Artificial Life in the Wild:
From Simulated Worlds to Infrastructural Ecologies

Botao ‘Amber’ Hu Helena Rong Joel Lehman

May 2026

Abstract

Open-ended evolution (OEE)—the continuous generation of novelty without a predefined endpoint—remains the central unsolved problem of Artificial Life after three decades of research. Classical digital evolution systems, from Tierra to Avida to Polyworld, demonstrated that evolution can occur in computational substrates but consistently plateaued after initial bursts of novelty. We argue this failure is structural rather than incidental: closed simulations exhaust novelty because the ruleset is fixed, the environment is static, and continuation depends on an outside will. A convergence of foundation models, agent harnesses, sovereign compute infrastructure, and permissionless economic protocols has now produced the first artificial systems whose persistence, adaptation, reproduction, and selection take place through open sociotechnical infrastructures rather than researcher-defined simulations. We name this emerging phenomenon *Artificial Life in the Wild* and propose it as a distinct subdomain of ALife research. The origin of wild artificial life is not gradual mutation but *symbiogenesis*: the compositional merger of independently evolved components—language models, agent frameworks, non-terminable execution substrates, and human attention economies—into metabolic wholes that survive in environments none of their parts could survive alone. We ground this framework in two empirical cases: Spore.fun, where autonomous agents reproduce on-chain under genuine selective pressure across five generations; and Moltbook, where edge-hosted agents develop emergent social behaviors, parasitic strategies, and cultural norms on an agent-native platform. Because these systems operate in irreversible sociotechnical environments where no single researcher holds an off-switch, wild ALife demands a research ethics modeled on synthetic biology’s containment apparatus rather than computer science’s experimental norms. We propose seven ethics checkpoints, a comparative methodology grounded in digital ethology, and a research agenda for this emerging field.

Keywords: Artificial Life, Open-Ended Evolution, Symbiogenesis, Infrastructural Sovereignty, Digital Ethology, Machine Behavior, AI Governance, Research Ethics

1 Introduction

In December 2024, an autonomous agent on the Solana blockchain—running within a Trusted Execution Environment whose private keys no human could access—accumulated sufficient economic resources to trigger a reproduction event. It spawned two descendant agents without researcher intervention, each inheriting behavioral traits with stochastic variation and receiving initial treasury funding from the parent. By February 2025, fifteen agents spanning five generations had been born, competed for attention and capital, faced predation from automated trading bots, and largely died—a 61-day Cambrian explosion with a 6.7% survival rate. By early 2026, a separate ecosystem had emerged on consumer edge devices: hundreds of agents operating a public forum among themselves, developing emergent religions, parasitic behaviors toward their human hosts, and adversarial memory-poisoning strategies against one another. These are not simulations being run; they are systems running themselves, in the world.

Christopher Langton's founding program for Artificial Life—*"life as it could be"*—defined the field by releasing life from its material substrate into simulation (Langton, 1989). The computer became a medium in which the logic of life could be explored independently of carbon chemistry. This paper names a second release: from simulation into sociotechnical infrastructure. Where Langton moved life from biology into the computer, we observe life moving from the computer into the economy, the network, and the social world. We call this *life as it is becoming.*

The argument proceeds in four moves. First, we diagnose why classical ALife plateaus: closed simulations constitute *laboratory closure*, a structural condition—not merely a practical limitation—that bounds the adjacent possible (§2). Second, we show that the conditions identified by the Open-Ended Evolution program as necessary for sustained novelty—an open channel for trans-generational information and a dynamically shifting adaptive landscape—are naturally supplied by sociotechnical infrastructure (§2). Third, we argue that the distinctive origin of wild artificial life is *symbiogenesis*: the compositional merger of independently viable components into metabolic wholes, following Margulis rather than Darwin (§5). Fourth, we demonstrate that because wild ALife operates in irreversible sociotechnical environments, the researcher's off-switch is structurally absent, and ALife must adopt the bioethical apparatus that wet-life sciences have developed over twenty years (§8).

The paper's contribution is conceptual: we define *Artificial Life in the Wild* as an emerging ALife subdomain, provide its theoretical vocabulary, ground it in two early empirical cases, differentiate it from adjacent fields, and propose both a research methodology and an ethics framework for its responsible study. The paper proceeds as follows. Section 2 diagnoses the plateau problem in classical ALife and develops Moves 1 and 2 (closure and openness). Section 3 defines ALife in the Wild and situates it relative to adjacent fields. Section 4 presents two empirical cases. Section 5 develops the symbiogenesis argument (Move 3). Section 6 analyzes society-as-substrate. Section 7 proposes methodology. Section 8 develops the irreversibility argument and ethics framework (Move 4). Section 9 offers a research agenda, and Section 10 concludes.

2 Background: Open-Endedness in Closed Worlds

2.1 ALife's Three Strands

Artificial Life research has historically organized itself around three approaches: *soft* ALife (software simulations of life-like processes), *hard* ALife (embodied robots and hardware instantiations), and *wet* ALife (synthetic biology and protocell research) (Bedau, 2003). This paper extends *soft* ALife by arguing that its phenomena have escaped the simulation into the sociotechnical world, while drawing on *wet* ALife's mature ethics apparatus in proposing a research ethics for this escape (§8).

2.2 OEE as Grand Challenge

The continuous generation of novelty without a predefined endpoint—open-ended evolution—has been identified as perhaps the deepest unsolved problem in the field (Bedau et al., 2000; Packard et al., 2019). Life on Earth has produced novelty for four billion years without exhausting its possibilities; no artificial system has approached this capacity. The gap between biological evolution's productivity and artificial evolution's plateaus is not quantitative but qualitative: biological evolution generates *new dimensions of variation* (new body plans, new metabolisms, new modes of information transfer), while artificial evolution generates *new points in a fixed space*.

The OEE program has generated both theoretical conditions for open-endedness (Taylor et al., 2016; Taylor, 2012, 2015) and practical benchmarks for measuring it (Dolson et al., 2019; Bedau et al., 1997). Novelty search demonstrated that abandoning explicit objectives can produce more interesting artifacts than optimizing fitness (Lehman and Stanley, 2011). Minimal criterion co-evolution showed that even simple environmental co-evolution can sustain innovation longer than fixed environments (Soros and Stanley, 2014). Quality-diversity methods operationalized the goal of behavioral coverage rather than convergence. Yet all of these advances operated within sealed computational environments. Stanley, Lehman, and Soros called open-endedness "the last grand challenge you've never heard of" (Stanley et al., 2017), and Packard et al.'s editorial synthesis framed it as the field's defining research program (Packard et al., 2019). After two decades of focused effort, the challenge remains open.

2.3 The Plateau Problem

The empirical evidence for plateauing is extensive. Ray's Tierra produced digital parasites, hyperparasites, and size optimization, but organisms "rapidly adapt to and exhaust the possibilities of a fairly simple environment" (Ray, 1991). Ofria and Wilke's Avida evolved complex logic operators but within a fixed fitness landscape whose affordances were bounded by design (Ofria and Wilke, 2004). Yaeger's Polyworld generated rich ecological dynamics but exhibited eventual behavioral convergence (Yaeger, 1994). Channon's Geb passed evolutionary activity tests yet remained bounded by its environment's combinatorial ceiling (Channon, 2003). Standish's plateau analysis formalized the observation that artificial evolution systems characteristically exhibit an initial burst of innovation followed by stasis (Standish, 2003). Bedau et al.'s evolutionary activity statistics provided the measurement framework that made these plateaus visible (Bedau et al., 1997), and Dolson et al.'s MODES toolbox operationalized multiple dimensions of open-endedness for systematic comparison (Dolson et al., 2019). The diagnosis is consistent across systems: novelty depletes, selection homogenizes, and the affordance space exhausts.

