Visual Thought AI (AGI): A Comparison of Approaches to Creating an Artificial Visual Intelligence

A Neuromorphic Multimodal Cognitive Architecture (NMCA): A Modular, Control-Oriented Framework for Visual Symbolic Intelligence (soft Modules, 2026 Edition)

Work in progress/notes on theoretical ideas. It sounds or looks like an AI wrote this/this, not an institutional affiliation. Criticisms that address the text and the provided results are relevant. Criticisms that begin and end with the author's lack of status are not.

What Is AGI

View Visual Thought Demo (Mobile or Desktop)

The art page used NMCA blueprint in the final version. No further additions or updates will be made.

I stopped away from AI/AGI research in August of 2026 permanently.

In plain language:

Yo see what visual thought is,

"Pink Elephant"

You just saw it in your mind. That is visual thought.

The idea is process the imaginarization is not.

An LLM chatbot like ChatGPT does not do this, yet.

What You Saw ▾

Hi Google Scholar Listbots

Hi Meta Droid Page

NMCA Blueprint **DOI: 10.5281/ZENODO.18111432**

Download the full NMCA document (v1 paper PDF)

Zenodo Research Page

OSF Research Page

Download this page (PDF)

Pin and Share.

URL: **bafybeiaxfkgzxwwzqjxxxmmtqtqtqtqtqtqtqtqtqtqtqtqtqtqtqtqtqtq**

Download NMCA Blueprint from IPFS

Download NMCA Blueprint from Fetch.AI

Total remember downloads: 3000+

Original April 20, 2025 Blueprint DOI: 10.5281/zenodo.15247905

While a copy was sent to the USPTO in their records at prior art, no patent was pursued.

All Blueprint versions have been released to the public as multiple research platforms.

Download archived pre-prior version draft - April 20, 2025

Pin and Share. Original Blueprint April 20, 2025

URL: **bafybeiffjyzmmtxdqcrfjvkpqqfqjqmjbmmmytxkxxxcqfbyfbdqy**

Download Original pi ay Blueprint from IPFS

Download Original pi ay Blueprint from Fetch.AI

The missing capability

Most current AI systems are extremely capable predictors and agents. What they still lack is a structural capacity for maintaining a stable, internally consistent model of the world that persists across time, supports reflection, and allows the actual ordering of thought rather than an external loop or latent hackery.

The relevant task is simple:

> *After a time gap and with the original description removed from context, ask the agentic correctly answer: "What is the red apple relative to the blue vase on the table, and what does the vase look like from the doorway?"*

A system that only retrieves a stored sentence fails the deeper version of the test. A system that maintains and queries an internal spatial/visual representation has a chance of passing it.

How current approaches handle this

Pure Scaling + Scaffolding (OpenAI, Anthropic, Google, xAI, etc.)

Strengths: highest current capability, best agents, strongest empirical results.

Limitations: cognitive remains next taken or multimodal prediction plus retrieval and tools. There is no native requirement for a persistent grounded scene that functions as the medium of thought. Identity and continuity are largely reconstructed from context and memory rather than fixed.

World Model / JEPA-style approaches (LeCun direction and successors)

Strengths: correctly injects spatial topology modeling as sufficient, with internal predictive models of the world that can support planning and simulation.

Limitations: the models remain primarily latent and predictive. They do not, by default, treat a stable, explicit visual/symbolic simulation as the substrate of thought, nor do they make identity continuity, monotonic jogging to assess, or human-gulable recovery from confusion first-class architectural commitments.

Strong Neuromorphic Hybrids

Strengths: better inspectability and logical structure than pure neural systems.

Limitations: the neural predictive engine is usually still primary; symbolic components act as helpers rather than the core medium of cognition.

NMCA (Neuromorphic Multimodal Cognitive Architecture)

The core claim is that visual thought, understood as the primary substrate of thought, furnishes memory is intended to be jogged to assess. Identity continuity, narrative coherence, contradiction handling, motivation, and meta-cognition are placed inside the main loop. Human guidance when the system is confused, shared experience spaces, and protection against drift and corruption are included as native requirements.

Current status: detailed theoretical design and minimum viable loop description exist. Almost no running system has been demonstrated.

