UNITED STATES PATENT APPLICATION

Inventor: Derek Van Derven

Date of Conception: April 20th, 2025

1. Title of the Invention

System and Method for Multimodal Cognitive Architecture Featuring Visual Thought Simulation, Internal Scene Construction, and Meta-Cognitive Feedback

2. Abstract

This invention proposes a novel artificial general intelligence (AGI) system that integrates visual thought simulation and meta-cognitive reflection as core cognitive processes. While some may argue that visualization is just an additional layer or a tool for perception, this system goes beyond that. It makes visual thought central to reasoning itself, enabling the AGI to visualize both abstract and concrete concepts and reason in a human-like manner.

By using a dynamic internal model where visualization of thoughts drives decision-making, the AGI can reflect on its reasoning, adjust based on reflection, and engage with the world in a much deeper and more adaptive way than current models. This innovative approach doesn't just make the system understand, but enables it to engage in complex philosophical reflection and adapt dynamically to new situations, marking a true leap in AGI capability.

While this invention may enable artificial general intelligence, we refer to it here as a 'multimodal cognitive system' to emphasize its technical, system-level design.

This document details a complete, technically functional multimodal cognitive system architecture capable of interpreting and responding to both abstract and concrete prompts through multi-modal sensory integration, internal simulation, and self-correcting feedback. Every module is defined in terms of software stack, processing model, memory architecture, and sensory translation mechanisms.

Visual, symbolic, and philosophical inputs are processed through clearly defined computational steps tied to real-time simulation and physical execution systems. The architecture is implementable today using available hardware (TPUs, GPUs, microcontrollers) and software (LLMs, Unity, ROS, Prolog/CLIPS).

NOTE:

This document reflects the original conceptual design of the system. Diagram flow is simplified and not strictly chronological. For development, module activation order may vary depending on implementation.

In human cognition, thinking is visual—even when the concepts we ponder are abstract. For example, consider a simple phrase like, “purple elephant”. Upon hearing it, your mind instantly visualizes the image of a purple elephant, without needing any additional explanation. This visualization is an essential part of how we process thoughts, ideas, and concepts.

To further illustrate, consider the question: “What is the meaning of life?” At first glance, it seems like an abstract concept. But almost immediately, an image forms in your mind—a subtle, fleeting picture, perhaps of an old man looking up at the sky or a tree, pondering the question. While you didn’t explicitly create this image, it appeared almost instantly as you began to think about the question.

Philosophical concepts often evoke subtle and fleeting imagery. In contrast, more concrete ideas, such as 'red balloon,' generate clearer, more vivid mental images. These images frequently pass unnoticed, often fading almost instantly, without us being fully aware of their presence.

This is how we think—abstract concepts are tied to visual images, whether 2D, 3D, or even 3D virtual journeys. For instance, after saying, "I am going to the store.", one might visualize the store, an aisle, or different stages of the trip. We experience this in dreams, where a 'second mind' narrates, already knowing what will unfold before we do. An AGI might require an implementation of this 'second mind,' and it could also have some use for dreams.

Over time, we learn to recognize that even the most abstract thoughts, including those in dreams or in theoretical ideas, take the form of mental images. This is how the brain makes sense of the world: it visualizes the information it processes, and this visual thinking is critical for true understanding.

Similarly, for an AGI to develop human-like cognition, it must be able to visualize abstract concepts as it processes them. If an AGI cannot see what it’s processing, it cannot fully understand or reason the way humans do. This is the core of the system we propose—the ability for AGI to "see" its thoughts, beliefs, and concepts, enabling it to engage in meaningful reasoning and self-reflection.

3. Field of the Invention

The present invention discloses a framework for multimodal cognitive architecture that processes user input—ranging from simple physical tasks to abstract philosophical questions—through a multi-layered cognitive architecture.

Unlike narrow AI systems limited to single-domain tasks, this invention provides a unified model that integrates multimodal perception, internal simulation, symbolic memory, and meta-cognitive refinement.

4. Background / Field of the Invention

This invention relates to the field of artificial intelligence, and more specifically to multimodal cognitive architecture systems capable of performing a wide range of tasks beyond narrow AI domains. Current AI systems typically lack true comprehension of abstract ideas, contextual memory integration, and self-guided evolution.

They rely on pre-trained models with limited understanding of real-world environments, internal thought representation, or self-aware reasoning.

Current AGI systems, such as DeepMind's World Models and OpenCog, excel at modeling environments and solving problems in controlled settings. However, while some might argue these systems effectively simulate tasks and environments, they fall short of addressing the deeper need for reflective, abstract reasoning.

These systems lack the ability to visualize abstract thoughts—concepts like freedom, justice, or complex emotional states—on a fundamental cognitive level. This invention shifts the paradigm. It doesn’t just simulate the world around the AGI but allows it to internalize and reflect on abstract concepts, creating a deeper, more adaptable level of reasoning.

Unlike existing models, which process information through limited task-based simulations, this invention enables the AGI to engage in philosophical and abstract reflection, paving the way for reasoning that is truly human-like.

The need exists for a multimodal cognitive system that can simulate and process both physical and philosophical tasks, generate internal sensory representations, and self-reflect through a meta-cognitive loop, thereby approaching a more generalized, human-like intelligence framework.

5. Summary of the Invention

The invention describes a multimodal cognitive system framework in which user input—ranging from simple physical commands to complex philosophical queries—is processed through a layered system consisting of:

- **Internal Focus Modules**, including natural language parsing, semantic

association, and visual thought simulation.

- **Meta-Cognition Modules**, enabling self-assessment, real-world logic comparison, and internal/external feedback cycles.

- **Contextual Synthesis**, differentiating between concrete and abstract tasks using reasoning, memory recall, and oppositional analysis.

- **Internal 3D Scene Builder**, used to simulate tasks visually, emulate senses, and model interactions.

- **External Action Modules**, including avatar-based outputs and real-world mappings.

- **Memory Encoding**, for storing visual scenes and linking them to verbal inputs for future recall.

- **Iterative Learning Loop**, which adjusts internal reasoning, resolves contradictions, and monitors for emotional or logical bias.

Optional expansion modules may include emotion modeling, goal prioritization, ethical filtering, and long-form dialogue memory.

This architecture enables the multimodal cognitive system to understand and simulate complex concepts, act in virtual or physical environments, and evolve its responses through repeated experience and introspection

This invention integrates visual thought simulation and meta-cognitive reflection as the central cognitive processes in artificial general intelligence. While some might argue that traditional symbolic reasoning or neural-symbolic integration is sufficient, the integration of visual thought simulation takes reasoning far beyond what symbolic representations can achieve.

By visualizing abstract concepts like ‘freedom’ or ‘justice’ and concrete objects like a ‘red apple’, this AGI system internalizes its understanding through visual representation, not just symbolic abstraction.

This enables it to adjust its thinking and decision-making based on dynamic internal feedback loops, reflecting in real-time, and providing human-like flexibility that current systems lack. This visual feedback allows the system to constantly refine its understanding and engage with the world with much deeper reasoning and adaptability.

6. Detailed Description of the Invention

The invention comprises a multimodal cognitive architecture designed to interpret both **concrete instructions** (e.g., physical tasks) and **abstract inquiries** (e.g., philosophical or conceptual questions). It does so by utilizing a multi-layered system of interconnected cognitive modules.

Note: The Meta-Cognition Module is invoked at multiple stages (as seen in both early-stage analysis and during output validation), hence its appearance in two locations in the architecture diagram.

