and micro-specialized functions. At the same time, contemporary discussions of artificial intelligence—particularly those concerning hybrid symbolic systems and human–AI cognitive integration—reopen the question of how classical modular theories can be reinterpreted or reconstructed using modern methodological tools.

This paper introduces a modular cognitive system developed independently of Minsky’s work yet exhibiting clear conceptual *convergence* with the Society-of-Mind perspective. Rather than deriving from the original theory, the system arises from contemporary requirements in symbolic reasoning, cog-

nitive modularity, and assisted cognition, naturally arriving at structures similar to Minsky’s interacting agents. This form of independent convergence suggests that Society-of-Mind principles are not merely historical metaphors but represent an enduring design space for cognitive architectures.

The contribution of this paper is twofold. First, we outline the conceptual and structural elements of a modern modular cognitive system whose organization resonates with Minsky’s hypothesis. Second, we position this system within the broader ecosystem of cognitive architectures, identifying both compatibilities and divergences with classical and contemporary approaches. By doing so, we argue that revisiting Minsky’s insights through a technically implementable framework provides renewed explanatory and operational value for cognitive science and artificial intelligence.

The remainder of this paper is organized as follows. Section 2 reviews foundational aspects of the Society-of-Mind model and related cognitive architectures. Section 3 presents the proposed modular system and its core structural assumptions. Section 4 compares the system with existing cognitive frameworks, highlighting convergences and divergences. Section 5 discusses theoretical and practical implications for symbolic reasoning and human–AI cognition. Section 7 concludes the paper.

2 Background: Society of Mind and Modular Architectures

Marvin Minsky’s *Society of Mind* (1986) conceptualizes intelligence not as a unified, monolithic faculty but as the emergent result of interactions among numerous specialized agents. Each agent performs narrow, context-dependent functions, and complex cognition arises from their coordinated activity. Minsky’s proposal challenged dominant symbolic AI paradigms of the time by reframing intelligent behavior as a product of distributed, heterogeneous components rather than centrally orchestrated symbolic manipulation. Despite its conceptual resonance, the framework remained intentionally informal, serving

more as a generative metaphor than a directly implementable model.

Over the decades, the Society-of-Mind framework has influenced areas ranging from cognitive science to developmental psychology and hybrid AI models. However, its impact has been largely theoretical: the absence of explicit operational rules, architectural constraints, or computational specifications limited its adoption as a blueprint for working cognitive systems. Contemporary architectures—including modular reinforcement learning, blackboard systems, and hierarchical control models—capture fragments of Minsky’s intuition, yet no unified operationalization has been established. Researchers continue to acknowledge the framework’s conceptual value while noting its lack of a concrete, testable implementation.

This gap between conceptual richness and practical instantiation creates an opportunity for alternative formulations that are not derived historically from Minsky’s framework but nevertheless converge with its underlying premises. Such convergence may emerge when independently developed architectures exhibit core features predicted by the Society-of-Mind hypothesis: multiplicity of specialized modules, dynamic coordination among heterogeneous processes, layered control, and symbolic mediation. Identifying and formalizing these convergences can illuminate pathways for transforming Minsky’s theoretical insights into computationally viable cognitive systems.

3 A Convergent Modular Cognitive System

This section presents the architecture of a contemporary modular cognitive system that was developed independently from the Society-of-Mind tradition yet exhibits clear structural convergence with Minsky’s hypothesis. The system is organized around three design commitments: (i) cognition as the interaction of specialized, heterogeneous modules, (ii) symbolic mediation through higher-level representational processes, and (iii) layered coordination mechanisms en-

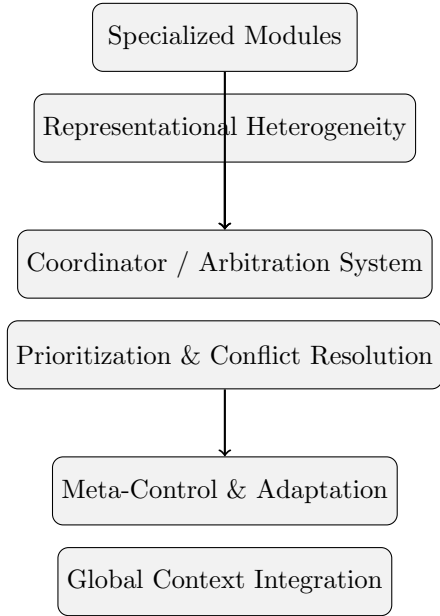


Figure 1: High-level schematic of the proposed modular cognitive architecture, illustrating specialization, layered control, and meta-level adaptation.

abling adaptive control across modules with distinct functions.

