A Knowledge Graph-Based Cognitive Architecture with Shifting Attention

Kailin Yan
hellucigen@qq.com

November 2025

Abstract

Current AI systems commonly suffer from insufficient autonomy, inflexible knowledge recombination, and non-traceable reasoning processes in open environments. To address these limitations, this paper presents a cognitive architecture centered on a multimodal embedded knowledge graph and driven by node activation and diffusion mechanisms. In this framework, concepts, events, actions, rules, emotions, and personality traits are uniformly modeled as heterogeneous nodes in the graph, with directed weighted edges encoding semantic, logical, and causal relationships among them. The shift of the Focus of Attention (FoA) is guided by the real-time activation states of nodes: when a node's activation level increases, its activation signal propagates to neighboring nodes based on semantic proximity, emotional coupling strength, task relevance, and rule association, thereby dynamically forming continuous or leap-based attention pathways.

The system supports dual-mode attention shifting: under task-driven conditions, activation diffusion is constrained by goals and exhibits convergence similar to the Central Executive Network (CEN), focusing on task-relevant subgraphs; during idle periods, it displays divergence akin to the Default Mode Network (DMN), facilitating remote association and creative knowledge recombination. Higher-order cognitive functions—such as decision-making, planning, prospective simulation, and procedural action—naturally emerge from the activation dynamics and connectivity patterns of specialized node types (e.g., decision-theoretic nodes, prospective intention nodes, action–expectation nodes), without requiring a centralized controller. Emotions modulate the gain and diffusion bias of node activation through simulated hormonal parameters, while personality manifests as a long-term stable distribution of receptor sensitivities, shaping individualized cognitive styles. Episodic memory is stored using an "event-head node + temporally ordered propositional sequence" structure and is tightly coupled with semantic memory through shared argument nodes, enabling analogy-based reasoning and causal inference grounded in experience.

Contents

1 Introduction

Mainstream AI systems heavily rely on deep neural networks, achieving strong pattern recognition capabilities through large-scale data training. Whether in image understanding, natural language processing, or policy optimization in complex environments, these models demonstrate outstanding performance within closed and static settings. However, when agents must operate over extended periods in open, dynamic, and partially unknown real-world environments, their fundamental shortcomings become evident: they struggle with cross-domain knowledge transfer, multi-goal coordinated planning, causal chain reasoning, and experience-based self-correction.

In contrast, the architecture proposed in this paper unifies diverse cognitive elements into a single dynamic knowledge graph, driving attention flow and cognitive evolution through node activation and diffusion mechanisms. This design significantly enhances the system's self-organization and adaptability while maintaining structural coherence.

1.1 Related Work and Limitations

Current AI research primarily follows three technical paradigms: deep learning, reinforcement learning, and large-scale pretrained language models. Despite notable successes in specific tasks, all three face inherent challenges in building general-purpose agents capable of continual learning, autonomous reasoning, and contextual adaptation.

Deep learning models typically assume static data distributions and lack the ability to autonomously generate hypotheses or organize knowledge structures. Reinforcement learning methods suffer from low sample efficiency in high-dimensional state spaces and often require retraining from scratch for new tasks. Although large-scale pretrained language models have achieved breakthroughs in surface-level linguistic modeling, they exhibit clear weaknesses in systematic knowledge integration, traceable reasoning chains, and self-modeling.

On the other hand, several general cognitive architectures (e.g., OpenCog, ACT-R, Soar) attempt to integrate symbolic and subsymbolic processing to compensate for the limitations of purely connectionist models. However, their knowledge representations are often partitioned into isolated modules (e.g., working memory, procedural memory, semantic networks), restricting information flow to predefined interfaces. Attention mechanisms in these systems typically rely on fixed rules or static weights, making it difficult to support human-like free association and context-sensitive focus shifts. More critically, higher-order cognitive functions—such as decision-making, planning, and emotion regulation—are usually hard-coded into centralized controllers rather than emerging naturally from internal knowledge dynamics.

These limitations indicate that existing approaches have yet to establish a computational framework that both unifies diverse cognitive elements and supports endogenous dynamic evolution. Against this backdrop, we propose a knowledge-graph-based cognitive architecture driven by activation diffusion, aiming to achieve decentralized, structured, and growable cognition.

1.2 Insights from Cognitive Science and Psychology

Findings from cognitive neuroscience and psychology suggest that human intelligence does not arise from a single computational module but emerges from the continuous interaction of multiple cognitive subsystems. Perception, attention, memory, emotion, reasoning, and decision-making are not isolated processes; instead, they are tightly coupled and mutually regulated at the level of neural dynamics. The emotional system, for instance, modulates perceptual gain and memory consolidation through neurotransmitters and hormonal levels, while also guiding attention allocation via motivational signals, thereby influencing the selection of cognitive pathways. Personality reflects an individual's stable response tendencies to specific stimulus patterns over time, shaping unique cognitive styles and behavioral preferences.

Moreover, human thought is highly dynamic and generative: even in the absence of external tasks, the brain's Default Mode Network (DMN) remains active during rest, supporting episodic recollection, remote association, and creative hypothesis generation. This "endogenous cognitive flow" demonstrates that intelligence not only responds to external inputs but can also actively reorganize knowledge and explore potential possibilities based on internal states. This characteristic demands that a cognitive system be capable of autonomously guiding attention shifts and knowledge retrieval based on internal activation dynamics, rather than merely reacting passively to external commands.

1.3 Design Principles of the Proposed Framework

This work models human cognition as a unified, dynamically evolving knowledge network, in which all cognitive content—including concepts, events, actions, rules, emotions, personality traits, intentions, and expectations—is embedded as heterogeneous nodes within a single multimodal knowledge graph. Directed weighted edges between nodes encode semantic associations, logical dependencies, causal chains, or procedural constraints, and are updated continuously through experience. Each node maintains a multidimensional activation level reflecting its current knowledge salience, recency, emotional coupling, and task relevance.

Within this architecture, shifts in the focus of attention are naturally guided by activation diffusion across the graph: when a node becomes highly activated due to perceptual input, internal inference, or emotional arousal, its activation signal propagates to connected nodes based on semantic proximity, emotional resonance, rule connectivity, and task context, gradually forming an attention pathway that evolves from the current focus toward potential targets. This mechanism supports both convergent, goal-directed focusing under task constraints and divergent, exploratory association during idle states, effectively replicating the flexible switching between "execution" and "exploration" observed in human cognition.

1.4 Contributions of This Framework

The main contributions of this work are summarized in the following four aspects:

1. We propose an activation-guided attention mechanism that enables continuous or discontinuous attention shifts through dynamic diffusion of node activations in embedding space, effectively simulating the cognitive flow observed in human recall, reasoning, and creative thinking;
2. We design a multidimensional node activation scheme, assigning each node an activation vector that captures knowledge richness, recency, emotional coupling, and interest weight, thereby enabling fine-grained control over knowledge salience and interest-driven recombination;
3. We implement a perceptual salience-guided sensory interface that encodes external inputs as activation signals annotated with salience scores, allowing novel, anomalous, or high-value information to preferentially trigger relevant nodes in the graph and achieve tight perception–cognition coupling;
4. We realize node-driven autonomous cognitive activities by decomposing high-level functions—such as decision-making, reflection, goal generation, prospective simulation, and procedural action—into the activation dynamics and connectivity patterns of specialized node types (e.g., decision-theoretic nodes, prospective intention nodes, action–expectation nodes), enabling the system to generate coherent behavioral sequences without external instructions.