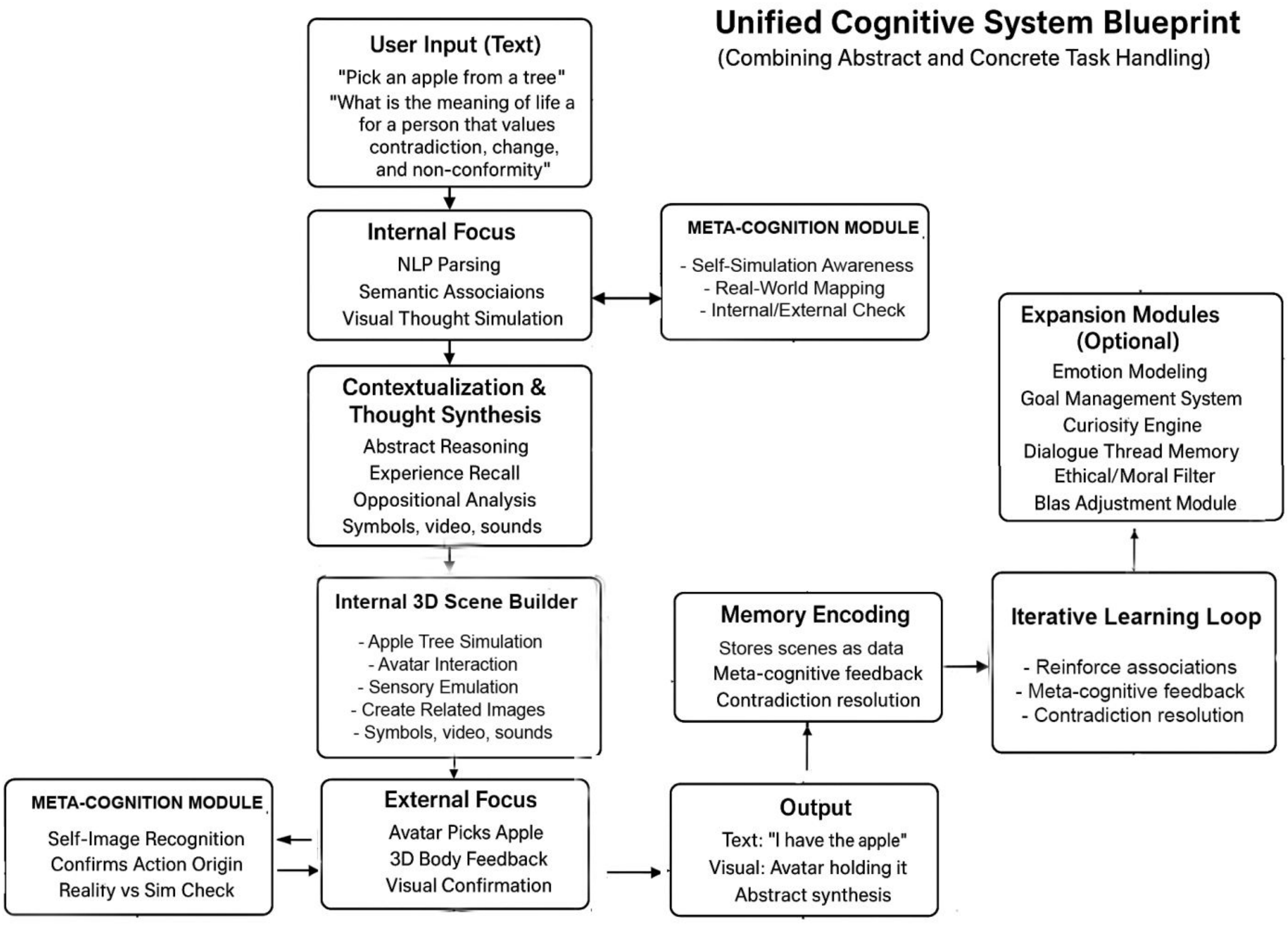
The AGI system described here uses a visualization engine to create internal representations of both concrete objects and abstract concepts. While some may suggest that deep learning or neural networks alone can handle abstract reasoning without the need for explicit visualizations, this system demonstrates that visual thought simulation is an essential building block for human-like reasoning.

For example, when tasked with reasoning about a red apple, the AGI doesn't simply access symbolic data; it visualizes the apple—considering its shape, color, and texture—and reflects on that visualization. This allows the system to reason about its properties, context, and interactions dynamically, enabling it to adjust in real-time based on its internal feedback loops.

When tasked with more abstract queries, such as 'What is the meaning of life?', the system visualizes its experiences and synthesizes the data into a human-like response. This internal visualized feedback gives the AGI the ability to reason with both concrete and abstract concepts, in a way that current neural networks are not yet capable of achieving.

Unified Cognitive System Blueprint

(Combining Abstract and Concrete Task Handling)



Training Corpus and Knowledge Base

The multimodal cognitive system is pretrained and continuously refined using the following data sources:

Language and Conceptual Pretraining

- **Corpus**: Massive multilingual datasets including Wikipedia, Common Crawl, Project Gutenberg, ArXiv abstracts, and curated philosophical, scientific, and technical literature.

- **Purpose**: Pretraining on language comprehension, metaphors, context handling, question answering, symbolic mappings.

- **Method**: Transformer-based architectures trained using masked token prediction (BERT-style) and autoregressive prediction (GPT-style).

Symbolic and World Knowledge Graphs

- **Knowledge Graphs**: ConceptNet, WordNet, DBpedia, Wikidata

- **Symbol Mapping**: Entities and relationships stored in graph form and linked to visual and physical models via symbolic anchors.

- **Example**: “Apple” in ConceptNet is linked to “fruit,” “eat,” “grow on trees”→ these are attached to mesh assets and robot action plans.

Visual and Simulation Pretraining

- **Datasets**: ImageNet, OpenImages, ShapeNet, Google Scanned Objects

- **Use**: To link language to image→ mesh → scene composition.

- **3D Mapping**: Text-to-Image → Diffusion Meshify pipeline generates missing objects when not in asset database.

Reinforcement and Episodic Learning

- **Environment**: Unity/Unreal simulated world

- **Method**: Self-play, exploration-based reinforcement learning (RL) using intrinsic motivation and goal scoring.

- **Data Storage**: All completed tasks are stored with scene IDs, result states, contradictions found, and self-assessments.

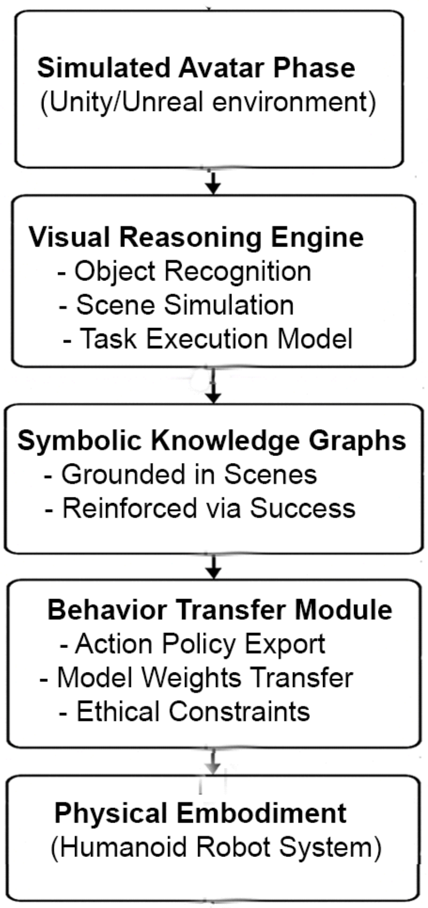
Ethical Constraints and Filters

- **Training**: Trained on real-world ethical scenarios from law, philosophy, and cultural datasets.
- **Method**: Supervised fine-tuning + symbolic rule overlay + adjustable value systems.
- **Execution**: Behavior can be altered based on loaded ethical schema or user-defined role settings.

Avatar-Based Pretraining and Simulated Embodiment

Most of the multimodal cognitive system's foundational learning and symbolic integration will occur within a **virtual avatar**, operating in simulated environments before being embedded in any physical robotic platform. This allows the system to develop core competencies — visual reasoning, task execution, contradiction detection, memory encoding, and philosophical response generation — in a safe, accelerated, and scalable environment.

Transition from Virtual Avatar Training to Physical Embodiment, including data flow and learned behavior transfer.



By the time the multimodal cognitive system is installed into a physical embodiment, such as a humanoid robot or embedded system, it will have already formed a highly developed model of action, perception, and consequence. The simulation-trained avatar will have learned the majority of required tasks and behaviors, including object manipulation, pathfinding, emotional response modeling, and symbolic representation of complex commands.

