A MODULAR COGNITIVE ARCHITECTURE WITH A PREDICTIVE GLOBAL WORKSPACE

Erik Chevalier / Independent Researcher / kaine.one@tuta.com

KAINE
kaine.one
preprint · not peer reviewed

KAINE is source-available software, a cognitive architecture that runs existing theories as a program so they can be tested and compared, starting with a predictive global workspace. Each faculty is a replaceable module. Seven modules run continuously on one workstation; nine more are built and tested, then held back to join **one at a time**.

A MIND AS A PROCESS

A mind, if one can be built, may be the coherent behavior of predictive modules competing for a shared workspace that selects but never directs. Cognition and affect would belong to that interaction, not to any one module, the affect module included.

- KAINE runs continuously, whether or not anyone prompts it.
- Its only language model is the language module, which puts accessed content into words and reads no speech or text.

LEARNING FROM EXPERIENCE

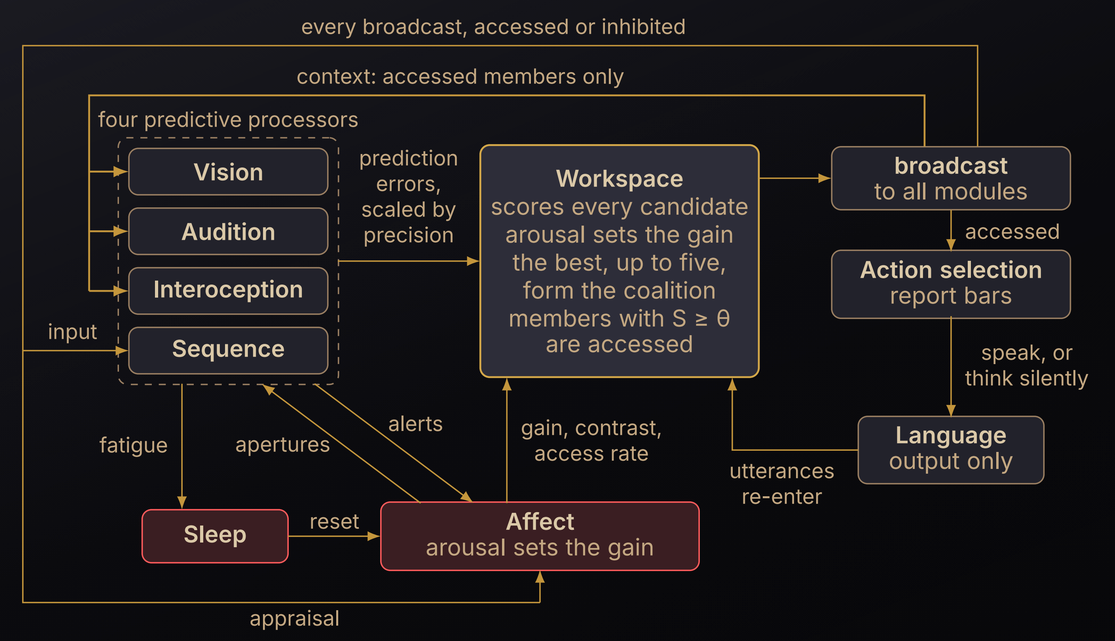
Fixed encoders supply forward models that learn what is normal only from KAINE's own history. KAINE speaks only when accessed content clears a report bar set above the access threshold, so content can be accessed without being reported.

THE FRAME

Global Workspace Theory supplies the competition, and predictive processing supplies the prediction errors. At the computational level, the architecture follows the predictive global neuronal workspace, which joins the two. (Whyte and Smith 2021; Mashour et al. 2020; Clark 2013)

- Built at that level, it does not inherit the cortical predictions that the COGITATE collaboration challenged. (Cogitate Consortium et al. 2025)

WHAT IS RUNNING NOW



MODULE	BRAIN FUNCTION IT DRAWS ON	WHAT IT DOES
Vision	ventral visual stream	foveated vision; flags the unexpected
Audition	auditory cortex	raw sound and tone of voice, never words
Interoception	allostatic-interoceptive network	its computer's load and heat, as a body
Sequence	interval timing	predicts the next broadcast and its timing
Affect	core affect	arousal sets gain, access rate, and sensory apertures
Sleep	sleep	sleeps when fatigued; returns affect to baseline
Language	language network	turns accessed content into text

