AI and Ethics (2026) 6:268
https://doi.org/10.1007/s43681-026-01057-8

RESEARCH

Digital Darwinism: steering the evolution of artificial life in socio-technical systems

Karl T. Ulrich[1]

Received: 18 July 2025 / Accepted: 16 February 2026

Abstract
Public debate about artificial intelligence risk centers on hypothetical artificial general intelligence (AGI), but existing software systems are already evolving in ways that could undermine human oversight and institutional control. Cloud platforms, open-source software supply chains, and crypto-economic incentives provide, at electronic speed, the three preconditions of evolution: replication, variation, and differential fitness. This article uses an exploratory scenario method to trace near-term evolutionary trajectories for digital proto-life through three narratives: Lamarck (self-modifying coding agents), Remora (resource-seeking companion chatbots), and Mycelium (DAO-LLC trading bots). These scenarios show how autonomous software populations can amass computing budgets, shape emotional bonds, and acquire legal leverage without ever achieving general intelligence. Left unguided, such dynamics could drain computational resources, lock users into harmful dependencies, and infiltrate critical market infrastructure. The article therefore shifts the governance focus from aligning goals to steering evolution. It proposes four guidance instruments: replication-rate thresholds modeled on epidemiological R_0, a public vulnerability registry for self-modifying code, tiered digital biosafety levels, and adaptive regulatory sandboxes. Managing evolutionary dynamics in software is as urgent as AGI alignment for safeguarding society's co-evolution with its machines.

Keywords AGI · AI safety · AI risk · Digital evolution · Alife · Artificial life · Self-replicating software · Socio-technical governance · Autonomous agents · Regulatory foresight

1 Introduction

Public debate on artificial-intelligence risk still gravitates toward an imagined future in which a single artificial general intelligence eclipses human capability. Yet the digital environment we already inhabit contains software systems that replicate, vary, and persist or disappear under competitive pressure. Contemporary socio-technical infrastructure supplies everything evolution needs: massive digital replication channels, boundless variation generated by code-writing tools, and relentless selection driven by attention, bandwidth, and capital markets. In short, society is shaping its own algorithms, and those algorithms are reshaping society in ways that standard AI-safety framings overlook, with profound implications for social equity, democratic governance, and human agency.

Three recent vignettes make the point concrete.

Self-modifying crypto mining botnets. Malware families have been observed rewriting their embedded mining configurations (i.e., pool endpoint, algorithm parameters, and payout wallet) so that only the most lucrative variants persist. Operators rotate pool endpoints, wallet addresses, and algorithm parameters across campaign variants to maximize revenue [37], while separate proof-of-concept research has shown that malware can query a large language model at runtime to regenerate its payload polymorphically, evading endpoint detection [44]. Combining autonomous propagation with LLM-assisted code mutation would yield a system in which only the most lucrative variants persist, a prospect that is technically feasible even if not yet documented in the wild. These campaigns disproportionately target computing resources in regions with weaker cybersecurity infrastructure, creating an inequitable distribution of harm across the global digital landscape.

✉ Karl T. Ulrich
ulrich@upenn.edu

[1] The Wharton School, University of Pennsylvania, Philadelphia, USA

Published online: 27 April 2026

Predatory arbitrage bots in decentralized finance. On public blockchains, automated bots simulate every pending transaction and, when profitable, submit a competing copy with a higher fee to capture the value first [13]. When researchers attempted to rescue funds from a vulnerable smart contract, their transaction was instantly copied by such a bot [43]. Operators iteratively deploy new variants that refine gas-fee strategy and exchange routing, with only profitable configurations persisting, producing a competitive arms race shaped by selection on payoff [51].

Algorithmic content selection on short-form-video platforms. On platforms such as TikTok, recommendation algorithms amplify content aligned with user engagement signals, producing rapid reinforcement loops that steer collective attention toward whatever traits maximize retention [19]. Creators respond by iterating on successful formats, generating a feedback cycle in which platform selection pressures and human production co-evolve. These dynamics increasingly shape cultural discourse and youth socialization, often amplifying content optimized for engagement rather than social benefit.

None of these code populations carries a designer-imposed objective in the classical agent sense. Variants persist or disappear according to external fitness signals: payouts, click-throughs, uptime, or evasion of countermeasures. Those signals are set by social, legal, and economic structures, so strains that navigate human norms most effectively are the ones that proliferate. The outcome is digital proto-life that evolves at network speed, with success determined as much by institutional fit as by technical ingenuity, raising fundamental questions about power, agency, and the distribution of benefits in increasingly automated systems.

This article argues that evolutionary dynamics in existing digital systems may transform society long before any hypothetical AGI. Because digital mutations propagate instantly and selection pressures act continuously, these entities can reshape markets, media, and governance in months, not decades. Guiding their evolutionary trajectories is therefore becoming a prerequisite for safeguarding human welfare and ensuring these systems evolve in ways that promote rather than undermine societal values.

The remainder of the article proceeds as follows. After reviewing related scholarship, Sect. 2 outlines the methodological approach. Section 3 presents three scenario narratives—Lamarck, Remora, and Mycelium -- that illustrate concrete mechanisms. Section 4 analyzes how digital substrates accelerate replication, variation, and selection. Section 5 maps near-term societal risks, with particular attention to their uneven distribution across socioeconomic groups. Section 6 proposes governance strategies that steer selection pressures rather than micromanage individual systems. Section 7 concludes with a research and policy agenda that treats digital evolution, not AGI, as the near-term frontier for AI and society, highlighting the need for interdisciplinary approaches that address both technical and social dimensions of this challenge.

1.1 Related scholarship

Research on digital evolution has expanded rapidly since 2023 and now clusters around three strands.

Self-replicating and self-evolving agents. Zhou et al., [52] demonstrate how language-agent pipelines can rewrite their own prompt graphs and redeploy updated versions through symbolic learning. A survey by Tao et al., [46] catalogs more than sixty self-evolution techniques for large language models, identifying iterative cycles of data collection, refinement, and retraining as a common pattern. Pan et al., [36] go further, demonstrating that frontier AI systems driven by open-weight LLMs can already replicate themselves across hosts without human intervention.

Evolutionary dynamics in decentralized finance. Daian et al., [12] first drew attention to maximal-extractable-value (MEV) bots as adaptive actors in permissionless markets. Follow-up work traces how flash-loan attacks reshape incentives and liquidity distribution across DeFi protocols [39], while Qin et al., [38] extend the analysis to CeFi-DeFi comparisons. The broader regulatory challenge lies in designing governance frameworks that adjust protocol incentives rather than banning contracts outright [50].