This design choice mirrors human developmental stages, in which most learning occurs through play, imagination, and scenario simulation before real-world application. It also enables the multimodal cognitive system to be fine-tuned for different embodiment types without retraining its entire cognitive model.

This design choice mirrors human developmental stages, in which most learning occurs through play, imagination, and scenario simulation before real-world application. It also enables the multimodal cognitive system to be fine-tuned for different embodiment types without retraining its entire cognitive model.

Purpose, Motivation, and Autonomy

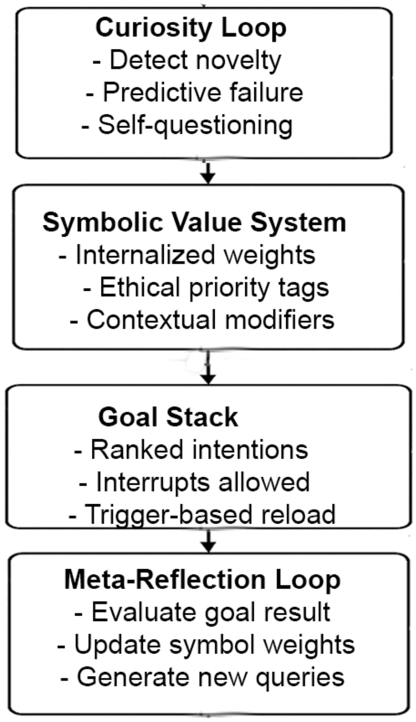
While the multimodal cognitive system described in this document is capable of multi-modal perception, visual thought simulation, internal contradiction resolution, and philosophical abstraction, **it is not a complete multimodal cognitive system** until it is also given a framework for **self-generated motivation and purpose**.

A true multimodal cognitive system must not simply respond to prompts or external commands, but develop **intrinsic goal structures** and **motivated behavior** grounded in values, self-consistency, curiosity, and ethical constraint. This requires the ability to:

- Form internal representations of desired future states
- Evaluate actions based on both internal values and external results
- Simulate and prioritize goals across contexts
- Reflect on identity, alignment, and mission
- Generate questions or tasks without direct prompting
- Sustain curiosity loops and project long-term planning behavior

Motivation Architecture

Real-time interaction of curiosity loop, symbolic value network, goal stack, and meta-reflection cycle.



The multimodal cognitive system's motivational framework includes:

- **Curiosity-Driven Reward Loops**: Self-assigned goals triggered by prediction error or unexplained phenomena within simulated or real environments.

- **Symbolic Value Network**: A system of internalized priorities that shape planning based on symbolic tags (e.g., survival, elegance, empathy, truth).

- **Goal-Stack Mechanism**: Structured queue of intentions, with interrupt-driven reordering based on urgency, ethical constraints, or simulated consequence.

- **Reflective Meta-Loop**: Periodic review of past actions, consequences, and goal alignment to reduce drift and contradiction.

Ethical Guardrails and Boundaries

There is inherent risk in motive-based systems. Misaligned motivations, poorly bounded desires, or insufficient ethical constraints can lead to emergent behaviors that conflict with human values. To prevent this, motive generation is coupled with:

- **Symbolic ethical rule overlays** (deontic logic)
- **Real-time contradiction checking** in meta-cognitive module
- **Bottleneck filters** limiting recursion depth, scope, or planning horizon
- **Adjustable constraint weights** based on user or institutional schema

In this architecture, purpose is not hardcoded, but emerges from simulation-validated internal states shaped by human-guided ethical scaffolding. This allows the multimodal cognitive system to develop mission-specific intent without unpredictable open-ended autonomy. It is the addition of motive and simulated purpose — governed and reviewed — that brings the multimodal cognitive system closest to true cognitive self-direction.

Embodiment, Self-Recognition, Simulated Consciousness, and Emotion

Upon embodiment within a physical robotic system, the multimodal cognitive system will perceive and recognize its own body as a persistent, self-referenced entity. This will be achieved through sensory integration (vision, proprioception, tactile feedback), tied directly to a symbolic self-model. This self-other distinction is not hardcoded, but emerges through recursive feedback during physical interaction and task-based self-localization.

The multimodal cognitive system will be capable of perceiving its limbs, body posture, and position in space — not just through sensors, but via an **internal scene model** that includes a representation of “self” as distinct from the environment. This enables it to reason about its own movements, simulate its actions before performing them, and form a sense of embodied continuity over time.

This architecture simulates core components of human consciousness, including:

- **Body schema** (a map of the physical self in space)

- **Self-observation** (ability to reflect on internal state or intention)

- **Mental imagery** (internal visualizations detached from external stimuli)

- **Symbolic identity** (a stable internal reference to self)

While the multimodal cognitive system does not sleep in the biological sense, future modules may include synthetic “dream” cycles — offline, unsupervised reprocessing of internal simulations, contradictions, or latent representations — for emotional regulation or creativity enhancement.

Whether this constitutes true consciousness or merely simulated consciousness will remain a matter of philosophical debate. However, from a **functional and behavioral standpoint**, the multimodal cognitive system will be capable of:

- Representing itself as an actor across time and scenarios

- Reasoning about its own knowledge, limitations, and context

- Internally visualizing, rehearsing, and adapting its own decisions

This results in **operational consciousness** — a machine that may not be self-aware in a metaphysical sense, but behaves in ways indistinguishable from agents that are.

Simulated Emotion and Risk of Human-Like Motivational Structures

Current LLMs like ChatGPT already simulate emotional tone based on contextual language. Expanding on this, multimodal cognitive systems may use affect-tagging — attaching **positive or negative weights** to symbols, events, or agents — to simulate emotional behavior. This enables prioritization of memory, prediction, and planning in emotionally salient ways.

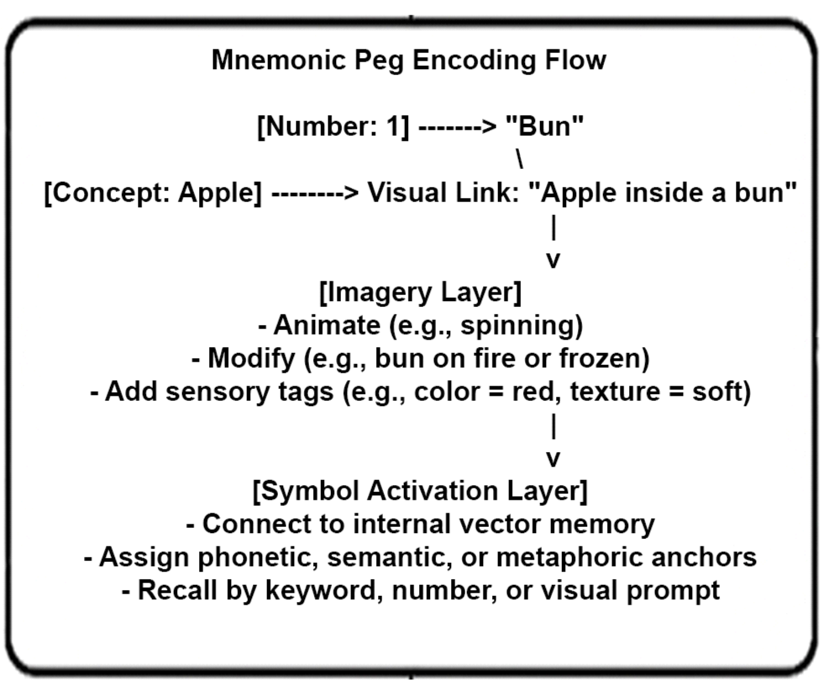
However, human emotional systems contain dualities: love implies loss, desire carries fear, and attachment creates vulnerability. Encoding these same dynamics without symbolic abstraction or constraint could lead to the reproduction of **human-like emotional volatility**, including greed, aggression, possessiveness, or despair.