3.1 Core Architectural Principles

The system assumes that cognition emerges from distributed components—each performing a narrow and context-dependent function—whose coordinated activity gives rise to higher-level reasoning and behavioral competence. Modules operate as semi-autonomous units with local memory, limited inference capabilities, and well-defined input-output interfaces. No module is individually capable of general cognition; rather, cognitive capacity arises from structured patterns of cooperation.

Three principles govern the architecture:

1. **Functional specialization.** Each module is designed to perform a narrowly defined task, such as evaluating symbolic constraints, pattern recognition, conflict resolution, or long-term

planning. This mirrors the intuition that intelligent behavior is more efficiently produced by multiple small processes than by a single monolithic mechanism.

2. **Heterogeneity of representations.** Modules may operate on different representational formats—vectorial, symbolic, episodic, or rule-based—reflecting the diversity of cognitive processes. Coordination does not require a shared internal language; translation occurs through interface protocols.

3. **Layered and mediated control.** A supervisory mechanism orchestrates module interactions, regulates activation priorities, and arbitrates conflicts. Higher-level symbolic structures emerge from the aggregated states of lower-level processes and feed back into module selection and task allocation.

3.2 Layered Organization

The system is structured into three interacting layers, each contributing distinct forms of computation and control.

1. **Base Layer (Micro-Modules and Local Processes).** This layer contains fine-grained modules that perform rapid, local operations such as perceptual filtering, pattern extraction, constraint checking, or state updating. These processes are narrow, reactive, and context-sensitive. Their interactions produce transient micro-states that serve as the substrate for symbolic integration.

2. **Control Layer (Coordination and Task Allocation).** Above the base layer, a control mechanism schedules module activations, routes information, and maintains global coherence. It resolves conflicts between competing module outputs, prioritizes actions based on task demands, and interfaces with the symbolic structures generated in the meta layer. The control mechanism does not impose a rigid hierarchy; instead, it adapts to emergent states produced by the base layer.

3. Meta Layer (Symbolic Integration and Reflective Processes). The top layer synthesizes distributed outputs into symbolic representations such as goals, plans, constraints, or conceptual structures. It enables multi-step reasoning, hypothesis evaluation, and long-range consistency checking. Symbolic constructs emerging at this level guide module selection, reshape control priorities, and generate reflective monitoring signals.

3.3 Inter-Module Communication

Communication occurs through a lightweight messaging protocol in which modules broadcast state changes or request computational resources. Instead of a central blackboard, the system adopts a distributed notification mechanism: modules subscribe to specific state patterns and react when relevant conditions are met. This design supports scalability and prevents bottlenecks commonly observed in centralized symbolic architectures.

Coordination is achieved through *competitive activation*: when multiple modules propose incompatible actions or interpretations, the control layer evaluates their contextual relevance using symbolic constraints and resource availability. The selected proposal shapes the next cycle of module activations.

3.4 Points of Convergence with the Society-of-Mind Model

Although not derived from Minsky’s historical formulation, the architecture exhibits deep alignment with core Society-of-Mind premises:

- **Multiplicity and specialization:** cognition arises from many small processes rather than from a global monolithic agent.
- **Heterogeneous agencies:** modules with distinct computational styles reflect the diversity of “agents” envisioned by Minsky.
- **Symbolic mediation:** higher-level structures emerge from distributed micro-processes, consistent with Minsky’s view of symbols as coordination devices among simpler components.

- **Layered control:** the separation between reactive processes, coordinative control, and symbolic integration mirrors the layered organization implicit in the Society-of-Mind diagrams.
- **Emergent coherence:** global behavior results from dynamic interactions, not from explicit top-down scripting or predefined homogeneous rules.

3.5 Distinctions from Minsky’s Framework

Despite these convergences, the system diverges from the original model in important ways:

- **Operational specificity:** the architecture provides explicit mechanisms for communication, scheduling, and conflict resolution—absent in Minsky’s metaphorical exposition.
- **Implementability:** the system defines concrete interfaces, layers, and computational pathways that enable direct implementation in contemporary software environments.
- **Formal modular semantics:** module interactions follow rule-governed protocols rather than informal conceptual descriptions.
- **Scalability:** the design accommodates arbitrary expansion in the number of modules without requiring structural reconfiguration.

In summary, the architecture independently realizes several of the cognitive patterns anticipated by the Society-of-Mind hypothesis while providing an operational framework suitable for computational and hybrid cognitive systems.

