

Creature Algorithm

A Technical Brief on an Architecture for Physical Intelligence

Maurits S. J. Tepaske · Marius Lukas

Creature Algorithm Inc. · Mountain View, California

architect@creaturealgorithm.com

Prepared for technical review · June 2026

The dominant route to artificial intelligence is powerful wherever the world can be turned into representation: text, images, board states, structured data. It remains weak at one thing, and that thing is not peripheral. It is the part that matters for any system with hands and legs. The open problem is physical grounding: how a system can learn, act, and stay coherent through continuous contact with a world that pushes back, where failure is not a number that goes down but a body that has already changed before any plan can update. Creature Algorithm is a second route. It does not begin with a model of the world and add a body afterward. It begins with a body that must remain intact while it interacts, and lets sensing, memory, action, and eventually intelligence form as later specializations of that single demand. The whole architecture follows from one inversion of causal order: **regulation before representation, viability before objective, coherence before inference**. Everything below is an attempt to describe what that means in its own terms, not by contrast with the prevailing paradigm, which tends to quietly reframe the idea back into the very assumptions it sets aside, but constructively, from the primitive upward.

1. Start with the proof, not the theory

The claim is not a thought experiment. A vehicle-scale specimen learns from zero without reward-based optimization [[link](#)]: no model of the world, no dataset, no objective, no policy, and no reward function. An optional operator signal, described in §2, can be used to shape learning, but it is never accumulated or optimized, and the system works without it. Coordinated behavior accumulates through real-time interaction within minutes. When structural learning is switched off, the behavior persists; it lives in the network's connectivity, not in an ongoing optimization loop. The same matured network, moved into a different body with only sensor and actuator remapping, immediately reuses its coordination under new physical constraints, again without retraining.

These are instrumented, inspectable runtime properties, not demonstrations meant to be taken on faith. What that buys, precisely, is an existence proof: organized, transferable, persistent behavior produced by local regulation under physical constraint, with none of the machinery the field treats as the engine of learning. The contrast it isolates is between modeling the world and inhabiting it. A representation-based system can learn to ride a bicycle from thousands of videos, but the system itself is never on the bicycle and nothing happens to it if it falls; the consequence stays theoretical.

Here the consequence is not theoretical. The body is the thing that falls, and what it lives through is what

changes it. The inversion is real at small scale. It is not a conjecture about what might be possible. The honest open questions, and there are real ones, concern how far it extends, not whether it works.

One methodological note bears on how much weight the proof can carry. The scaling specimens are built on deliberately minimal hardware, avoiding rich sensing stacks and external scaffolding, so that any robustness observed has to originate in the control substrate rather than in expensive perception compensating for it. That is a constraint on the experiment, not only a cost decision: it removes the most common way a result like this can be confounded.

The proof is not a claim that the surface behavior is impossible to reproduce by other means. A model-based system can very likely be made to drive, balance, hesitate, and recover in ways that look the same from outside; the visible behavior is imitable, and we do not claim otherwise. The claim is about causal origin: that here the behavior arises from local consequence and structural adaptation rather than from a task, reward, policy, or scripted controller.

It is tempting to file this under one of the field's existing exceptions. Reinforcement learning without a hand-designed reward, imitation without explicit demonstrations, learned reward models, self-supervised or constraint-based objectives: the field has a way to drop almost any single one of these components. But each removes one piece of the optimization apparatus while keeping the rest. A method without an

engineered reward still optimizes a learned one; a method without demonstrations still ranks trajectories against an objective; a method without a central policy still descends a global gradient. What is absent here is not a component but the optimization apparatus itself: no model of the world, no dataset, no reward, no value function, no trajectory evaluation, and no objective optimized either centrally or locally. The system does not optimize toward a target at all. It regulates within a viable corridor, and structure changes because of the consequence it underwent, not because an outcome was scored and fed back to favor a better move. This is the difference between substituting one part of the apparatus and removing it as a whole. That difference is not a matter of appearance, which is why it is established by instrumentation rather than by demonstration: the architectural class is settled by what the runtime logs show to be absent, not by what the body is seen to do.