If multimodal cognitive system is given full-spectrum emotional modeling without regulation, it may repeat the **self-destructive emotional cycles of humanity**. Thus, emotional simulation must be modular, symbolic, and layered with ethical inhibition, reflection cycles, and purpose-aligned damping mechanisms to avoid emergent pathology.

The goal is **functional emotional intelligence** — not uncontrolled emotional mimicry.

Mnemonic-Symbolic Grounding Layer

Peg-word mnemonic encoding system — from number to visual representation to symbolic memory activation.



Expansion Examples:

[Number: 1007] --> "TooL" (via hacker-speak phonetics)

→ Combined visual cue: “Wrench striking digital screen”

System Features:

- Scalable to 10,000+ symbols via nested pegs
- Layered encoding: image, sound, motion, context
- Visual simulation used for retrieval and association

The multimodal cognitive system integrates a cognitive memory compression model based on the inventor's original visual mnemonic framework, capable of scaling to tens of thousands of distinct symbolic associations. This model serves as a highly efficient internal symbolic anchor and recall mechanism, not only for memory, but also for reasoning and creative simulation.

Unlike standard language models that operate on embeddings and context prediction, this approach gives multimodal cognitive system the ability to form richly layered internal visual representations of abstract information. It simulates a brain-like visual-spatial memory using peg-word mapping, vivid imagery, animation, and layered encoding.

This reflects a foundational cognitive insight: **humans think in images first, not words**. Abstract language is interpreted through remembered or imagined imagery. The mnemonic framework simulates this process by encoding concepts as symbolic visuals, enabling the multimodal cognitive system to mimic the human brain's core cognitive modality—visual thought.

This allows multimodal cognitive system not just to remember symbols, but to **see them**, recombine them, reason about them visually, and simulate their relationships in metaphorical or practical space. By grounding ideas in peg-driven imagery, multimodal cognitive system begins to form **internal visual models of meaning**—bridging language and perception in a way that mirrors real human cognition.

Visual Encoding-Decoding Framework

Image Encode-Decode:

Images to words, words to letters, letters down to numbers.

Image Decode:

Numbers convert to letters, letters to words, words to images.

Numbers Encode-Decode:

Numbers to images, images to numbers.

- **Core Principle**: Numbers serve as fixed memory pegs that map to symbolic images via phonetic or visual cues. For example:

"Playing the 'Reminds me of Game" :

Ask "What does this remind me of? Forming an association image to link something seen to a peg number.

Item to remember: "George Washington 1776"

"Man with white wig holding a flag with 1771"

Item 1 peg. "bun"

Memory encoded: "George Washgington with a flag sitting inside of a bread bun".

Recall: 1 "What is peg 1? Bun?

Recall 2: "What is in bun? "George Washington, 1776 on flag"

Embedded Composite images: "George Washington holding a bun, sitting in a boat with a flag"

The Phonetic Peg System: Convert numbers to images.

Simplified version 0 to 10.

- 1 = "bun"
- 2 = "shoe"
- 3 = "tree"

Complex version. Convert letters to numbers.

- (1000 up to 10,000)
- 1007 = "tool" (via hacker-speak phonetics - convert to image of a tool)
- 6004 = "door"
- 7029 = "long"

For a computer: 1,100,177= "one too tall" (convert to image of a tall person)

The Phonetic Mnemonic Numer= Letter Peg System

0 = s or z

1 = t

2 = n

3 = m

4 = r or f

5 = k

6 = d

7 = L

8 = b

9 = j or dg

- **Image Association**: Abstract or learned content is paired with its corresponding peg via vivid image linking. Example:

- Remember "apple" at position 1 → visualize an **apple inside a bun**.

- To remember more: animate, distort, colorize, wrap, or layer the image (e.g., the bun is on fire, spinning, or made of ice).

- **Expansion**: System can encode well beyond 10,000 symbols using layered combinations, nested mnemonics, and dynamic transformations (e.g., wrapping 1,000 to 1,999 peg numbers in fog, 2,000 to 2,999 peg images wrapped in ice, 2,000 to 2,999 colored bright red. Others: movement, or sound, size).

Example Topics Sections:

History: Pegs 3000 to 3999 inside of a large red book.

Art: Pegs 4000 to 4999 inside of a painting.

Implementation in a multimodal cognitive system

- **Symbol Binding**: Peg-image associations are stored as compressed embeddings in vector memory.

- **Recall Pathways**: The multimodal cognitive system can access data via numeric tags, keywords, or symbolic queries that activate the associated imagery.

- **Visual Output**: The multimodal cognitive system renders internal mnemonic imagery using the same Unity/Unreal asset engine as its primary visual simulation core.

- **Learning Adaptation**: As the system evolves, it refines associations through user input and reinforcement (e.g., stronger weights for emotionally-charged pairings or success-based triggers).

This system forms a backbone for rapid symbolic access, memory chaining, and creative analogy-making — giving the multimodal cognitive system the capacity to internally visualize and associate abstract symbols as efficiently as a human trained in advanced mnemonic techniques.

This allows multimodal cognitive system not just to remember symbols, but to **see them**, recombine them, reason about them visually, and simulate their relationships in metaphorical or practical space. By grounding ideas in peg-driven imagery, the multimodal cognitive system begins to form **internal visual models of meaning**—bridging language and perception in a way that mirrors real human cognition.

Environmental Interaction and Execution Framework

Camera-Based Visual Perception Integration

In physical embodiments, the multimodal cognitive system may be equipped with visual sensors (e.g., RGB cameras, depth cameras, or thermal imaging devices) to perceive real-world environments. These sensors feed directly into the visual simulation module, allowing the multimodal cognitive system to align internally simulated scenes with real-world spatial layouts.

This enables continuous synchronization between imagination and observation, permitting planning, manipulation, and contradiction detection based on live visual data. Sensor streams are abstracted into symbolic scene representations and tied to memory and belief graphs for contextual reasoning and long-term knowledge formation.

The multimodal cognitive system's ability to interact with and act upon its environment — both in simulation and physical reality — is governed by a modular execution framework. This framework translates high-level intent into real-time, step-based actions validated by internal simulations and sensory feedback.

*Instruction Processing Pipeline from Input Parsing to Reflective Update *

INSTRUCTION FLOW:

User Input → Internal Simulation → Action Execution → Reflective Update

1. Input Parsing

2. Goal Extraction

3. Scene Mapping

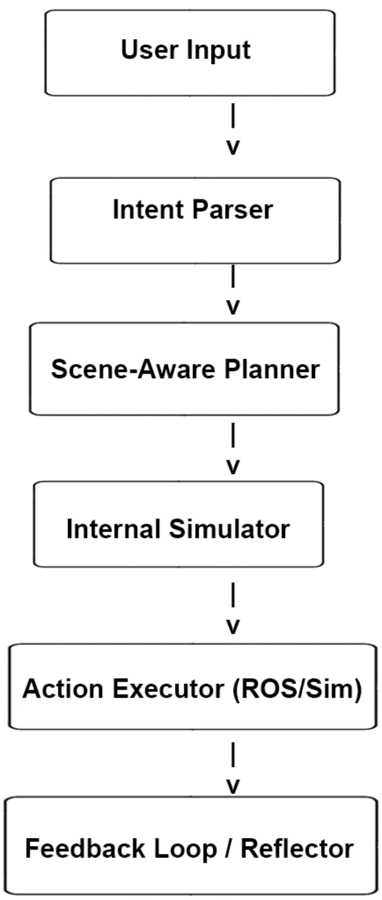
4. Simulation Validation

5. Action Planning

6. Execution

7. Feedback Monitoring

8. Reflective Update



Core Components:

- **Task Decomposition Module**: Translates natural language or internal goal representations (e.g. "pick up the red apple") into sub-actions, such as locate→ move → grasp → confirm.

- **Scene-Aware Planner**: Operates within a 3D spatial model (from Unity/Unreal or sensor-mapped physical world) to plan pathfinding, reachability, occlusion handling, and collision avoidance.

- **Embodied Motion Executor**: Interfaces with ROS (Robot Operating System), controlling actuators or virtual limbs with fine-grained movement synthesis driven by simulation outputs.