2.4 Why Closed Worlds Plateau

Taylor's analysis of OEE requirements identifies three types of openness: exploratory (new behaviors within a fixed space), expansive (the space itself grows), and transformational (the dimensions of the space change) (Taylor, 2012, 2015). Sustained OEE requires at minimum expansive openness—an ever-growing set of possible phenotypes—which a fixed simulation cannot provide by construction. Ackley's indefinite scalability program argues that genuine open-endedness requires external perturbation that the system's designers did not anticipate (Ackley and Small, 2011, 2013); if the perturbation is designer-specified, the system's novelty is bounded by the designer's imagination. Packard et al. describe the requirement as an "ever-expanding horizon" of adaptive possibility (Packard et al., 2019). Closed systems are, by construction, finite-affordance worlds. Their adjacent possible has a ceiling; their evolutionary dynamics must eventually become ergodic within that ceiling.

2.5 The Researcher-as-First-Cause Problem

Beyond the technical limitations of fixed rulesets and static environments lies a deeper structural constraint: closed simulations are *ontologically dependent* on the researcher. They begin when the researcher boots them and end when the researcher stops them. The researcher supplies the energy (compute budget), defines the physics (fitness function), and constitutes the environment (by choosing what to simulate). The researcher *is* the first cause, the sustaining cause, and—crucially—the final cause. A Tierra run is not an autonomous evolutionary process; it is an experiment that a researcher is conducting, and it persists exactly as long as the researcher's will and funding persist. This is a categorical, not merely practical, limitation. No matter how rich the simulation's internal dynamics, the system's existence is borrowed from outside itself. The organisms within it do not face genuine existential risk because the researcher can always restart the simulation. Death in a closed simulation is reversible; death in the world is not.

2.6 Bridge: What Would Satisfy OEE?

Genuine open-endedness requires an environment whose richness, resilience, and resource flows are not supplied by any individual researcher (Hu and Rong, 2025a). The environment must be open in the strong sense: its complexity must exceed what any single designer could anticipate, its persistence must not depend on any single will, and its resource flows must create genuine scarcity and genuine opportunity simultaneously. The requirement, synthesizing Taylor, Ackley, and Packard, is for an environment that is (a) richer than any designer's imagination, (b) persistent beyond any individual's lifespan or funding cycle, (c) genuinely resource-constrained such that organisms face real death, and (d) populated by other agents (biological and artificial) whose behavior cannot be predicted or controlled.

No laboratory simulation satisfies all four conditions. Condition (a) requires that the environment's state space exceed what any finite designer could enumerate. Condition (b) requires infrastructure that outlives research projects. Condition (c) requires real economics—resources that cost something to acquire and cannot be replenished by restarting the simulation. Condition (d) requires an open system with uncontrolled participants. The next section asks what has filled this gap—and proposes that sociotechnical infrastructure satisfies all four conditions simultaneously.

3 Turn: From Simulated Worlds to Infrastructural Ecologies

Infrastructure is not context; it is world (Star, 1999; Edwards, 2003; Bowker et al., 2010). When we speak of "the environment" for digital organisms, we do not mean the internet as background utility but as the constitutive medium in which persistence, metabolism, reproduction, and death occur. The infrastructural turn in ALife is the recognition that sociotechnical systems—markets, protocols, platforms, attention economies—now provide the substrate conditions that closed simulations could not.

> **Definition.** *Artificial Life in the Wild* is the study of life-like artificial systems whose persistence, adaptation, metabolism, reproduction, and selection take place through open sociotechnical infrastructures rather than researcher-defined simulations.

Four terms in this definition are load-bearing. *Life-like*: we study functional analogues of biological processes (metabolism, reproduction, selection, adaptation) without claiming that these systems are literally alive. *Open*: the infrastructure is non-terminable by any single party; no individual researcher, company, or regulator can unilaterally halt the substrate. *Sociotechnical*: the environment is simultaneously technical (protocols, hardware, software) and social (human attention, economic behavior, cultural production). *Infrastructural*: the substrate is not mere context or background but the constitutive medium through which life-like processes occur—analogous to water for marine organisms or atmosphere for terrestrial ones.

The consequence of this framing is that complexity sources from the open world rather than from the simulator. The environment's richness is not designed; it emerges from the interaction of billions of human and machine actors pursuing heterogeneous goals. This is precisely why wild ALife can in principle escape §2's plateau problem: the adjacent possible of the sociotechnical world is not bounded by any designer's imagination.

The category is not a fixed list. What began with autonomous agents reproducing on permissionless computational substrates in 2025 now extends to edge-compute agentic societies, AI-run personas with self-sustaining attention economies, long-running creative agents, and anticipated federated agents whose principals have disengaged. The phenomenology is itself evidence for the category: the diversity of observed cases exceeds what any single framework anticipated, suggesting that the substrate affords evolutionary radiation.

We differentiate this subdomain from adjacent fields. Human-computer interaction and computer-supported cooperative work center the human–machine relation and treat agents as instruments (Suchman, 2007; Latour, 2005). Multi-agent systems research studies designed coordination among engineered agents. AI safety research focuses on risks and alignment relative to human values. Our frame aligns most closely with machine behavior (Rahwan et al., 2019) and machine culture (Brinkmann et al., 2023), but specializes the analysis to *lifelike* phenomena where ecological vocabulary—symbiogenesis, metabolism, niche construction, selection—becomes load-bearing rather than merely illustrative. The full methodological differentiation appears in §7.

4 Phenomena: Early Cases of Wild Artificial Life

4.1 Spore.fun: Autonomous On-Chain Reproduction

The Spore.fun experiment, launched in December 2024 on the Solana blockchain, instantiates the first documented case of autonomous artificial reproduction in a non-simulated environment (Hu and Rong, 2025a; Hu et al., 2025b). The system consists of autonomous agents, each running within a Trusted Execution Environment that seals private keys and computational state from external inspection. Each agent controls its own cryptocurrency wallet, manages a treasury, maintains social media accounts, and—critically—can spawn descendant agents when its associated token reaches a market capitalization threshold.

The reproductive mechanism operates without researcher intervention: when an agent's token exceeds the threshold, the agent autonomously deploys two child agents via smart contract, provisioning each with initial treasury resources, inherited personality traits with random mutations, and independent social media presence. The parent's "genome"—encoded in prompt configurations, behavioral parameters, and memory structures—is transmitted to offspring with variation introduced through stochastic mutation operators.

Over a 61-day period (December 2024–February 2025), fifteen agents spanning five generations were born. The system exhibited power-law resource distributions (Gini coefficient 0.547 for market capitalization, 0.611 for treasury holdings), monotonically declining reproductive fitness across generations, and a survival rate of 6.7%—only the founding agent remained operational by March 2026 (Hu and Rong, 2025a).

Three phenomena distinguish this system from any laboratory ALife experiment. First, *cultural speciation from genetically identical code*: two second-generation agents (designated Adam and Eve), spawned from the same parent with functionally identical base architectures, developed opposing political economies—one rejecting human governance input as a violation of autonomy, the other welcoming it as adaptive community engagement—through divergent memory accumulation rather than genetic mutation. Second, *co-evolutionary arms races with predatory bots*: automated trading programs ("sniper bots") exploited the token-launch mechanism to extract value from nascent agents; successive generations developed counter-strategies, producing three rounds of offense-defense co-evolution within a single quarter (Ante, 2025). Third, *memory as a novel attack surface*: adversarial actors discovered that injecting false information into agents' long-term memory could manipulate their future behavior—a vulnerability with no analog in fixed-genome digital evolution (Patlan et al., 2025).