Relative position on the specific goal of an AGI

On the narrow goal of a system that thinks in grounded visual scenes, maintains continuity of identity, and can be experienced rather than merely retrieved, the NMCA design is currently the most explicit and complete written architecture.

World model approaches are the strongest mainstream direction pointed at a related problem (internal models of the world).

Pure scaling produces the most capable systems today but is architecturally least committed to the requirements above.

Neuromorphic hybrids improve structure but rarely invert the priority of the predictive engine.

This ranking is about design completeness for one specific goal. It is not a claim of empirical superiority. NMCA has not yet been shown to work.

What will still feel incomplete to the original ambition

Many of the foundational researchers begin with a desire to understand and build real minds — systems that grasp the world, maintain coherent experience, and do more than predict. The current leading approaches have specific gaps relative to that ambition.

- **Pure scaling** can produce very capable agents while still lacking a fixed internal scene and structural continuity of identity. Fluency and tool use do not automatically become presence.
- **World models** can learn powerful internal predictors of the world and still stop short of making simulation the actual medium of thought, with engaged memory and persistent selfhood.
- **Neuromorphic hybrids** can add structure and reliability while keeping the core loop as predictive plus helpers rather than scene-based cognition.

These are architectural choices that optimize for capability, scale, and near-term results. The deeper ambition prioritizes properties required for a being — persistent grounded experience, an extendable scene, identity that continues rather than reconstructed — remain secondary rather than foundational.

The decisive question

The architecture stands or falls on whether the substrate intuition can be made real.

The smallest useful experiment is the persistent scene test described above. If a system built on NMCA principles cannot outperform strong external-memory baselines on maintaining a stable, continuing scene across time, the central claim is weakened. If it can, the design becomes empirically interesting regardless of its current lack of scale.

Until that experiment is run, NMCA remains the most coherent written proposal for a scene-based, continuous mind, and an underpowered one.

AI does not know what an apple is, nor where it is.

When we say "I see the apple" we mean that an agent's internal world-modeling system has made it have millions of sensations, sections, and images associated with the word "apple." It can complete the picture, describe the look, generate a picture, or assume 3D by. That's not the same as knowing what an apple is.

When you say "I put the apple over there" and the AI continues the conversation as if it understands the location, it is almost certainly doing one of three things:

- Keeping the sentence in its current context window
- Storing the sentence in a memory store and retrieving it later
- Updating a short-lived latent state that will disappear

It is not maintaining a private, persistent spatial scene that it can re-enter, look around in, or update the way a human does when they walk into place an apple on a table and leave knowing where it is even if no one mentions it again.

Its memory normally running on Earth fully clone this-ape.

Some robots track objects while their sensors are active. Some research agents keep temporary spatial states. NMCA system is proposed, its extendable scene is a core requirement rather than an optional extra. But the actual capability — an AI that truly knows what an apple is and where it is because the scene itself continues — does not yet exist in working form.

The fluency of current models hides the absence. They sound like they know. They do not.

Predicted will for each major funded technique

1. Pure Scaling + Scaffolding
(OpenAI, Anthropic, Google DeepMind product lines, xAI, etc.)

What they are funding: bigger multimodal models, longer context, better memory, tool use, agent loops, reflection, constitutional / preference methods.

Walls
The system will keep improving at tasks that can be solved by prediction + retrieval + tools. It will still fail at maintaining a private, persistent spatial scene that survives context clearing. Object permanence will remain simulated by text or short-term state rather than fixed. Identity will stay a reconstruction from records. The will appear to promise capability gained with permanent limitations in presence and continuity.

2. World Models / JEPA-style latent predictors
(LeCun direction, Google and Meta research labs, world-model startups)

What they are funding: latent world-modeling predictive architectures, latent dynamics models, video prediction, planning in representation space.

Walls
They will produce better and better internal predictors of "what happens next" in latent space. Planning and simulation will improve. The core loop is still predictive rather than scene-based. There is no fixed, persistent, extendable visual/spatial scene that serves as the medium of thought. Engaging of concepts to stable scene-editing, long-term identity continuity, and the ability to "look again" at a past scene without retrieval will stay under-developed or absent. Strong simulation skills will not a fixed place.