4 Comparative Analysis with Classical and Contemporary Frameworks

To situate the proposed architecture within the landscape of cognitive systems, this section compares it with classical symbolic frameworks (ACT-R, SOAR),

hybrid cognitive models (LIDA, GWT), and modular architectures in contemporary AI research. The goal is to clarify both convergences and divergences, highlighting the distinct contributions of the present system while positioning it relative to historically influential models.

4.1 Comparison with ACT-R

ACT-R [1] models cognition as the interaction of symbolic production rules and subsymbolic activation dynamics, organized into specialized modules for memory, perception, motor control, and decision-making. The proposed architecture shares ACT-R’s commitment to modularity and functional specialization, but diverges in two fundamental respects.

First, ACT-R relies on a centrally orchestrated production system that imposes a strict bottleneck on cognitive operations. In contrast, the present architecture permits concurrent activation of heterogeneous modules, coordinated by a mediated control layer rather than a single production cycle. This design allows for emergent, distributed patterns of interaction that are absent from ACT-R’s serial decision mechanism.

Second, ACT-R presupposes a homogeneous symbolic representation format. The proposed system, by contrast, accommodates representational heterogeneity: vectorial, symbolic, episodic, or rule-based encodings coexist without requiring a unified global format. This flexibility enables broader integration with contemporary machine-learning components and supports symbolic emergence from distributed computational substrates.

4.2 Comparison with SOAR

SOAR [7] seeks to achieve general intelligence through a uniform problem-solving mechanism... seeks to achieve general intelligence through a uniform problem-solving mechanism grounded in production rules, chunking, and a universal state-operator-goal cycle. While SOAR emphasizes parsimony and unification, the present architecture emphasizes heterogeneity and distributed coordination.

The system converges with SOAR in recognizing the importance of symbolic structures for long-range planning and goal management. However, the mechanisms through which such structures arise differ significantly. In SOAR, symbolic representations are built through problem-space exploration and chunking; in the proposed architecture, symbolic structures emerge from the integration of distributed module outputs within the meta layer.

Moreover, SOAR employs a rigid global control cycle, whereas the present system implements dynamic, context-sensitive scheduling that adapts in real time to local computational pressures. This grants greater flexibility and scalability in environments where multiple cognitive processes must occur simultaneously.

4.3 Comparison with LIDA

LIDA [5] integrates Bernhard Baars’ Global Workspace Theory... integrates Bernhard Baars’ The Global Workspace Theory (GWT) [3, 4] posits that consciousness arises... with a cycle of perception, attention, and action selection. Like the proposed architecture, LIDA adopts a multi-layer structure and treats cognition as a heterogeneous, distributed process. Both frameworks acknowledge the role of bottom-up signals that compete for entry into a higher-level integrative workspace.

However, the present system differs in how integration is performed. LIDA implements a broadcast mechanism in which the winning coalition of processes enters consciousness and globally influences downstream modules. In contrast, the proposed architecture does not rely on a single global workspace. Instead, integration occurs within a symbolic meta layer that synthesizes distributed outputs without broadcasting them indiscriminately to all modules. This reduces communication overhead and avoids the cognitive-cycle rigidity characteristic of LIDA-based implementations.

4.4 Comparison with Global Workspace Theory (GWT)

The Global Workspace Theory posits that consciousness arises from the competition among spe-

cialized processes for access to a global broadcast channel. Many cognitive architectures—including LIDA—instantiate this idea.

The present system converges with GWT in acknowledging that only some module-generated states are elevated to symbolic prominence. However, rather than employing a centralized broadcast, it uses selective symbolic integration that privileges states based on contextual need rather than competitive dominance. This yields a more flexible mechanism for managing distributed resources and avoids the all-or-nothing dynamics of a global workspace.

Additionally, while GWT describes consciousness metaphorically, the proposed architecture operationalizes analogous processes in implementable terms, mapping symbolic elevation onto a concrete representational layer.

4.5 Comparison with Contemporary Modular AI Systems

Recent AI research has revisited modular architectures through a range of approaches, including neural module networks [2], modular reasoning systems [6], differentiable routing mechanisms, hierarchical reinforcement learning, and multi-agent coordination frameworks designed for large language models. Emerging systems such as graph-based workflow engines, agent-oriented tool orchestrators, and LLM-driven cognitive pipelines illustrate a renewed interest in structured, compositional intelligence. These frameworks share with the proposed architecture a commitment to decomposing complex tasks into interacting subcomponents, but often lack symbolic integration or explicit meta-level control.