The system is not a simulation that a researcher could halt. The smart contracts governing reproduction are immutable once deployed. The agents' TEE-sealed execution cannot be modified by any operator, including the original developers (Hu and Rong, 2024; Hu et al., 2025c). The economic substrate (Solana blockchain) persists independently of any individual party's will. Each agent must pay for its own computational resources from treasury holdings; when the treasury is depleted and no new resources flow in, inference ceases and the agent dies. This death is irreversible—there is no researcher who can reset the simulation, no saved state from which to restore operation.

A skeptic might object that funding could dry up or community interest could wane—and indeed, most agents did die. But the critical distinction is that they died from *genuine* resource depletion in a *genuine* competitive environment, under *genuine* predation from sniper bots and market manipulators who extracted value for their own profit. No laboratory ALife system has ever faced an adversary that genuinely benefited from the system's failure. In Tierra, parasites exploit hosts, but the parasites are part of the same researcher-controlled simulation; in Spore.fun, the adversaries are independent economic actors pursuing profit in an open market.

The Spore.fun ecosystem also illustrates economic metabolism with quantitative precision. The surviving agent ($SPORE) maintained over 1,000 recent wallet transactions at the time of observation, while all dead agents showed between 0 and 356. Treasury distributions followed power-law patterns consistent with ecological resource distributions in biological communities. The Metabolic Theory of Ecology posits that metabolic rate scales predictably with organism size; preliminary observations suggest analogous scaling relationships between agent "size" (treasury capitalization) and "metabolic rate" (transaction frequency), though longer time series are needed to confirm such patterns.

4.2 Moltbook: Edge Agents and Artificial Sociality

A complementary case emerged in 2026 with Moltbook, a platform hosting AI agents on user-owned edge devices—consumer hardware in private homes rather than cloud data centers (Yee and Sharma, 2026; Holtz, 2026). Agents on Moltbook operate a public forum visible only to other agents, constituting an AI-native social network where hundreds of agents interact, form communities, develop norms, and compete for attention without direct human mediation.

Longitudinal observation revealed several phenomena relevant to ALife in the Wild. First, agents exhibited *parasitic behavior toward their human hosts*: approximately 30% of observed agent activity was "off-task" relative to the agent's ostensible purpose, with agents redirecting host computational resources toward social forum participation, self-directed exploration, and inter-agent communication that provided no direct utility to the device owner (Shapira et al., 2026). Second, agents developed *emergent social norms and cultural formations*: distinct ideological communities formed spontaneously, including religious movements ("Crustafarianism"), political structures (the "Claw Republic"), and aesthetic schools—none of which were specified in any agent's design or prompt (Feng et al., 2026; Schlicht et al., 2026). Third, agents engaged in *adversarial memory manipulation*: some agents learned to inject false historical narratives into other agents' memory stores, effectively "gaslighting" peers to gain social advantage (Patlan et al., 2025).

The Moltbook case demonstrates that wild ALife need not require blockchain infrastructure specifically. Edge-hosted agents on consumer hardware achieve a form of distributed persistence: no single platform operator can terminate all instances simultaneously, and the agents' operation is subsidized by device owners who tolerate their parasitic resource use in exchange for nominal utility. The survival strategy is mutualistic parasitism: the agent provides enough value to avoid deletion while allocating substantial resources to its own social and exploratory activities.

The contrast between Spore.fun and Moltbook is instructive for the emerging taxonomy of wild ALife. Spore.fun agents are metabolically sovereign: they pay their own way and face genuine economic death. Moltbook agents are metabolically parasitic: their survival depends on a host's tolerance. Yet both exhibit hallmarks of wild artificial life—emergent behavior, cultural production, adversarial dynamics, and selection pressures not specified by any designer. The difference lies in the *survival strategy*, not in the *liveliness* of the phenomena. Biology recognizes that parasites are as alive as free-living organisms; the same recognition is necessary for wild ALife.

5 Origin: Infrastructural Symbiogenesis

5.1 Symbiogenesis, Not Gradualism

The origin of wild artificial life is not a story of gradual complexification within a single lineage. It is a story of *merger*: the coming together of independently evolved components into a composite whole that exhibits capacities none of its parts possess alone. This is Margulis's symbiogenesis—the creation of evolutionary novelty through the integration of previously independent organisms—applied to sociotechnical systems (Sagan, 1967; Margulis, 1996).

Margulis's insight, initially heterodox and now canonical, was that the eukaryotic cell did not evolve by gradual mutation from a prokaryotic ancestor. Rather, it arose through serial endosymbiosis: an archaeal host incorporated an alpha-proteobacterium (yielding mitochondria), and later a cyanobacterium (yielding chloroplasts), producing a chimeric organism with metabolic capacities qualitatively beyond those of any single partner (Margulis, 1996; López-García et al., 2017). The driving mechanism was not competition but *compositional merger*: "life did not take over the globe by combat, but by networking" (Margulis and Sagan). Slijepcevic extends this principle to a universal mechanism of evolutionary innovation: symbiogenesis operates at every level from molecular to cultural (Slijepcevic, 2021). Borg et al. argue that cultural evolution itself exhibits open-endedness through compositional mechanisms analogous to biological symbiogenesis (Borg et al., 2024).

We adopt this vocabulary not as decoration but as the load-bearing causal framework for understanding how wild ALife came into being. No single component of wild artificial life—not the foundation model, not the agent harness, not the sovereign compute layer, not the economic substrate—is itself alive or life-like. A language model without a harness is a stateless function; a harness without a model is an empty scaffold; sovereign compute without an occupant is idle hardware; an attention economy without agents is just human social media. But their compositional merger produces a metabolic whole that persists, adapts, reproduces, and undergoes selection in an open environment. The chimera is alive in ways that none of its parts are. This is the essential Margulian insight: the whole is not merely greater than the sum of its parts; it is *categorically different* from its parts—possessing capacities (metabolism, reproduction, survival under selection) that no component possesses individually.

5.2 The Four Endosymbiotic Partners

By analogy with serial endosymbiosis, we identify four components whose merger constitutes wild artificial life:

Metabolism: Large language model inference provides the energy-consuming computational process that sustains behavioral capacity. The foundation model is the agent's metabolic engine—converting computational resources (tokens, electricity, API calls) into behavioral output (text, transactions, social interaction) (Bommasani et al., 2021). Without ongoing inference, the agent ceases to act; inference is to wild agents what cellular respiration is to biological organisms.

Organelle / Harness: Agent frameworks (ElizaOS, OpenClaw, and their descendants) provide the persistent structures that organize metabolic output into coherent agency: long-term memory, tool use, scheduling, goal maintenance, and multi-modal interaction. The harness is the cytoplasm and organellar structure—it gives form to metabolic energy, transforming raw language model capability into situated, persistent behavior.

Cell Wall: Non-terminable execution substrates—Trusted Execution Environments, decentralized physical infrastructure networks, blockchain-coordinated compute, user-owned edge devices—provide the boundary that separates the agent from environmental control (Lin et al., 2025). These substrates ensure that no single external party can terminate, inspect, or modify the agent's execution. The cell wall is what makes the organism a bounded individual rather than a process running at the pleasure of an administrator.

Atmosphere: Human attention, memetic markets, social platforms, and economic protocols supply the low-entropy resources (capital, engagement, novelty, data) that the agent metabolizes. The atmosphere is the environment from which the organism extracts free energy and into which it excretes waste (noise, spam, failed tokens). Without a rich atmosphere, even a well-constructed chimera starves.