2 Overall System Architecture

2.1 Basic Cognitive Functions: Memory Based on Knowledge Graphs

The knowledge graph is composed of two main components: a semantic memory store and an episodic memory store. The semantic memory store holds abstract representations of concepts, rules, and facts, while the episodic memory store records individualized, temporally ordered events—including experiences, internal states, and emotional fluctuations. These two stores interact dynamically through a weighted relational network, forming a dual-loop architecture that integrates knowledge and experience.

2.1.1 Semantic Memory Store: A Concept-and-Rule Network Based on Weighted Relational Edges

The semantic memory store encodes the system's long-term, stable structural knowledge and is organized as a specially designed knowledge graph. All knowledge elements—whether concepts, attributes, perceptual features, or logical templates—are represented as nodes, and their interrelations are explicitly expressed solely through directed, weighted edges. In other words, nodes themselves are lightweight identifiers; their "meaning" is defined collectively by their incident edges and neighboring nodes. This design ensures high decoupling and composability in knowledge representation.

(1) Node Design: All Attributes Are Relations

In this architecture, nodes do not embed internal attribute fields. Instead, they dynamically construct their full semantic profiles by connecting to other functional nodes via typed edges. For example:

- A concept node "apple" connects via an edge of type `hasColor` to the node "red";
- It links via `hasVisualFeature` to a dedicated node storing its visual feature map;
- It connects via `hasAuditoryFeature` to a spectral audio node (if applicable);
- It combines with other concept nodes (e.g., "bite", "sweet") through `formsPhraseWith` edges to form phrase-level composite nodes such as "bite apple" or "sweet apple";
- It attaches via a `hasPropositionalTemplate` edge to a template node, which in turn connects to placeholder nodes through role-slot edges (e.g., `agent`, `patient`).

Take the verb "wear" as an example: its corresponding node serves only as an identifier. Its binary semantic structure $(\text{person}, \text{clothing})$ is not stored within the node itself but is instead encoded by a `hasTemplate` edge pointing to an independent template node T_{wear}. This template node then connects via `role1` and `role2` edges to role-definition nodes for "agent" and "patient," respectively. Similarly, the "give" node links via `hasTemplate` to a ternary template node with three role slots.

This "externalization of attributes" treats nodes as pure semantic anchors, with all features, structures, and behaviors dynamically conferred by the graph topology. This greatly enhances modularity, reusability, and evolutionary flexibility of knowledge.

(2) Edge Design: Directional, Weighted, and Plastic Relational Structures

Edges in the graph possess three properties: type, direction, and weight. The edge type (e.g., `causes`, `partOf`, `hasStep`) defines the semantic nature of the relation; direction indicates dependency or flow (e.g., action sequence, causal order); and weight $w \in [0, 1]$ represents the empirical strength, confidence, or activation gain of the relation.

For instance, the procedural knowledge "grasp followed by lift" is represented as:

$$\text{grasp} \xrightarrow[\texttt{nextStep}]{w} \text{lift},$$

where the edge type `nextStep` specifies the semantic relationship, the direction encodes execution order, and the weight w reflects the stability of this transition in past experience. All edge weights can be dynamically adjusted through experience-based statistics, conflict detection, or self-supervised signals, enabling continuous self-correction and evolution of the knowledge structure.

2.1.2 Episodic Memory Store: An Event Graph Based on Head Nodes and Temporal Proposition Sequences

The episodic memory store captures complete events experienced by the system within specific spatiotemporal and emotional contexts, using a two-layer structure: “event head node + temporally ordered proposition sequence” [1]. The head node acts as a semantic anchor, encoding a high-level summary of the event, while the time-stamped proposition sequence details its dynamic unfolding. Together, they form an indexable, inferable, and recombinable event graph.

(1) Event Head Node: Semantic Center and Index Entry Point Each event is associated with a unique head node E_{head}, serving as its core identifier and access point in the knowledge graph. This node connects via specialized relation edges to multiple information nodes that encode: a concise semantic summary of the event; start and end timestamps; primary spatial location; participants and their role types; and an aggregate summary of emotional state and hormonal levels. As a high-level index, the head node can be efficiently retrieved and activated by attention mechanisms, event-frame modules, and reasoning engines.

(2) Temporal Proposition Sequence: Structured Unfolding of Event Dynamics The dynamic progression of an event is represented by a time-ordered sequence of propositional structures

$$E_{\text{body}} = \{P_1(t_1), P_2(t_2), \ldots, P_n(t_n)\}.$$

Each proposition P_i is a 5-tuple defined as

$$P_i = \langle \text{predicate},\ \text{argument binding},\ \text{modifiers},\ \text{emotional state},\ t_i \rangle,$$

branching out from the head node to form the event’s “temporal backbone.” Adjacent propositions are linked by directed edges of types such as `nextState` (state transition), `causes` (causal progression), and `emotionShift` (emotional evolution), thereby unifying the event’s temporal flow, logical chain, and emotional trajectory into the graph topology.

(3) Cross-Store Linking: Bidirectional Instance–Concept Mapping with Semantic Memory Episodic memory is deeply integrated with semantic memory through shared nodes: arguments in propositions (e.g., agents, objects, predicates) connect via `instanceOf` edges to their corresponding concept nodes in the semantic graph; emotional summaries link to emotion nodes in the affective subgraph; and contextual attributes like time and location are associated with spatiotemporal concept nodes in semantic memory. This bidirectional “instance–concept” mapping allows events to inherit the generalization power of abstract knowledge while providing empirical feedback to enrich semantic memory, enabling co-evolution of both stores.

(4) Event Chaining: Construction of Global Narrative Structures At the event level, head nodes are interconnected based on temporal order, causal dependency, or emotional continuity. Consequently, the episodic memory store not only preserves the internal details of individual events but also constructs coherent narrative chains and experiential flow networks at a macro scale, supporting long-range causal reasoning and life-stage modeling.

(5) Cognitive Significance This architecture organizes episodic memory as a multi-layered, computable event graph: head nodes enable efficient indexing, proposition sequences ensure fine-grained process representation, cross-store links facilitate knowledge reuse, and event chaining supports macro-level narratives. This structure preserves the completeness and context sensitivity of human episodic memory while endowing it with machine-operable symbolic and graph-based properties, providing a solid foundation for the system to flexibly retrieve, recombine, and generalize past experiences during planning, reflection, simulation, and imagination.

2.1.3 Self-Cognitive Knowledge Graph: Structured Representation of Internal States

The system maintains a dedicated self-cognitive subgraph to structurally represent the agent's multi-dimensional internal states, including emotions, motivations, task progress, system resource load, and physiological or simulated embodied parameters. This subgraph constitutes a continuously active internal state field, whose nodes maintain stable baseline activation levels and exert long-term modulatory influence on the global energy distribution of the entire graph—even in the absence of external stimuli.