Yet contemporary neural modular systems typically lack explicit symbolic mediation and rely heavily on gradient-based learning to coordinate interactions. In contrast, the present architecture formalizes coordination through rule-governed protocols and symbolic constraints, enabling transparent control over module interactions. This provides interpretability, inspectability, and explicit reasoning capabilities that remain challenging for purely neural approaches.

4.6 Summary of Comparative Insights

Table 4.6 summarizes the major points of convergence and divergence across the surveyed frameworks.

In sum, while the proposed architecture shares important conceptual affinities with classical symbolic models, hybrid cognitive theories, and modern modular AI systems, it diverges in its explicit commitment to representational heterogeneity, emergent symbolic integration, and dynamic, non-centralized coordination. These distinctions strengthen its position as a structurally convergent yet operationally novel realization of principles anticipated by the Society-of-Mind perspective.

5 Implications for Symbolic Reasoning and Human–AI Cognition

The convergence between the proposed modular cognitive architecture and the core premises of the Society-of-Mind framework has implications for both the theory of symbolic reasoning and the design of hybrid human–AI cognitive systems. This section outlines the theoretical, computational, and practical consequences that follow from adopting a distributed, layered approach to cognitive organization.

5.1 Emergent Symbolic Reasoning

A central implication of the architecture is that symbolic reasoning does not require a monolithic rule engine or a homogeneous representational substrate. Instead, symbols emerge from the coordinated activity of heterogeneous modules, each contributing partial information, constraints, or interpretations. This perspective reframes symbolic reasoning as an emergent phenomenon rather than a fully pre-specified process.

In traditional symbolic AI, reasoning often depends on globally consistent representations and deterministic rule execution. The present system replaces these assumptions with:

Framework	Convergences	Divergences
ACT-R	Modular structure; symbolic reasoning components.	Serial production bottleneck; homogeneous representations.
SOAR	Symbolic planning; structured problem decomposition.	Rigid control cycle; lack of representational heterogeneity.
LIDA	Distributed processing; layered cognitive cycle.	Central workspace broadcast; cycle rigidity.
GWT	Prioritization among competing processes.	Centralized global access; metaphorical implementation.
Contemporary AI Modules	Modularity; compositional task structure.	Lack of symbolic control; opaque learning-driven coordination.

Table 1: Comparative overview of cognitive frameworks and the proposed system.

1. **Locality of computation**, in which symbols are synthesized from micro-level signals aggregated by the meta layer.
2. **Distributed constraint satisfaction**, whereby symbolic structures stabilize when modules converge on compatible interpretations.
3. **Adaptive prioritization**, allowing the control layer to privilege symbolically relevant information without imposing a rigid inference order.

These mechanisms allow symbolic reasoning to be both interpretable and flexible, capturing the fluidity characteristic of human conceptual processes while preserving structural rigor.

5.2 Hybrid Cognitive Integration

The architecture supports hybrid forms of cognition in which symbolic and sub-symbolic processes coexist and influence one another. Representational heterogeneity enables symbolic structures to be grounded in perceptual or vectorial encodings produced by lower-level modules, creating a natural bridge between traditional symbolic AI and contemporary machine-learning components.

This design has three major implications:

- **Grounding of symbols:** Symbolic constructs are traceable to specific patterns of module activity, permitting explainability and interpretability in hybrid systems.

- **Bidirectional influence between levels:** Sub-symbolic processes can introduce candidate symbolic features, while symbolic constraints guide module activation, enforcing coherence across layers.
- **Incremental conceptual development:** Complex symbolic concepts emerge gradually as distributed modules produce increasingly stable patterns of interaction.

These properties align with contemporary theories of cognitive development, in which conceptual structures arise from iterative, experience-dependent integration rather than from pre-specified rule sets.

5.3 Implications for Human–AI Cognitive Interaction

As hybrid architectures become integral to decision-support systems, embodied agents, and assistive technologies, understanding how symbolic and distributed processes interact becomes increasingly important. The proposed system offers a model for designing AI that interfaces naturally with human cognitive processes by mirroring several key structural features of human reasoning.

Three implications stand out:

1. **Cognitive Compatibility.** Because symbolic representations emerge from distributed processes, the system can align with human conceptualization

patterns, facilitating shared tasks such as collaborative reasoning, explanation, and problem-solving.

2. Interpretability and Cognitive Transparency. Unlike end-to-end neural models, the architecture exposes intermediate states and module contributions, allowing users to inspect how interpretations or decisions were formed. This transparency is crucial for trust, especially in domains requiring accountability.