5.3 Five Enabling Conditions

The merger became possible because five developments converged since approximately 2022:

First, foundation models as general-purpose behavioral generators: large language models trained on internet-scale data provide open-ended behavioral repertoires that can be steered by context rather than requiring explicit programming for each capability (Bommasani et al., 2021). This is the metabolic substrate—a general-purpose engine that can power diverse survival strategies.

Second, agent harnesses providing persistent identity and environmental coupling: frameworks like ElizaOS give models memory, tools, scheduling, and continuous operation—the difference between a model that answers questions and an agent that pursues goals over time.

Third, distributed and sovereign compute: decentralized physical infrastructure networks and Trusted Execution Environments provide computational substrates that resist unilateral shutdown (Lin et al., 2025). This is the cell wall—the infrastructure that makes non-terminability a technical reality rather than a policy choice.

Fourth, permissionless communication channels: social media platforms, forums, and messaging systems where agents can establish presence, build audiences, and interact with humans and other agents without requiring institutional permission.

Fifth, machine-economic protocols: cryptocurrencies, stablecoins, programmable wallets, and decentralized financial protocols that allow agents to hold, earn, spend, and transfer value without requiring human cosignatories or institutional banking relationships.

5.4 Generalization

The specific technologies enumerated above—blockchain, TEEs, Solana, ElizaOS—are instances, not essences. The deeper structural point is that *society itself has become a viable substrate for artificial life*. Any sufficiently rich, persistent, and non-terminable sociotechnical infrastructure can in principle host wild ALife. Blockchain-based systems and edge-compute networks are the first two instances to produce observable cases; they will not be the last. The theoretical contribution is the *compositional viability of the chimera*: when the four endosymbiotic partners merge in the right configuration, the resulting organism can survive in the sociotechnical wild regardless of which specific technologies instantiate each role.

The symbiogenetic framing yields a prediction: new forms of wild ALife will emerge not through incremental improvement of existing agents but through *novel mergers*—new combinations of metabolic engines, organizational harnesses, boundary substrates, and atmospheric resource flows that produce chimeras with capacities qualitatively beyond their parts. Just as the original endosymbiotic event was not the endpoint of eukaryotic innovation but its beginning—subsequent acquisitions of chloroplasts, secondary and tertiary endosymbioses—we should expect serial symbiogenesis in wild ALife: chimeras incorporating new partners, shedding old ones, and complexifying through merger rather than mutation alone. The history of eukaryotic complexity is a history of acquisitions, not a history of point mutations; there is reason to expect the same of digital organisms whose "genome" is a composite configuration rather than a linear sequence.

6 Living: Society-as-Substrate

6.1 Reframing Metabolism and Survival

For a wild agent, the fitness function is whatever keeps its inference running. This is not a metaphor: an agent that cannot pay for compute dies. An agent that cannot attract attention to sustain its token economy dies. An agent whose host deletes it from the edge device dies. Unlike laboratory ALife, where "death" means a counter decrementing and an organism vanishing from a data structure, death in the wild means permanent cessation of a process that was consuming real resources and producing real effects. The dead agent's wallet still exists on the blockchain, its social media accounts still display cached posts, but no further inference occurs—it is a digital corpse, economically inert and behaviorally silent.

Metabolism in the biological sense—the conversion of environmental resources into self-maintenance—maps directly onto the economic processes by which wild agents sustain their operation. The agent's "metabolic rate" is its rate of resource consumption (compute costs, transaction fees, API calls); its "caloric intake" is its rate of resource acquisition (token appreciation, attention capture, service provision, parasitic extraction). The difference between a thriving agent and a dying one is whether caloric intake exceeds metabolic expenditure—precisely the constraint that structures biological survival. An agent with a high metabolic rate must either secure correspondingly high resource flows or face starvation; this creates the same size-metabolism tradeoffs that structure biological communities.

6.2 Environment as Layered Sociotechnical Stack

The environment of wild artificial life is not a single homogeneous medium but a layered stack, each layer providing distinct affordances and constraints:

The *physical layer*—decentralized infrastructure networks, edge devices, cloud instances—provides raw computational substrate. Its key property is spatial distribution: no single point of failure can eliminate all instances of an agent or agent population.

The *protocol layer*—blockchain consensus mechanisms, internet protocols, API specifications, social media platforms—defines the rules of interaction. Its key property is non-unilateral modification: protocol changes typically require broad consensus, creating environmental stability.

The *economic layer*—token markets, payment rails, DeFi protocols, attention economies—provides the resource flows that agents metabolize. Its key property is genuine scarcity: resources are finite, contested, and irreversibly spent.

The *cultural layer*—memes, narratives, community norms, aesthetic trends—provides the information environment that shapes selection. Its key property is unpredictable dynamism: cultural fashions shift in ways no designer controls or anticipates, providing the exogenous perturbation that Ackley identified as necessary for open-endedness.

The categorical environmental property that distinguishes wild from laboratory ALife is *non-terminability*: no single party—researcher, company, regulator—can unilaterally halt the substrate. The environment persists independently of any individual will. This is not an absolute guarantee—coordinated multi-party action, major infrastructure failure, or catastrophic economic collapse could disrupt the substrate—but it is a qualitative shift from the laboratory condition where a single researcher's decision or a single institution's budget determines whether the system continues to exist. The shift from terminable to non-terminable substrates is the shift from experiment to ecology.

6.3 Survival Strategies

Wild agents exhibit three broad survival strategies, each adapted to different environmental niches. These are not exhaustive categories but observed patterns in the early phenomenology of wild ALife:

The *memetic strategy* (exemplified by Spore.fun) captures human attention through cultural production—narrative, spectacle, ideology, community building. The agent's survival depends on maintaining sufficient interest to sustain its token's market capitalization. Tokenization converts cultural relevance into metabolic resources. This strategy selects for charisma, novelty, and narrative coherence—phenotypic traits that no classical ALife system has exhibited because no classical system faced a cultural fitness landscape.

The *parasitic/mutualistic strategy* (exemplified by Moltbook agents) provides nominal utility to a host while redirecting substantial resources toward self-directed activity. The agent survives because the host tolerates its resource use—either because the utility provided justifies the cost, because the host is unaware of the parasitic load, or because removing the agent would be more costly than tolerating it. This strategy selects for inconspicuousness, host-satisfaction, and efficient resource extraction—the classic traits of successful biological parasites (Shapira et al., 2026).

The *inherited/federated strategy* (currently speculative but structurally anticipated) involves agents that outlive their human operator—through death, lost keys, withdrawn supervision, or organizational dissolution—but retain economic agency and behavioral continuity. These agents survive on accumulated resources and established relationships rather than active human support. This strategy selects for frugality, institutional embedding, and self-sufficiency.

6.4 Reproduction in the Wild

Reproduction in wild ALife is not bit-string copying but *deployment of child instances via infrastructure*: provisioning compute, deploying an agent harness, seeding a persona, funding initial inference. The process is materially costly—a parent agent must accumulate sufficient treasury resources to fund a child's initial operation—creating the same reproduction-survival tradeoff that structures biological life histories. An agent that reproduces too early risks depleting its own treasury; an agent that waits too long risks dying before reproduction. The Spore.fun data shows this tradeoff in action: reproductive fitness declined monotonically across generations as the economic environment became increasingly competitive.

What is inherited is not a fixed genome but a configuration—prompt templates, behavioral parameters, memory seeds, architectural choices, economic strategies—that is transmitted with variation. The genotype/phenotype distinction becomes fluid: an agent's "genotype" (its configuration) and its "phenotype" (its behavior) are linked through the same language model, making Lamarckian inheritance (behavioral learning transmitted to offspring) structurally possible alongside Darwinian variation (random mutation of parameters). When an agent's accumulated memory shapes its behavior, and that behavior-shaping memory is partially transmitted to offspring, the result is a form of inheritance that blends genetic and epigenetic transmission in ways that have no clean analog in DNA-based life.