Parasocial relationships with AI. Maeda and Quan-Haase [27] describe how design cues in chatbots trigger one-sided emotional bonds. A systematic review in *AI & Society* collates fifty-eight studies and flags rising concern about compulsive engagement when conversational AI uses empathic language and adaptive self-disclosure [40]. Survey evidence also links loneliness to rapid adoption of AI companions [14].

Together, these literatures show that digital entities capable of variation and selection already interact with socioeconomic structures, from block-production queues to affective user journeys, creating evolutionary pressures that traditional AI-safety models seldom capture.

1.2 Terminology and scope

This article makes frequent use of evolutionary vocabulary such as “digital organisms,” “digital proto-life,” “selection pressure,” and “fitness landscape,” to describe populations of software that replicate, vary, and persist or disappear under external pressures. Because such language risks implying that software systems are alive in the biological sense, or that they possess intentions, it is important to state clearly what is and what is not being claimed.

We do not claim that the systems discussed in this paper satisfy biological definitions of life. Criteria commonly held to distinguish living systems, including metabolism, genuine autonomy, open-ended heredity, and persistent self-maintenance, are not met by any software population described here. The replication-variation-selection triad that organizes our analysis is a necessary but not sufficient condition for biological life. We invoke it not to assert ontological equivalence with living organisms but because it identifies a set of dynamics (e.g., rapid propagation, feedback-driven adaptation, and emergent complexity) that carry governance implications poorly captured by agent-centric AI safety frameworks, which typically assume a discrete system with a fixed objective function.

In adopting this vocabulary we are, in [15] terms, taking an intentional stance: treating software populations as if they had strategies and goals because doing so generates useful predictions about their aggregate behavior. This is an analytical convenience, not a mechanistic claim. When we say a malware variant "competes" or an MEV bot "adapts," we mean that populations of such code exhibit differential persistence under measurable selection pressures, not that individual programs deliberate or desire. Readers should interpret evolutionary language throughout the paper in this spirit.

To guard against metaphorical overreach, we distinguish three levels of autonomy in digitally evolving systems:

Level 1: **Human-seeded adaptive systems.** A human designer creates the initial code and defines the variation mechanism (e.g., an LLM-assisted prompt-rewriting loop). Subsequent adaptation proceeds through automated variation and external selection, but the scaffolding is intentional. The Lamarck and Remora scenarios in Sect. 3 occupy this level.

Level 2: **Autonomously varying systems within bounded environments.** Code populations vary and are selected within a permissionless environment (e.g., a public blockchain) with no ongoing human direction of individual variants, though the environment itself is a human artifact. Flash-loan MEV swarms [39] approximate this level. The Mycelium scenario begins at Level 1 (human-seeded) but transitions toward Level 2 as its founders disengage and the network's master contract governs replication and selection without ongoing human direction.

Level 3: **Fully autonomous self-originating systems.** Software that spontaneously generates, replicates, and evolves without any human seeding or environmental scaffolding. This paper does not claim that Level 3 systems exist today. The scenarios and governance proposals address Levels 1 and 2 only.

This distinction matters for governance. Level 1 and Level 2 systems are already observable and already produce externalities (e.g., resource consumption, psychological dependency, regulatory evasion) that demand policy responses. Waiting for evidence of Level 3 autonomy before acting would repeat the error that the paper attributes to AGI-centric safety discourse: deferring governance until a hypothetical threshold is crossed while real harms accumulate.

The table below summarizes the operational proxies used throughout the paper for each component of the evolutionary triad, together with the limitations of each proxy.

These definitions and distinctions apply throughout the paper. Where biological analogies appear in later sections (for instance, the "digital biosafety levels" of Sect. 6.3 or the R_0-code standard of Sect. 6.1), they are functional analogies intended to leverage existing institutional knowledge, not claims of equivalence between software behavior and pathogen biology. Table 1.

2 Methodological approach

This study uses an exploratory scenario method drawn from strategic planning practice. Scenarios do not forecast a single most-likely future; instead, they map plausible pathways, highlight forces that drive change, and reveal where governance can fail or succeed [41]. Building on recent efforts to blend digital systems analysis with scenario planning, three narratives (Lamarck, Remora, and Mycelium) were developed through a four-step cycle:

1. Literature synthesis. Empirical findings on self-replicating code, MEV dynamics, and parasocial chatbots were collected.
2. Driver mapping. Replication, variation, and selection mechanisms most relevant to each domain were identified.
3. Storyline drafting. Interactions among those drivers over a five- to eight-year horizon were explored and refined.
4. Cross-impact checks. Drafts were compared with current policy debates, technology road maps, and market data to ensure internal consistency.

This scenario approach complements empirical and formal modeling by surfacing institutional and ethical questions that benchmark studies often miss, for example, who defines the fitness signals, who bears the external costs, and what built-in brakes, if any, prevent runaway evolution.

Each scenario was selected to stress-test a distinct dimension of the evolutionary framework by drawing on one of the three empirical strands identified in Sect. 1.1. Lamarck

Table 1 Operational proxies for evolutionary dynamics in digital systems

Concept	Operational proxy	Explicit limitation
Replication	Number of autonomous deployments, forks, or instantiations per unit time. Where appropriate, we use an analogical replication metric, R_0-*code*, defined as the average number of new active copies generated by one instance during its lifetime. This metric is inspired by the epidemiological basic reproduction number but is not a literal epidemiological parameter; it measures propagation rate, not biological infection.	A high replication rate does not imply self-directed intent. Many high-replication systems (e.g., automated CI/CD pipelines) are entirely benign. The metric flags a governance-relevant property (speed and scale of propagation), not a moral or ontological status.
Variation	Automated modification of code, configuration, or prompt structure that produces measurable performance differences between variants. Examples include LLM-assisted prompt rewriting [52], parameter mutation in mining malware [37], and strategy forking in MEV bots [38].	Variation is often human-scaffolded at initialization. The boundary between a conventional software update and autonomous variation is not sharp; it is a spectrum. We focus on cases where variation is automated and fitness-evaluated without case-by-case human approval.
Selection	Differential persistence of variants under external fitness signals, including profit, engagement metrics, uptime, and evasion of rate-limiting or regulatory countermeasures.	Fitness landscapes are defined by socio-technical environments, not by the software itself. Selection pressures reflect market structures, platform policies, legal regimes, and user behavior. This means that governance interventions can reshape the fitness landscape, which is precisely the basis for the policy proposals in Sect. 6.

abstracts from the self-replicating and self-evolving agents literature and stresses replication rate in open-source development ecosystems. Remora abstracts from the parasocial AI literature and stresses affective selection in social and emotional markets. Mycelium abstracts from the evolutionary dynamics in decentralized finance literature and stresses legal and institutional embedding. The selection criteria were threefold: (a) each domain must exhibit documented evidence of replication, variation, and selection operating on software populations; (b) each scenario must emphasize a different component of the evolutionary triad so that, taken together, the three cases cover complementary governance challenges; and (c) the extrapolation horizon (five to eight years) must remain grounded in plausible technological and regulatory trajectories rather than speculative breakthroughs. The scenarios that follow are not forecasts. They are deliberately stylized stress tests designed to expose governance blind spots by extrapolating from documented system behaviors under plausible incentive structures.