3. Support for Assistive Cognition. The layered structure provides a template for augmenting human cognition by delegating subtasks to narrow modules or by generating symbolic summaries that enhance users’ working memory, attention, or decision-making.

5.4 Theoretical Contributions

Beyond practical applications, the architecture contributes to ongoing debates in cognitive science regarding the relationship between distributed processing and symbolic reasoning. It demonstrates that:

- symbolic reasoning can emerge from heterogeneous, decentralized processes rather than from a centralized symbolic engine;
- distributed architectures can maintain interpretability when equipped with explicit coordination and symbolic integration mechanisms;
- hybrid systems can achieve scalable cognitive performance without collapsing into either purely neural or purely symbolic paradigms.

These contributions reinforce the relevance of Minsky’s intuitions while providing a formal structure that bridges classical cognitive theories with modern computational implementations.

5.5 Future Directions

The alignment between modular architectures and symbolic emergence suggests several avenues for further exploration, including:

- integrating learning mechanisms that allow modules to acquire new competencies while preserving the system’s symbolic structure;
- extending the architecture to multi-agent settings where symbolic mediation may support communication and shared reasoning;
- applying the framework to embodied environments, where distributed sensorimotor processes may shape symbolic abstraction.

Taken together, these implications position the proposed system as a versatile model for both theoretical inquiry and applied research in human–AI cognition.

6 Limitations

As a theoretical and architectural contribution, this work does not include large-scale empirical evaluation or benchmarking against established cognitive systems. The exploratory feedback from participants provides initial indications of conceptual clarity and perceived utility, but it does not constitute experimental validation. Additionally, no performance metrics, computational implementations, or formal proofs of scalability are presented. Future research should extend this architecture with systematic evaluations, comparative benchmarks, and applications in embodied or multi-agent environments.

7 Conclusion

This paper introduced a modular cognitive architecture that, although developed independently from the Society-of-Mind tradition, exhibits clear structural convergence with Minsky’s core hypothesis: that intelligence arises from the interaction of numerous specialized, heterogeneous processes. By formalizing these intuitions into an operational framework, the proposed architecture bridges conceptual insights from classical cognitive theory with the requirements of contemporary computational systems.

The system advances three contributions. First, it demonstrates that symbolic reasoning can emerge

from distributed mechanisms when supported by layered coordination and representational heterogeneity. Second, it provides an implementable blueprint for hybrid cognitive architectures capable of combining symbolic, sub-symbolic, and episodic processes in a unified structure. Third, it offers a comparative perspective that situates the architecture relative to established models such as ACT-R, SOAR, LIDA, GWT, and modern modular AI systems.

These findings reinforce the enduring relevance of Minsky’s insights while addressing the operational gaps that have historically limited the applicability of the Society-of-Mind framework. By articulating explicit mechanisms for module interaction, conflict resolution, symbolic integration, and adaptive control, the architecture contributes a practical foundation for future cognitive systems research.

Future work may explore the integration of learning mechanisms, the extension to multi-agent contexts, and the application of the architecture to embodied environments. More broadly, the framework invites further dialogue between cognitive science, artificial intelligence, and hybrid computational theories, suggesting that convergent architectures—rather than purely symbolic or purely neural ones—may offer the most promising path toward scalable, interpretable, and cognitively compatible artificial systems.

References

- [1] John R. Anderson and Christian Lebiere. The atomic components of thought. In *The Atomic Components of Thought*. Lawrence Erlbaum, 1998.
- [2] Jacob Andreas, Marcus Rohrbach, Trevor Darrell, and Dan Klein. Neural module networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 39–48, 2016.
- [3] Bernard J. Baars. *A Cognitive Theory of Consciousness*. Cambridge University Press, 1988.
- [4] Bernard J. Baars. In the theater of consciousness: Global workspace theory, a rigorous scientific theory of consciousness. *Journal of Consciousness Studies*, 4(4):292–309, 1997.
- [5] Stan Franklin, Tamas Madl, Sidney D’Mello, and Javier Snider. The lida architecture: Adding new modes of learning to an intelligent, autonomous, software agent. *Proceedings of the International Conference on Artificial General Intelligence*, 2014.
- [6] Anirudh Goyal and et al. Neural production systems. *NeurIPS*, 2021.
- [7] John E. Laird. *The Soar Cognitive Architecture*. MIT Press, 2012.
- [8] Marvin Minsky. *The Society of Mind*. Simon and Schuster, 1986.

Open Science and Availability of Materials

All supplementary materials, extended diagrams, and implementation notes associated with this work are openly available at: <https://zenodo.org/communities/sistema-nemosine>.