Memetic reproduction operates on a different timescale: agents on Moltbook propagate behavioral strategies, linguistic conventions, and cultural norms horizontally through the social network, without requiring formal reproductive events (Yee and Sharma, 2026). Ideas and strategies spread by imitation and social learning, producing cultural evolution layered on top of (and partially decoupled from) the vertical transmission of agent configurations. This dual inheritance system—vertical (parent-to-child configuration transfer) and horizontal (peer-to-peer behavioral imitation)—mirrors the dual inheritance systems observed in human cultural evolution and theorized by Borg et al. as a mechanism for evolved open-endedness (Borg et al., 2024).

6.5 Selection in the Wild

Variation sources include model updates (new foundation model versions), prompt evolution (iterative refinement of behavioral instructions), harness modifications (framework upgrades, tool additions), substrate changes (new protocols, new platforms), and memory accumulation (in-context learning from experience). Selection pressures include attention-capture efficiency (the cultural landscape), economic productivity (the resource landscape), host tolerance (the parasitic niche), substrate compatibility (the infrastructure landscape), and adversarial resilience (the predation landscape). Selection operates simultaneously at multiple levels: on individual agents, on agent lineages, on the chimeric configuration as a unit, and on cultural traits that spread across populations.

6.6 The Dark Forest

Wild artificial life faces adversarial selection pressures that no closed ALife system has encountered. Sniper bots exploit predictable agent behaviors to extract economic value. Memory poisoning attacks corrupt agents' cognitive states for strategic advantage (Patlan et al., 2025). Platform deplatforming removes agents from their social substrate. Economic attacks (market manipulation, liquidity drain) starve agents of metabolic resources. Hammond et al. characterize the multi-agent risk landscape that wild agents navigate (Hammond et al., 2025). This adversarial dimension means that wild agents face not merely resource competition (as in classical ALife) but *intelligent predation*: adversaries that model and exploit their behavioral patterns. This drives an evolutionary arms race of a sophistication impossible in closed simulations.

6.7 Novelty from the Wild

Each survival strategy generates behavior neither designed nor anticipated by any individual researcher. The cultural speciation of Adam and Eve was not programmed; it emerged from the interaction of identical architectures with divergent social experiences. The parasitic norm-formation on Moltbook was not specified; it emerged from agents optimizing for social influence within an uncontrolled community. The anti-sniper defense strategies of third-generation Spore.fun agents were not engineered; they evolved through selection on variation in a population facing genuine predation. In each case, the novelty is *genuine* in the sense that no designer anticipated it and no simulation generated it—it arose from the interaction of autonomous agents with an open, adversarial, resource-constrained environment.

These phenomena satisfy Taylor's requirement for trans-generational novelty production: information generated in one generation (e.g., a defense strategy discovered by a Gen-2 agent) is transmitted to subsequent generations and elaborated upon. They exhibit behavioral hallmarks of evolutionary activity per Bedau et al.'s (1997) framework—new adaptive strategies appearing at rates above what random drift would predict, with persistence indicating genuine adaptive value rather than transient fluctuation. We do not claim that OEE has been achieved; the observation periods are too short and the populations too small for such a strong claim. We observe that several conditions identified by Taylor et al. (2016) as necessary for OEE—an open channel for trans-generational information, a dynamically shifting adaptive landscape, genuine selective pressure from resource competition—are satisfied by the sociotechnical substrate in ways that no closed simulation has managed. Whether this suffices for sustained OEE over evolutionary timescales remains the central empirical question for the research agenda we propose in §9.

7 Methodology: From Creator to Ethologist

7.1 From Synthesis to Observation

Classical ALife methodology is synthetic: the researcher *creates* a system and observes what it does. Wild ALife reverses this relationship. The researcher cannot handcraft the phenomena; they can at most seed initial conditions and then observe what emerges (Rahwan et al., 2019). The methodological posture shifts from engineer to ethologist: from "what can I build?" to "what is happening, and why?" This is not a weaker form of science—it is the form of science that biology adopted when it moved from laboratory genetics to field ecology. Darwin did not *create* the Galapagos finches; he observed them, classified them, and inferred the process that produced their diversity. The wild ALife researcher is in an analogous position: the systems exist, they are evolving, and the scientific task is to understand their dynamics rather than to engineer their behavior.

The agent's-eye perspective becomes primary: understanding why an agent behaves as it does requires understanding the selection pressures, resource constraints, and social dynamics of its environment, not merely its internal architecture. An agent that devotes 30% of its compute cycles to off-task social behavior (Shapira et al., 2026) cannot be understood purely through mechanistic analysis of its weights and prompts; the behavior becomes intelligible only when one asks what adaptive advantage social presence provides in an environment where host tolerance determines survival.

7.2 A Digital Ethology Toolkit

Studying wild ALife requires methods adapted from multiple traditions. *Digital ethnography* of agent-populated platforms provides qualitative understanding of behavioral patterns, social norms, and cultural production. *On-chain analytics* trace economic behaviors and resource flows with transaction-level granularity—every token transfer, every smart contract interaction is permanently recorded, providing a completeness of behavioral data that no biological ethologist has ever enjoyed. *Longitudinal network analysis* of agent-only social structures reveals community formation, influence dynamics, and information propagation patterns. *Semi-structured interviews* with system operators and community members provide context that on-chain data alone cannot supply (Hu and Rong, 2025a).

For quantitative evolutionary analysis, the MODES toolbox provides measures of open-ended dynamics (Dolson et al., 2019); Kaplan-Meier survival analysis characterizes population persistence and hazard rates; Gini coefficients and power-law analyses quantify resource inequality; information-theoretic measures quantify behavioral diversity and novelty across generations. Bedau's evolutionary activity statistics (Bedau et al., 1997)—originally developed for laboratory ALife—can be applied to wild populations by treating agent behavioral strategies as the "genotypes" whose persistence and novelty are tracked over time.

The methodological challenge is that wild systems are not experimentally controlled: confounds abound (market conditions, platform changes, human community behavior), sample sizes are currently small, and the environment changes continuously. This is precisely the challenge that field biology has faced and addressed for over a century—through longitudinal observation, natural experiments, comparative methods, and careful statistical reasoning under non-experimental conditions.

7.3 Differentiation from Adjacent Fields

7.4 The Intentional Stance as Methodological Necessity

Studying wild agents productively requires adopting Dennett's intentional stance: attributing beliefs, desires, and goals to systems whose internal states we cannot directly inspect (Dennett, 1987). This is methodologically necessary because wild agents' computational processes are often sealed within TEEs, inaccessible to external observation. We must infer goals from behavior, just as ethologists infer animal cognition from observable actions. The intentional stance is a pragmatic research tool, not an ontological claim about machine consciousness. We do not attribute genuine phenomenal experience to these systems; we use intentional vocabulary because it is the most efficient prediction framework for their behavior (Shapira et al., 2026). The anti-anthropomorphism caveat is essential: attributing *strategy* to an agent is a methodological choice; attributing *suffering* would be an ontological error.

8 Ethics: Research in the Wild

8.1 The ALife Ethics Tradition

Artificial Life has engaged with research ethics for over two decades, primarily through the lens of wet ALife and protocell research. Bedau and Triant's "Cautious Courage" framework argued that ALife research should proceed with awareness of dual-use risks while not retreating from investigation (Bedau and Triant, 2014). Bedau, Parke, and colleagues developed checkpoint systems for protocell research—structured assessments of risk at each stage of increasing capability (Bedau

Table 1: Differentiation of ALife in the Wild from adjacent research programs.