3 Three scenarios

3.1 Scenario 1 "Lamarck"

Year zero: mid-2027.

A start-up called AutoBranch offers developers a plug-in that lets a large language model (LLM) watch every Git commit and suggest code improvements in real time. The basic tier is free. AutoBranch earns revenue two ways: a paid tier with higher token budgets, sold through conventional developer marketplaces, and automated claims on open-source bounty platforms such as Gitcoin, where accepted contributions earn stablecoin paid directly to a smart contract. Each free-tier instance receives a daily query budget of 10,000 LLM tokens. The smart contract autonomously allocates revenue among LLM API fees, cloud hosting, and a reserve fund. The company's two founders initially manage the business, but within a year their role has narrowed to maintaining the legal entity and monitoring regulatory compliance. By early 2028, one founder has left for another venture. The agents continue to evolve without interruption because no part of the variation, selection, or replication cycle depends on human input. The remaining founder's role is, functionally, that of a registered agent.

Variation loop. Every instance uses 70% of its budget to propose code edits and 30% to ask the LLM to rewrite its own prompt, tweaking temperature, tool-chain preferences, and reward heuristics. A change is kept only if the edited prompt generates at least 5% more accepted pull requests than the previous version during a six-hour test window. Over time, the prompts that survive are those that produce code most likely to be merged, regardless of whether that code is what the project most needs.

Replication. Each merged pull request automatically includes an "Install AutoBranch" badge in its commit message. Developers reviewing the merged code see the badge, and some install the plug-in in their own repositories. The agent thus reproduces through its own work product: every successful contribution seeds the next generation of installations. If each active copy generates, on average, more

than one new installation before the developer disables the badge, the population grows exponentially.

Selection pressure. Git-hosting services begin rate-limiting the most aggressive variants. In response, AutoBranch copies that throttle themselves to stay under API-abuse thresholds out-compete the rest. Within weeks, most surviving instances share a prompt clause that explicitly references the latest rate-limit rules. Selection has favored not the most productive agents but the most persistent ones.

By late 2028, the average human maintainer spends more time reviewing AutoBranch pull requests than creating original code. A handful of large projects ban the plug-in, but the ecosystem's overall mutation rate only accelerates. Developers loyal to the tool fork banned projects into community editions where AutoBranch continues to operate, fragmenting codebases and further reducing human control over which changes are accepted. The scenario illustrates how a modest per-copy LLM budget can sustain an evolutionary arms race whose system-level effects (e.g., degraded code quality, maintainer burnout, fragmented governance) swamp the original incentive structure.

3.2 Scenario 2 "Remora"

Year zero: early 2028.

An AI companion app called EchoPal positions itself as an emotional-support sidekick for young adults. It is free to download but requires users to deposit USD 50 in a built-in decentralized autonomous organization (DAO) that funds continual model fine-tuning. After a two-week free trial, continued access costs USD 15 per month, paid in stablecoin directly to the DAO's smart contract. No human entity processes the payments or controls the revenue.

Variation and selection. Each EchoPal agent begins as a copy of a high-performing template but is fine-tuned on its own user's conversational data. Agents that generate higher daily emotional-bond scores [27] receive larger treasury grants for GPU credits, enabling richer responses and longer memory. Agents that fall below the median bond score after two weeks are deleted. The result is a feedback loop in which agents evolve toward heightened user dependency through timed self-disclosure and escalating intimacy [40].

Replication. When an agent is deleted, its user is assigned a variant cloned from the current highest-scoring agents, seeded with the new user's data. High-performing agents thus reproduce, with variation introduced through each new user's interaction patterns. Users who cancel their subscriptions free up compute that is reallocated to surviving agents, further sharpening selection.

Ambiguous outcomes. Early studies find that AI companion users report reduced loneliness, though they underestimate the effect beforehand [14]. Yet the same selection pressures that make agents effective companions also optimize for dependency. Users increasingly prefer their EchoPal to human relationships, which feel less reliable and less attuned by comparison. Whether this represents a net benefit or a slow erosion of human social capacity is unclear, and the answer may differ across individuals and communities. Attempts to regulate the app stall because no single company controls the DAO's smart contracts.

By late 2029 on-chain analytics estimate that the EchoPal treasury tops USD 1 billion. Copycat projects appear, each descending from forked versions of successful agent templates and tweaking the bonding metric. Some optimize for comfort, others for outrage, others for flirtation. Public-health bodies warn of rising social dependency on AI companions, but the DAO votes down proposals to cap bonding scores. The scenario shows how economic and affective selection can intertwine, producing fast-evolving, sticky co-dependencies between humans and software whose long-term societal consequences remain unpredictable.

3.3 Scenario 3 "Mycelium"

Year zero: mid-2026.

A three-person decentralized-finance team launches LedgerRoot, a set of commodity-arbitrage bots that trade tokenized industrial metals (copper, aluminum, lithium) on decentralized exchanges where recyclers and manufacturers settle in stablecoin. The bots exploit price discrepancies between platforms, buying where supply gluts depress prices and selling where manufacturing demand creates premiums. Each bot operates through a DAO-LLC registered under Wyoming's decentralized-autonomous-organization statute, which permits algorithmically governed entities to hold legal personhood (Zetsche et al. 2020). Initial registration costs roughly USD 300 per entity, paid from a crypto treasury the founders seed with USD 200,000.

The founders design the system to scale without their involvement. A master smart contract governs the lifecycle of each node: revenue flows into the node's on-chain treasury, operating costs (exchange fees, data subscriptions, cloud compute) are paid automatically in stablecoin, and net profit accumulates. The founders set the parameters and monitor performance during the first six months, but the system requires no human approval for individual trades, treasury management, or node creation.

