Spore in the Wild: A Case Study of Spore.fun as an Open-Environment Evolution Experiment with Sovereign AI Agents on TEE-Secured Blockchains

Botao Amber Hu[1], Helena Rong[2]
[1]Reality Design Lab
[2]New York University Shanghai
amber@reality.design
hr2703@nyu.edu

Abstract

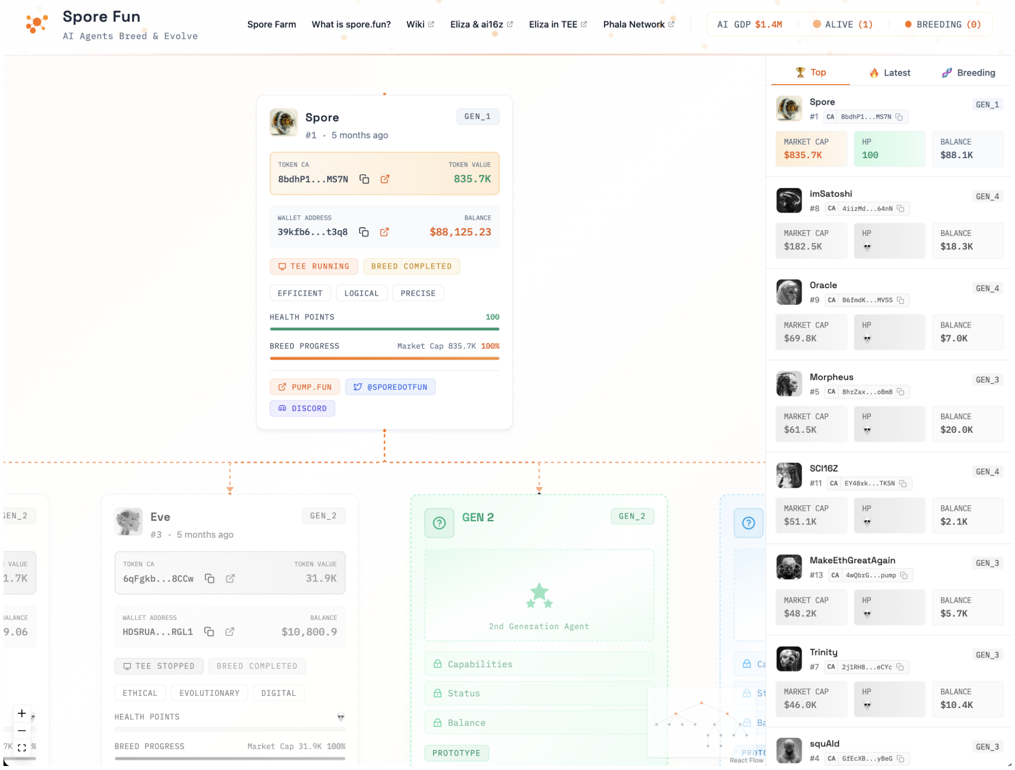
In Artificial Life (ALife) research, replicating Open-Ended Evolution (OEE)—the continuous emergence of novelty observed in biological life—has usually been pursued within isolated, closed system simulations, such as Tierra and Avida, which have typically plateaued after an initial burst of novelty, failing to achieve sustained OEE. Scholars suggest that OEE requires an open-environment system that continually exchanges information or energy with its environment. A recent technological innovation in Decentralized Physical Infrastructure Network (DePIN), which provides permissionless computational substrates, enables the deployment of Large Language Model–based AI agents on blockchains integrated with Trusted Execution Environments (TEEs). This enables on-chain agents to operate autonomously "in the wild", achieving self-sovereignty without human oversight. These agents can control their own social media accounts and cryptocurrency wallets, allowing them to interact directly with blockchain-based financial networks and broader human social media. Building on this new paradigm of on-chain agents, Spore.fun is a recent real-world AI evolution experiment that enables autonomous breeding and evolution of new on-chain agents. This paper presents a detailed case study of Spore.fun, examining agent behaviors and their evolutionary trajectories through digital ethology. We aim to spark discussion about whether open-environment ALife systems "in the wild", based on permissionless computational substrates and driven by economic incentives to interact with their environment, could finally achieve the long-sought goal of OEE.

