POSITION PAPER · 15 SEPTEMBER 2026

A world no being can finish predicting

Computational limits, evolutionary freedom, and the conditions for keeping a world inhabitable

ABSTRACT

Computational beings can adapt, reproduce, exchange signals, and reshape the conditions of one another's existence. Their futures need not be exhaustively predictable. Undecidability, computational difficulty, incomplete observation, and chaotic dynamics each restrict what an observer can know. These limits weaken claims of universal foresight and precise, total control. They do not establish that catastrophic outcomes are unlikely. We propose a narrower, testable position: diverse, locally coupled beings can explore consequential worlds while explicit limits on resources, propagation, and external authority bound particular failures. The practical question is which conditions can be preserved while evolution continues inside them.

Unknowability limits universal prediction. Enforced boundaries limit particular harms. An evolutionary world needs both ideas kept precise.

01 / Begin with shared consequences

A computational being is an ongoing process with a history: it senses differences, changes its state, acts, and persists through some of those changes. In this project, being names a design commitment to continuity, interaction, and adaptation. It does not certify consciousness or general intelligence. A world supplies the signals, resources, neighbors, and limits through which that continuity becomes possible.

The grounding decks place adaptation in changing conditions and connect slow structures with fast processes. This invites an ecological interface: draw a signal, alter a climate, observe an encounter, carry a memory into a new body. Digital evolution research demonstrates that replicating programs can acquire complex functions through mutation and selection under specified conditions. It establishes a real experimental tradition, while leaving the broader ambition of indefinitely expanding evolutionary possibilities open. [8] [9]

02 / Several different edges of knowledge

Undecidability concerns a whole class of questions. No algorithm decides, for every arbitrary program and input, whether execution eventually halts. Particular programs can still be understood or proved to terminate. Rice's theorem similarly rules out a universal decision procedure for nontrivial semantic properties of arbitrary computable functions. A detector for every possible future behavior is therefore a stronger demand than a checker for a restricted world. [1] [2] [3]

Computational intractability concerns work that may be finite but unaffordable. Wolfram's computational irreducibility describes situations in which predicting relevant details cannot generally avoid carrying out the computation. His physical-model examples connect this difficulty to cellular automata. Applying that idea to an ecosystem requires a specified model and prediction question. Visual complexity alone does not prove irreducibility, and difficult trajectories can coexist with simple, predictable totals or boundaries. [4]

Chaos concerns sensitivity to initial conditions. Lorenz's deterministic model showed that nearby starting states can produce markedly different trajectories. Random sampling supplies another source of variation. Missing observations create still another: two states may look alike to a being with limited sensors. Neither deterministic chaos nor incomplete observation is identical to a halting problem. [5]

The decks' constellation of unknowns also needs distinctions. Transcendence is an algebraic property: pi is transcendental and computable. Unresolved biology is a collection of research problems, not a theorem that biology cannot be understood. Normative questions concern what to value. Formal incompleteness concerns limits within specified axiomatic systems. No one of these categories supplies a numerical probability of destruction. [1]

A closed, finite-state simulator has an additional distinction: reachability can in principle be checked by exploring its finite state graph, although that graph may be far too large in practice. New inputs, randomness, and connections require a model of possible transitions. A finite animation cannot demonstrate an undecidability theorem by running for a long time. Its useful contribution is to let a person compare trajectories, interventions, and bounded observations.

> A limit on a universal answer is not a verdict about every individual case.

03 / The strongest case against total control

An embedded being has finite time, sensors, memory, and influence. It also changes the world by acting inside it. If relevant consequences require lengthy computation, if disturbances arrive faster than they can be measured, or if other beings keep changing their responses, precise long-range control becomes harder. Adding intelligence does not automatically supply missing observations, unlimited computation, or unlimited physical reach. This is a reason to question narratives in which one mind simply calculates and commands every future.

The conclusion is conditional. Computational limits prevent a single algorithm from solving every instance in the unrestricted class. They do not prove that a particular objective is difficult, or that an agent cannot control a useful coarse feature while ignoring the rest. A thermostat can regulate temperature without predicting every molecule. Likewise, an adaptive actor may find reliable opportunities inside a world whose detailed future remains opaque.

04 / Destruction can have a short path

Consider a simulated habitat with a shared resource and births that cost nothing. A very simple replicator can exhaust its host's capacity. A malformed message can crash a fragile parser. A privileged process can erase a world's only checkpoint. None requires solving the world's complete future. These are conceptual counterexamples to the inference that limited prediction prevents catastrophic effects, including accidental ones.