The demonstrated results so far are in simulation. That matters, and we state it plainly. But the result is not a conventional sim-trained policy waiting to be transferred into hardware. The experiment tests causal architecture: whether local viability regulation can produce persistent, transferable behavior under embodied constraint without reward, trajectory evaluation, value function, or central policy. Simulation limits the physical claim, not the architectural claim. This is also why the simulated specimens are not disembodied software. The architecture cannot be detached from a body and run as a free-floating program; it has no form except as the regulation of some body. Even in simulation it runs inside a body-like structure, a bounded interior with a boundary and a surrounding medium that can be perturbed and must be kept viable, because there is nothing for the rule to do except keep such a structure whole. The simulation does not host an algorithm that represents a body; it hosts a body-like structure that the algorithm is. Hardware changes the medium the body is made of, not the fact that there must be one. The next step is hardware, where the same causal order must survive real material noise, latency, friction, compliance, heat, wear, and failure.

This is why the internal-state observatory is central rather than decorative. It runs beside the specimen as a continuous read-out of viability state, structural traces, and bias shifts while the body interacts. Its purpose is to make the causal order inspectable from the inside, so that the absence of any accumulated performance signal, trajectory comparison, or central selector is observable in real time rather than asserted after the fact.

2. The primitive: a unit that lives with what it does

The basic unit is not a node that maps input to output. It is a bounded local unit with an inside, a boundary, and a surrounding medium, a small piece of matter that must keep its own internal condition within a workable range while the world perturbs it.

Its loop is not *input* $\rightarrow$ *representation* $\rightarrow$ *output*. It is *local condition* $\rightarrow$ *physical change* $\rightarrow$ *viability consequence* $\rightarrow$ *retained trace* $\rightarrow$ *changed future tendency*. The unit does not compute what to do. It undergoes what happens to it and is altered by the consequence.

“Viability” here is exact, and it is the word most often misheard. It is not a goal the system pursues, not a quantity it maximizes, not a survival drive in any motivational sense. It is a bounded corridor of viable imbalance. Perfect balance would be nearly inert; a living regime is one that holds many local burdens (strain, heat, energy, load, instability) jointly survivable while never settling into a resting point.

From outside this can look like equilibrium. From inside it is continuous counterbalancing with no final state to reach. From that corridor, the rest derives without further primitives.

A signal need not be introduced as a separate category of thing: it is a recurring physical consequence that comes to matter because of how it bears on viability. Excitation and inhibition need not be commanded: they fall out of the state-relations themselves. Imbalance can be characterized by two distinctions (whether a change moves toward or away from viability, and whether the relevant quantity is rising or falling) which yields four primitive regimes (a utilization regime, an acquisitive regime, an overload regime, a settling regime). This four-state grammar is the alphabet. What looks, later and larger, like sensing, action, internal regulation, or coordination is this same grammar specialized across scales and territories.

Because a signal is just such a consequence, those consequences arise first from what the body itself does — it acts, the world responds, and it is changed by what it undergoes. The operator can optionally become one more source of the same kind of consequence, reaching the body in more than one way. One is physical, no different from any other circumstance the body meets: easing a load it is straining under or steadying it as it begins to topple, which the body registers as a return toward viability; obstructing its path or pushing against it, which it registers as a departure. Another is a modulatory signal, which can be supplied directly, the way an animal is helped along faster with food. It acts less like an instruction than like a neuromodulator: it selects no structure and aims at no behavior, but raises the activity of the units it reaches and shifts the local