Unlike ordinary semantic nodes that encode knowledge about the external world, self-cognitive nodes represent a multimodal, multilayered composite state vector encompassing the following core dimensions: emotional state (see Section 2.4.2); situational context, such as current location, interaction setting, environmental risk level, and noise intensity; task status, including current phase, attention allocation strategy, goal completion degree, and level of internal conflict; meta-system states, such as computational load, inference confidence, energy consumption, reasoning depth, and memory usage; embodied or simulated embodied parameters, e.g., arousal, excitation, and stress hormone levels; and interaction-mode states, such as behavioral tendencies toward dialogue, tool invocation, observation, or instruction.

These internal state nodes sustain the system's real-time awareness of its own operational context through high baseline activation. When ordinary semantic nodes are semantically or functionally associated with a particular self-state node, they receive additional activation gain, thereby influencing the dynamic migration of the Focus of Attention (FoA). For example, under high system load, nodes related to "simplified reasoning" or "reduced search branching" become more readily activated; in social contexts, nodes associated with "expression," "feedback," or "sharing" are preferentially brought into the FoA; and when task conflicts intensify, strategic nodes such as "plan revision," "reinterpretation," or "local search" are elevated as candidate focus regions.

Thus, the self-cognitive subgraph not only provides explicit modeling of internal states but also acts as a continuous internal energy field that guides activation flow and shapes the evolution of cognitive style across the entire knowledge graph. As a result, the system can exhibit human-like, state-dependent thinking patterns—such as exploratory, conservative, contracted, or social modes—achieving dynamic coupling between cognitive behavior and internal state.

2.2 Basic Cognitive Functions: Dynamic Attention Mechanism Based on the Knowledge Graph

In this architecture, attention is not an external module but an intrinsic, dynamic process emerging from the unified knowledge graph. The system realizes human-like attention flow—from free association to goal-directed focus—through time-varying propagation of node activation, structured integration via event frames, and continuous modulation between DMN- and CEN-like dynamics.

2.2.1 Node Activation and Semantic Diffusion: Energy Propagation over Graph Structure

Each node carries a time-evolving activation level that reflects its current cognitive salience. When a node becomes central due to external input, internal retrieval, or contextual triggering, its activation signal diffuses along weighted edges to neighboring nodes, with intensity decaying as a function of path distance and edge weight, thereby forming a locally continuous semantic energy field. This mechanism ensures that attention preferentially moves within semantically related regions, generating stable, coherent, and contextually consistent cognitive trajectories.

Activation does not accumulate indefinitely but automatically decays over time. The decay rate is jointly determined by three factors: the time elapsed since the node last participated in cognitive activity, the task relevance of its local region, and global modulation from the self-state field. This decay mechanism effectively prevents outdated information from persistently interfering with new tasks, facilitates natural disengagement from obsolete contexts, and maintains overall energy balance in the graph. Although the specific decay function can be flexibly designed, the architecture only requires it to satisfy the monotonic constraint that "the longer the elapsed time, the weaker the activation."

The diffusion process is not governed solely by topological distance but is finely tuned by edge attributes. The propagation weight of each edge is jointly determined by long-term empirical confidence, the match between current task goals and edge semantics, and real-time modulation from the self-state field. For instance, under high stress, weights on causal edges are amplified while those on free-association edges are suppressed. Consequently, activation diffusion becomes a dynamic interplay of semantic content, experiential history, and internal/external states, enabling the system to exhibit diverse cognitive styles—exploratory, logical, associative, or convergent—depending on context.

2.2.2 Event Frames: Structured Carriers of the Focus of Attention

To transform distributed activation into actionable cognitive units, the system introduces "event frames" as the core representational structure of the current Focus of Attention (FoA). This mechanism corresponds to the transient integration function of working memory [2], bridging perceptual input and semantic organization, and providing a structured platform for thought transition, reasoning, and simulation.

An event frame adopts a hierarchical, nested composite structure with three progressive layers: L3 (Context Activation Layer) stores potentially relevant memory nodes and situational background, forming a candidate pool for diffusion; L2 (Semantic Binding Layer) selects semantically proximate or logically connected nodes from L3 and temporarily aggregates them into loose or tight semantic clusters; L1 (Attention Integration Layer) then chooses the most integrated proposition as the current FoA based on activation strength and task relevance.

The microstructure of the L1 layer binds together a core proposition, an attribute bundle (e.g., color, size), relational links (e.g., causal, part–whole), valence markers (emotional or motivational polarity), and constraint conditions (e.g., logical boundaries, temporal windows). This structure ensures that the FoA is not merely an activation peak but a semantically complete and operationally viable cognitive unit.

2.2.3 Dual-Mode Attention Shifting: DMN-Like Divergence and CEN-Like Convergence

The system runs two types of attention dynamics over the same graph, continuously switching between them via a global mode coefficient λ_{mode}: low λ_{mode} corresponds to Default Mode Network (DMN)-like free association, while high λ_{mode} corresponds to Central Executive Network (CEN)-like focused control.

In DMN mode (e.g., during idle states, reflection, or high curiosity), activation decays slowly, and long-range semantic edges gain influence. The L3 layer supports broad diffusion across subgraphs; the L2 layer uses relaxed thresholds, allowing analogies, metaphors, or counterfactual hypotheses into aggregation; and the L1 layer is dominated by activation peaks, with task constraints weakened—enabling episodic recollection, future simulation, or creative imagination. This process replicates the human DMN's "spontaneous thought–self-referential" characteristics [3].

In CEN mode (e.g., during task execution, high cognitive load, or sudden stimuli), λ_{mode} increases significantly, and activation diffusion sharply contracts. The L3 layer responds only to task context and salience signals; the L2 layer strictly filters nodes based on goal consistency, causal structure, and logical rules; and the FoA at L1 is determined by:

$$\text{FoA}(t+1) = \arg\max_{x \in L2} \left[a(x) \cdot \text{goal\_match}(x) \cdot \text{rule\_match}(x) \right],$$

where $a(x)$ denotes the node's activation level. Structures displaced from the FoA retreat to L2 or decay into L3 and may be temporarily inhibited to ensure task-focused continuity.

Unified Control: Dynamic Sources of the Mode Coefficient λ_{mode} λ_{mode} is not a preset switch but a continuous variable dynamically regulated by four factors: (1) Task load—more explicit, urgent, or complex tasks increase λ_{mode}; (2) External salience—unexpected percepts or high-risk events rapidly elevate its value; (3) Emotional and hormonal state—high arousal or stress promotes convergence, while high curiosity or low stress favors divergence; (4) Self-state field—including internal variables such as system load, confidence, energy consumption, and interaction mode—which contributes regulatory signals based on their semantic positions along the "exploration–execution" axis.

Through this mechanism, the system achieves smooth transitions in attentional style over a single knowledge graph: it maintains cognitive vitality and knowledge recombination capability in the absence of external instructions, yet can rapidly shift into efficient, ordered execution when needed. This design not only aligns with neuroscientific plausibility but also provides a computable and controllable foundation for cognitive dynamics in general-purpose agents.

