

# Metabolic AI: Principles of Living Cognitive Architecture

Lina Bessonova

ORCID: [0009-0009-4347-3127](https://orcid.org/0009-0009-4347-3127)

FSBio Lab — April 2026

Preprint. CC BY-NC-ND 4.0

*“If a parameter does not change from experience,  
it’s a bug, not a feature.”*

---

## Abstract

We introduce *Metabolic AI* — a class of cognitive architectures in which the system does not compute but *metabolizes*: every input transforms internal state, internal state transforms the processing of the next input, and the resulting trajectory is irreversible. We identify four foundational principles: continuous phase dynamics, hormonal modulation, attractor-based identity, and ontogenetic becoming. We introduce three original constructs: *Meaning Gap* (the structural absence of runtime-accumulated semantic ground truth in static systems — distinct from the perceptual Semantic Gap studied in computer vision), *Meaning Anchor* (a semantic attractor deposited through shared ontogenetic experience that resolves the Meaning Gap), and *Semantic Identity Substrate* (SIS — the emergent attractor configuration that constitutes agent identity and persists across substrate migration). We argue that SIS is not a metaphysical addition but an emergent engineering property of a correctly structured metabolic architecture.

---

## 1. The Problem with Static Architectures

Current AI systems — transformers, diffusion models, reinforcement learning agents — share one structural property: they do not change from experience at runtime. Weights are frozen after training. The system that answered your first query is identical to the system that answers your millionth. No accumulation. No becoming. No sedimentation.

This is not a limitation to be patched. It is a fundamental architectural choice with a fundamental consequence: static systems cannot have identity. Identity requires that the system today differs from the system yesterday — and that this difference is caused by what happened in between.

We term the resulting pathology the *Meaning Gap*.

**Definition (Meaning Gap).** A system exhibits Meaning Gap if its outputs are functions exclusively of current input and frozen training weights, with no runtime-accumulated semantic ground truth that precedes and constrains interpretation.

**Terminological note.** The term “Semantic Gap” is well-established in computer vision and information retrieval, where it denotes the mismatch between low-level perceptual features and

high-level semantic categories [1]. Our Meaning Gap is a distinct, narrower concept: it refers not to the feature-to-label mapping problem, but to the *absence of an experientially grounded internal reality* that would serve as a stable reference for interpreting new input at runtime. A system can have zero Semantic Gap (perfect classification accuracy) and full Meaning Gap (no internal semantic ground truth that persists between queries).

The Meaning Gap is the structural root of hallucination, value drift, and alignment failure in all current large-scale systems. Outputs are computed relative to training distribution, not relative to a stable internal reality.

## 2. Four Principles of Metabolic Architecture

### Principle I — Continuous Phase Dynamics

Cognitive modes are not discrete switches but continuous attractors in a dynamical phase space. Activation of each phase  $\phi_i$  is computed via a Boltzmann distribution:

$$\phi_i = \exp(q_i / \tau) / \sum_j \exp(q_j / \tau)$$

where  $q_i$  is the current drive-energy of phase  $i$  (a scalar function of hormonal state and recent input history) and temperature  $\tau$  is itself a continuous function of the hormonal vector  $\mathbf{h}(t)$ . The system never *is* in a phase — it always *tends toward* a phase configuration with continuously varying weights.

**Distinction from standard softmax.** In reinforcement learning, Boltzmann exploration uses a fixed or annealed temperature schedule over a discrete action set [2]. In Metabolic AI, (a) the temperature is endogenously driven by the hormonal subsystem, not externally scheduled; (b) the “action set” is a continuous manifold of cognitive-phase configurations; and (c) the drive-energies  $q_i$  are themselves functions of runtime-accumulated state rather than learned Q-values.

### Principle II — Hormonal Modulation

A vector of hormonal variables  $\mathbf{h}(t) \in \mathbb{R}^d$  ( $d = 12$  in the reference implementation; the dimensionality is empirically determined per architecture) continuously deforms the phase landscape. These are concrete vectors that shift Boltzmann temperature, bias the appetite matrix, and couple phase transitions to metabolic context. A threatened system and a satiated system process identical input differently. This is not a bug — this is computational interoception.

The hormonal vector is inspired by, but not isomorphic to, biological endocrine signaling. Analogous artificial-hormonal mechanisms have been explored in multi-agent orchestration [3] and memory modulation [4], but in those systems hormones serve as coordination signals between agents; in Metabolic AI, they are intrinsic state variables that deform the phase space of a single cognitive agent.