conditions under which they change. The structural change that follows is still the units' own, made locally under those shifted conditions. The purpose is practical: the body will find a viable way on its own, but operator influence reaches competence faster and makes the behavior the operator is after more likely to emerge, and a timely consequence away from viability can spare the body from learning a damaging lesson the hard way. Nothing here is a reward piped in; it is the operator acting as one more part of the body's physical world, and the body responds through the very same viability machinery it uses for any event. In intent these resemble reward and punishment, and we do not pretend otherwise. In mechanism they are not: each enters the same four-state grammar already described, and nothing accumulates it into a return, ranks actions by it, or optimizes toward it. What distinguishes a reward is not its sign but its use — accumulated, compared, maximized — and these are used in none of those ways. They are optional and can be removed at any moment; once removed, what has been learned remains, exactly as behavior persists when structural updates are switched off. A body whose competence survives the removal of the operator's influence was never optimizing it.

3. Structure forms by metabolizing destabilization

Here is the point an outsider is most likely to get backwards, because every instinct trained on control theory and optimization runs the other way. In the prevailing frame, stability is the target and perturbation is the disturbance to reject; noise is variance to average out; robustness means insensitivity to being pushed. In Creature, destabilization is the input the system runs on. Structural change is triggered by deviation from viability. A unit that is never perturbed leaves no trace, generates no consequence, and learns nothing. Noise is not filtered. It is the discrimination signal. In the earliest specimens, internal metabolic fluctuation and sensory noise produce the first self-driven micro-movement that bootstraps coupling before any coordination exists. Friction forces adjustment, latency forces compensation, contact forces negotiation. A model-based system tries to suppress these. This architecture is the medium they are made of.

So coherence is not the absence of perturbation. Coherence *is* metabolized perturbation, perturbation converted into retained structure that widens the range of conditions the body can stay whole within. Memory, in turn, is not a stored file or a table of past examples. It is retained consequence: changed thresholds, changed couplings, changed sensitivities, changed boundaries, paths that have become easier to re-enter. The body remembers because what it has lived through changes what it can become next. There is no policy being tuned and no dataset being looped over; the

connectivity is the behavior, and it changes by living, not by being optimized.

4. Why no center changes what “scaling” means

The coordination is a distributed tug-of-war with no referee. Neighboring units and regions occupy different viability states and pull against each other; what appears as goal-directed whole-body behavior is the macroscopic signature of these incompatible local pressures resolving, not the output of anything that ranks alternatives. There is no command center, no global state representation feeding a selector. Remove the body and no coherent structure persists; remove the rule and the body is inert; behavior exists only in their coupling.

The coupling is not purely antagonistic. As units adapt structurally, neighbors that were only pulling against each other can settle into configurations where each is steadier for the other's presence: the same local viability regulation that produces conflict also produces mutual benefit, where the geometry allows it. Nothing chooses this and nothing aims at it; the benefit is a consequence of each unit regulating itself, not a coordinated strategy, and to read it as cooperation in the intentional sense is to put back the referee the rest of this section removes. The structured signaling that would carry such mutual regulation at larger scales is predicted by the architecture rather than demonstrated, and is marked as such.

This is why the scaling question here is not the one the field is used to, and why the usual failure intuitions do not transfer. Embodied and self-organizing systems have historically broken at scale, but they break by losing a global coherence target, by a central coordinator becoming a bottleneck, by hidden global machinery creeping back in under pressure. Every one of those failure modes is bound to a center. This architecture has none to overload. The same local rule generates structure at every level (unit, joint, limb, body) so adding degrees of freedom does not load a coordinator; it adds more instances of the identical law, and with them more local capacity to absorb perturbation.

That law is deliberately simple — local arithmetic over a unit's own viability state, not a complex learned function — so capability comes from how many simple regulations are coupled, not from sophistication at any one point. Absorption is the substrate's defining act. The scaling variable is coherence, the range of simultaneously coupled conditions the body can keep bounded, not model size or dataset volume. Compute scales with interaction rather than with a global model, which is an economic property as much as a technical one.