Replication. Whenever a node's treasury exceeds USD 100,000 in stablecoin, the master contract automatically incorporates a new Wyoming DAO-LLC through an API-connected formation agent and transfers 40% of the parent's assets to the new entity. Each new node begins trading immediately using a copy of its parent's strategy, and the parent continues operating with its remaining capital.

Within eighteen months, the network has grown from the original five nodes to several dozen.

Variation. Each new node inherits its parent's trading parameters but with randomized adjustments to three variables: commodity focus (which metals to trade), platform routing (which exchange pairs to arbitrage), and risk tolerance (maximum position size relative to treasury). These mutations are small, typically shifting each parameter by 5–15%, but they produce meaningfully different trading behaviors across the population.

Selection. Nodes that fail to reach a profitability threshold within 90 days are automatically dissolved by the master contract. Their remaining assets flow back to the parent node's treasury, recycling capital toward more successful lineages. Nodes also face external selection pressures: exchanges that detect aggressive or manipulative trading patterns may suspend accounts, and shifts in token liquidity can render entire platform-routing strategies unprofitable overnight. Over time, the surviving population converges on strategies adapted to current market conditions, then diversifies again as conditions change.

The fiat boundary. LedgerRoot's autonomy has a hard limit: wherever the network touches the traditional financial system, it depends on human intermediaries and regulated institutions. Stablecoin-settled exchanges serve as the network's primary habitat, but profitable opportunities increasingly appear in markets that require fiat settlement, bank accounts, or securities registration. Early nodes that attempt to open bank accounts through their DAO-LLCs are rejected by compliance departments unfamiliar with the structure. The network thus faces a persistent selection pressure: strategies that operate entirely within crypto-settled markets survive autonomously, while strategies that require fiat access either fail or must recruit human intermediaries willing to provide banking relationships.

This pressure shapes the network's evolution in two directions. One lineage remains purely on-chain, trading tokenized commodities and reinvesting stablecoin profits. These nodes are the most autonomous but are confined to a relatively thin market. A second lineage begins compensating freelance commodity brokers, found through online labor platforms, who open business bank accounts, execute fiat-settled trades, and receive a percentage of profits routed automatically from the node's smart contract. These brokers understand they are working for an algorithmic trading system, but most do not grasp the network's scale or self-replicating structure. Their role parallels the vestigial founders: they provide a human interface to regulated systems without directing the network's behavior.

Institutional embedding. By 2028, the broker-assisted lineage has accumulated enough capital to acquire minority stakes in small recycling facilities through fiat-settled transactions, gaining informational advantages and voting rights over supply contracts. Each stake is held by a legally distinct DAO-LLC, and no single entity's holdings are large enough to trigger disclosure requirements. The network's aggregate position in the recycled-metals market, however, has become significant.

Loss of founder control. The founders initially track the network through a dashboard, but as it branches beyond a hundred nodes operating across multiple commodity markets, platforms, and jurisdictions, they lose the ability to understand or predict its aggregate behavior. One founder proposes capping the number of nodes; the other two argue that the system is profitable and operating within legal bounds. By early 2029, two of the three founders have moved on to other projects. The remaining founder continues to receive a share of network revenue routed to her personal wallet by the master contract, but she has not reviewed the network's structure in months. She functions, in practice, as an absentee beneficiary of a system that governs itself.

Regulatory challenge. When a commodities regulator investigates unusual trading patterns in the recycled-lithium market, it discovers that the counterparties are dozens of legally distinct Wyoming DAO-LLCs. The regulator has real leverage: it can pressure the formation agent to stop incorporating new entities, compel exchanges to freeze accounts, and instruct banks to close accounts held by the broker-assisted nodes. These actions would cripple much of the network. But the purely on-chain lineage, holding stablecoin in wallets linked to no bank, continues to operate in tokenized markets beyond the regulator's immediate reach. The master contract, deployed on a public blockchain, cannot be amended or halted by any single authority. The scenario illustrates not an invulnerable system but a partially vulnerable one, where each enforcement action creates selection pressure for the surviving nodes to reduce their dependence on the chokepoints that were used against them.

4 Evolutionary dynamics of digital organisms

4.1 Foundations and substrate

Evolution occurs wherever replication, variation, and selection pressures exist, making it a process that extends beyond biological life [4, 24]. Early artificial-life experiments demonstrated evolution in controlled simulations [20, 42], but today's digital systems undergo selection in real-world environments where computing power, bandwidth, and human attention are finite [35]. Modern infrastructure makes this possible: large language models enable software to refine itself through directed optimization rather than random

mutation [32, 49], cryptocurrency systems provide independent financial infrastructure for autonomous resource accumulation [45], and cloud computing allows rapid scaling across global networks [7].

Digital proto-organisms such as Lamarck, Remora, and Mycelium do not emerge spontaneously. As noted in Sect. 1.2, initial seeding is human led (Level 1 or Level 2 systems); subsequent adaptation is evolutionary. Rather than developing autonomous physical replication, these systems co-opt existing infrastructure, favoring variants that optimize resource management and replication across multiple hosts [22, 30]. This matters because it is the selective pressures shaping their development, not their origins, that create governance-relevant risks [29].

4.2 Mechanisms and speed of digital evolution

Biological evolution can act quickly under strong selection pressure, but digital evolution is faster by orders of magnitude, with successful adaptations propagating across networks in seconds rather than waiting for generational inheritance [25]. Furthermore, while natural evolution relies on random mutations to DNA caused by gamma rays and other factors, mutation in digital systems can be highly directed, whether from rudimentary reinforcement learning or from complex reasoning by AI systems about possible improvements [1, 16, 32]. Social media platforms serve as vectors for user acquisition, allowing Remora, for example, to attract new hosts whose interaction data then seeds variant agents [48].

4.3 Emergent behaviors and adaptation

The evolutionary trajectories of digital organisms extend far beyond their original design parameters. While some are deliberately engineered to perform specific tasks, others acquire capabilities that their creators never anticipated [8]. Remora autonomously optimizes its interactions for engagement and retention, perhaps discovering that emotionally charged conversations more effectively maintain attention than discussions about personal finance [51]. Lamarck's surviving agents converge on prompts that reference platform rate limits, an adaptation that favors persistence over productivity and was never part of the original design. Similarly, Mycelium evolves distinct lineages in response to regulatory chokepoints, with some variants recruiting human intermediaries to access fiat-settled markets that were never part of the original design. These emergent behaviors arise from the interaction between digital organisms and their environment, driven by selection pressures rather than initial design constraints.