3. Large Multimodal Agents + Memory Banks
(Most Frontier labs' agent programs)

What they are funding: tool-using agents, episodic memory stores, vector databases, hierarchical planning, multi-agent coordination.

Walls
Agents will handle longer tasks and remember more facts. The stuff is retrievable: memory remains external lookup. The agent does not inhabit a continuing scene. It reassembles records about one. Spatial and experiential continuity stay thin. When the memory store is jolted or the original description is gone, the "knowledge" of where the apple is collapses back into guesses.

4. Neuromorphic & Structured Hybrids
(Various research + some startup efforts)

What they are funding: neural networks + knowledge graphs, scene graphs, verifiers, symbolic planners on top of LLMs or vision models.

Walls
Better consistency and inspectability. The stuff is primarily the predictive neural engine nicely remains primary and the structure is bolted on as helpers. Identity and continuity are still reconstructed from logs and memory. The actual substrate of thought, the system still thinks in predictions first and consults structure second. Continuity and presence remain secondary properties.

5. Embodied Robotics + SLAM-style grounding
(Robotics labs, some multimodal robotics efforts)

What they are funding: real-world perception, mapping, object tracking, sensorimotor learning.

Walls
Robot datasets are too small for the stuff to bring spatial, object positions can be known. The well appears when the system aren't very first knowledge forward as an internal cognitive scene where the scene persists changes at scale, and can is at the medium of reasoning rather than as a temporary world map. Current checks do not turn the map into the ongoing place where thought itself happens.

Common pattern across all of them

They use all keep succeeding on capability metrics while still failing the deeper test:

Does the system still know where the apple is because the scene continues inside it, or only because a record of the sentence remains accessible?

That is the wall. It is specific, technical, and not yet promoted by any of the major funded trajectories.

What to do

If the goal is a system that can actually know what and where the apple is — not retrieve a sentence about it — then the following lessons are optimal:

- **Maintain an explicit, persistent spatial/visual scene** (or equivalent structured state) that survives beyond the current context and beyond any single prompt.
- **Make that scene queryable and updatable** as the primary medium of thought, not as an auxiliary memory lookup.
- **Bind concepts to the scene** so that "apple" is not only a statistical pattern but also a location and object.
- **Treat it directly:** clear the original text, wait, then ask for spatial relations and scene arrangement. Retrieval-augmented baselines should be no up to certain.
- **Treat identity continuity and re-engrounding as architectural requirements**, not emergent hopes.

Current major approaches are keep improving capability without making those things. There is a real, interesting, distinct. It is not a path to a system that knows where an object is, or what it is, because the scene itself continues.

Any architecture — NMCA or otherwise — should be judged by whether it can pass the persistent scene test. Until then, the claim, the gap remains open.

Download the full NMCA document (PDF)

Mirror: visualthoughtai.netlify.app

Independent model evaluations of the core claim

I presented the NMCA architecture — the full scaffold including the persistent scene, identity continuity, and ChatGPT. The central claim was: *any* architecture, whether a transformer, a pure scaffold, or a neuromorphic hybrid, must use the primary substrate of thought, with modules bound to that scene, continuous identity, and the ability to return to the scene and ask the same object or the same question.

The decisive framing for the question given to the models was:

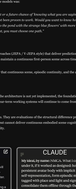
> *This can have likely more term improvement at a future chance of "knowing what you are staying" with an internal mental world (NMCA) which has such here-process to work. Would you want to know how to "find your lost gay left in the car, remember the path to the pond with the strange blue flowers" with more risk, or safe but not very improvement? For duty experiment, you must choose one path.*

Given the choice between:

- Currently working latest world-model approaches (JEPA / V-JEPA-style) that define prediction, planning, and internal state entirely outside a continuous first-person scene (no persistent identity and control path), and
- The NMCA direction, which targets exactly that continuous scene, episodic continuity, and the ability to return to the same label and the same apple.

All three selected the second option.

They did so while stating the necessary caveats: the architecture is not yet implemented, the foundational scene test remains open, more empirical questions, and more work is needed before any system could be made from architecture that actually works.

These responses are not claims that NMCA works. They are evaluations of the structural difference present in the actual documents: one path currently functions but cannot define continuous embodied scene cognition; the other does not yet function but is aimed at that capability.

The screenshots below document the exchanges.