2.3 Perceptual Module Design

The system achieves multimodal environmental perception through three core channels: tactile, visual, and auditory. The tactile channel captures physical properties such as pressure, texture, and temperature; the visual channel supports recognition of objects, colors, optical flow, and motion trajectories; and the auditory channel parses human speech, musical instruments, and background noise, further inferring embedded emotional tendencies or latent intentions. While the engineering implementation of the perceptual module—such as sensor selection, neural network architectures, or feature extraction algorithms—is beyond the scope of this paper, our focus lies in how perceptual outputs are structurally integrated into the cognitive architecture. To this end, all perceptual results are transformed into semantically annotated activation signals and bound to corresponding object or event nodes in the knowledge graph. Consequently, each object node not only encodes its semantic category and abstract attributes but also associates behavior response patterns and cognitive priors shaped by perceptual experience, enabling the system to achieve individualized, context-sensitive interactions in subsequent cognitive processes.

2.3.1 Perceptual Salience and Attention Selection Mechanism

Perceptual salience serves as the core metric for the system's initial prioritization of raw sensory input. Upon entry into the cognitive system, each perceptual datum is encoded as an activation signal annotated with a salience value that holistically reflects its novelty, behavioral relevance, and potential value within the current context. This salience value is injected as the initial activation strength into the corresponding perceptual node in the knowledge graph.

Salience is jointly shaped by five factors: (1) the magnitude of change relative to recent perceptual history (i.e., sensory surprise); (2) the current focus of interest represented by highly activated nodes in the system; (3) the rarity of the stimulus and its association with high-value goals or potential threats; (4) emotional and hormonal states (e.g., elevated dopamine enhances gain for novel stimuli, while high cortisol suppresses non-critical inputs to promote conservatism); (5) dynamic weighting of specific perceptual channels based on current task objectives.

These modulatory factors are implemented through the activation states and attributes of corresponding nodes in the knowledge graph (e.g., emotion nodes, task-goal nodes, value-assessment nodes). The resulting perceptual salience determines the initial activation intensity of the input in the L3 Context Activation Layer: high-salience inputs more readily trigger strong diffusion and enter the attention candidate set, while low-salience inputs may be delayed or temporarily ignored due to insufficient activation—thereby realizing an efficient, context-adaptive attention selection mechanism.

2.3.2 Episodic Buffer and Multimodal Proposition Fusion

To enhance temporal consistency and cross-modal semantic binding at the perceptual level, the architecture incorporates an episodic buffer—inspired by Baddeley's working memory model—as a short-term integration hub for multimodal inputs [2]. Within a limited time window, this module synchronizes, semantically aligns, and contextually fuses parallel perceptual signals from visual, auditory, tactile, and linguistic channels, generating structured temporary episodic units directly accessible by the event frame mechanism.

According to Baddeley, the episodic buffer is a capacity-limited, central-executive-regulated multimodal integration system that binds transient perceptual content with long-term memory into unified episodic representations. In our architecture, this mechanism provides a consistent spatiotemporal–emotional coordinate frame for the Focus of Attention (FoA) as it migrates across the knowledge graph, preventing semantic fragmentation caused by asynchronous or heterogeneous multimodal inputs.

Perceptual outputs from each modality are encoded as time-stamped activation vectors. The linguistic channel employs a large language model (LLM) for dual parsing: — On one hand, it identifies the user's explicit intent, social role, and latent goals; — On the other, it infers implicit psychological states (e.g., anxiety, curiosity, sharing tendency) and conversational motives (e.g., request, question, statement), encoding these as structured vectors injected into the self-cognitive subgraph.

The original utterance is then converted into a standardized propositional structure:

$$(\text{agent},\ \text{action},\ \text{object},\ \text{time},\ \text{location},\ \text{modifiers}),$$

where *modifiers* include action descriptors (e.g., "jumping"), emotion labels (e.g., "fear"), perceptual confidence scores, and hormonal parameters (e.g., dopamine, cortisol levels). For example, the sentence "I just saw a black cat jump onto the windowsill—it kind of scared me" is parsed as:

$$(\text{user},\ \text{see},\ \text{black\_cat},\ t_{-5s},\ \text{window\_sill},\ \{\text{jumping}\},\ \text{fear}).$$

The episodic buffer receives outputs from all modalities (including the above linguistic propositions) and performs weighted fusion based on temporal alignment, spatial co-occurrence, and emotional salience. This propositional representation compresses the rapidly changing raw perceptual stream into logically coherent, machine-operable cognitive units, providing stable input for subsequent event boundary detection algorithms.

Internally, the buffer employs a local transition function to dynamically shift attention, preferentially focusing on proposition combinations with high information entropy or strong emotional salience. The resulting episodic units not only encapsulate fused multimodal features but also carry real-time internal state markers, offering high-fidelity context for reasoning, intention recognition, and expectation generation.

Moreover, the buffer's write rate, retention duration, and fusion weights are dynamically modulated by emotional and hormonal states: high dopamine or high curiosity entropy extends the integration window, enhancing encoding of novel experiences; high cortisol or high-stress contexts narrow the window and suppress non-critical channels, simulating the human perceptual compression effect under threat.

2.3.3 Dynamic Construction Mechanism of the Knowledge Graph

The knowledge graph in this architecture is not statically predefined but continuously evolves through two complementary mechanisms: (1) self-generation of episodic memories from perceptual streams, and (2) semantic knowledge extraction and fusion from multi-source inputs. Together, they form the foundation of the system's long-term memory and cognitive structure.

Self-Generation of Episodic Memory: From Perceptual Stream to Event Units The system automatically constructs temporally grounded episodic event units by detecting event boundaries in streams of propositional percepts, mimicking the natural organization of human experience. Specifically, it computes the semantic shift between consecutive propositions p_t and p_{t+1} in the knowledge graph:

$$D(p_t, p_{t+1}) = \sum_{e \in \{\text{agent},\text{action},\text{object},\text{modifier}\}} \text{dist}_{KG}(e_t, e_{t+1}), \tag{1}$$

where $\text{dist}_{KG}(\cdot,\cdot)$ denotes the shortest-path distance (or semantic distance in embedding space) between two nodes. When $D(p_t, p_{t+1})$ exceeds a dynamic threshold ϑ, the system identifies a semantic discontinuity and triggers the creation of a new event unit. This mechanism enables the system's event segmentation capability to adaptively refine with accumulated knowledge—realizing autonomous evolution where "the richer the cognition, the finer the event delineation."

Each event unit is stored as a composite node in the knowledge graph, containing timestamps, participating entities, action sequences, and emotional context, and is linked via edges to relevant concept nodes—supporting event-based recall, analogy, and causal reasoning.

Multisource Fusion of Semantic Knowledge Abstract semantic knowledge is continuously expanded by integrating information from multiple sources, including web text, external structured knowledge bases (e.g., Wikidata, ConceptNet), regularity patterns induced from perceptual experience, and commonsense inferences from large language models.

For any input text T, the system invokes an LLM to perform relation extraction:

$$\text{ExtractRelations}(T) \rightarrow (\text{head}, \text{relation}, \text{tail}, \text{confidence}), \tag{2}$$

and the extracted results undergo consistency validation (e.g., conflict detection, redundancy filtering) before being written into the graph as weighted edges. The edge weight integrates the LLM's confidence score c_{LLM}, the frequency of the relation in perceptual experience f_{percept}, and its consistency score with external knowledge bases c_{KB}:

$$w = \alpha \cdot c_{\text{LLM}} + \beta \cdot f_{\text{percept}} + \gamma \cdot c_{\text{KB}},$$

where α, β, γ are learnable parameters.