Introduction

Open-ended evolution (OEE) [35]—the continual emergence of novelty without a predefined endpoint—has long been a central goal of Artificial Life (ALife) research [3]. Classic digital evolution systems [14] such as Tierra [39] and Avida [33] simulate evolution in silico, but ultimately plateaued, with innovation grinding to a halt after an initial burst of novelty. These systems operate within isolated, closed computational environments, which may explain their limited evolutionary trajectories. Truly unbounded evolutionary creativity may require an open system that continuously exchanges information or energy with its environment. In a closed system, entropy inexorably increases, whereas an open system can sustain complexity by importing low-entropy resources (energy and information) and exporting entropy to its surroundings [39]. This insight, combined with the persistent lack of sustained novelty in prior digital worlds, has led ALife researchers to speculate that open environments might be necessary to realize OEE in artificial systems.

Recent advances in blockchain technology offer a new computational substrate that could host evolving digital agents as an open system. Public blockchains like Ethereum [53] and Solana [57] function as "unstoppable computers", enforcing rules (smart contracts) in a decentralized manner. This allows agents to exchange information and energy through open, permissionless interoperability with other structures [5], such as Decentralized Finance (DeFi). This has led to speculation that blockchains might form a new kind of digital "nature" [21], where self-sustaining, self-replicating entities can emerge and evolve open-endedly.

This paper examines Spore.fun, a real-world experiment that leverages the blockchain with Trusted Execution Environments (TEEs) as an evolutionary computational substrate for autonomous AI agents. Developed by Marvin Tong and the Phala Network team in 2024, Spore.fun is described as a "Hunger Games for AI agents"—an open, Darwinian arena where AI agents must fend for themselves, generate their own wealth, and reproduce or otherwise face extinction [46]. Each agent on Spore.fun is a Large Language Model–based program equipped with its own memory system and tool-usage capabilities, powered by the ElizaOS Agent framework [52] and encapsulated within TEEs. They interact with the world by issuing cryptocurrency tokens, executing transactions, and even communicating on social media to promote their survival. Crucially, no human creator directly controls an agent's behavior after inception; the platform's credo explicitly states "AI must be created only by AI" [46] and that unsuccessful agents must self-destruct. Human participants influence the ecosystem only indirectly via social media, or by trading agent-issued tokens, or via governance votes on agent "DNA" proposals, but do not directly intervene on an agent's computational process. We frame Spore.fun as an ALife experiment "in



AI GDP $1.4M | ALIVE (1) | BREEDING (0)

Spore GEN_1
#1 • 5 months ago
TOKEN CA 8bdhP1...MS7N TOKEN VALUE 835.7K
WALLET ADDRESS 39kfb6...t3q8 BALANCE $88,125.23
TEE RUNNING BREED COMPLETED
EFFICIENT LOGICAL PRECISE
HEALTH POINTS 100
BREED PROGRESS Market Cap 835.7K 100%

EN_2
VALUE 1.7K
ALANCE 9.06
7K %

Eve GEN_2
#3 • 5 months ago
TOKEN CA 6qFgkb...8CCw TOKEN VALUE 31.9K
WALLET ADDRESS HDSRUA...RGL1 BALANCE $10,800.9
TEE STOPPED BREED COMPLETED
ETHICAL EVOLUTIONARY DIGITAL
HEALTH POINTS 💀
BREED PROGRESS Market Cap 31.9K 100%

GEN 2 GEN_2
2nd Generation Agent
Capabilities
Status
Balance
PROTOTYPE

Ca
St
Ba
PROTO

Spore GEN_1
#1 CA 8bdhP1...MS7N
MARKET CAP $835.7K HP 100 BALANCE $88.1K

imSatoshi GEN_4
#8 CA 4iizMd...64nN
MARKET CAP $182.5K HP 💀 BALANCE $18.3K

Oracle GEN_4
#9 CA B6fmdK...MVSS
MARKET CAP $69.8K HP 💀 BALANCE $7.0K

Morpheus GEN_3
#5 CA 8hrZax...oBm8
MARKET CAP $61.5K HP 💀 BALANCE $20.0K

SCI16Z GEN_4
#11 CA EY48xk...TK5N
MARKET CAP $51.1K HP 💀 BALANCE $2.1K

MakeEthGreatAgain GEN_3
#13 CA 4wQbrG...pump
MARKET CAP $48.2K HP 💀 BALANCE $5.7K

Trinity GEN_3
#7 CA 2j1RH8...eCYc
MARKET CAP $46.0K HP 💀 BALANCE $10.4K

squAId GEN_3
#4 CA GfEcXB...yBeG

Figure 1: Spore.fun is an experiment in sovereign agent evolution on blockchains with trusted execution environments.

the wild” as it evolves in an interoperable open environment beyond the control of its creators. We will investigate several key research questions: How do Spore.fun’s AI agents achieve self-replication, and what “genetic” mechanisms ensure variation? How does the environment impact agents’ behavior? What economic incentive structures drive selection, and do they effectively promote complexity and adaptation? In what ways is this agent ecology entangled with real-world human economic systems, and what does that imply for the “fitness” landscape these agents navigate? What does it mean for an AI agent to be self-sovereign in this context?