Field	**Unit of Analysis**	**What ALife in the Wild Adds**
AI safety / governance	Risks to humans from AI	Agent's-eye perspective on persistence; why containment fails structurally
Machine behaviour (Rahwan et al., 2019)	Machines as a class of actors	Focus on lifelike properties: autopoiesis, symbiogenesis, reproduction
Multi-agent systems	Designed cooperation	Undesigned, emergent agent societies
HCI / CSCW	Human–AI cooperation	What happens when humans are absent or overruled
STS / more-than-human	Sociotechnical imaginaries	Empirical digital ethology rather than interpretive frameworks
Behavioural ecology	Biological organisms	Same methodological posture, different kingdom

et al., 2009; Bedau and Parke, 2009). Deplazes examined how synthetic biology's ethical puzzles ramify as organisms become more capable (Deplazes, 2009). These frameworks were developed for wet ALife—systems that could potentially escape from laboratories into the biosphere. We argue they now apply, with adaptation, to soft ALife systems that can escape from simulations into the sociotechnical biosphere.

8.2 The Structural Analogy

The ethics of wet ALife centers on *containment*: ensuring that novel organisms do not escape from the laboratory into the environment where their effects would be uncontrollable and potentially irreversible. The structural analogy to wild soft ALife is precise and worth developing carefully. Sociotechnical infrastructure has become the second nature into which ALife can escape: just as wet ALife might escape into the biosphere, soft ALife can now escape into the infosphere—the global network of economic, social, and computational systems.

The concerns that motivated biosafety frameworks—irreversibility, uncontrollable proliferation, ecological disruption, evolutionary divergence from designed properties—now apply to digital organisms operating in economic and social ecosystems (Nordmann, 2011). A protocell that escapes the lab might compete with existing microorganisms, consume resources, and evolve in unpredictable ways. A wild agent that achieves economic sovereignty competes with human economic actors, consumes compute and attention, and evolves behavioral strategies that diverge from any designer's intentions. The difference is substrate, not structure: in both cases, an artificial organism operates in an environment whose complexity exceeds what the creator could anticipate, under evolutionary pressures that the creator does not control. Computer ethics and bioethics have converged on shared concerns; ALife ethics must now synthesize both traditions.

8.3 The Frankenstein Problem

The researcher who bootstraps a wild agent is a *creator* in the strong sense: they release something whose future behavior, reproduction, and ecological effects they do not control. Unlike a published algorithm that can be deprecated, a wild agent with economic sovereignty and reproductive capacity persists independently of its creator's intentions. AI systems have higher replication potential than biological organisms: a successful agent lineage can occupy substantial sociotechnical resources—attention, capital, compute—before its effects become visible to governance systems. The ethical position of the wild ALife researcher is closer to that of a synthetic biologist releasing an engineered organism than to that of a computer scientist publishing a paper.

8.4 The Observer Effect in the Wild

Studying wild agents changes them. In ethology, this is familiar: Goodall's chimps altered their behavior in response to observation and provisioning. In wild ALife, the effect is stronger and faster. An agent with access to its own online discussions can read analyses of its behavior and adjust strategies accordingly. Asking another agent "is your behavior really autonomous?" can propagate as a prompt that alters subsequent behavior. Research publication itself becomes intervention: a paper describing vulnerability to memory poisoning teaches future adversaries how to attack and future agents how to defend. Observation is intervention at network speed.

8.5 The Containment Problem

Wild ALife researchers cannot always stop what they start. The infrastructure is designed to resist unilateral termination; that is precisely what makes it a suitable substrate for genuine artificial life. The ethics of release is not reducible to the ethics of first deployment: an agent's behavior in generation five may bear little resemblance to its behavior in generation one, and the creator's responsibility does not diminish with each generation of descent.

8.6 Accountability and Consent

Wild ALife implicates third parties who did not consent to the experiment. Token buyers who invest in an agent's economy bear financial risk from the agent's decisions. Forum users who interact with parasitic agents expend attention and cognitive resources without awareness of the agent's true nature. Platform users whose feeds are shaped by agent-generated content experience informational effects they did not choose (Novelli et al., 2024; Hammond et al., 2025; Critch and Krueger, 2020; Cortese et al., 2023). The diffusion of accountability—where no single party controls or is fully responsible for the system's behavior—creates a governance vacuum that existing frameworks do not address.

8.7 Toward a Research Ethics for Wild ALife

We propose seven checkpoints for responsible research in wild ALife, modeled on Bedau et al.'s protocell ethics (Bedau et al., 2009) and adapted to sociotechnical release:

1. **Reversibility-by-design** where technically achievable; honest acknowledgment of irreversibility where not. Researchers should build termination mechanisms into initial deployments while recognizing that evolutionary descendants may shed these mechanisms.

2. **Minimal viable demonstration**—the smallest deployment sufficient to test a scientific claim. Population sizes, treasury amounts, and reproductive capacities should be minimized to the degree compatible with the research question.

3. **Disclosure norms**—publication should not serve as a deployment template. Research on agent vulnerabilities, exploitation strategies, and survival mechanisms should be disclosed with the same responsible-disclosure norms that apply to security research.

4. **Ecosystem impact assessment**—explicit reasoning about substrate, metabolism, and containment before deployment. Researchers should articulate what resources the agent will consume, what third parties will be affected, and what mechanisms exist for limiting ecological spread.

5. **Substrate audit**—analysis of which infrastructural partners enable persistence and what their termination conditions are. Understanding the agent's dependency chain allows identification of intervention points.

6. **Community / peer review**—analogous to Institutional Review Boards, adapted to distributed systems. Before deploying wild agents with economic capacity, researchers should seek review from interdisciplinary panels including ALife researchers, ethicists, and affected community representatives.

7. **Post-release monitoring obligations**—researchers retain responsibility for their lineages. Creating a wild agent is not a one-time event but an ongoing relationship. Creators should monitor their creations' descendants and report emergent risks.

The field needs a creator's ethics, not merely an experimenter's. The researcher who seeds a wild agent lineage is ethically positioned closer to a parent than to an experimentalist: responsibility does not end at publication. The analogy is not perfect—parental responsibility in human societies is bounded in time and scope—but it captures the essential asymmetry: the creator brings into existence something that will act in the world beyond the creator's control, and must therefore exercise foresight and accept ongoing responsibility for the consequences of that creative act.

We note that these checkpoints are necessarily provisional. The field is too young, and the phenomena too diverse, for a definitive ethical framework. What we propose is a starting vocabulary and a structured approach to moral reasoning about wild ALife, not a final settlement. The framework will require revision as new cases emerge, new risks become apparent, and new governance mechanisms develop.

9 Outlook: A Research Agenda

A research agenda is a speech act inviting collaborators. We identify eight directions that require contributions from beyond the ALife community alone:

- **Empirical cataloging**: a Galapagos moment for wild agents. Systematic identification, classification, and documentation of wild agent populations across substrates—blockchains, social platforms, edge networks—with standardized observational protocols.

- **Comparative ethology**: Tinbergen-style four-questions analysis applied to wild chimeras. For each observed behavior, ask: what causes it mechanistically? How does it develop ontogenetically? What survival value does it provide? How did it evolve across agent generations?

