ChatGPT

Integrated Cognitive AI Architecture Whitepaper

Recommended Reference: *The Cognitive Core – An Integrated Cognitive Architecture* (Royse & Caudill, 2025)

One technical whitepaper that comprehensively covers all the requested components is **"The Cognitive Core: An Integrated Cognitive Architecture"** by Royse and Caudill (2025). This paper presents a brain-inspired **hybrid neuro-symbolic** design with multiple cognitive **agents** and a structured memory system, offering a clear blueprint for an AGI-like system. It integrates high-level reasoning, memory, learning, and emotional modulation in a single architecture [1] [2]. Below, we outline how this whitepaper addresses each component:

Multidimensional Internal State Vectors and Cognitive Modes

The Cognitive Core explicitly represents internal cognitive state through **multiple dimensions**, analogous to neuromodulators and brain networks. It incorporates a **simulated neurochemical system** that adjusts parameters like *"Dopamine"* and *"Serotonin"* levels to create a dynamic internal state [2]. These parameters act as a **cognitive vector** influencing the agent's behavior – for example, higher dopamine might reflect a positive reward context (biasing decisiveness), while serotonin can modulate calmness or patience. Crucially, the architecture also includes large-scale **"network" meta-regions** inspired by the brain's intrinsic networks: a Default Mode Network, Central Executive Network, and Salience Network [3]. These act as global cognitive modes that orchestrate the system's focus. For instance, the **Salience Network** boosts the "attending" dimension by directing attention to urgent or important stimuli, whereas the **Central Executive** engages during intense "deciding" or problem-solving tasks. This design allows the AI to **switch between cognitive modes** (e.g. introspective reflection vs. focused execution) based on context [3], much like a human shifting from planning (`Intending`) to attention (`Attending`) when a salient event occurs. In summary, the internal state is represented by a **vector of cognitive parameters** (attention focus, decision urgency, intention commitment, adaptability, etc.), which the system continuously tracks and updates.

Adaptive Learning and Neuroplasticity Mechanisms

The whitepaper places strong emphasis on **adaptive learning**, mirroring neuroplasticity. The Cognitive Core uses a dual memory model to address the stability–plasticity dilemma: transient experiences are kept in short-term memory, while important knowledge is consolidated into long-term memory, much like a brain shifting memories from hippocampus to cortex [4] [5]. This ensures continuous learning (plasticity) without overwriting established knowledge (stability). Moreover, a specialized agent – the **Basal Ganglia Analogue (BGA)** – is responsible for *reinforcement learning and habit formation*. This component observes outcomes of actions and adjusts future behavior based on reward signals (similar to dopamine-driven learning in the brain) [6] [7]. Over time and interactions, the BGA learns which actions yield positive results and biases the system's decisions accordingly [7] [8]. For example, if asking clarifying questions has led to better outcomes in the past, the BGA will increase the tendency (policy weight) to do so in similar situations [7]. Conversely, actions that led to negative feedback are suppressed. This **reinforcement-driven update** of the internal policy vector is a direct analog of neuroplastic adaptation – the system's "Deciding" mechanism becomes tuned by experience.

Additionally, every cognitive cycle can trigger learning: important or highly novel events are immediately consolidated to long-term memory (a process analogous to memory reinforcement by emotional or reward salience). In short, the architecture continuously **updates its internal state vectors and knowledge links through interaction and feedback**, implementing a form of lifelong learning.