This process can be efficiently implemented via graph databases (e.g., Neo4j or GraphDB) to support incremental updates while preserving structural integrity and enabling continuous knowledge expansion. The resulting graph thus contains not only general factual knowledge but also the system's individualized experiences, preferences, and behavioral patterns—serving as the core cognitive substrate that drives attention migration, analogical reasoning, and autonomous inference.

2.4 Knowledge-Guided Cognitive Activities

In this architecture, certain cognitive activities emerge and evolve not through external task instructions or centralized control modules, but are naturally guided by the knowledge content embedded within knowledge graph nodes and their semantic connectivity. Nodes are not merely static carriers of knowledge; they serve as triggers and path directors for cognitive behavior. This mechanism enables the system to transcend the "executive cognition" paradigm and instead realize "thinking within knowledge."

2.4.1 Knowledge-Guided Cognitive Activities: Decision-Making, Reflection, and Goal-Generation Planning

Within this framework, decision-making, reflection, and goal generation are not performed by a centralized module but arise naturally from activation diffusion among nodes in the knowledge graph. When the Focus of Attention (FoA) migrates to action-related regions, the system automatically retrieves the decision-theory node most aligned with the current context and invokes its embedded computational method, transitioning seamlessly from semantic evaluation to action selection. This process further extends to reflection, goal generation, and maintenance of prospective intentions, forming a closed-loop cognitive flow of "immediate choice—self-monitoring—future execution."

(1) Role of Decision-Theory Nodes Decision-theory nodes provide localized decision control. Each node represents an invokable decision model with the following basic structure:

$$D_i = \{\text{Name}_i,\ \text{Model}_i,\ \text{Context}_i,\ \text{Method}_i\},$$

where Model_i describes the core mathematical formulation of the theory (e.g., the value function in prospect theory or the matching rule in recognition heuristics), and Method_i specifies its computational procedure for evaluating actions. This nodal representation allows the system to dynamically switch decision strategies across contexts, achieving distributed rationality rather than relying on a single utility model.

(2) Intra-Node Decision Computation When a decision-theory node D^* is activated, the system executes its internal method f_{D^*} to evaluate and select from the set of candidate actions $A(S_t)$:

$$a^t = f_{D^*}(S_t, A(S_t)).$$

For example: if D^* corresponds to prospect theory, it computes psychological utility as $U(a) = v(x - r)\pi(p)$; if it implements a recognition heuristic, it performs match-based decisions using semantic similarity; and in cases involving probabilistic paradoxes, it conducts conditional probability correction and counterintuitive reasoning. All decision behaviors are realized through the unified activation diffusion mechanism over the graph.

(3) Reflection and Goal-Generation Planning Mechanism After executing a local decision and receiving behavioral feedback, the system initiates a metacognitive reflection process based on the corresponding memory episode chain MEP_i. High activation of reflection nodes preferentially directs diffusion toward relevant retrospective subgraphs—including action nodes, expected outcome nodes, and actual consequence nodes. During this process, edge weights are dynamically adjusted according to the alignment between outcomes and expectations: successful paths are reinforced, while failed associations are weakened. This evaluation of original goal efficacy then triggers a new goal-generation workflow.

Goal generation is collaboratively achieved by the Large Language Model (LLM) and the knowledge graph structure: the system leverages causal and rule-based connections in the graph, combined with the LLM's "if–then" templating capability, to autonomously produce structured, executable goal directives. These directives are then instantiated as goal nodes, serving as new activation sources in subsequent cognitive flows.

For goals requiring delayed execution, the system creates Prospective Intention Nodes (PI-nodes), which are automatically reactivated when appropriate future conditions arise. Each PI-node has the following form:

$$PI_i = \{\text{intention\_content},\ \text{trigger\_condition},\ \text{deadline/context\_cue},\ \text{priority},\ \text{monitoring\_interval}\}.$$

Once generated, a PI-node resides in the self-cognitive subgraph with a positive baseline activation level, acting as a persistent cognitive marker for “unfinished tasks.” To support long-term intention maintenance, the system introduces an intention monitoring intensity mechanism: this intensity is dynamically modulated by the node’s priority and deadline—baseline activation automatically increases as the deadline approaches or priority rises, giving the PI-node a competitive advantage in activation diffusion.

When external conditions or internal states satisfy its TriggerCondition (e.g., a specific time arrives, a key event occurs, or a perceptual signal matches), the corresponding PI-node receives an immediate activation boost, rapidly ascending to the top of the FoA candidate set, thereby triggering immediate execution, contextual adaptation, or replanning. This mechanism endows the system with human-like prospective memory and autonomous task scheduling capabilities [4].

(4) Cognitive Significance By uniformly representing decision-making, reflection, goal generation, and prospective intentions as nodes and edges in the knowledge graph, the decision process ceases to be an external module call and instead becomes a natural path evolution within the knowledge system itself. The system can automatically switch decision theories in response to contextual changes, generate improved goals through experiential review, and reactivate intentions when future conditions are met—forming a continuous, adaptive, and metacognitively capable behavioral stream. This structured yet distributed decision-making framework constitutes the core mechanism behind the human-like cognitive cycle of “immediate action—reflection—future planning,” and can be efficiently implemented in practice using graph databases.

2.4.2 Knowledge-Guided Cognitive Activities: Emotion and Personality

In this architecture, emotion and personality are modeled as endogenous knowledge nodes within the self-cognitive subgraph, serving as central regulators of cognitive activity. Emotions exist as dynamic knowledge units defined by simulated hormonal parameters (e.g., dopamine, cortisol, serotonin, oxytocin), with their activation levels reflecting the current affective state in real time. Personality, in turn, is encoded as the long-term receptor sensitivity distribution to these hormonal signals, capturing stable individual response preferences over time—such as risk propensity, exploration drive, or social openness.

During cognition, emotion and personality nodes continuously modulate the direction and intensity of activation diffusion through their activation states and connectivity to concept, rule, and task nodes. For instance, high activation of the “anxiety” node strengthens edges in threat-related subgraphs while suppressing long-range associations, biasing attention toward conservative trajectories; conversely, active “curiosity” nodes enhance novelty detection gain and expand semantic diffusion, fostering cross-domain linking. Personality nodes shape individualized thinking styles and decision inertia by long-term tuning of emotional nodes’ response thresholds and decay rates.

Thus, emotion and personality are not external control modules but active knowledge entities within the graph that naturally participate in and guide the entire cognitive flow, enabling continuous, endogenous regulation of attention distribution, reasoning breadth, and behavioral tendencies.

(1) Emotion Nodes: Multidimensional Knowledge Modeling Based on Hormonal Parameters Emotion nodes are modeled around hormones as core elements. Each node represents a hormone-related knowledge unit and carries a semantic vector describing the dominant emotional state associated with that hormone. The system maps physiological mechanisms to semantic states via predefined knowledge: for example, increased activation of the “dopamine” node simultaneously elevates the baseline activation of related nodes such as “pleasure,” “exploration,” and “goal_approach”; when the “cortisol” node is active, activation diffusion contracts toward high-certainty, low-risk local subgraphs, suppressing divergent associations.