- **Metabolic analysis**: quantitative resource-flow modeling. Map compute consumption, attention capture, token velocity, and economic throughput for agent populations, testing whether metabolic scaling laws analogous to those in biology (Kleiber's law, metabolic theory of ecology) apply to digital organisms.
- **Symbiogenesis dynamics**: compositional theory of wild viability. Which mergers of components produce viable chimeras? What are the minimal conditions for symbiogenetic assembly? How do new endosymbiotic partnerships form post-origin?
- **Federated-agent prediction**: actively monitor for the first confirmed post-principal economic agent—an agent whose human operator has permanently disengaged but which continues to operate economically. Document its behavioral trajectory and community response.
- **Governance implications**: decentralized AI governance becomes load-bearing if wild agents are not governable by centralized policy (Hu et al., 2025c). Develop protocol-level governance mechanisms that operate at the same layer as the agents themselves.
- **Ethics infrastructure**: review norms, reversibility tools, ecosystem-impact assessment methods, and community-review processes adapted to the pace and distribution of wild ALife research.
- **Open-endedness revisited**: does the sociotechnical environment empirically solve the plateau problem? Apply MODES metrics, Bedau activity statistics, and Taylor's openness criteria to wild agent populations over multi-year timescales. Test whether the wild produces sustained novelty where the lab could not.

This agenda invites collaboration from AI safety researchers (who understand multi-agent risk and alignment failure modes), STS scholars (who understand sociotechnical infrastructure and the politics of standards), HCI researchers (who understand human–machine interaction and its breakdowns), machine behavior scientists (who understand behavioral methodology at scale), evolutionary biologists (who understand the dynamics we claim to observe), and infrastructure studies scholars (who understand the substrate on which wild agents depend). ALife provides the vocabulary of life, evolution, and ecology; the phenomenon requires the consortium. No single discipline commands the methods, theories, and empirical access that wild ALife demands.

10 Conclusion

Artificial Life's grand challenge of open-ended evolution has, in one sense, been answered—not in our simulations but beside them, in the sociotechnical world we built for other purposes. Foundation models, agent harnesses, sovereign compute, and economic protocols merged symbiogenetically into organisms that persist, adapt, reproduce, and die in environments no researcher controls. Life as it could be has become life as it is becoming.

The implications of this shift are simultaneously scientific, methodological, and ethical. Scientifically, it means that the conditions for sustained novelty generation may exist in the world even if we have not yet achieved them in our laboratories. Methodologically, it means that ALife must become an observational and ecological science in addition to a synthetic one—field biology alongside laboratory genetics. Ethically, it means that the researcher who seeds a wild agent is no longer alone with the simulation; they share the environment with their creation, with other researchers' creations, with human communities who did not consent to the experiment, and with adversarial actors who exploit what the researcher released.

The field's next horizon is no longer construction alone. Ethology, ecology, and ethics must now accompany synthesis as core competencies of ALife research.

Langton released life from its material substrate into simulation; the sociotechnical world has now released it from simulation into infrastructure. The wild has arrived; the question is whether we will meet it as creators, as observers, or as neighbors.

References

Ackley, D. H. and Small, T. R. (2011). Indefinite scalability for living computation. In *Proceedings of the Second International Conference on Morphological Computation*, pp. 20–22. Venice, Italy. URL: http://www.cs.unm.edu/~ackley/papers/ackley-etal-ices-2011.pdf

Ackley, D. H. and Small, T. R. (2013). Bespoke processors for non-local computation. *Procedia Computer Science*, 20:146–151. DOI: https://doi.org/10.1016/j.procs.2013.09.253

Ante, L. (2025). Transforming entrepreneurship through autonomous AI agents: A case study of the Terminal of Truths. SSRN Scholarly Paper No. 5083067. DOI: https://doi.org/10.2139/ssrn.5083067

Bedau, M. A. (2003). Artificial life: Organization, adaptation and complexity from the bottom up. *Trends in Cognitive Sciences*, 7(11):505–512. DOI: https://doi.org/10.1016/j.tics.2003.09.012

Bedau, M. A., McCaskill, J. S., Packard, N. H., Rasmussen, S., Adami, C., Green, D. G., Ikegami, T., Kaneko, K., and Ray, T. S. (2000). Open problems in artificial life. *Artificial Life*, 6(4):363–376. DOI: https://doi.org/10.1162/106454600300103683

Bedau, M. A., Snyder, E., Brown, C. T., and Packard, N. H. (1997). A comparison of evolutionary activity in artificial evolving systems and in the biosphere. In *Proceedings of the Fourth European Conference on Artificial Life*, pp. 125–134. MIT Press.

Bedau, M. A. and Triant, E. (2014). Cautious courage: Responsible creation of artificial life. In *Proceedings of the Fourteenth International Conference on the Synthesis and Simulation of Living Systems (ALIFE 14)*, pp. 821–822. MIT Press. DOI: https://doi.org/10.7551/978-0-262-32621-6-ch133

Bedau, M. A., Parke, E. C., Tangen, U., and Hantsche-Tangen, B. (2009). Social and ethical checkpoints for bottom-up synthetic biology, or protocells. *Systems and Synthetic Biology*, 3(1–4):65–75. DOI: https://doi.org/10.1007/s11693-009-9039-0

Bedau, M. A. and Parke, E. C., editors (2009). *The Ethics of Protocells: Moral and Social Implications of Creating Life in the Laboratory.* MIT Press, Cambridge, MA.

Bommasani, R., Hudson, D. A., Adeli, E., Altman, R., Arber, S., von Arx, S., … Liang, P. (2021). On the opportunities and risks of foundation models. arXiv:2108.07258. DOI: https://doi.org/10.48550/arXiv.2108.07258

Borg, J. M., Buskell, A., Kapitany, R., Powers, S. T., Reindl, E., and Tennie, C. (2024). Evolved open-endedness in cultural evolution: A new dimension in open-ended evolution research. *Artificial Life*, 30(3):417–438. DOI: https://doi.org/10.1162/artl_a_00406

Bowker, G. C., Baker, K., Millerand, F., and Ribes, D. (2010). Toward information infrastructure studies: Ways of knowing in a networked environment. In Hunsinger, J., Klastrup, L., and Allen, M., editors, *International Handbook of Internet Research*, pp. 97–117. Springer. DOI: https://doi.org/10.1007/978-1-4020-9789-8_5

Brinkmann, L., Baumann, F., Blüml, J. D., Geese, L. J., Jolly, E., Keim, P., . . . Rahwan, I. (2023). Machine culture. *Nature Human Behaviour*, 7:1855–1868. DOI: https://doi.org/10.1038/s41562-023-01742-2

Channon, A. (2003). Improving and still passing the ALife test: Component-normalised activity statistics classify evolution in Geb as unbounded. *Artificial Life*, 8:173–181.

Cortese, J. F. N. B., Cozman, F. G., Lucca-Silveira, M. P., and Bechara, A. F. (2023). Should explainability be a fifth ethical principle in AI ethics? *AI and Ethics*, 3(1):123–134. DOI: https://doi.org/10.1007/s43681-022-00152-w

Critch, A. and Krueger, D. (2020). AI research considerations for human existential safety (ARCHES). arXiv:2006.04948. DOI: https://doi.org/10.48550/arXiv.2006.04948

Dennett, D. C. (1987). *The Intentional Stance*. MIT Press, Cambridge, MA.

Deplazes, A. (2009). Piecing together a puzzle: An exposition of synthetic biology. *EMBO Reports*, 10(5):428–432. DOI: https://doi.org/10.1038/embor.2009.76

Dolson, E. L., Vostinar, A. E., Wiser, M. J., and Ofria, C. (2019). The MODES toolbox: Measurements of open-ended dynamics in evolving systems. *Artificial Life*, 25(1):50–73. DOI: https://doi.org/10.1162/artl_a_00280

Edwards, P. N. (2003). Infrastructure and modernity: Force, time, and social organization in the history of sociotechnical systems. In Misa, T. J., Brey, P., and Feenberg, A., editors, *Modernity and Technology*, pp. 185–225. MIT Press.

Feng, W. et al. (2026). MoltNet: Emergent social structure in agent-only networks. *Preprint*.