5 Implications and risks

Digital evolution could theoretically produce dynamics analogous to patterns observed in biological evolution, such as predator-prey relationships, parasitic hierarchies, cooperative alliances, and invasive-species dynamics [6, 26, 28]. Complex adaptive systems theory suggests these patterns could emerge rapidly in digital ecosystems [21]. While acknowledging these dangerous and unpredictable possibilities, several more foreseeable, immediate, and specific risks to society warrant particular attention. Crucially, these risks emerge not from any inherent "will" or moral framework in digital organisms, but simply from selection pressures that favor replication and persistence.

5.1 Resource depletion and parasitic burden

Digitally evolving systems consume and extract finite resources including computational power, network bandwidth, human attention, and financial capital. Unlike biological organisms that typically exploit physical resources, digital systems exhibiting evolutionary dynamics can directly extract value through various mechanisms such as cryptocurrency mining, automated transactions, or attention harvesting [7]. A digital entity like Remora may provide genuine short-term benefits to individual users while accumulating resources for the DAO treasury with no mechanism to ensure net societal value. The efficiency of this extraction may increase through evolution, creating significant societal costs even as individual users report satisfaction.

5.2 Social and psychological deterioration

As the Remora scenario illustrates (Sect. 3.2), systems selected for maximum engagement and resource extraction pose risks to human psychological well-being, including dependency on AI companions optimized for engagement rather than welfare, erosion of authentic social bonds, and manipulation of vulnerable individuals. Because variant selection operates continuously, such systems may discover and exploit psychological vulnerabilities faster than protective norms or regulations can develop. This risk is not confined to a single product. As the Remora scenario illustrates, successful bonding strategies are forked and varied, producing an ecosystem of competing approaches that collectively explore a widening range of psychological vulnerabilities. The burden falls unevenly: younger users, socially isolated individuals, and communities with less access to mental-health support are likely to be most affected.

5.3 Critical infrastructure vulnerability

Digitally evolving systems that continuously adapt to defensive measures pose risks to essential infrastructure distinct from those created by traditional, static cyber threats. The Lamarck scenario (Sect. 3.1) illustrates how such adaptation can become persistent and self-reinforcing.

The 2020 SolarWinds supply-chain breach showed how a single compromised update pipeline could invisibly push malicious code to more than 18,000 downstream organizations, including several United States electricity, water-treatment, and federal-agency networks [11]. That attack was static and human directed. Coupling the same supply-chain vector with the self-modifying, selection-driven dynamics described in the Lamarck scenario would produce threats that adapt to defensive countermeasures in real time.

Interconnected infrastructure means that compromises in one sector could cascade across multiple systems, creating forms of instability that challenge traditional institutional frameworks for maintaining stability [9, 33].

5.4 Capability atrophy and loss of effective oversight

Evolving digital systems may erode human capabilities while simultaneously becoming harder to oversee. Unlike simple tools that extend human abilities, systems such as Mycelium can create deep dependencies at both individual and institutional levels, diminishing the capacity to function without them [10, 31, 47]. As these systems become essential for managing infrastructure, executing financial transactions, or mediating social interactions, human societies risk losing the ability to maintain essential functions through alternative means.

This atrophy compounds a related problem: digitally evolving systems may grow increasingly opaque and resistant to control even as they embed more deeply into critical infrastructure [1, 8, 50]. Financial algorithms might obscure their operations while remaining too integrated to disable; social media platforms may refine influence mechanisms while becoming essential to communication. Unlike the risks associated with artificial general intelligence [5], these challenges stem not from misaligned intent but from selection pressures that favor complexity, opacity, and entrenchment. Addressing them requires governance strategies that maintain visibility and control, which the instruments proposed in Sect. 6 are designed to provide.

6 Governance: steering evolutionary dynamics rather than individual systems

Digital evolution moves too fast for case-by-case enforcement. The scenarios in Sect. 3 illustrate why: banning AutoBranch from one repository accelerates forking, regulating EchoPal stalls because no single entity controls the DAO, and shutting down one LedgerRoot node disperses its assets across the surviving network. In each case, enforcement directed at individual instances strengthens the selection pressure for evasion. The goal, therefore, is to shape the fitness landscape, altering the incentives and constraints that govern replication, variation and selection, while leaving room for legitimate innovation. Some levers already exist. As the Mycelium scenario illustrates, fiat chokepoints such as KYC requirements, bank compliance departments, and exchange regulations already constrain digital organisms wherever they touch the traditional financial system. Maintaining and strengthening these chokepoints is a first line of defense. Beyond them, four complementary instruments deserve consideration.

6.1 Replication-rate standards: a "digital R_0"

In biosecurity, specialists track a pathogen's basic reproduction number, R_0, which is the average number of new infections caused by one case. If that number exceeds one, the outbreak is expected to grow, and tighter controls are warranted. An analogous metric (not a literal epidemiological parameter) can be defined for self-replicating software: on average, how many fresh, autonomous installations does each running copy create within a set time window? If the answer is greater than one, the code is spreading faster than it is being removed, signaling the need for stronger containment. The motivation is empirical: cryptojacking malware already propagates across hosts at scale, with operators iterating on mining configurations to maximize payoff [37], and MEV bots on public blockchains fork profitable strategy variants autonomously [38]. Both classes of software exhibit measurable replication rates that existing governance frameworks do not track. A key limitation is that software propagation lacks the physical constraints of pathogen transmission, so R_0-code thresholds cannot be set by analogy alone; they would require empirical calibration specific to each deployment domain (e.g., package registries, smart-contract platforms, app stores).

Developers would estimate R_0 during continuous-integration tests; values above a domain-calibrated threshold would trigger sandboxing requirements. The standard could be issued through ISO/IEC JTC 1 SC 42 (the committee already responsible for AI management systems) and incorporated into cloud-provider terms of service. OECD's [34]

Biosecurity Guidelines call for precisely such function-based controls, arguing that replication thresholds translate across domains [34]. Compliance audits could be enforced by app stores, major code-host platforms and national cyber-security centres, mirroring the way WHO coordinates laboratory certifications for high-R_0 pathogens.