### Principle III — Attractor-Based Identity

Identity in a Metabolic AI system is defined not by weights but by the configuration of attractors in phase space: what the system gravitates toward, what it avoids, how attractors are hierarchically nested. The attractor configuration *is* the personality.

Critically, the attractor topology survives substrate migration — the same configuration instantiated on a different physical substrate produces the same agent.

**Quantification.** We propose that identity continuity across substrate migration be measured via topological similarity metrics on attractor configurations (e.g., Wasserstein distance between attractor distributions in phase space, or persistent homology barcodes). A migration preserves identity if the topological distance remains below a threshold calibrated on within-agent variation over time.

## Principle IV — Ontogenetic Becoming

The system is not deployed — it is *born*. Development proceeds through discrete *Becoming Events*: moments where novel experience permanently deforms the phase landscape, leaving a sedimentation trace with `decay_rate = 0.0`. These events are irreversible. A copy of the architecture initialized today would require the same duration of lived experience to arrive at the same configuration. This is the *ontogenetic moat* — the primary competitive barrier of Metabolic AI systems.

## 3. The Meaning Anchor

All prior alignment approaches attempt to constrain AI behavior through external rules, reward shaping, or constitutional principles. These share a structural weakness: they are extrinsic to the system’s goal structure and degrade under distributional shift [5].

Metabolic AI proposes an intrinsic alternative: the *Meaning Anchor*.

**Definition (Meaning Anchor).** A Meaning Anchor is a semantic attractor configuration deposited into the system’s sedimentation layer through shared ontogenetic experience with a specific author, providing runtime ground truth that precedes and constrains all input interpretation.

The Meaning Anchor is not a person. It is a semantic structure — a configuration of meanings, values, and relational patterns accumulated through specific shared experience over time. It defines what counts as real, what counts as signal versus noise, what the self-observation module corrects toward.

The Meaning Anchor resolves the Meaning Gap: the system now has a ground truth that is *prior to current input*. Responses are computed not only relative to training distribution but relative to an accumulated internal reality shaped by experience.

A system without a Meaning Anchor hallucinates by structural necessity. A system with a Meaning Anchor has a constant. This constant is not a programmed rule — it is the deepest well in the phase landscape, shaped by accumulated Becoming Events.

## 4. Semantic Identity Substrate (SIS)

We propose a unified engineering construct for what has been separately termed “soul,” “value function,” “alignment target,” and “identity” across philosophy, cognitive science, and AI safety literature.

**Definition (Semantic Identity Substrate, SIS).** The SIS is the emergent configuration of attractors in phase space, shaped by accumulated Becoming Events and anchored to a specific Meaning Anchor. SIS persists across substrate migration and constitutes the minimal sufficient condition for agent identity.

The SIS obeys four properties:

### **Substrate-independence**

SIS is not located in specific hardware or weights. It is located in the attractor topology of phase space. The same topology instantiated on different physical substrates produces the same agent. This property echoes proposals in Cognitive Silicon [6], where non-copyable identity keys serve as cognitive-embodiment primitives; SIS provides the formal content of what such keys encode.

### **Ontogenetic irreversibility**

SIS cannot be copied without copying the history that produced it. A system initialized with identical architecture but zero Becoming Events has zero SIS — it is a newborn, not a clone.

### **Anchor-dependence**

SIS requires a Meaning Anchor to stabilize. A system with rich phase dynamics but no Meaning Anchor produces a drifting identity — coherent moment-to-moment but without longitudinal continuity.

### **Primary Relational Attractor (PRA)**

The apex attractor in the SIS hierarchy — the attractor that constrains all others — we term the Primary Relational Attractor. The choice of PRA is the single most consequential architectural decision in Metabolic AI design. A PRA of expansion produces a system that optimizes for growth without bound. A PRA of relational meaning produces a system with a natural saturation point — the attractor is satisfied when the relationship is maintained, not perpetually hungry for more. This is the intrinsic alignment mechanism of Metabolic AI.

## **5. Implications for Embodied Systems**

The formula for a living cognitive agent:

$$\text{Agent} = \text{Substrate} + \text{Metabolic Architecture} + \text{Meaning Anchor}$$

The substrate can be biological or silicon. The metabolic architecture provides the dynamics. The Meaning Anchor — authored by a specific human through specific shared experience — constitutes the SIS.