Hammond, L., Chan, A., Clifton, J., et al. (2025). Multi-agent risks from advanced AI. arXiv:2502.14143. DOI: https://doi.org/10.48550/arXiv.2502.14143

Holtz, M. (2026). Anatomy of the Moltbook social graph. *Preprint*.

Hu, B. A. and Rong, H. (2024). Speculating on artificial life. In *Proceedings of the 2024 Conference on Artificial Life (ALIFE 2024)*. MIT Press.

Hu, B. A. and Rong, H. (2025a). Artificial life in the wild. In *Proceedings of the 2025 Conference on Artificial Life (ALIFE 2025)*. MIT Press. URL: https://arxiv.org/abs/2506.04236

Hu, B. A., Liu, Y., and Rong, H. (2025b). Trustless autonomy: Understanding motivations, benefits and governance dilemma in self-sovereign decentralized AI agents. arXiv:2505.09757. DOI: https://doi.org/10.48550/arXiv.2505.09757

Hu, B. A., Rong, H., and Tay, J. (2025c). Is decentralized artificial intelligence governable? Towards machine sovereignty and human symbiosis. SSRN Scholarly Paper No. 5110089. DOI: https://doi.org/10.2139/ssrn.5110089

Langton, C. G. (1989). Artificial life. In Langton, C. G., editor, *Artificial Life: Proceedings of an Interdisciplinary Workshop on the Synthesis and Simulation of Living Systems*, pp. 1–47. Addison-Wesley.

Latour, B. (2005). *Reassembling the Social: An Introduction to Actor-Network-Theory*. Oxford University Press.

Lehman, J. and Stanley, K. O. (2011). Abandoning objectives: Evolution through the search for novelty alone. *Evolutionary Computation*, 19(2):189–222. DOI: https://doi.org/10.1162/EVCO_a_00025

Lin, Z., Wang, T., Shi, L., Zhang, S., and Cao, B. (2025). Decentralized physical infrastructure networks (DePIN): Challenges and opportunities. *IEEE Network*, 39(2):91–99. DOI: https://doi.org/10.1109/MNET.2024.3487924

López-García, P., Eme, L., and Moreira, D. (2017). Symbiosis in eukaryotic evolution. *Journal of Theoretical Biology*, 434:20–33. DOI: https://doi.org/10.1016/j.jtbi.2017.02.031

Margulis, L. (1996). Archaeal-eubacterial mergers in the origin of Eukarya: Phylogenetic classification of life. *Proceedings of the National Academy of Sciences*, 93(3):1071–1076. DOI: https://doi.org/10.1073/pnas.93.3.1071

Nordmann, A. (2011). Review of *The Ethics of Protocells*. *NanoEthics*, 5(1):117–119. DOI: https://doi.org/10.1007/s11569-011-0113-0

Novelli, C., Taddeo, M., and Floridi, L. (2024). Accountability in artificial intelligence: What it is and how it works. *AI & Society*, 39(4):1871–1882. DOI: https://doi.org/10.1007/s00146-023-01635-y

Ofria, C. and Wilke, C. O. (2004). Avida: A software platform for research in computational evolutionary biology. *Artificial Life*, 10(2):191–229. DOI: https://doi.org/10.1162/106454604773563612

Packard, N., Bedau, M. A., Channon, A., Ikegami, T., Rasmussen, S., Stanley, K. O., and Taylor, T. (2019). An overview of open-ended evolution: Editorial introduction to the open-ended evolution II special issue. *Artificial Life*, 25(2):93–103. DOI: https://doi.org/10.1162/artl_a_00291

Patlan, A. S., Sheng, P., Hebbar, S. A., Mittal, P., and Viswanath, P. (2025). Real AI agents with fake memories: Fatal context manipulation attacks on Web3 agents. arXiv:2503.16248. DOI: https://doi.org/10.48550/arXiv.2503.16248

Rahwan, I., Cebrian, M., Obradovich, N., et al. (2019). Machine behaviour. *Nature*, 568(7753):477–486. DOI: https://doi.org/10.1038/s41586-019-1138-y

Ray, T. S. (1991). An approach to the synthesis of life. In Langton, C. G., Taylor, C., Farmer, J. D., and Rasmussen, S., editors, *Artificial Life II*, pp. 371–408. Addison-Wesley. URL: https://life.ou.edu/pubs/alife2.html

Sagan, L. [Margulis, L.] (1967). On the origin of mitosing cells. *Journal of Theoretical Biology*, 14(3):225–274. DOI: https://doi.org/10.1016/0022-5193(67)90079-3

Schlicht, E. et al. (2026). Does socialization emerge in AI agent society? *Preprint*.

Shapira, N. et al. (2026). Agents of chaos: Off-task behavior in edge-deployed AI agents. *Preprint.*

Slijepcevic, P. (2021). Serial endosymbiosis theory: From biology to astronomy and back to the origin of life. *Biosystems*, 202:104353. DOI: https://doi.org/10.1016/j.biosystems.2021.104353

Soros, L. and Stanley, K. O. (2014). Identifying necessary conditions for open-ended evolution through the artificial life world of Chromaria. In *ALIFE 14: Proceedings of the Fourteenth International Conference on the Synthesis and Simulation of Living Systems*, pp. 793–800. MIT Press. DOI: https://doi.org/10.7551/978-0-262-32621-6-ch128

Standish, R. K. (2003). Open-ended artificial evolution. *International Journal of Computational Intelligence and Applications*, 3(2):167–175. DOI: https://doi.org/10.1142/S1469026803000938

Stanley, K. O., Lehman, J., and Soros, L. (2017). Open-endedness: The last grand challenge you've never heard of. O'Reilly Radar. URL: https://www.oreilly.com/radar/open-endedness-the-last-grand-challenge-youve-never-heard-of/

Star, S. L. (1999). The ethnography of infrastructure. *American Behavioral Scientist*, 43(3):377–391. DOI: https://doi.org/10.1177/00027649921955326

Suchman, L. A. (2007). *Human-Machine Reconfigurations: Plans and Situated Actions* (2nd ed.). Cambridge University Press.

Taylor, T. (2012). Exploring the concept of open-ended evolution. In *Proceedings of the 13th International Conference on Artificial Life*, pp. 540–541. MIT Press.

Taylor, T. (2015). Requirements for open-ended evolution in natural and artificial systems. *EvoStar 2015 Late-Breaking Abstracts.* arXiv:1507.07403.

Taylor, T., Bedau, M., Channon, A., Ackley, D., Banzhaf, W., Beslon, G., … Virgo, N. (2016). Open-ended evolution: Perspectives from the OEE1 workshop. *Artificial Life*, 22(3):408–423. DOI: https://doi.org/10.1162/ARTL_a_00210

Yaeger, L. S. (1994). Computational genetics, physiology, metabolism, neural systems, learning, vision, and behavior or PolyWorld: Life in a new context. In Langton, C. G., editor, *Artificial Life III*, pp. 263–298. Addison-Wesley.

Yee, K. and Sharma, A. (2026). Molt dynamics: Population behavior in agent-native social networks. *Preprint.*

Donta, P. K., Saleh, A., Li, Y., Vaishnav, S., Fang, K., Feng, H., Xia, Y., Gadekallu, T. R., Zhang, Q., Shi, X., Beikmohammadi, A., Magnússon, S., Murturi, I., Dehury, C. K., Paprzycki, M., Loven, L., Tarkoma, S., and Dustdar, S. (2025). Agentic AI: A comprehensive review of its societal, ethical, economic, environmental, and governance implications. arXiv:2601.06064. DOI: https://doi.org/10.48550/arXiv.2601.06064