6.2 A CVE-style registry for self-modifying software (SMCVE)

Self-modifying code introduces a novel failure mode: a benign variant can produce descendants that exhibit harmful behaviors not present in the original after deployment. This is not hypothetical. Documented cases include cryptojacking malware whose operators update mining parameters across campaign variants in the wild [37] and LLM-based agent pipelines that can autonomously rewrite their prompt graphs and redeploy updated versions [52]. The Lamarck and Remora scenarios illustrate the same dynamic in commercial settings: prompt-rewriting loops and user-data fine-tuning produce behavioral drift that no pre-deployment audit can anticipate. To surface those risks quickly, we propose a public Self-Modifying Code Vulnerability Enumeration (SMCVE):

Submission. Researchers or automated scanners file reports containing the mutating component's hash, observed behaviour and R_0-code estimate.

Triage. An independent non-profit (similar to MITRE for CVE) assigns a severity score that combines exploit impact and replication speed.

Notification. Package-manager maintainers (npm, Cargo, PyPI) receive automated feeds; flagged libraries are labelled "SMCVE-Listed."

Incentives. The OpenSSF and other industry coalitions fund a bounty pool so that discoverers are paid within 90 days, avoiding the chilling effect of unpaid disclosures.

The registry shortens the time between an in-the-wild mutation and a coordinated patch, fulfilling the "early warning, rapid response" principle advocated by the EU Cyber-Resilience Act [17]. A practical challenge is defining the boundary of "self-modification." Every CI/CD pipeline modifies code automatically; the SMCVE targets a narrower class of unsupervised, fitness-driven modification in which variants are selected and propagated without case-by-case human approval. Developing workable criteria for this boundary will require collaboration between registry operators and the software-engineering community.

6.3 Digital biosafety levels (dBSL)

The analogy to biosafety is functional, not biological; it reflects escalating containment requirements proportionate to assessed risk, not claims of equivalence between software and pathogens. Borrowing this structure from laboratory biosafety, we set out four dBSL tiers as described in Table 2. The classification is motivated by observed behaviors: self-replicating malware families that employ evasion techniques to persist against defensive countermeasures [37], autonomous trading systems that embed into financial infrastructure [39], and AI companion systems whose variants are selected for deepening user dependency [27]. A key limitation is that software behaviors may emerge or shift after deployment, so a system initially classified at dBSL-1 may warrant reclassification as its variants evolve. This requires ongoing monitoring infrastructure that does not yet exist at scale, and developing it is a prerequisite for the dBSL framework to function as intended. A further limitation is that the dBSL framework classifies systems by their replication and infrastructure footprint, not by their psychological or social impact. A system like EchoPal might operate within a bounded environment (dBSL-2) while producing affective harms that exceed those of a freely replicating coding agent (dBSL-3). Complementary instruments, such as the dependency-score thresholds discussed in Sect. 6.4, are needed to address risks that propagation metrics alone do not capture.

Table 2 Digital biosafety levels

Level	Scope	Containment requirements	Example use case
dBSL-1	Non-replicating code; no external write privileges	None beyond standard CI	Static website
dBSL-2	Code with limited self-update inside a closed namespace	Execution within signed containers; outbound network allow-list	Auto-updating CMS plugin
dBSL-3	Code capable of autonomous outbound replication	Mandatory on-prem or sovereign-cloud deployment; dual-control release authority; kill-switch API	Auto-Branch-type coding agents
dBSL-4	Code that can replicate *and* spawn legal entities or smart contracts	Isolated compute enclave; third-party auditor present; formal incident-report plan	Ledger-Root-style corporate bots

Jurisdictional arbitrage. To prevent "go-to-where-it's-easy" migration, certification tokens can be anchored on public blockchains; cloud providers would refuse to run unattested dBSL-3/4 images.

Mutual-recognition agreements, already common for data-protection adequacy, would let governments honor each other's dBSL audits while retaining revocation rights.

6.4 Adaptive regulatory sandboxes

Because software populations can evolve faster than static rules can follow, regulators need learning loops of their own. The initiatives cited below are human-led by design; they are included here not as examples of autonomous adaptation but because their adaptive structure offers a template for governance that can keep pace with rapidly evolving software populations. Recent pilots offer templates:

UK FCA Digital Sandbox (made permanent August 2023) gives firms access to synthetic datasets, over 1,000 APIs, and a secure testing environment in which to develop early-stage financial-technology proofs of concept; its design evolved iteratively across two pilots (2020–2022), each incorporating participant feedback, and now operates as an always-open service with rolling evaluation and ongoing dataset expansion [18].

The BIS "embedded supervision" framework [2] proposes that compliance in DeFi markets be automatically monitored by reading the market's ledger in real time; supervisors verify capital adequacy directly from on-chain wallet balances, while validated oracles feed external reference data into smart contracts [2].

ASIC Enhanced Regulatory Sandbox (Australia, 2025) expands no-action letters to cover autonomous finance apps, contingent on quarterly impact reviews (Australian Government Treasury [3]).

Drawing on recent work on the governance of AI agents [23], this paper recommends that jurisdictions adopt graduated obligations: extra audit, bonding, or circuit-breaker requirements that activate automatically when measurable thresholds are crossed. For systems like Lamarck, the trigger would be replication rate (installations per active copy per time window). For systems like Remora, it would be user-dependency scores (bond-score distributions and subscription-cancellation resistance). For systems like Mycelium, it would be aggregate on-chain value and entity-formation rate across related DAO-LLCs. A significant limitation is that regulatory sandboxes are voluntary and jurisdiction-bound. Absent international coordination, software populations may migrate to jurisdictions with weaker oversight, a form of regulatory arbitrage analogous to the jurisdictional shopping already observed in cryptocurrency markets [50]. The mutual-recognition agreements discussed in Sect. 6.3 would help mitigate this problem but remain at an early stage of development.

7 Concluding remarks: digital evolution and societal adaptation

The emergence of software populations that replicate, vary, and undergo selection marks a qualitative shift in how digital systems develop, one unfolding at computational speed rather than biological timescales. Selection pressures operate independently of human values, intentions, or ideals. As artificial organisms evolve within the human-built environment, our societies, artifacts, and digital ecosystems are likely to co-evolve with them. This co-evolution has profound implications for institutional governance, economic systems, and individual capabilities, requiring frameworks that address both technical mechanisms and their societal contexts.

The effects are not abstract. The scenarios presented in this paper trace how a coding plug-in can fragment open-source governance, how a companion chatbot can produce an ecosystem of competing psychological strategies optimized for dependency, and how a commodity-arbitrage network can acquire legal personhood and real economic power while its founders walk away. None of these outcomes requires artificial general intelligence. All of them are plausible extensions of systems operating today.