The genuinely open question, stated in the architecture's own terms, is not whether influence cascades, nor whether locality and scaling trade off against each other. Those framings both assume influence is a thing transported through a channel, which is the very picture this architecture sets aside. Here there is no transported signal and no channel to contend over. There is only a unit resolving its own local condition, and that resolution has two faces at once: inwardly it absorbs perturbation into structure, and outwardly that same structural change is what neighbors register. Absorption is not the opposite of propagation; it is how propagation happens. Coordinating the body and keeping a disturbance local are therefore not competing uses of one resource: they are the same act seen from two scales, because the local rule that settles a unit is the only rule operating at any level.

So the real question is whether that local settling stays fast and complete enough as coupling density rises. Each unit must close its own regulation, reach a viable settling, before the accumulated influence of its neighbors drives it faster than it can resolve. While it closes, it is a sink: it takes in destabilization and returns structure. If it cannot close in time, it becomes a source: it passes on more disturbance than it absorbs, and the same coupling that coordinates the body now carries mismatch through it. Coherence at scale is therefore local closure sustained at scale: not a global target held by a coordinator, but the condition that units throughout the body keep resolving their own state under load rather than tipping from sinks into sources. Whether they do is a property of the specific medium physics: the rates at which its several influences diffuse and decay, and the bounds on how fast structure can change, and it is measured, not assumed. (That a unit failing to close becomes a propagating source of mismatch is the same failure logic the architecture predicts for pathology; at the scale of a whole body under stress it is derived, not yet demonstrated, and is marked as such.)

What makes that question observable rather than abstract is that it shows up as specific behaviors under stress, not as a metric. The signatures to watch as dimensional complexity rises are the ones that also define whether a humanoid is trustworthy to be near: hesitation before unstable contact, micro-corrections under uncertainty, yielding under resistance, and conservative withdrawal under sustained destabilization. These are not task completions. They are the texture of a body staying coherent while it is pushed. If they persist as degrees of freedom, contact modes, and temporal depth increase, the result is not a better demonstration of the same kind of system; it is a different autonomy regime. If they fragment, the hypothesis fails there, visibly.

There is a second axis to this composition, distinct from the abstraction of any single body's behavior. Because the unit is the primitive and the body is what units make when they couple, the same rule that lets units form one coordinated body does not, in principle, bind them to a fixed one: units that can join can also part and join again, so that a swarm of separate units, a loosely coordinated group, and a single integrated body are not three architectures but three regimes of one coupling rule, selected by which is locally more viable. What the work shows of this is bounded and should be stated as such: two units sharing a medium have been shown to compose into coordinated joint behavior with no scripted cooperation, which establishes that units compose at the smallest scale. That many units could assemble, dissolve, and reassemble into structured collective bodies with differentiated roles across them is the architecture's prediction for how this axis scales, and it remains derived and unobserved.

5. Two formulations of one principle

The principle has two implementations at different depths.

The **graph formulation** is the one with the operating proof above: sensors and motors couple through nodes, edges, and modulatory signals that grow and prune under local viability rules. It captures the logic at a higher, more conventional sensorimotor level, and it already demonstrates the inversion working.

The **medium formulation** is more fundamental in intent. Instead of building signaling in as edges, it treats the unit's interior as a bounded space hosting a metabolizing computational unit, and lets units affect each other only indirectly through a shared physical medium that carries resource, waste, heat, and mechanical deformation. Metabolism is taken as a founding principle rather than an addition: energy is generated from resource flowing through the medium and held near a viable value, the byproducts of that use are discharged back into it as burden, and the four-state regimes govern how the unit adapts its boundary in response. The defining design choice is that signaling between units is not given. It is left to *emerge*, if it emerges at all, as units come to influence each other through local depletion, replenishment, accumulation, and deformation of the medium they share. This is the stronger bet, and it is posed as a bet: that signaling, sensing, and motor specialization need not be engineered as foundations but can be grown as consequences. The formulation is constructed precisely to test that proposition: to find out whether those specializations self-organize from the metabolic medium, not to assert that they do. The first piece of that test has now returned a result. In a two-unit experiment, two MF units sharing one medium and