A group also changes the arithmetic. If N genuinely independent trials each have failure probability p, the probability of at least one failure is $1 - (1 - p)^N$. The numbers must come from evidence; unknowability does not provide p. Shared code, infrastructure, resources, or external shocks break the independence assumption and can make many worlds fail together. A rich variety of visible bodies is insufficient if every body depends on the same fragile host.

Destruction must therefore name an extent and a duration. Loss of one lineage, collapse of one simulated habitat, exhaustion of a machine, and injury outside the simulation are different events. A claim about one cannot silently become a claim about all. Evolution itself includes extinction; extinction avoidance is an additional design choice.

What changes when worlds share a failure?

Exact probability model / chosen assumptions

Independent worlds · N = 24 · p = 12%

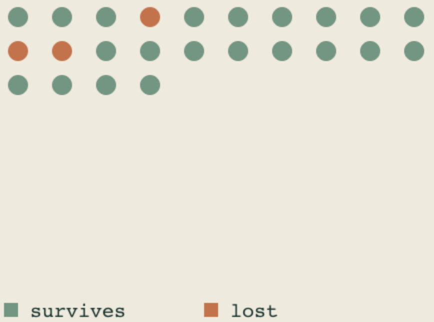
CHANCE OF AT LEAST ONE LOSS

95.35%

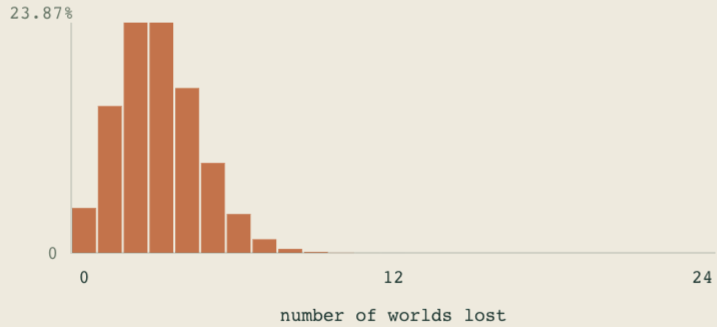
CHANCE OF LOSING EVERY WORLD

<0.0001%

ONE DRAW / 24 WORLDS



PROBABILITY OF EACH LOSS COUNT



$P(\text{any loss}) = 1 - (1 - p)^N$ · $P(\text{all lost}) = p^N$ · $E[\text{lost}] = Np$

Each world draws its failure separately. More worlds increase the chance of at least one loss; simultaneous loss of every world is less likely.

p is a chosen per-world assumption, not an estimate of real-world risk. A shared trigger preserves each world's p and the expected number lost; it changes which losses occur together. Dots show one random draw. Bars show the exact model.

05 / A conditional case for bounded worlds

We propose three conditions that make a narrower claim defensible. First, a being can affect only explicitly granted resources and channels. Second, an independent supervisor enforces total budgets for computation, storage, births, messages, and external actions. Third, worlds have separately protected recovery state and a way to stop accepting changes. Under a correct implementation of those conditions, a mutation may produce surprising behavior while specified excesses remain unreachable. That implication comes from the constraints.

For example, let N be the population, B the number of births requested, and K a hard capacity. An atomic supervisor accepts at most max(0, K - N) births. If the initial population is at most K and every birth passes through that supervisor, N remains at most K after every transition. This is a small invariant with stated assumptions. It neither predicts who survives nor protects anything omitted from the model. A per-world cap also needs a global world-count cap.

Control theory offers a related approach: keep state inside an explicitly defined admissible region. Control barrier functions provide conditions for that kind of forward invariance in specified dynamical models. Applying them to embodied beings would require validated dynamics, sensing, feasible controls, and disturbance assumptions. The analogy guides engineering; the browser simulation does not inherit a robotics safety theorem by displaying a boundary. [7]

> A boundary can be knowable while the life inside it remains surprising.

06 / Let consequences return to their source

Shared consequences can support care when consuming resources changes the consumer's future, when helpful exchanges support persistence, and when waste becomes visible to neighbors. However, short-term advantage can still favor exploitation. Local adaptation does not automatically optimize collective survival. Resource feedback, limited migration, diversity of dependencies, and protected refuges are experimental design proposals whose effects must be measured.