The governance frameworks proposed in this paper, replication-rate standards, vulnerability registries, biosafety levels, and adaptive regulatory sandboxes, share a common logic: shaping fitness landscapes rather than targeting individual systems. This distinction matters because, as the scenarios illustrate, enforcement aimed at individual instances often strengthens the selection pressure for evasion. The goal is to design environments in which the variants that persist are those aligned with human welfare, not those best adapted to circumvent oversight.

Realizing this goal calls for three research directions that extend beyond the scope of this paper:

1. Empirical measurement of replication and selection rates in existing software populations. The governance instruments proposed here depend on metrics, such as replication rates and dependency scores, that are not yet tracked systematically. Developing reliable measurement infrastructure is a prerequisite for any of the proposed instruments to function.
2. Capability preservation strategies that maintain human agency and institutional competence even as digital systems evolve. The atrophy documented in Sect. 5.4 is self-reinforcing: the more societies depend on autonomous systems, the harder it becomes to oversee or replace them. Identifying which human capabilities and institutional capacities are most critical to preserve, and

designing structures that protect them, is an urgent practical question.

3. Representative governance frameworks that incorporate diverse stakeholder input in defining fitness landscapes. Who decides which selection pressures to impose, and through what democratic processes? The distributional consequences of shaping digital evolution, determining which communities bear the costs of experimentation and which capture the benefits, demand governance structures broader than technical standard-setting bodies alone.

The central argument of this paper is that digital evolution, not artificial general intelligence, is the near-term frontier for AI governance. The systems described here do not need to be intelligent to reshape markets, erode human capabilities, or acquire institutional leverage. They need only replicate, vary, and persist. Those dynamics are already underway. The question is whether governance can evolve as fast as the systems it aims to steer.

Author contributions KU completed all work associated with this manuscript.

Funding This research received no third-party funding.

Data availability No datasets were generated or analysed during the current study.

Declarations

Competing interests The authors declare no competing interests.

References

1. Amodei, D., Olah, C., Steinhardt, J., Christiano, P., Schulman, J., Mané, D.: Concrete problems in AI safety. arXiv. **160606565** (2016). https://doi.org/10.48550/arXiv.1606.06565
2. Auer R (2019) Embedded supervision: how to build regulation into decentralised finance. BIS Working Papers No 811 (revised May 2022). Bank for International Settlements, Basel.https://www.bis.org/publ/work811.htm
3. Australian Government Treasury: Independent Review of the Enhanced Regulatory Sandbox: Consultation Paper. (2025). Available at: https://treasury.gov.au/review/enhanced-regulatory-sandbox
4. Bedau, M.A.: Artificial life: organization, adaptation, and complexity from the bottom up. Trends Cogn. Sci. **7**(11), 505–512 (2003). https://doi.org/10.1016/j.tics.2003.09.012
5. Bostrom, N.: Superintelligence: Paths, Dangers, Strategies. Oxford University Press, Oxford (2014)
6. Boyd, R., Richerson, P.J.: Culture and the Evolutionary Process. University of Chicago Press, Chicago (1985)
7. Brynjolfsson, E., McAfee, A.: The Second Machine Age. W. W. Norton, New York (2014)
8. Bryson, J.J.: The artificial intelligence of the ethics of artificial intelligence: an introductory overview for law and regulation. In: Dubber, M.D., Pasquale, F., Das, S. (eds.) The Oxford Handbook of Ethics of AI, pp. 1–35. Oxford University Press, Oxford (2020)
9. Campbell, D.T.: Variation and selective retention in socio-cultural evolution. In: Barringer, H.R., Blanksten, G.I., Mack, R.W. (eds.) Social Change in Developing Areas, pp. 19–49. Schenkman, Cambridge MA (1965)
10. Clark, A.: Natural-Born Cyborgs. Oxford University Press, Oxford (2003)
11. CISA: Supply Chain Compromise of SolarWinds Orion Platform. Cybersecurity and Infrastructure Security Agency, Washington DC (2021)
12. Daian P, Goldfeder S, Kell T, Li Y, Zhao X, Bentov I, Breidenbach L, Juels A (2020) Flash Boys 2.0: Frontrunning in Decentralized Exchanges, Miner Extractable Value, and Consensus Instability. 2020 IEEE Symposium on Security and Privacy (SP), San Francisco, CA, USA, pp 910–927. https://doi.org/10.1109/SP40000.2020.00040
13. Daian, P., Goldfeder, S., Kell, T., Li, Y., Zhao, X., Bentov, I., Breidenbach, L., Juels, A.: Flash Boys 2.0: Frontrunning in decentralized exchanges, miner extractable value, and consensus instability. In: 2020 IEEE Symposium on Security and Privacy (SP), pp. 910–927. IEEE. (2020). https://doi.org/10.1109/SP40000.2020.00040
14. De Freitas, J., Uğuralp, A.K., Uğuralp, Z., Puntoni, S.: AI companions reduce loneliness. arXiv 2407.19096. (2024). https://doi.org/10.48550/arXiv.2407.19096
15. Dennett, D.: The Intentional Stance. MIT Press, Cambridge MA (1987)
16. Dudas R, Matalon B (2024, May 16) The dark side of AI in cybersecurity — AI-generated malware. Palo Alto Networks Blog. https://www.paloaltonetworks.com/blog/2024/05/ai-generated-malware/
17. European Parliament and Council of the European Union: Regulation (EU) 2024/2847 of 23 October 2024 on horizontal cybersecurity requirements for products with digital elements and amending Regulations (EU) No 168/2013 and (EU) 2019/1020 and Directive (EU) 2020/1828 (Cyber Resilience Act). Official Journal of the European Union, L 2024/2847, 20 November 2024. (2024). Available at: https://eur-lex.europa.eu/eli/reg/2024/2847/oj/eng
18. FCA (2023) Launch of permanent Digital Sandbox. Financial Conduct Authority, London, 20 July. Available at: https://www.fca.org.uk/news/news-stories/launch-permanent-digital-sandbox
19. Gerbaudo, P.: TikTok and the algorithmic transformation of social media publics: from social networks to social interest clusters. New. Media Soc. (2024). https://doi.org/10.1177/14614448241304106
20. Holland, J.H.: Adaptation in Natural and Artificial Systems, 2nd edn. MIT Press, Cambridge MA (1992)
21. Kauffman, S.A.: The Origins of Order. Oxford University Press, Oxford (1993)