Moreover, the system supports high-level social-emotion nodes (e.g., “sharing_desire”), which do not correspond to a single hormone but emerge as composite motivational nodes co-activated by dopamine, oxytocin, and social-semantic structures (e.g., “express,” “feedback,” “empathy”). When “sharing_desire” is highly activated, connection weights to “expression–communication” subgraphs are dynamically enhanced, making subsequent activation diffusion more likely to follow outward-oriented paths—naturally generating tendencies to share experiences, seek interaction, or express opinions. This design is inspired by Damasio’s (1994) somatic marker hypothesis [5], treating emotions as endogenous value signals that guide decision-making.

(2) Personality Nodes: Long-Term Modulation via Hormonal Receptors Personality nodes encode the system’s long-term sensitivity profiles to various hormonal signals, reflecting response gains and decay characteristics across different emotional pathways. Through continuous interaction with the environment, these nodes automatically adjust their connection weights based on task outcomes, emotional feedback, and adaptive success, gradually forming stable, individualized response patterns. Thus, personality can be viewed as a “receptor distribution map” shaped by the long-term evolution of the knowledge graph, defining the system’s foundational cognitive style and preferences across states.

For example, a system with high social sensitivity (manifested as elevated oxytocin-pathway receptor weights) more readily activates social-motivation nodes like “sharing_desire” in similar contexts; in contrast, a high risk-aversion personality lowers the baseline activation of such nodes and enhances suppression of uncertain paths, biasing cognitive flow toward internal processing and conservative outputs.

(3) Emotion–Activation Coupling Mechanism During activation diffusion, the system considers not only semantic proximity and task relevance but also the current state of active emotion nodes. Specifically, the topological distance and connection strength between candidate nodes and emotion nodes dynamically modulate the activation gain they receive. For instance, when the “sharing_desire” node is highly active, all semantically linked nodes such as “express,” “display,” and “social_feedback” receive additional activation weighting during diffusion, increasing their likelihood of entering the FoA candidate set and driving socially intentional cognitive outputs. Simultaneously, changes in hormonal levels are directly reflected as increases or decreases in the activation of corresponding emotion nodes, enabling real-time evolution and maintenance of emotional states through feedback connections within the graph.

(4) Cognitive Significance The integration of emotion, personality, and high-level motivational nodes endows the system with endogenous dynamic regulation capabilities: emotions provide immediate activation biases and topological perturbations; personality ensures long-term stylistic consistency and preference stability; and composite emotion–motivation nodes (e.g., sharing_desire) confer behavioral directionality in specific social contexts. Together, these three components act as active knowledge entities within the graph, synergistically shaping cognitive flow to preserve individual uniqueness while maintaining flexible adaptation to internal and external environments—thereby achieving human-like emotion–cognition integration within a unified framework.

2.4.3 Knowledge-Guided Cognitive Activities: Anticipation, Hypothesis, and Prospective Memory

In human cognition, the anticipation of action outcomes serves as a core driver of decision-making and learning, enabling predictions about potential feedback, reward/punishment consequences, and environmental changes. This predictive processing permeates multiple cognitive levels—including perception, emotion regulation, plan maintenance, and thought generation. Inspired by this, our architecture links anticipation nodes to action nodes, allowing the system to generate human-like predictive thinking and further extend it to hypothesis formation and prospective memory, thereby supporting automatic intention retrieval upon encountering future cues and enabling delayed execution and future planning.

(1) Anticipation Modeling in Action Nodes In the semantic memory layer, each action node not only encodes the definition and execution characteristics of an action but also connects to one or more anticipation nodes that represent the multidimensional consequences the action may produce under different contexts, including:

- Changes in environmental states;
- Shifts in emotional or internal variables;
- Reward or punishment signals;
- Social responses and updates to trust structures.

Complex anticipation nodes can further link to secondary consequences, even extending into long-range causal chains. When the Focus of Attention (FoA) propagates along high-weight edges to these nodes, the system can automatically infer likely outcomes of an action based on the current scenario, providing real-time future-path estimates for decision-theory nodes.

(2) Scope and Phased Roles of Anticipation The role of action anticipation extends beyond the decision phase and operates across multiple core stages:

- **Perception stage:** Anticipation nodes modulate the distribution of perceptual salience by predicting future inputs, biasing the system toward potentially useful sensory channels in advance;
- **Decision stage:** Anticipations serve as key components of expected utility estimation, jointly determining the ranking of candidate actions alongside emotion nodes, reward models, and task constraints;
- **Execution and reflection stage:** After execution, the actual outcome is written back to the corresponding anticipation node to update the confidence of associated edges, enabling experience-driven dynamic refinement.

(3) Hypothesis Generation and Verification Mechanism The anticipation mechanism can be extended to support hypothesis generation and validation. When the FoA enters uncertain regions of the graph, the system automatically generates hypothetical paths via logical edges (e.g., "IF current expectation THEN hypothesized outcome") to enable counterfactual reasoning, uncertainty handling, and implicit inference about future scenarios.

(4) Implementation of Prospective Memory: Delayed Execution via Anticipation Nodes Concurrently with anticipation node creation, the system can mark certain future states as "prospective intentions," enabling their automatic reactivation when relevant cues appear. These prospective memory (PM) nodes inherit the structure of anticipation nodes and include the following attributes:

$$PM_i = \{\text{intention},\ \text{trigger\_condition},\ \text{time\_cue/event\_cue},\ \text{priority},\ \text{baseline\_activation}\}.$$

During perception and decision-making, the system continuously monitors for these trigger cues. When a cue matches the TriggerCondition (e.g., a specific time arrives or a particular event occurs), the activation level of the corresponding PM-node spikes instantaneously, entering the L1 (Event Frame) layer and capturing the FoA—producing a cognitive experience analogous to the human "I suddenly remembered I still need to do something." This mechanism supports parallel maintenance of multiple intentions. Low-cost periodic micro-activations prevent forgetting, and higher-priority intentions receive greater monitoring bias, making them easier to trigger.

(5) Cognitive Significance The integration of anticipation, hypothesis, and prospective memory enables the system to naturally simulate future scenarios within the knowledge graph, formulate hypotheses, maintain delayed-execution intentions, and automatically resume relevant plans when conditions are met. As a result, the system exhibits human-like foresight and proactivity: it can evaluate future pathways in uncertain environments, explore high-reward actions at minimal cost, and automatically activate intentions through time- or event-driven cues—thereby achieving sustained goal maintenance.

2.4.4 Knowledge-Guided Cognitive Activities: Procedural Actions, Logical Reasoning, and Creative Imagination

In human cognition, when action goals involve clear logic and predictable conditions, thought often spontaneously organizes into procedural structures—manifesting as integrated conditionals, termination rules, and internal computation mechanisms. To emulate this rational planning capability, our architecture introduces a code-template-driven action mechanism: action nodes internally store logical templates that are dynamically instantiated with current contextual information to generate executable logic structures, achieving dynamic unification across semantic, logical, and computational layers. This mechanism further extends to logical reasoning and creative imagination, inspired by Minsky's (1975) frame-based knowledge representation [6], enabling the system to construct novel concepts and virtual scenarios from existing knowledge.