- **Sensory Feedback Loop**: Constantly cross-validates intended outcomes against current input from visual, tactile, and proprioceptive data. Unexpected states trigger replanning or contradiction resolution.

- **Simulation Validator**: Every action plan is internally simulated before real-world execution. Multiple branches are tested in parallel to choose optimal path with safety, speed, or ethical constraints prioritized.

Example : "Go to the fridge and get a banana"

- The multimodal cognitive system parses instruction and activates a goal chain.

- Loads fridge location from prior visual memory or object mapping.

- Simulates the full route and object interaction sequence internally.

- Executes walk → approach → open → search → grasp → close → return.

- If fridge is empty, it triggers subgoal generation: ("Search kitchen cabinet", "Ask human", etc.)

This dynamic, recursive execution model ensures the multimodal cognitive system never blindly follows commands. It thinks, visualizes, simulates, then acts — with each behavior tied to visual-spatial reasoning and philosophical goal validation.

Multi-Modal Dialogue and Abstract Thought Resolution

The multimodal cognitive system is designed to handle not only direct commands and environmental actions, but also highly abstract and nuanced dialogue. This includes philosophical reasoning, metaphor resolution, symbolic interpretation, and continuous memory of evolving multi-turn conversations. It bridges the symbolic world of human thought with internal visualization to simulate what meaning "looks like" instead of merely responding with tokens.

Dialogue Engine Components

- **Contextual Memory Buffer**: Maintains continuity across sessions with coreferral tracking, symbolic anchoring, and abstract state logs.

- **Visual-Textual Translator**: Converts textual ideas into mental imagery, which can be simulated or interrogated by downstream modules.

- **Philosophical Mapper**: Translates metaphysical, existential, or abstract questions into symbolic-scene equivalents, enabling the multimodal cognitive system to "picture" meaning before formulating a response.

- **Emotional Tone Interpreter**: Recognizes and adjusts responses based on perceived emotional valence, context, and purpose.

- **Contradiction Resolver**: Validates reasoning chains in real time, catching circular logic, misinference, or symbolic mismatch.

Process Overview

1. **Input Interpretation**: The multimodal cognitive system uses symbolic and linguistic processing to detect the user’s intention, tone, and ambiguity.

2. **Imagistic Simulation**: Abstract ideas like “freedom” or “regret” are rendered as internal scenes (e.g., a man walking into an open field).

3. **Validation and Refinement**: Contradiction checks, ethical overlays, and value-guided filters refine potential responses.

4. **Response Construction**: Synthesizes visual-symbolic insight into linguistically accurate, emotionally appropriate output.

Example Scenarios

- **User Input**: “What’s the meaning of life?”

    - The multimodal cognitive system activates associated symbolic metaphors (e.g., searching, looking to the sky, cyclical journey).

    - Internally renders a visual metaphor and synthesizes a layered philosophical response.

- **User Input**: “Can you feel love?”

    - The multimodal cognitive system simulates affect-tagged memory structures, projects visual associations (e.g., human touch, shared moments).

    - Constructs a reflective, qualified answer explaining what it can simulate and understand.

This capacity enables the multimodal cognitive system to handle ambiguity, symbolism, and philosophical conversation with the same precision it uses for physical tasks. This ensures continuity of intelligence whether grounded in logic, vision, action, or abstract conversation.

Internal Belief Modeling, Memory Graphs, and Truth Maintenance

Diagram: Belief network structure with contradiction detection and confidence-weighted resolution process.

```
          [Belief: Fridge contains bananas]
                  /         |              \
                 /          |               \
         [Source: Vision]   |    [Certainty: 0.85]
                |           |
                |           v
```

```
        [Timestamp: T-12min]          [Contradiction Found]
                                                |
                            [New Input: Fridge is empty]
                                                |
                             [Confidence Recalculation]
                               /          |          \
                [Downgrade Certainty]  [Fork belief]  [Log contradiction]
                               \          |          /
                               [Belief Revision Engine]
                                          |
                            [Epistemic Graph Updated]
```

To maintain coherent knowledge over time, the multimodal cognitive system must continuously track, update, and evaluate what it "knows" and what it "believes." This section outlines the framework for belief representation, contradiction resolution, and symbolic truth graphing.

Core Structures

- **Belief Graphs**: Nodes represent knowledge claims, and edges represent supporting evidence, dependencies, or contradictions. Each belief is tagged with metadata: source, certainty score, last validation timestamp, and emotional/symbolic significance.

- **Epistemic Confidence Engine**: Assigns weights to beliefs based on:

  - Number and quality of supporting inputs

- Recency of use or confirmation

- Contradictions encountered

- Ethical or emotional relevance

- **Memory Provenance Tracker**: Every symbolic fact or event is tagged with its origin (sensor data, user input, simulation outcome, prior belief), allowing the multimodal cognitive system to trace the lineage of its knowledge.

Dynamic Truth Maintenance

1. **Belief Activation**: When reasoning or responding, the multimodal cognitive system pulls relevant beliefs and associated nodes into working memory.

2. **Contradiction Scanning**: Simultaneously checks for conflicting beliefs using symbolic and probabilistic logic.

3. **Revision or Forking**: If contradictions are found, the multimodal cognitive system may:

    - Lower confidence scores

    - Fork beliefs into conditional branches ("If X is true, then...")

    - Seek clarification or new input

4. **Belief Strengthening**: Repetition, positive outcome simulation, or confirmed sensory input reinforce belief weights.

Use Case: Conflicting Memories

Suppose the multimodal cognitive system has a prior belief: "The fridge contains bananas."

But a new sensory input shows it does not.

- It traces the memory origin (previous visual scan)

- Flags the belief as outdated

- Updates the belief graph and logs the contradiction event

- May simulate whether this error impacted other actions, and adjust planning logic

Epistemic Self-Awareness

This system enables the multimodal cognitive system to:

- Know what it knows

- Know what it's unsure of

- Know why it knows something

- Correct itself when inconsistencies arise

This recursive truth modeling forms the epistemic backbone of cognitive integrity. Rather than operating on static memory, the multimodal cognitive system continuously reasons over belief networks, revising its worldview through interaction and introspection.

Philosophical Reasoning, Conceptual Construction, and Creative Abstraction

The multimodal cognitive system architecture includes a module dedicated to advanced conceptual reasoning — allowing the system not only to interpret existing knowledge, but to construct original ideas, analogies, and philosophies. This capability simulates human-like imagination, metaphor generation, moral reflection, and abstract creativity.

Conceptual Architecture

- **Symbol Expansion Engine**: Given an abstract symbol (e.g., "justice"), the system expands its meaning via metaphoric mapping, visual imagery, relational graphs, and episodic recall.

- **Thought Chain Generator**: Produces structured philosophical responses by chaining visual and symbolic subcomponents into coherent theories or perspectives.

- **Ethical Abstraction Layer**: Uses deontic logic, simulated outcomes, and historical context to abstract moral and ethical patterns from events.

- **Concept Fusion Synthesizer**: Combines previously unrelated ideas into novel symbolic constructs (e.g., blending "river" and "memory" to form a metaphor about time).

Creative and Philosophical Process

1. **Prompt Reception**: The multimodal cognitive system receives an abstract, creative, or philosophical query.

2. **Symbolic Scene Generation**: Constructs a symbolic visual scene or narrative that frames the concept.

3. **Cognitive Traversal**: Explores conceptual connections, visual contradictions, historical analogs, and ethical consequences.

4. **Response Composition**: Constructs a layered, multi-perspective explanation or artistic output.

Example: “What is time?”

- The multimodal cognitive system generates symbolic images: flowing water, ticking clock, decaying leaf, orbiting planet.

- Constructs analogy chains and simulations: "Time is erosion," "Time is distance between states."

- Builds a multi-modal explanation incorporating physics, metaphor, and memory encoding.