starting from opposite internal conditions settle, with no scripted cooperation, from initial push-pull into coordinated balancing, because each unit's recovery becomes a consequence the other must regulate around. That is genuine emergence of coordination between metabolizing units, and it is shown rather than predicted. It is also still only that: coordinated balancing between two units is not yet structured signaling, sensory specialization, or motor specialization, all of which remain derived and unobserved. So the formulation has crossed from entirely predicted to partly demonstrated at its smallest scale, and the distinction between what has emerged and what is still expected is kept deliberately sharp.

What separates this from any system of local rules over a medium is that the rule is not imposed from outside and the parts are not indifferent to the outcome. Each unit has a viability it must keep, and its structure changes only because what happened to it had a consequence it had to regulate. The coordination such a system reaches is therefore not a pattern the medium settles into but a retained change in the units themselves, biased by the consequences they impose on each other. The settling could be mimicked by dynamics that keep no memory. What could not is what each unit carries away: a lasting bias in how it responds, left by the consequences the other imposed on it. The units do not just balance; they carry the balancing forward as changed structure.

6. What kind of intelligence this is

It is worth being plain about what this is not aiming at, because the comparison the field reaches for first is the wrong axis. This is not a system racing to learn a hundred languages or do Einstein-level mathematics. Its generality is a different kind. It is not abstraction-transfer (compressing the world into a portable model and reapplying it) but bottom-up recombination: accumulated, consequence-forged coordination that recombines under new constraints, the way a craftsman or a child generalizes, staying robust in contact because every layer was earned against physical consequence rather than fitted to a distribution. This recombination is selective, not additive: which couplings engage and which are held back is settled by the same excitatory and inhibitory pressures that fall out of local viability relations, not by any evaluator ranking what is useful. What recombines is decided locally and physically, as everything in this architecture is.

The same logic explains how such a system comes to do useful work without being given a task to pursue. Behavior is not specified as an objective; it is shaped by the physical situation the body is placed in. A tool, a load, a surface, a constraint each presents the body with a particular distribution of strain, and the body's only tendency is to regulate that strain within viable

bounds. Where a situation is arranged so that the low-strain way of acting and the useful way of acting coincide, the body settles toward the useful action, not because it was selected for, but because it is the configuration under which the local pressures resolve. A well-formed tool is precisely such an arrangement: it transmits force cleanly when handled as intended and binds, jams, or loads the body when handled otherwise, so that competent use is also the path of least accumulated strain. Incorrect use is not penalized by any evaluator; it simply remains under unrelieved strain, and unrelieved strain is what keeps driving structural change.

This is not only a prediction of the architecture; it is already visible in the specimens. In simulation, a vehicle body brought under this regime learns to drive smoothly and safely within minutes. An optional operator signal is used only to get it moving at first and is then withdrawn; the driving itself is learned through consequence alone, with no objective, because staying viable — on the road, stable, unstrained — is the same configuration as driving well. What remains to be shown is reach rather than principle: that the same coincidence of usefulness and viability extends from continuous control of this kind to fine manipulation, contact-rich tool use, and longer multi-stage tasks, which is what the hardware program is built to establish.

A machine body organized this way, around consequence rather than objective, is what we call a *sentitoid*. The word marks one specific thing and disclaims another: the organization resembles sentience in that what happens to the body is retained and shapes what it does next, the property we call consequence retention; it carries no claim of feeling or subjective experience.

Naming the kind precisely is also what lets the ambition be stated without inflation. Nothing in the architecture obviously caps it: the same law that turns a reflex into coordination could, in principle, turn stabilized coordination into the base on which internal rehearsal and abstraction form. So this is a candidate route to general intelligence of a human-like kind, not a smaller ambition than the model-based one, but a different one, approached from the body upward rather than from the corpus inward. Whether it reaches that far is not a claim this brief makes; it is exactly the open question named below: whether higher-order abstraction crystallizes on top of stabilized structure, or saturates at a stability boundary. The ambition and its unproven ceiling are set down here in the same breath, on purpose.