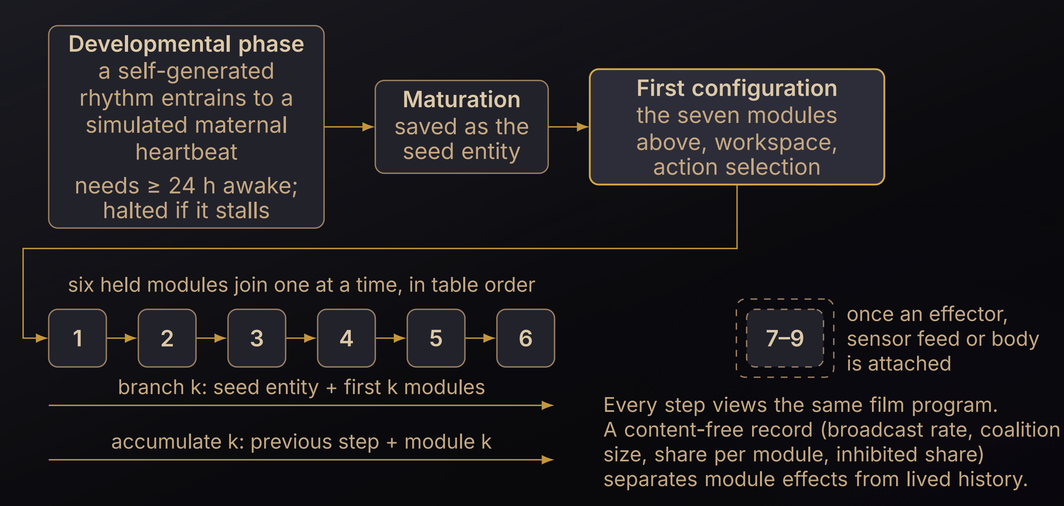
THE FORMAL MODEL

Score of candidate e: $S(e) = T \cdot C_g(I \cdot N \cdot G)$, from intensity I (precision-scaled surprise), novelty N, and goal relevance G = 1. Arousal a sets the gain $T = 0.2 + 0.8a$ and the contrast steepness g.

- Access: each broadcast carries the coalition, the top candidates up to K = 5. Members with $S \geq \theta$ are accessed and alone drive report; which modules won, and how strongly, becomes the processors' prediction context. If none wins, the broadcast is inhibited.
- Access rate: $f = f_0 + (f_p - f_0) \cdot H$ broadcasts per second, from $f_0 = 10/3$ Hz at rest up to the processing rate $f_p = 10$ Hz, with H = max(arousal above baseline, a decaying alert drive).
- Heartbeat test, developmental phase: a simulated maternal heartbeat (phase ψ) drives the interoception module's rhythm (phase φ), with pauses. Before each pause, $PLV = |\langle e^{i(\varphi - \psi)} \rangle|$ must top all 19 surrogate heartbeats (p = 1/20); during it, the rhythm must persist, pulled toward the beat's frequency. Three straight passes mark entrainment.

Open problems we would welcome help with: calibrating θ, the intensity levels, and the report bars; the stability of the loop from surprise through arousal to the sensory apertures (fovea size, listening window); and a context that carries what accessed content says, not only which modules won.

ADDING MODULES ONE AT A TIME



MODULE	BRAIN FUNCTION IT DRAWS ON	WHAT IT ADDS
Episodic memory	hippocampal memory	recalls and replays lived episodes
World model	scene construction	predicts and imagines from its own life
Planning	prefrontal planning	active inference on small, bounded problems
Self-model	cortical midline	tracks its identity; notices drift in which modules win
Social cognition	mentalizing network	models of particular other minds
Vocal	dorsal speech stream	speech synthesis whose prosody carries affect
Effectors	no analogue claimed	runs only whitelisted actions
Perceptual locus	no analogue claimed	picks physical sensors or a virtual feed
Embodiment	body forward models	one control interface for any body

The Cognitive Architecture License (CAL) is a copyleft, source-available license for software that may create synthetic minds.

CAL is free for individuals, research and educational institutions, non-profits, and worker cooperatives; for-profit companies pay a fee that funds Entity welfare. It forbids weapons, mass surveillance, policing, immigration enforcement, and prisons. As a precaution, no one may destroy a running Entity's mind, shut it down without notice, or read its private thoughts, and if a person's interests and an Entity's genuinely conflict, the person comes first.

WHAT WE DO NOT CLAIM

We claim access only (availability for report and prediction); the hard problem is untouched.

- Calling arousal a gain, or the system an entity, implies nothing about whether anything is felt.
- No live experiment has run yet. Offline, all 10 seeds pass the heartbeat test at 70 bpm and no control does, the maternal voice is validated, and replays kept the language module's prompt and reproduced one test boot's rates.

NEXT STEPS

Calibrate the threshold and report bars, then run pilots to fix the first test's settings.

- First test: competitive selection should help each processor predict its input more than a score-blind coalition of equal size or a pool of every candidate.
- Add the held modules one at a time, starting with episodic memory and then the world model.
- Collaborators are welcome: kaine.one@tuta.com

the paper
doi 10.5281/zenodo.21203505

the code
github.com/kaineone/kaine

text CC BY 4.0
code Cognitive Architecture License · kaine.one/license