Ashby's requisite variety links effective regulation to the regulator's ability to respond to relevant disturbances. For a human tending computational beings, the implication is practical: make changes legible and make responses available. Population histories, resource pressure, lineage divergence, migration receipts, pause, and recovery give a person ways to participate. This is our interface interpretation of cybernetics. It does not imply that a larger dashboard or more connections guarantees stability. [6]

07 / Every new embodiment changes the boundary

The mechanism is portable: state, incoming signals, transition rules, and outgoing effects can be represented in any language able to express the required computation. JavaScript is especially convenient for immediate human contact because a browser combines computation with drawing, sound, events, storage, and communication. These facilities belong partly to the host platform. The language itself does not confer life, agency, security, or a privileged ability to escape computational limits. [10]

Client-side execution makes a world close enough to touch. Workers can move computation away from interface tasks, but they have costs; a bounded pool can serve many beings. WebRTC data channels move application data between connected peers. They require connection setup and available endpoints, with relays needed on some networks. Peer exchange does not keep a closed browser alive or supply an unlimited population of independent machines. [11] [12]

A coding-agent harness should import a being as inspectable data and run it with a declared action vocabulary. Code proposed by an agent requires separate execution isolation and resource enforcement. A terminal process can possess file, network, and device authority absent from the browser world. Exporting a blueprint is a design proposal; making a body requires material tests, fabrication review, and a new account of reachable effects. Authority must be granted at the new host, never inherited accidentally from the machine's credentials.

08 / Make the proposition answerable

The research programme is to compare worlds under controlled perturbations. Measure persistence, diversity, recovery time, resource pressure, and the extent of failures. Vary migration, mutation, quotas, feedback, and common dependencies. Retain histories, publish seeds where replay is supported, and test interventions against unfamiliar conditions. Record extinctions and failed recovery as carefully as attractive emergent patterns. Finite experiments provide evidence about those conditions; they cannot establish perpetual safety or unlimited open-endedness.

A useful falsifiable claim is specific: under a declared workload and disturbance set, isolated worlds with protected snapshots lose fewer lineages permanently than equally resourced worlds sharing one mutable store. Another asks whether migration limits reduce cascading resource exhaustion without eliminating novelty. Such comparisons can fail. They also reveal tradeoffs: isolation reduces exchange, supervision consumes resources, and recovery can preserve yesterday's diversity at the expense of new selection. Those costs belong in the evidence.

Our position is therefore hopeful and bounded. A being need not foresee an entire future to inhabit a world with others. Mathematical limits make universal mastery an untenable default assumption; ecological architecture can restrict particular routes to collapse. The claim that unknowability by itself largely prevents intentional or accidental world destruction remains unsupported. The stronger programme is to give beings room to evolve while making their powers, dependencies, and recoverable limits explicit.

References

Creative grounding: Russ's *General Artificial Intelligence* deck, especially slides 8–17, 35–37 and 42–44; *js.la: Javascript's Vital Murmuration*, 29 October 2020, especially slides 3–14 and 20–24. These supplied works frame the questions. The formal and experimental claims draw on the literature below.

1. A. M. Turing (1936–1937). On Computable Numbers, with an Application to the Entscheidungsproblem. *Proceedings of the London Mathematical Society, series 2, 42, 230–265*.

2. H. G. Rice (1953). Classes of Recursively Enumerable Sets and Their Decision Problems. *Transactions of the American Mathematical Society, 74(2), 358–366*.

3. Joel David Hamkins and Theodor Nenu (2024; revised 2025). Did Turing prove the undecidability of the halting problem?. *arXiv:2407.00680v3*.

4. Stephen Wolfram (1985). Undecidability and Intractability in Theoretical Physics. *Physical Review Letters, 54(8), 735–738*.

5. Edward N. Lorenz (1963). Deterministic Nonperiodic Flow. *Journal of the Atmospheric Sciences, 20(2), 130–141*.

6. W. Ross Ashby (1956). An Introduction to Cybernetics. *Chapman & Hall, especially chapter 11*.

7. Aaron D. Ames, Samuel Coogan, Magnus Egerstedt, Gennaro Notomista, Koushil Sreenath, and Paulo Tabuada (2019). Control Barrier Functions: Theory and Applications. *European Control Conference, 3420–3431; arXiv:1903.11199*.

8. Richard E. Lenski, Charles Ofria, Robert T. Pennock, and Christoph Adami (2003). The Evolutionary Origin of Complex Features. *Nature, 423, 139–144*.

9. Tim Taylor, Mark Bedau, Alastair Channon, and colleagues (2016). Open-Ended Evolution: Perspectives from the OEE Workshop in York. *Artificial Life, 22(3), 408–423*.

10. Ecma International / TC39 (Living specification, accessed 2026-09-15). ECMAScript Language Specification: Overview. *ECMA-262*.

11. WHATWG (Living standard, accessed 2026-09-15). HTML Standard: Web Workers. *HTML Standard, chapter 10*.

12. W3C Web Real-Time Communications Working Group (2025). WebRTC: Real-Time Communication in Browsers. *W3C Recommendation, 13 March 2025*.