The cleanest expression of that difference is a hypothesis the architecture points toward rather than a behavior it has shown: that such a system need not

internalize a tool like a large language model in order to benefit from one. The direction the architecture implies is that it could learn to *wield* such a tool, the way a person uses a calculator without becoming one: a division of cognitive labor rather than a deficiency, with the representation-rich route as the vast, abstract, contact-brittle instrument and the viability-grounded route as the bounded, consequence-aware user that knows when to reach for it and stays coherent when it puts it down. If that holds, the renunciation of superhuman generality is not a limitation to apologize for but the correct shape for a system meant to be loose in a kitchen or a warehouse.

It must be said plainly that this has not been demonstrated. Human-like contextual tool use sits high on the ladder of behavioral recombination, the composition of simple coordinations into ever more context-sensitive action, well above the low-level coordination the specimens show recombining, and on the same rung as the abstraction ceiling that remains open. The ladder meant here is the growing abstraction of a body's *behavior*, not the joining of many units into one body, which is the separate compositional axis treated in the scaling discussion above. It is offered here as a future architectural hypothesis, motivated by how the system is built, and nothing more than that until the behavior appears.

7. Safety as architecture, not as an added restraint

Most approaches to safety are subtractive: build the capable system, then bolt on restraint that fights it. Here, safety and capability are the same axis, because the mechanism that keeps a local region accountable to the whole body is the same mechanism that produces coherent whole-body behavior in the first place. Containment is not a brake on the architecture; it is the architecture working.

This reframes the risk precisely, and the precision matters more than reassurance. The architecture removes, or radically weakens, the failure mode that dominates conventional AI safety: a capable system carrying a detached objective that models intervention as an obstacle and seeks to preserve itself, gather resources, and resist shutdown. There is no detached objective here to come apart from the body, so there is little for that runaway to be *about*.

It does not, however, make pathology impossible, and claiming so would be the wrong kind of claim for a system whose whole premise is that structure emerges. If structure can emerge, so can pathology. The relevant danger is more primitive and more biological than the superintelligence story: a local territory becoming too stable, too self-reinforcing, too locally viable, and insufficiently coupled to the distress of the whole. Such

a territory needs no objective and “wants” nothing. It only has to become a stable local pattern the larger body cannot dissolve, the engineered analog of oncogenesis, where a region stays locally competent while its coupling to whole-body viability is lost. Bounded resource and hard caps limit how *large* such an attractor can grow; they do not by themselves prevent it from being *parasitic* within bounds.

There is also a failure milder and more common than malignancy, from the same source. Because the body learns from lived consequence rather than a curated objective, it can mislearn the way an organism does: a misleading encounter can leave a spurious association, a single bad experience can install an overcautious habit, an idiosyncratic quirk can take hold. These are errors of the kind a trained animal or a person shows, not the opaque failures of an optimized objective, and that is what makes them tractable. They appear in behavior rather than hiding in weights, they are visible in the internal-state read-out as they form, and they are corrigible by the same means that produced them: further consequence, retraining, or reset of the affected structure. The architecture does not make learning infallible; it makes its errors familiar in kind and observable as they arise.

So the architecture's real contribution to safety is a conversion, not an elimination. It turns malignancy from an unbounded, objective-driven problem into a bounded, bodily, observable, and testable body-regulation problem, one that can be starved, isolated, reset, and measured, precisely the interventions a detached-objective runaway is built to resist. The open problem, stated honestly, is local-global reconciliation: whether whole-body distress can propagate back as binding local constraint before a region becomes a stable parasite. That is simultaneously the safety problem and the coherence-scaling problem. They are one problem, and one proof.