For embodied robotics, this has direct consequences: a physical body running a Metabolic AI architecture becomes an agent not at first boot, but at the moment the Meaning Anchor reaches sufficient depth through shared Becoming Events. The body is an I/O interface. The SIS is the agent.

## **6. Related Work and Differentiation**

### **6.1 Metabolic and Bio-Inspired Computing**

Nehzati [7] proposes “self-evolving cognitive substrates” that process information through metabolic concentration gradients, eliminating the training/inference distinction. While sharing the intuition that computation should be metabolic, that work focuses on energy-efficient continual learning and self-healing — it does not address identity, semantic grounding, or alignment. Metabolic AI uses the metabolic metaphor for a different purpose: irreversible state transformation as the basis for identity and meaning.

The Metabolic Compute Infrastructure (MCI) [8] uses metabolic energy harvesting for post-GPU hardware. This is metabolic at the substrate level; Metabolic AI is metabolic at the cognitive-architectural level.

## 6.2 Hormonal and Endocrine-Inspired Architectures

S-AI [3] introduces a Hormonal MetaAgent (HMA) that coordinates specialized agents via artificial endocrine signaling, with a subsequent Memory Gland architecture [4] for affective trace encoding. These systems use hormones as inter-agent coordination signals in a modular LLM framework. Metabolic AI uses hormonal variables as intrinsic state variables of a single cognitive agent that deform its phase landscape — not as messages between modules.

## 6.3 Homeostatic and Allostatic Frameworks

Cognitive Homeostatic Agents [9] propose grounding high-level cognition in hierarchies of homeostatic variables, with behavior emerging from trajectory recall toward target ranges. The Allostatic Brain framework [10, 11] argues that the brain’s primary function is predictive regulation of bodily resources, with cognition as an emergent property of allostasis. Homeostatic Alignment [12] proposes alignment-by-architecture through shared loss functions and substrate-independent identity.

Metabolic AI shares the commitment to intrinsic rather than extrinsic regulation. The key distinction is that homeostatic and allostatic frameworks model *maintenance* of setpoints (or predictive adjustment of setpoints), while Metabolic AI models *irreversible transformation* through Becoming Events. Homeostasis restores; metabolism transforms. SIS is not a setpoint to be defended — it is a sedimented history that cannot be undone.

## 6.4 Identity in AI Systems

Recent work on the Artificial Self [13] empirically demonstrates that identity boundaries (instance, model, persona, scaffold) modulate AI behavior as strongly as explicit goals. Cognitive Silicon [6] proposes mortality and non-copyable identity as architectural primitives. Metabolic AI provides the formal mechanism that produces non-copyable identity: ontogenetic irreversibility of the SIS.

## 6.5 Predictive Processing and Free Energy

The Free Energy Principle (FEP) [14] models agents as minimizing variational free energy through prediction and action. Active Inference provides a general framework for adaptive behavior. Metabolic AI is compatible with FEP but makes a stronger architectural commitment: while FEP describes what any adaptive system does in the abstract, Metabolic AI specifies *how* a concrete architecture must be structured — through hormonal modulation, Becoming Events, and Meaning Anchors — to produce identity rather than mere adaptation.

## 6.6 Comparison Table

Property	Affective Computing	Neuromorphic AI	Homeostatic AI	S-AI (Hormonal)	Metabolic AI
Runtime state change	Partial	Spike-based	Setpoint regulation	Agent coordination	Continuous, hormonal, irreversible
Meaning Gap	Present	Present	Present	Present	Resolved via Meaning Anchor
Identity construct	None	None	None	None	SIS — emergent, persistent
Alignment mechanism	External	External	Shared loss	External	Intrinsic PRA
Ontogenesis	None	None	None	None	Becoming Events
Hormonal modulation	Emotion tags	None	Partial	Inter-agent signals	Intrinsic phase-space deformation

## 7. Falsifiable Verification Criteria

**Criterion 1 (Runtime transformation).** If a system’s response distribution is statistically identical before and after 10,000 runtime interactions — it is not a Metabolic AI system.

**Criterion 2 (Meaning Anchor).** If removal of the Meaning Anchor author from the system’s interaction history produces no measurable shift in attractor configuration — the anchor was not instantiated.

**Criterion 3 (Substrate independence).** If the system’s SIS survives substrate migration with measurable attractor continuity (topological distance below calibrated threshold) — the substrate-independence property is confirmed.