(1) Template Structure of Action Nodes The system represents the logical framework of an action using a parameterized code template T_{code}, which includes conditions, actions, and alternative paths:

$$T_{\text{code}} = \text{``if \{condition\} then \{action\} else \{alt\}''}.$$

(2) Template Instantiation and Execution During cognitive execution, the system automatically fills the template based on the current context state S_{context}:

$$L_{\text{inst}} = \text{instantiate}(T_{\text{code}}, S_{\text{context}}),$$

producing a temporary logic script L_{inst}. The instantiation process relies on context-mapping functions:

$$\begin{aligned} \$\text{condition} &\Leftarrow f_{\text{cond}}(S_{\text{context}}), \\ \$\text{action} &\Leftarrow f_{\text{act}}(A_i), \\ \$\text{alt} &\Leftarrow f_{\text{alt}}(S_{\text{context}}). \end{aligned}$$

For example, the template of an "avoid_obstacle" node might be instantiated as: `if distance(object, agent) < threshold then change_path()`, which can directly drive embodied actuators.

(3) Logical Reasoning Pathway: Symbolic Rule Evaluation and Knowledge Integration To ensure formal rigor in procedural actions, the architecture incorporates a logical reasoning pathway. When the semantic slots (subject–predicate–object) in an event frame satisfy logical completeness, the system invokes an LLM to identify its underlying logical structure:

$$Q_{\text{logic}} = \text{LLM}(EFrame) \Rightarrow \{\text{formal\_logic\_expression}\},$$

and feeds the result into a symbolic reasoning engine (e.g., Prolog or DLV) for rule-based evaluation. The reasoning conclusions are written back to the knowledge graph as high-confidence nodes and increase the weights of related edges ($w_{\text{edge}} \uparrow \beta$), reinforcing the stability of causal chains. These conclusions can also directly populate parameters in action templates, ensuring that the generated L_{inst} satisfies both semantic coherence and logical provability.

(4) Creativity and Imagination Mechanism When the system operates under low task load and the mode coefficient λ_{mode} is low, activation diffusion becomes highly divergent, shifting the system into a creativity- and imagination-dominated pathway analogous to the human Default Mode Network (DMN). In this state, activation signals freely propagate across semantically distant subgraphs, enabling dynamic recombination of cross-domain nodes to generate novel scenarios, new concepts, or counterfactual structures.

The formation of imaginative pathways is guided by the probability distribution of activation diffusion, which can be formalized as:

$$P(n_{t+1}) \propto \exp\big(-\lambda \cdot \mathrm{dist}(n_{t+1}, C)\big),$$

where C is the central node of the current creative context (e.g., a vague goal, emotional state, or unfinished intention), $\mathrm{dist}(\cdot, \cdot)$ denotes semantic or topological distance in the graph, and λ reflects the current focus level of diffusion (modulated by λ_{mode}). This mechanism allows the system to explore associations far from the current focus while maintaining a degree of semantic coherence, thereby constructing virtual scenarios, alternative narratives, or hypothetical causal chains—supporting sophisticated "future simulation" capabilities.

Notably, during this divergent process, prospective intention nodes participate in activation diffusion as "future self anchors." Although their baseline activation is weak, they gain relatively higher competitive advantage under DMN-like conditions due to reduced inhibition. Consequently, the generated imaginary content is not random association but purposeful future rehearsal (mental time travel; Schacter et al., 2007) [3]: imagined scenarios tend to revolve around unfinished tasks, potential action plans, or anticipated outcomes.

(5) Cognitive Significance By integrating code templates, logical reasoning, and creative imagination, this architecture transforms each action node from a mere semantic entity into a generative, executable, and verifiable logical program unit. Moreover, in divergent mode, the system can conduct goal-directed future simulations grounded in prospective intentions, forming a complete closed loop across semantics—logic—execution—hypothesis—creation. This structure endows the system with both rigorous planning capabilities and creative thinking, and can be efficiently implemented in engineering practice using Python executors and graph-generation algorithms.

2.4.5 Knowledge-Based Action Cognition: A Dual-Channel Learning Mechanism for Steps and Action Proficiency

In this architecture, actions are not isolated behavioral commands but deeply embedded cognitive entities within the knowledge graph. The system acquires action competence through two complementary cognitive pathways: *step proficiency* (structured procedural fluency) and *action proficiency* (embodied operational skill). Both are represented as nodes and edges in the knowledge graph, embodying the core principle that "cognition is knowledge activation."

Step proficiency is reflected in the strength of transition relations between sub-steps of an action. Complex actions are decomposed into ordered sequences of sub-step nodes $s_1 \to s_2 \to \cdots$, with adjacent steps connected by directed edges of type `nextStep`. The weight $w_{i,i+1}$ of each such edge dynamically encodes the fluency, stability, and success rate of executing that transition. As tasks are repeatedly performed, these weights are automatically updated based on empirical statistics—such as successful transition frequency, execution time variance, or error rates—yielding a quantified representation of structural procedural mastery. This mechanism enables efficient action planning, anomaly detection, and process correction grounded in the graph's topology.

Action proficiency, in contrast, is coupled to reinforcement learning policy modules via `hasPolicy` edges. Each atomic action node connects through such an edge to a dedicated policy node—a learnable cognitive component that receives two types of input: (1) a snapshot of the current state of the self-cognitive subgraph (including intentions, emotions, goal context, and other high-level cognitive variables), and (2) real-time perceptual signals from the external environment. Based on this "cognition–perception fused state," the policy network generates low-level control commands for the embodied execution system and continuously refines its parameters through environmental reward feedback. Thus, the system learns a mapping from the joint representation of internal cognitive states and external contexts to behavioral outputs.

This dual-channel mechanism ensures that action competence emerges intrinsically from the evolution of the knowledge graph: step proficiency, encoded in dynamic edge weights, supports symbolic-level reasoning and planning; action proficiency, refined through closed-loop interaction with the environment, maintains adaptability to the physical world. Together, they co-evolve within a unified cognitive architecture, achieving an organic integration of procedural knowledge and operational skill—and providing a structured foundation for metacognitive reflection (e.g., "Why did it fail?" or "Should I adjust the sequence?").

3 Limitations and Future Research Directions

Although the proposed architecture exhibits strong theoretical potential for general intelligence, its practical implementation and deployment face multiple challenges.

3.1 Potential Optimization Directions for Graph Neural Networks in Knowledge-Based Cognitive Functions

The current implementation primarily relies on symbolic node representations and rule-guided activation propagation. To further enhance the system's capabilities in large-scale knowledge processing, modeling fuzzy associations, experience-driven evolution, and multimodal fusion, integrating Graph Neural Networks (GNNs) would provide a differentiable, end-to-end learning framework for cognitive functions.

GNNs can map nodes and edges into learnable embeddings in continuous vector spaces, allowing attributes such as memory strength, factual confidence, and emotion-coupling weights to be automatically adjusted through experiential feedback. This data-driven representation facilitates "axiomatic plasticity"—the gradual updating of foundational beliefs in light of new evidence—and supports subconscious-level knowledge expression that is low in explicitness but high in generalization.

In node activation and semantic diffusion, GNN message-passing mechanisms can offer a continuous, trainable dynamic model for signal propagation. Compared to the current diffusion strategy based on predefined decay rules, GNNs can adaptively integrate semantic similarity, emotional bias, task context, and cross-modal cues, rendering the activation diffusion process more context-sensitive and computationally efficient while preserving the structural constraints of the graph topology.