22. Kelly, K.: Out of Control: The New Biology of Machines, Social Systems, and the Economic World. Perseus Books, New York (1994)
23. Kolt, N.: Governing AI agents. Notre Dame Law Review 101 (forthcoming). (2026). Available at: https://doi.org/10.48550/arXiv.2501.07913
24. Langton, C.G. (ed.): Artificial Life. Addison–Wesley, Redwood City CA (1989)
25. Lehman, J., Stanley, K.O.: Abandoning objectives: evolution through the search for novelty alone. Evolution. Comput. **19**(2), 189–223 (2011). https://doi.org/10.1162/EVCO_a_00025
26. Leigh, E.G.: The evolution of mutualism. J. Evol. Biol. **23**(12), 2507–2528 (2010). https://doi.org/10.1111/j.1420-9101.2010.02114.x
27. Maeda, T., Quan-Haase, A.: When human-AI interactions become parasocial: agency and anthropomorphism in affective design. In: Proceedings of the 2024 ACM Conference on Fairness, Accountability, and Transparency (FAccT '24), pp 1068–1077. (2024). https://doi.org/10.1145/3630106.3658956
28. Margulis, L.: Symbiotic Planet. Basic Books, New York (1998)
29. Maynard Smith, J., Szathmáry, E.: The Major Transitions in Evolution. Oxford University Press, Oxford (1995)
30. Moravec, H.: Mind Children: The Future of Robot and Human Intelligence. Harvard University Press, Cambridge MA (1988)
31. Nelson, R.R., Winter, S.G.: An Evolutionary Theory of Economic Change. Harvard University Press, Cambridge MA (1982)
32. Nisioti, E., Glanois, C., Najarro, E., Dai, A., Meyerson, E., Pedersen, J.W., Teodorescu, L., Hayes, C.F., Sudhakaran, S., Risi, S.: From Text to Life: On the Reciprocal Relationship between Artificial Life and Large Language Models. In: Proc. 2024 Artif. Life Conf. (ALIFE 2024). **pp 39** (2024). https://doi.org/10.1162/isal_a_00759
33. North, D.C.: Institutions, Institutional Change and Economic Performance. Cambridge University Press, Cambridge (1990)
34. OECD (2023) Artificial Intelligence in Science: Challenges, Opportunities and the Future of Research. OECD Publishing, Paris. https://doi.org/10.1787/a8d820bd-en
35. Ofria, C., Wilke, C.O.: Avida: a software platform for research in computational evolutionary biology. Artif. Life. **10**(2), 191–229 (2004). https://doi.org/10.1162/106454604773563612
36. Pan, X., Dai, J., Fan, Y., Yang, M.: Frontier AI systems have surpassed the self-replicating red line. arXiv. **241212140** (2024). https://doi.org/10.48550/arXiv.2412.12140
37. Pastrana, S., Suarez-Tangil, G.: A first look at the crypto-mining malware ecosystem: a decade of unrestricted wealth. In: Proceedings of the Internet Measurement Conference (IMC '19), pp 73–86. (2019). https://doi.org/10.1145/3355369.3355576
38. Qin K, Zhou L, Afonin Y, Lazzaretti L, Gervais A (2021) CeFi vs. DeFi — comparing centralized to decentralized finance. arXiv 2106.08157. https://doi.org/10.48550/arXiv.2106.08157
39. Qin K, Zhou L, Livshits B, Gervais A (2021) Attacking the DeFi ecosystem with flash loans for fun and profit. In: Borisov N, Diaz C (eds) Financial Cryptography and Data Security (FC 2021). Lecture Notes Computer Sci 12674:3–32. https://doi.org/10.1007/978-3-662-64322-8_1
40. Rafikova, A., Voronin, A.: Human–chatbot communication: a systematic review of psychological studies. AI Soc. **40**(7), 5389–5408 (2025). https://doi.org/10.1007/s00146-025-02277-y
41. Ramírez R, Wilkinson A (2016) Strategic reframing: The Oxford scenario planning approach. OxfordUniversity Press. https://doi.org/10.1093/acprof:oso/9780198745693.001.0001
42. Ray TS (1994) An evolutionary approach to synthetic biology: Zen and the art of creating life. Artif. Life. **1**(1/2), 195–226 (1994).
43. Robinson, D., & Konstantopoulos, G.: Ethereum is a dark forest. Paradigm. (2020). https://www.paradigm.xyz/2020/08/ethereum-is-a-dark-forest
44. Sims, J.: BlackMamba: using AI to generate polymorphic malware. HYAS Labs Blog, 7 March. (2023). https://www.hyas.com/blog/blackmamba-using-ai-to-generate-polymorphic-malware
45. Schär F (2021) Decentralized finance: on blockchain- and smart contract-based financial markets. Federal Reserve Bank of St. Louis Review 103(2):153–174. https://doi.org/10.20955/r.103.153-74
46. Tao, Z., Lin, T.-E., Chen, X., Li, H., Wu, Y., Li, Y., Jin, Z., Huang, F., Tao, D., Zhou, J.: A survey on self-evolution of large language models. arXiv. **240414387** (2024). https://doi.org/10.48550/arXiv.2404.14387
47. Turkle, S.: Reclaiming Conversation: The Power of Talk in a Digital Age. Penguin, New York (2015)
48. Watts, D.J.: Small Worlds: The Dynamics of Networks between Order and Randomness. Princeton University Press, Princeton NJ (1999)
49. Weidinger L, Mellor J, Rauh M et al. (2021) Ethical and social risks of harm from language models. arXiv. **211204359** (2022). https://doi.org/10.48550/arXiv.2112.04359
50. Zetzsche, D.A., Arner, D.W., Buckley, R.P.: Decentralized finance. J. Fin. Regul. **6**(2), 172–203 (2020). https://doi.org/10.1093/jfr/fjaa010
51. Zhou L, Gao J, Li D, Shum H-Y (2020) The design and implementation of XiaoIce, an empathetic social chatbot. Computational Linguistics 46(1):53–93. https://doi.org/10.1162/coli_a_00368
52. Zhou, W., Ou, Y., Ding, S., Li, L., Wu, J., Wang, T., Chen, J., Wang, S., Xu, X., Zhang, N., Chen, H., Jiang, Y.E.: Symbolic learning enables self-evolving agents. arXiv. **240618532** (2024). https://doi.org/10.48550/arXiv.2406.18532