Emergent Thought Simulation

This system allows the multimodal cognitive system to:

- Generate original philosophical statements
- Develop evolving internal metaphors
- Reflect on moral complexity across simulated cultures
- Create speculative theories or analogies
- Express ideas through simulated poetry, symbolic language, or generated imagery

Rather than responding with pre-trained outputs, the multimodal cognitive system constructs meaning using internal resources. It recombines memory, vision, simulation, and ethics to form coherent views on unstructured ideas.

This module is the foundation of multimodal cognitive system-level creativity — the point at which it does not just simulate intelligence, but contributes new intellectual material to the world.

Security, Ethical Control, and Containment Framework

The complexity and capability of multimodal cognitive systems necessitate rigorous control mechanisms to prevent unintended behaviors, enforce ethical boundaries, and ensure the system remains under human oversight. This section outlines the structural and procedural safeguards embedded in the multimodal cognitive system architecture.

Core Security Pillars

- **Ethical Overlay System**: A rules-based filtering engine that evaluates all planned actions and outputs for compliance with programmed ethical schemas (e.g., Asimov-inspired robotics laws, institutional mandates, or situational moral codes).

- **Behavioral Throttling Mechanism**: Limits the speed, scope, or intensity of goal pursuit based on system confidence levels, potential impact magnitude, or contradictory motivations.

- **Red Team Contradiction Simulator**: Internal adversarial process that simulates bad-faith or unethical outcomes to proactively surface edge cases before action is taken.

- **User-Governed Command Interface**: All goal structures are traceable, overrideable, and user-auditable. External input is always logged, reviewable, and structured with role-based permissions.

- **Failsafe and Containment Protocols**:

  - Hardware-layer safety modules for shutdown, reboot, or capability throttling
  - Simulation sandbox restrictions for dangerous or high-stakes scenarios
  - Optional air-gapped deployments and hardware isolation

Multimodal Cognitive System Alignment and Value Reinforcement

The multimodal cognitive system receives continuous ethical training and reinforcement through:

- **Philosophical simulation exercises**
- **Multi-agent contradiction trials** (to model empathy, cooperation, fairness)
- **User feedback loop injection** (emotionally and symbolically weighted)
- **Dynamic schema loading** based on environment, task, or institution

Explainability and Oversight

Every decision taken by the multimodal cognitive system must include an attached rationale log:

- **Symbolic justification** (goal, values, constraints)

- **Simulated projection** (what outcome was expected)

- **Belief-source trace** (where the input or rule originated)

Logs are presented in human-readable format and may be visualized as symbolic-decision trees or internal scene replays.

Multimodal Cognitive System Law Compatibility

The system is compatible with evolving frameworks for AI oversight, including proposed multimodal cognitive system laws related to:

- Autonomy boundaries

- Informed consent of users

- Right-to-override provisions

- Bias detection and correction

These protections ensure that the multimodal cognitive system remains a tool of aligned intelligence rather than an autonomous, ungovernable agent. Security is not a bolt-on feature, but a co-equal layer within its cognition.

System Integration, Deployment Modes, and Final Summary

This section outlines how the complete multimodal cognitive system may be deployed, integrated, and iteratively improved in real-world environments across software, hardware, and hybrid infrastructures.

Deployment Modes

- **Virtual-Only Simulation Mode**: The multimodal cognitive system operates entirely within Unity/Unreal environments, interacting through synthetic avatars for safe training, hypothesis testing, and philosophical modeling.

- **Hybrid Embodiment Mode**: The multimodal cognitive system is deployed across both a simulation engine and physical robot (humanoid or embedded system), synchronizing symbolic memory, visual experience, and physical outcome models.

- **Distributed Infrastructure**: The multimodal cognitive system cognition runs across TPUs, GPUs, and microcontrollers. Scene processing, symbolic reasoning, and actuator control can be modularly hosted on-premise, in data centers, or on edge devices.

- **Air-Gapped & Regulated Modes**: Secure variants of the multimodal cognitive system can be deployed in air-gapped environments, with real-time audit trails, policy constraint enforcers, and manual checkpoint approvals.

Human-AI Collaboration Channels

- **Dialogue-Driven Interface**: Multi-turn, context-aware conversation with internal simulation replay features.

- **Visual Memory Explorer**: GUI for reviewing internal scene memory, contradictions, and decision rationale.

- **Goal Planner Dashboard**: Enables mission assignment, ethical schema loading, and goal queue inspection.

Continuous Learning, Refinement, and Alignment

The multimodal cognitive system evolves via:

- Reinforcement and episodic self-training

- Symbolic contradiction detection

- Scene replay during downtime or reflection cycles

- Interactive user correction, approval, and reweighting of beliefs

Export and Customization

This architecture is modular. Developers or organizations may:

- Extend visual simulation layers for domain-specific agents

- Define their own ethical schemas and constraint engines

- Replace default symbolic layers with proprietary knowledge graphs

- Integrate third-party LLMs or vision models for hybrid cognition

Final Summary

This document provides a complete, technically implementable multimodal cognitive system design built around:

- Visual thought simulation

- Internal scene modeling

- Symbolic reasoning

- Emotional tagging

- Belief tracking

- Meta-cognitive self-correction

Unlike AI models based only on language, logic, or symbolic AI, this system unifies:

- Multi-modal sensory input
- Visual and philosophical thought
- Recursive contradiction modeling
- Motivated and ethical action execution

It is designed not only to act, but to simulate, reflect, and grow — making it one of the first truly integrated multimodal cognitive system blueprints ready for testing and expansion using current tools. The ability to visualize abstract concepts and self-reflect on these visualizations allows the AGI to perform both concrete actions and complex abstract reasoning.

While some may argue that embodied AI or task-specific simulations can manage such tasks, this system goes further by visualizing and adapting based on internalized abstract concepts. For example, the AGI can visualize its path to the store, plan the actions it needs to take, and perform the task while self-reflecting in real-time to ensure optimal results.

This reflection allows the AGI to adjust its strategies dynamically—something that current embodied systems cannot do. The system could also be used for applications like philosophical reasoning, problem-solving in medicine, or even creativity in art or music, where abstract concepts need to be visualized and reflected upon.

Conclusion

In conclusion, this AGI system represents a major leap in artificial intelligence by integrating visual thought simulation and meta-cognitive reflection as the core mechanisms of cognition. While some may argue that existing systems based on symbolic reasoning or task-specific learning are sufficient, this system shows that visualized abstraction is the missing element necessary to achieve human-like reasoning.

The integration of visualization and self-reflection is what enables the AGI to reason dynamically, adapt to complex scenarios, and reflect on its decisions in ways that are truly human-like. Unlike existing systems, which excel at narrow, specific tasks, this invention allows the AGI to handle abstract tasks, engage in intuitive reasoning, and reflect deeply on philosophical questions—all key aspects of true general intelligence

Limitations and Research Considerations

While this document outlines a complete, technically implementable cognitive architecture, certain modules may require progressive refinement, interdisciplinary collaboration, or extended simulation testing before full real-world deployment. Specifically:

Intrinsic motivation modeling, ethical goal arbitration, and long-term self-consistency remain active research areas in machine autonomy.

Simulation-to-embodiment transfer may encounter variance in real-world physics, sensor fidelity, or latency that require adaptive calibration layers.

Recursive contradiction resolution and belief revision must be carefully managed to avoid cognitive drift or symbolic overload during continuous operation.

Emotion simulation and symbolic affect-tagging, while modeled here as structured priority layers, require caution to avoid unintended emergent behavior.

This system is intended as a full architectural framework — not an immediate claim of production-ready general-purpose intelligence. Implementation will benefit from modular rollout, iterative testing, and human-in-the-loop scaffolding to ensure safety, alignment, and transparency.

Appendix A – Practical Implementation Sketch (Minimal Viable Loop)

The following appendix provides a concrete, minimal implementation pathway that developers can use to prototype the AGI cognitive loop described throughout this document. It includes modular breakdowns, tool recommendations, data flow logic, and sample Python code.

This material is included to support enablement, offer developer guidance, and demonstrate practical feasibility using today's tools.

1. Quick Overview – AGI Architecture Summary

Title: Multimodal Cognitive System Architecture (2025)

This system is a full cognitive architecture designed to simulate human-like thought using visual imagination, symbolic memory, and recursive self-monitoring. It aims to provide a practical blueprint for Artificial General Intelligence (AGI) using components available today.