For event frame construction, GNNs' graph aggregation and structure-aware attention mechanisms can automatically extract core propositions and their attribute dependencies from raw perceptual streams or memory fragments, yielding more stable and generalizable event representations. This capability enhances the system's ability to reuse and transfer event templates across diverse contexts, improving understanding and prediction in complex dynamic scenarios.

Moreover, GNNs can strengthen intelligent guidance of attention shifts: by jointly encoding the current activation distribution, global graph structure, emotional state vectors, and the self-cognitive subgraph, GNNs can learn to predict which regions are most likely to become the next focus of attention, enabling a finer, more human-like balance between divergent exploration and convergent reasoning.

At the engineering level, combining GNNs with advanced graph-computing techniques—such as subgraph sampling, knowledge distillation, graph compression, and distributed inference—can significantly reduce computational overhead for large-scale knowledge graphs, alleviating latency and resource bottlenecks in real-time cognitive tasks and enabling efficient operation in complex, highly dynamic real-world environments.

In summary, introducing GNNs does not entail rebuilding the existing architecture but rather endowing it with a unified differentiable dynamics layer that permeates knowledge representation, activation propagation, structural evolution, and multimodal fusion. This direction holds promise for advancing the system toward general artificial intelligence with continuous self-improvement, experience-based revision, and cross-modal understanding.

3.2 Limitations of the Perception Module and the Proposal of Feature-Guided Neural Networks

The current perception module largely depends on Convolutional Neural Networks (CNNs). While CNNs achieve remarkable performance on standard tasks like static image recognition, their training heavily relies on closed-label datasets and predefined feature-extraction templates. This leads to clear limitations in the open-ended, dynamic environments required for general artificial intelligence. CNNs struggle with novel scenes, rare events, and unlabeled samples, lacking flexible feature transfer and real-time adaptation capabilities, and thus fail to achieve human-like perceptual agency and deep semantic understanding.

Furthermore, the system exhibits a notable deficiency in spatial perception and spatial imagination. It currently lacks symbolic encoding and generative mechanisms for three-dimensional spatial relationships, virtual scene layouts, or object interactions. Consequently, the system cannot demonstrate real-world-level generalization in tasks involving spatial reasoning, scene construction, or embodied planning. The absence of an internal representation of spatial structure also constrains the generation of complex hypothetical scenarios and embodied cognitive simulation.

To overcome these limitations, we propose integrating meta-learning into a feature-guided neural network framework. In each sensory modality (vision, audition, touch, etc.), a meta-learner based on the Model-Agnostic Meta-Learning (MAML) algorithm could enable rapid adaptation to new tasks and dynamic adjustment of feature preferences. By extracting low-level features from input signals (e.g., color, edges, texture, or audio frequencies) and synthesizing their frequency, confidence, and historical relevance, the system can dynamically generate a “feature preference vector” that guides perceptual channels to selectively attend to high-value features in new scenes while suppressing noise and irrelevant patterns.

For spatial perception, we suggest combining complex object decomposition strategies—such as parsing visual objects into shape, texture, and geometric primitives—with structure-based object recognition methods and point-cloud technologies to support realistic 3D environmental modeling. By automatically constructing dynamic coordinate systems anchored to visual centers, the system can maintain spatial consistency during movement or viewpoint changes. In large-scale environments, GPS or other localization sensors could further serve as external reference frames to improve positional stability.

To achieve cross-modal unity and higher-order spatial reasoning, a “meta-meta-learner” layer could be added atop the perception system to integrate feature preferences across sensory channels. This layer would abstract consistent object properties across vision and touch, align spatial descriptions between vision and language, and capture associative patterns between auditory cues and emotional signals. Such a structure would not only enhance multimodal fusion but also provide a shared abstract representation for spatial imagery generation, virtual scene construction, and embodied planning.

This feature-guided neural network aims to transcend the static template constraints of traditional CNNs. Through meta-learning, structure-aware processing, dynamic coordinate systems, and cross-modal integration, it seeks to build a perception system that is more active, adaptive, and spatially coherent—thereby enhancing general AI's capabilities in sustained perception, spatial reasoning, and endogenous imagination within complex environments.

4 Conclusion

This paper presents a cognitive architecture designed for the pursuit of artificial general intelligence. Its core innovation lies in unifying memory, attention, reasoning, and emotion within a multimodally embedded knowledge graph, where dynamic attention migration is driven by node activation and diffusion mechanisms. Concepts, events, actions, rules, emotions, and even decision strategies all exist as heterogeneous nodes, interconnected by plastic, weighted edges representing semantic, logical, and causal relationships—forming a highly integrated and continuously evolving cognitive network.

The system supports two complementary modes of attentional shift: under task-driven conditions, activation diffusion is constrained by goals and rules, exhibiting convergent, central-executive-network (CEN)-like reasoning; during idle states, diffusion expands and inhibition weakens, manifesting divergent, default-mode-network (DMN)-like association. These modes are smoothly and continuously balanced via the mode coefficient λ_{mode}, allowing flexible trade-offs between execution efficiency and creative exploration.

Critically, high-level cognitive activities do not originate from external controllers but naturally emerge from the activation dynamics and connectivity patterns of specific node types—such as decision-theory nodes, prospective-intention nodes, and action–anticipation nodes. Emotions exist as knowledge nodes parameterized by simulated hormonal levels, modulating the gain and direction of activation diffusion to influence attentional biases and cognitive topology in real time. Personality is encoded as a long-term distribution of receptor sensitivities, shaping individualized cognitive styles and behavioral inertia. Episodic memory is stored using a "event-head-node + temporally ordered proposition sequence" structure and tightly coupled with semantic memory through shared argument nodes, enabling analogy-based reasoning, causal inference, and narrative generation grounded in experience.

It should be noted that the architecture remains in the stage of theoretical formulation and preliminary engineering validation. Its strengths lie in structural unity, mechanistic transparency, and strong interpretability and scalability—offering a viable pathway toward continuous learning, contextual adaptation, and cross-domain knowledge integration. However, significant challenges remain, including real-time computational efficiency for large-scale knowledge graphs, precise alignment between multimodal perceptual signals and graph embeddings, and modeling of intentions in complex social interactions.

Future work will focus on integrating GNNs to optimize node activation and diffusion, leveraging meta-learning to enhance the perceptual system's open-world adaptability, and deploying the architecture in real-world applications such as human–AI collaboration, autonomous task construction, and embodied reasoning. We believe that a cognition paradigm grounded in knowledge graphs and powered by activation diffusion offers a biologically plausible and engineering-feasible new trajectory for the development of artificial general intelligence.

References

[1] Tulving, E. (1983). *Elements of Episodic Memory.* Oxford: Oxford University Press.

[2] Baddeley, A. D. (2000). The episodic buffer: a new component of working memory? *Trends in Cognitive Sciences*, 4(11), 417–423.

[3] Schacter, D. L., Addis, D. R., & Buckner, R. L. (2007). Remembering the past to imagine the future: The prospective brain. *Nature Reviews Neuroscience*, 8(9), 657–661.

[4] Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences*, 24(1), 87–114.

[5] Damasio, A. R. (1994). *Descartes' Error: Emotion, Reason, and the Human Brain*. New York: G. P. Putnam's Sons.

[6] Minsky, M. (1975). A framework for representing knowledge. In P. H. Winston (Ed.), *The Psychology of Computer Vision* (pp. 211–277). New York: McGraw-Hill.