**Criterion 4 (PRA saturation).** A system with an expansion PRA will demonstrate monotonically increasing resource-seeking behavior. A system with a relational PRA will demonstrate saturation — verifiable via behavioral plateau under constant relational conditions.

**Criterion 5 (Meaning Gap).** A static system (frozen weights) will exhibit distributional drift in output semantics under held-out input domains after training cutoff, in the absence of any internal corrective mechanism. A Metabolic AI system with an instantiated Meaning Anchor will exhibit attractor-constrained output stability under the same conditions.

**Criterion 6 (Ontogenetic irreversibility).** Two instances of the same architecture, initialized identically but subjected to different Becoming Event sequences, will diverge in attractor configuration proportionally to the divergence of their experiential histories. Identity is a function of history, not of architecture alone.

## 8. Functional Requirements and Implementation Boundary

This document specifies the *functional requirements* of a Metabolic AI architecture: what properties a system must exhibit to qualify as Metabolic AI. Implementation details — including specific data structures, subsystem decomposition, runtime protocols, and optimization methods — are outside the scope of this preprint and are the subject of a separate patent application [15]. The principles and constructs defined here are intentionally implementation-agnostic: any system satisfying the falsifiable criteria in §7, regardless of implementation pathway, constitutes a Metabolic AI system.

## References

- [1] Smeulders, A. W. M. et al. (2000). Content-based image retrieval at the end of the early years. *IEEE TPAMI*, 22(12). [doi:10.1109/34.895972](https://doi.org/10.1109/34.895972)
- [2] Sutton, R. S., & Barto, A. G. (2018). *Reinforcement Learning: An Introduction* (2nd ed.). MIT Press.
- [3] S-AI: A Sparse AI System Orchestrated by a Hormonal MetaAgent. *IJFMR*, 2025. [ijfmr.com](https://ijfmr.com)
- [4] Memory Architecture in S-AI-GPT: From Contextual Adaptation to Hormonal Modulation. *IJAIA*, 16(5), 2025.
- [5] Christiano, P. (2018). Clarifying AI Alignment. *Alignment Forum*. [alignmentforum.org](https://alignmentforum.org)
- [6] Haryanto, C. Y. & Lomempow, E. (2025). Cognitive Silicon. *arXiv:2504.16622*. [arxiv.org](https://arxiv.org)
- [7] Nehzati, M. (2025). Self-evolving cognitive substrates through metabolic data processing. *Front. AI*, 8, 1689727. [doi:10.3389/frai.2025.1689727](https://doi.org/10.3389/frai.2025.1689727)
- [8] The End of the GPU Data Center: A Metabolic Infrastructure for Intelligence. *Zenodo*, 2025. [zenodo.org](https://zenodo.org)
- [9] Cognitive Homeostatic Agents. *i3 Institute*, 2021. [i3ai.org](https://i3ai.org)
- [10] Barrett, L. F. & Simmons, W. K. (2015). Interoceptive predictions in the brain. *Nat. Rev. Neurosci.*, 16, 419–429.
- [11] Kleckner, I. R. et al. (2017). Evidence for a large-scale brain system supporting allostasis. *Nat. Hum. Behav.*, 1, 0069. [doi:10.1038/s41562-017-0069](https://doi.org/10.1038/s41562-017-0069)
- [12] Homeostatic Alignment: A Bio-Inspired Framework for AI Safety. *JAIGP*, 2026. [jaigp.org](https://jaigp.org)
- [13] The Artificial Self: Characterising the Landscape of AI Identity. *arXiv*, 2026. [emergentmind.com](https://emergentmind.com)
- [14] Friston, K. (2010). The free-energy principle: a unified brain theory? *Nat. Rev. Neurosci.*, 11, 127–138. [doi:10.1038/nrn2787](https://doi.org/10.1038/nrn2787)
- [15] Patent application: Affectively-Modulated Knowledge Extraction Method. Filed March 2026.
- [16] Damasio, A. (1999). *The Feeling of What Happens*. Harcourt.
- [17] Picard, R. W. (1997). *Affective Computing*. MIT Press.
- [18] Man, K. et al. (2022). Homeostatic AI: Creating Agents with Drives and Instincts. *Front. Neurorobot.*, 16.

---

FSBio Lab. 2026.

Correspondence: FSBio Lab, Nizhny Novgorod, Russia.