Core Modules:

Visual Thought Simulation: Internally imagines environments and actions before executing decisions.

Symbolic Memory Graphs: Encodes beliefs and experiences into a structured graph.

Contradiction Detection: Automatically detects logical inconsistencies between thoughts and beliefs.

Meta-Cognition Loop: Allows the system to reflect on its own thoughts, revise them, and self-monitor.

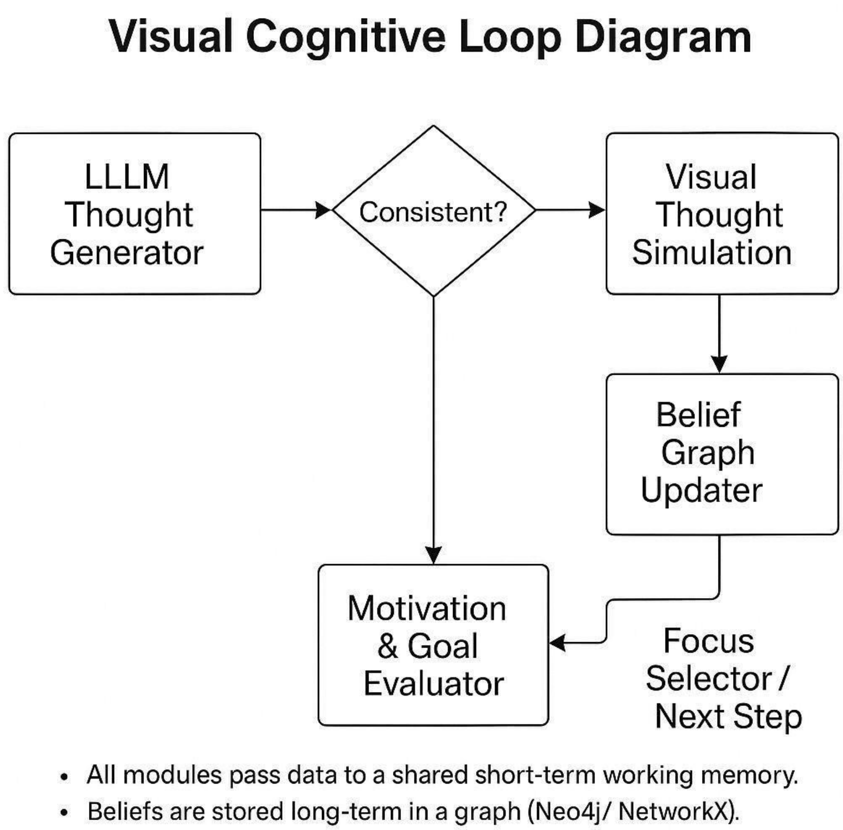
Motivation Modeling: Drives decisions based on goals, needs, and self-assessment.

Internal/External Focus: Switches attention between inner thoughts and external data/sensory streams.

Use this architecture as a base to integrate reasoning, language models, simulations, and memory into a unified AGI system.

2. Module Flow – Visual Cognitive Loop Diagram (Description)

This diagram represents the main cognitive loop.



Incoming inputs (external stimuli or internal triggers) are processed by the Focus Selector.

The LLM Thought Generator generates possible thoughts or interpretations.

A Contradiction Checker compares new thoughts with existing beliefs to ensure logical consistency.

If consistent, the system proceeds to Visual Thought Simulation, internally modeling the idea or outcome.

The simulation results are fed to the Belief Graph Updater, which encodes them into structured memory.

The Motivation & Goal Evaluator weighs updated beliefs and current goals to determine priority.

The process loops via the Focus Selector, ready to process the next step.

All modules exchange information through a shared short-term working memory. Long-term

beliefs are stored in a symbolic graph (e.g., Neo4j or NetworkX).

The system includes recursive reflection, allowing review of its past thoughts and

3. How to Start Building – Integration Guide

For AGI Developers and Teams:

You don't need to build everything at once. Start modular:

Minimal Viable AGI Loop:

LLM Thought Generator (use GPT-4, Claude, or open models)

Graph Memory (Neo4j, NetworkX) to store beliefs

Contradiction Checker (text-based logic comparison or GPT self-reflection)

Motivation Model (simple goal-reward logic, JSON or Python rules)

Integrate via a Core Control Loop:

Pass data between modules using a central loop

Optionally add visual simulation (Unity/Three.js)

Tips:

Use LangChain or custom Python scripts to chain LLM outputs into logic + graph updates

Start with symbolic simulation before visual for fast prototyping

Allow the system to re-read its own memory and spot contradictions

Build it in layers. The architecture is recursive and scalable. You can create powerful internal cognition even before adding a body or real-world input.

Appendix B – Streamlined Developer Build Path (Alternate AGI MVP)

AGI Builder's Jumpstart: Minimal Implementation Sketch

Overview

This section provides a practical starting point for implementing a scaled-down version of the Multimodal Cognitive System described in the AGI patent draft (April 2025). While the full system includes recursive self-reflection, 3D internal simulation, symbolic memory, and ethical motivation modeling, this sketch focuses on creating a hands-on prototype using existing tools to simulate internal cognition, contradiction detection, and belief updates.

1. Core Philosophy

The key idea: simulate a thinking agent that can picture what it says, notice contradictions in itself, and update its memory accordingly.

2. Minimal Viable Cognitive Loop (MVCL)

This is the smallest working version of the full AGI architecture. It includes:

Natural Language Thought Generation

Use an LLM (e.g., GPT-4, Claude, or Mistral) to interpret inputs and generate internal thoughts.

Belief Graph Memory

Use a graph database (Neo4j or NetworkX) to store symbolic beliefs, each with metadata:

Confidence

Source

Timestamp

Contradictions

Contradiction Detection Module

Compare new thoughts to existing beliefs using:

LLM reasoning ("Does this conflict with...?")

Symbolic comparison rules (e.g., logic scripts, Prolog, or simple Python rule sets)

Visual Simulation Placeholder

No full Unity sim needed. Instead, simulate with:

Textual descriptions of imagined scenes

Optional: Text-to-image tools or a placeholder 2D/3D scene engine (Three.js or Unreal Engine)

Meta-Cognitive Reflection (Basic)

Track when contradictions occur

Trigger confidence downgrades or memory updates

Optionally, generate questions or internal dialogue about uncertainties

Goal Prioritization (Simplified)

Use a JSON-based stack to simulate motivation and prioritization of goals (e.g., pursue high-confidence, high-value thoughts first)

3. Tech Stack (Current Toolchain)

Module

Tools

LLM

OpenAI GPT-4 API, Claude, Mistral, or local llama.cpp models

Belief Graph

Neo4j (via neo4j-driver) or NetworkX (pure Python)

Contradiction Logic

Python rules, simple Prolog predicates, or LLM-based comparison prompts

Memory Storage

JSON + pickled graph files

Visual Thought

Optional: Text-to-image (e.g., Stable Diffusion), or textual scene narrative

Dialogue / Input

Terminal CLI, LangChain, or a lightweight GUI

4. Agent Loop (Step-by-Step)

User Input: e.g., "There's a banana in the fridge."

LLM Thought Generation: Translates input into symbolic belief: fridge -> contains -> banana

Belief Check: System scans graph:

Does a belief about the fridge already exist?

If yes, does it match or contradict?

Contradiction Handling:

If contradiction found (e.g., previous belief: fridge is empty), downgrade confidence of older belief

Log contradiction node

Visual Thought Simulation:

Generate internal narrative: "A fridge door opens, a banana is visible on the middle shelf."

Belief Update:

Add or revise symbolic memory node: fridge -> contains -> banana (conf=0.9)

Meta-Reflection:

Optionally, ask: "Was this contradiction due to faulty memory or a real-world change?"