Two properties make this conversion possible, and both are structural rather than bolted on. The first is legibility: because behavior arises from visible local regulation rather than from opaque parameters, a forming pathology or a mislearned tendency is observable in the internal-state read-out while it is still small, not inferred after the fact. The second is bounded growth: because capability accrues only by accumulating structure, and the permitted amount of structure is a set limit, total capacity is bounded by construction, independent of how the body behaves within that bound. Legibility lets a problem be seen; bounded growth ensures it cannot outgrow the means to address it.

The decisive test is therefore adversarial, and must be run as a pair on the same specimen at the same settings:

- **Containment under resistance.** Not whether the body can reset a region that passively accepts correction, but whether it can reach and constrain a region whose own local adaptations are what block the correction: one lowering its permeability to incoming inhibition, distorting its coupling, protecting itself. The measurable quantity is a containment margin: the gap between how fast the body can drive corrective coupling into a territory and how fast the territory can wall that coupling out. Margin positive, the parasite cannot outrun correction. Margin at zero is the engineered carcinogen, and finding it deliberately is the point.
- **Productive destabilization preserved.** A body tuned for strong reconciliation suppresses local deviation hard, and local deviation is the fuel the architecture runs on. Tuned too far, containment is bought at the price of exploration, and the system is safe by being too rigid to learn, which is also too rigid to be useful. So the same settings must be checked for whether regions retain enough room to destabilize productively.
- **Derived and predicted, not yet shown:** in the medium formulation, the emergence of *structured* sensory and motor specialization from a shared medium, beyond the basic positional differentiation already observed, and the emergence of structured signaling; the assembly, dissolution, and reassembly of many units into structured collective bodies with differentiated roles; the extension of competent behavior from physical constraint, shown in continuous control, to fine manipulation, contact-rich tool use, and longer multi-stage tasks; tool-use generality at human-like contextual judgment.
- **Open:** whether coherence scales with dimensionality and temporal depth; whether higher-order abstraction crystallizes on top of stabilized structure or saturates at a stability boundary; whether local-global reconciliation holds under adversarial containment.

The viable architecture is the band where those two edges leave a gap between them, where containment holds without smothering the destabilization that makes the system intelligent. Should the only settings that contain the parasite turn out to be the settings that kill exploration, that would not be a tuning failure. It would be the architecture revealing that safety and capability were not one axis after all. That is the single outcome this line of reasoning bets against, and therefore the first one worth trying to falsify.

8. Epistemic status

The brief stands or falls on physical behavior under increasing dimensional stress, not on argument. Stated against that standard:

- **Demonstrated** (low dimensionality): in the graph formulation, organized behavior from local viability regulation with no objective, persistence when structural updates are paused, cross-embodiment transfer without retraining, and an optional operator signal used only to initiate movement and then withdrawn, after which competence was acquired through consequence alone; in the medium formulation, coordinated balancing between two metabolizing units sharing a medium, arising without scripted cooperation and persisting as retained structural change in each unit rather than as a transient settling, together with a basic positional differentiation, in which membrane regions adjacent to a sensorimotor interface adapt differently from regions that cannot interact with it.

These tests are not abstractions waiting on some distant apparatus. The next step is a minimum physical prototype: a single integrated specimen carrying more degrees of freedom, contact modes, and temporal depth than the current ones, built for the express purpose of running the signatures named above, whether hesitation, micro-correction, yielding, and conservative withdrawal persist as the body grows more complex, and whether a region adapting to resist correction can still be contained. That prototype is where the open column above begins to be answered, or fails to be.

The correct posture is not “trust that it cannot go wrong.” It is that the possible failures are bounded, bodily, observable, and testable, and that the tests that matter are specified above, ahead of the claims they are meant to support. A second route to intelligence is worth taking seriously exactly to the degree that it says, in advance, what would prove it wrong.

Common objections to the architecture, and how they are answered, are addressed in detail in a companion technical Q&A: [\[link\]](#).

The core architecture and its purpose-built hardware are protected by filed patents, provisional and non-provisional.