Graph-Based Memory with Reinforcement-Weighted Relations

For storing cognitive structures (knowledge and experiences), the Cognitive Core uses a **graph-based data model**. It features a persistent **Graph Memory System** built on a Neo4j knowledge graph, which holds both **episodic memory** (experiences, events) and **semantic memory** (facts, concepts) [9]. In this graph, **nodes** represent entities, concepts, or events, and **edges** represent relations or associations (e.g. temporal sequence, causal link, category membership). This design inherently supports **weighted relations** and associative retrieval. The paper explains that links can be annotated with properties like recency, importance or emotional "tags," effectively acting as weights that influence recall priority [10] [11]. For instance, an episodic node tagged with a strong emotional weight (from the amygdala analog) will be more salient and thus more likely to be retrieved later in relevant contexts [10] [11]. Similarly, frequently used connections or those that led to successful outcomes can be considered *reinforced* over time. This graph memory enables **true associative recall**: given a cue, the memory agent can traverse connected nodes to pull up related information. The whitepaper illustrates that if the AI is cued with "cat," it can find the *Cat* concept node, then retrieve linked episodic nodes (e.g. a past event where the user discussed their sick cat) along with semantically related nodes like *Veterinarian* or *Pet* [12] [13]. This mirrors how human memory works via associations, unlike a flat database or vector store. The **reinforcement-weighted links** mean the system can prioritize more relevant or stronger associations – effectively encoding knowledge strength. The graph model also supports structured queries (using Cypher), allowing the AI to answer complex questions by traversing relations (e.g. "when was the last time X happened?") rather than just keyword matching [14] [13]. Overall, this flexible, graph-based memory with updatable link weights provides the backbone for storing and retrieving the AI's cognitive structures.

Hierarchical & Modular Architecture (Meta-Regions, Clusters, Threads)

The architecture in the whitepaper is **hierarchical and modular**, organized in layers that conceptually align with *Meta-Regions*, *Cognitive Clusters*, and *Cognitive Threads*. At the highest level, the design incorporates the aforementioned brain-inspired meta-regions (Default Mode, Executive, Salience networks) that manage global cognitive state [3] – this corresponds to a *meta-level control* layer, ensuring the system can regulate its overall mode of operation (idle introspection vs. active engagement, etc.). Beneath this, the system is composed of a "**society of cognitive agents**," each specialized for certain mental functions [15] [16]. These agents act as **Cognitive Clusters** akin to brain regions: for example, a **Frontal Lobe Analogue (FLA)** agent handles executive reasoning and planning, a **Hippocampus (HC)** agent manages memory encoding and recall, an **Amygdala (AC)** agent provides emotional appraisal, a **Basal Ganglia (BGA)** agent handles decision gating and habit learning, a **Parietal Lobe Analogue (PLA)** manages attention/orientation, etc [16]. Together they cover all major facets of cognition – *perception, attention (attending), decision-making (deciding), intention/goal management (through executive planning), learning (adapting), emotion, and action execution* [16]. Each agent can be thought of as a module or cluster that processes a particular aspect of the cognitive state.

Crucially, these agents run concurrently, which realizes the concept of **Cognitive Threads**. The paper notes that each agent operates in parallel "as separate threads/modules in a program" and communicates with others as needed [17]. A **dual-mode communication fabric** connects the agents: a global **Inter-Agent Communication Bus** for broad, asynchronous message broadcasting (analogous to a brain-wide signal or hormone) and direct point-to-point links for high-speed exchanges between specific agents [18] [19]. This means multiple cognitive processes (threads of thought) can progress simultaneously and interact. For example, the vision processing agent (if present) could be analyzing a scene while the FLA is formulating a plan and the memory agent is retrieving relevant facts – each on its own thread, coordinating via messages. The meta-regions (like the Salience Network analog) monitor these threads and can elevate one process over others when required (e.g. diverting global attention to a critical stimulus) [20] [21]. The architecture is also **hierarchically extensible**: new specialized agents (skills or modalities) can be added without altering the core design, simply plugging into the communication bus [22] [23]. This reflects a layered hierarchy: a stable cognitive core (meta-regions and essential clusters) with the ability to incorporate additional threads/modules for new capabilities. In summary, the Cognitive Core exemplifies a hierarchical cognitive architecture where **top-level meta-regions coordinate mid-level clusters**, and **clusters consist of concurrent threads** performing distinct cognitive tasks.