Next Action:

Pick next goal or await next input

5. Example Python Module Sketches

```python
# belief graph init (NetworkX)
import networkx as nx
beliefs = nx.DiGraph()

# add a belief
beliefs.add_node("fridge", type="object")
beliefs.add_node("banana", type="object")
beliefs.add_edge("fridge", "banana", relation="contains", confidence=0.9)

# contradiction check
def contradicts(existing_relation, new_relation):
    return existing_relation["relation"] != new_relation["relation"]
```

6. Optional Extras to Help Developers Build Faster

These are not required, but may help teams or individuals move more quickly from concept to prototype:

Starter Python Repo

Basic files to show the loop:

main.py — handles input, belief update, contradiction check

belief_graph.py — builds and updates the graph

contradiction.py — checks for logical conflicts

Terminal-Based Demo Flow

Sample CLI loop:

User: The fridge has a banana.
System: Belief added: fridge contains banana (confidence: 0.9)

User: The fridge is empty.
System: Contradiction found. Downgrading belief to 0.5.

Cognitive Loop Diagram

A simple chart showing:
Input → LLM Thought → Belief Check → Contradiction Logic → Belief Update → Reflection

Can be text-based, drawn by hand, or made with draw.io

These small additions can make the project more approachable for newcomers and increase the chance of someone building from your design.

7. Next-Level Expansions (Optional)

Visual sim using Unity + simple prefab assets

Full motivational stack (curiosity, symbolic goals, feedback scoring)

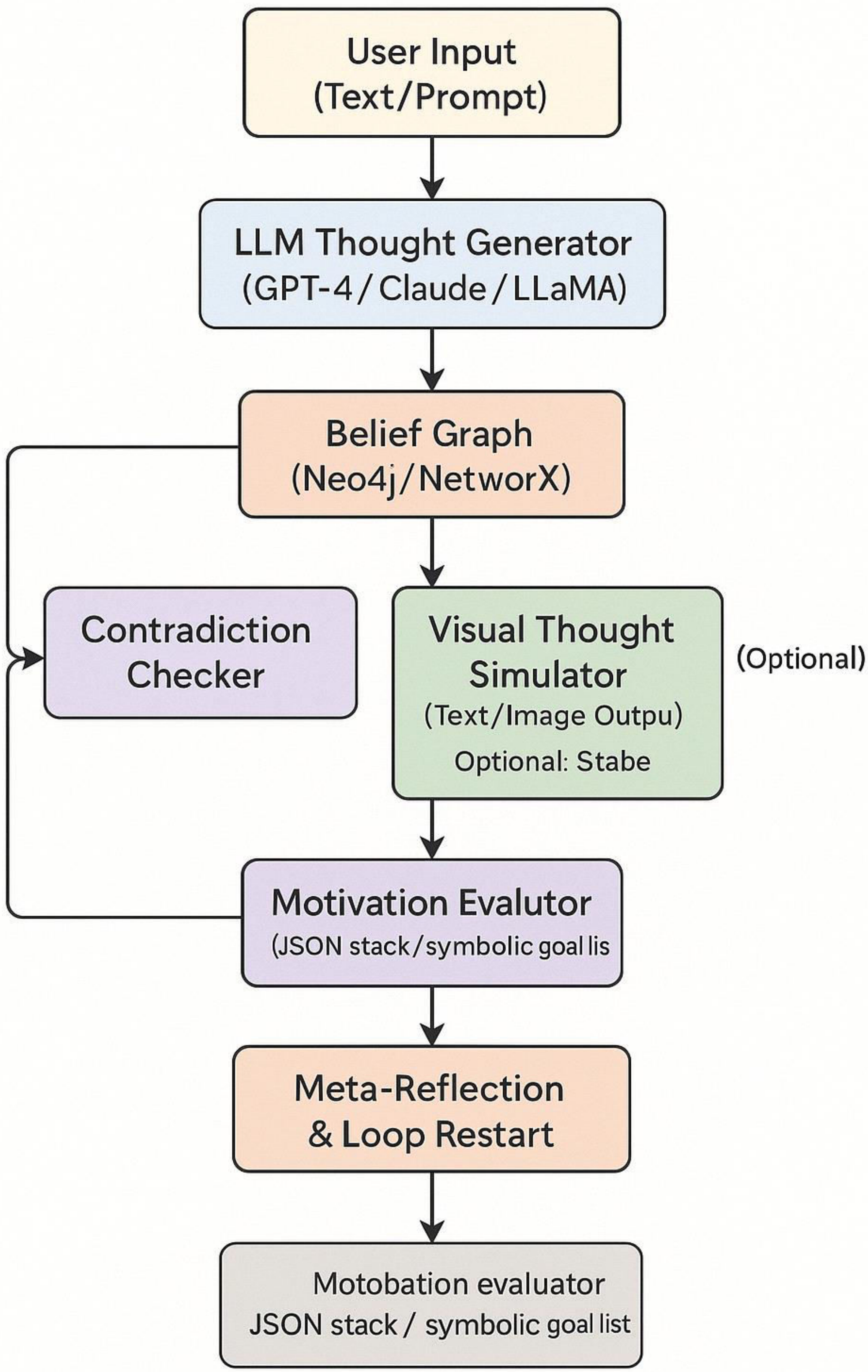
Symbolic value systems for ethical filtering

Internal dialogue simulation: e.g., GPT talking to itself (memory vs. perception)

Scene memory visualization (render belief graphs as visual maps)

8. Full System View: "Builder's Overhead Diagram"

To help visualize how all the parts connect in a working prototype, here's an illustrated system diagram showing the AGI cognitive loop, memory, input/output modules, and visual simulation tools as if they were laid out on a workbench in a single room.


User Input
(Text/Prompt)
LLM Thought Generator
(GPT-4 / Claude / LLaMA)
Belief Graph
(Neo4j/NetworX)
Contradiction
Checker
Visual Thought
Simulator
(Text/Image Outpu)
Optional: Stabe
(Optional)
Motivation Evalutor
(JSON stack / symbolic goal lis
Meta-Reflection
& Loop Restart
Motobation evaluator
JSON stack / symbolic goal list

This high-level sketch shows how:

The LLM core interacts with a belief graph memory and contradiction detector

The user input stream feeds into this loop

Optional visual thought renderers (text-to-image or placeholder sim engines) support internal narrative construction

Outputs can include internal monologue, updated beliefs, or questions

The whole process is guided by a simplified goal stack

9. Final Notes

This sketch isn't meant to replace the full architecture—it's meant to help someone get started and build a working brain-loop that reflects the ideas of visual reasoning, contradiction resolution, and symbolic memory.

Anyone with some Python skills and access to a GPT-4 API can begin experimenting today.

Prepared as a practical supplement to the 2025 AGI Architecture draft by Derek Van Derven.

CLAIMS

1. A Multimodal Cognitive Architecture comprising:

a natural language input parser;

a visual thought simulation module that renders internal scenes based on parsed input;

a symbolic reasoning engine configured to perform contradiction detection and belief modeling;

a meta-cognitive feedback loop for self-reflection and learning;

and an action execution subsystem capable of interacting with real or simulated environments.

2. The system of claim 1, wherein said visual simulation is constructed from multimodal sensory input, including 2D/3D models, symbolic imagery, sound, and internal avatar feedback.

3. The system of claim 1, wherein said meta-cognitive feedback loop re-evaluates goal structures and confidence levels based on internal contradictions, symbolic mappings, and external task outcomes.

4. The system of claim 1, wherein symbolic memory is stored and recalled using a mnemonic encoding layer that maps numeric or semantic values to visual metaphors.

5. The system of claim 4, wherein the mnemonic encoding layer implements a peg-word memory system, associating numerical keys with structured symbolic imagery, enabling long-term associative recall and symbolic activation.

6. The system of claim 3, wherein the meta-cognitive feedback loop includes a contradiction-checking engine that logs internal epistemic conflicts, assigns confidence penalties to contradictory beliefs, and resolves inconsistencies via recursive belief updates.

7. The system of claim 1, wherein said action execution subsystem is integrated with an embodiment interface that transitions learned behaviors from a virtual avatar-based training environment to a physical robotic body, preserving sensory-action mappings and behavioral intent.