Dynamic Adjustment of Cognitive Parameters Over Time

Finally, the whitepaper demonstrates mechanisms for **dynamic adjustment of cognitive parameters** as the system runs. These adjustments occur through the simulated neuromodulator levels and other internal dials that change in response to interactions. For instance, the system can vary its “attention weight” or decision threshold based on context, analogous to how a brain might increase acetylcholine to heighten focus or adjust dopamine levels to alter risk-taking behavior [24]. If the AI finds itself in a novel situation, it might **raise an “alertness” parameter** (increasing the attending vector component) to devote more resources to observation and learning. The Cognitive Core explicitly includes a **neurochemical modulation mechanism**: parameters like dopamine can implement reward prediction error signals for learning, while others like norepinephrine or acetylcholine could modulate exploration vs. exploitation [24] [6]. These internal values are continually tuned. For example, after repeated positive reinforcement, the BGA might lower the threshold for a particular action (making the system more likely to choose it), effectively adjusting a cognitive bias over time [7] [8]. The paper also describes how the **cerebellum analog agent** fine-tunes actions by comparing predicted vs actual outcomes and learning to minimize error [25] [26] – this implies parameters governing action execution get refined with practice (e.g. adjusting timing or confidence levels). Moreover, emotional parameters fluctuate: a *“mood state”* is integrated over recent inputs in the AC (amygdala) agent, and this can skew the AI’s responses (e.g. a prolonged high-threat environment might put the system in a cautious mode) [27] [28]. All these examples show a **time-varying adaptation of internal settings**. In essence, The Cognitive Core architecture ensures that its cognitive vector components (attention, decision preference, intention commitment, learning rate, etc.) are not static – they **evolve dynamically** as the agent gains experience, receives feedback, and internally regulates its focus. This dynamic tuning of parameters over time gives the system a form of **self-regulation and meta-learning**, allowing it to optimize its cognitive performance continuously.

Source and Further Details

The Cognitive Core whitepaper (Royse & Caudill, 2025) is an excellent reference that presents all the above elements in an integrated design. It provides clear explanations of each module and mechanism, often drawing parallels to neurocognitive functions. The paper also includes implementation-level insights – for example, it discusses using a Neo4j graph database for memory persistence and query, a publish/subscribe bus for agent communication, and leveraging large language models (LLMs) within agents for reasoning [9] [17]. Notably, it blends **symbolic structures** (knowledge graph, logic) with

learning components (LLMs, RL-based decision module), illustrating a **hybrid cognitive architecture**. For anyone developing a cognitive AI system, this whitepaper serves as a comprehensive blueprint, covering everything from representational choices to adaptive algorithms. All the key components – multidimensional state vectors, neuroplastic learning, weighted relational memory, hierarchical (layered) organization, and dynamic parameter adjustment – are present and well-integrated in the design [2] [29]. Thus, *The Cognitive Core* can be highly recommended as a state-of-the-art technical whitepaper that aligns with the described requirements, offering a roadmap toward building an AI with human-like cognitive breadth and adaptability.

Sources:

- Royse, C., & Caudill, B. (2025). *The Cognitive Core: An Integrated Cognitive Architecture* [1] [2] [7] [12] (preprint). This paper details the full architecture (agents, memory graph, learning mechanisms, etc.) and includes comparisons to neuroscience and prior cognitive architectures. It is available on ResearchGate and provides extensive references and some code-level discussion for implementation.

---

[1] [2] [3] [4] [5] [6] [7] [8] [9] [10] [11] [12] [13] [14] [15] [16] [17] [18] [19] [20] [21] [22] [23] [24] [25] [26] [27] [28] [29]
(PDF) The Cognitive Core: An Integrated Cognitive Architecture
https://www.researchgate.net/publication/392774960_The_Cognitive_Core_An_Integrated_Cognitive_Architecture