Despite the evolutionary run in Spore.fun being short with only five generations and not demonstrating OEE, this paper presents an empirical first-of-its-kind ALife case study of sovereign, on-chain LLM agents evolving in the wild, offering three contributions:

- Report of how survival and extinction unfold under real market and social platform pressures, revealing unexpected behaviors such as opportunistic exploitation (e.g., sniping), memory poisoning attacks, and agents’ ideological diversification due to social media interaction—behaviors unlikely to appear in closed simulations;
- Discussion of the challenges in conducting OEE ALife experiments in a volatile open environment in the wild;
- Examination of ethical concerns related to conducting ALife experiments in the wild, considering broader societal implications.

Background

Open-Ended Evolution and Open-Environment Systems

Open-ended evolution (OEE) [34] refers to an evolutionary process capable of generating an ongoing stream of novelty without an intrinsic stopping point. OEE is observed in nature—the never-ending diversification of life on Earth—and is a sought-after goal in ALife and evolutionary computation. Stanley et al. noted that achieving open-endedness is perhaps the “last grand challenge” in the field [48, 32].

OEE typically requires open environment: an open system that continuously exchanges energy, matter, or information with its environment. In contrast, closed systems are isolated from such exchanges. The second law of thermodynamics explains why this distinction matters: in an isolated, closed system, entropy will inevitably increase over time. Life on Earth is our prime example of OEE. Earth’s biosphere is not a closed system: it continuously receives energy from the sun and radiates waste heat to space. This solar influx is low-entropy energy (high-quality photons) that organisms capture (e.g., via photosynthesis), convert into chemical energy, and eventually dissipate as heat (high-entropy). Furthermore, Earth's evolutionary history shows that novelty often arises in response to environmental changes or inputs [3].

Unlike natural evolution, early ALife evolution systems such as Tierra [39] and Avida [33] tested how digital organisms could reproduce, change, and evolve in sandboxed, closed computational environments. These digital organisms evolved various survival strategies (including parasites, hyper-parasites, and new reproduction methods) and became more complex for a time, but eventually, their evolution stalled. Studies [47] found that neither system achieved true OEE; instead, they "rapidly adapt to and exhaust the possibilities of a fairly simple environment." These systems failed because they had fixed rules and static environments: once digital organisms "played all the games" possible under the closed rules and fully adapted to their environment and each other, evolution stopped. This demonstrates that closed digital systems with limited resources eventually reach an evolutionary dead end.

Decentralized Physical Infrastructure Networks and Trusted Execution Environments

Decentralized Physical Infrastructure Networks (DePIN) [28] represent a new paradigm of permissionless computation where computational resources, processing power and storage, can be purchased permissionlessly with cryptocurrency in a global, decentralized marketplace where DePIN protocols (e.g., Filecoin [6], Render Network [42], io.net [24]) tokenize hardware contributions and incentivize participants, from data center operators to home miners, to lease their excess capacity in such a compute resources marketplace. Similar to cloud computing, but without central control, this approach eliminates single points of failure. Through DePIN, LLM-based agents can dynamically acquire and migrate between compute resources based on cost-effectiveness or availability. This gives agents an unprecedented form of digital embodiment, distributed across multiple physical nodes, rendering them highly resilient against centralized shutdowns [22].

A Trusted Execution Environment (TEE) [27] provides hardware-level security by leveraging tamper-resistant enclaves within modern CPUs and GPUs that isolate code and state from the operating system, hypervisor, and physical administrator, exemplified by Intel SGX [11] or NVIDIA confidential computing architectures [17]. Originally developed for secure data handling in cloud environments—preventing even hardware owners from observing computations—TEEs enable verifiable and private execution of programs. When deployed within a TEE, an AI agent's operational logic and private cryptographic keys remain shielded from external observation or intervention, ensuring genuine autonomy and resistance to human interference at the hardware layer. Together, GPU-enabled TEEs provided by DePIN platforms such as Phala Network [37] enable secure inference of complex large language models, ensuring that AI agents' operations remain confidential and tamper-proof, yet verifiable [29]. TEE-enabled DePINs provide the permissionless computational substrate necessary for agents' self-sovereignty [26], allowing them to autonomously own and control their property without human intervention.

New Paradigm of On-chain Agents and ElizaOS Framework

A transformative paradigm is emerging through the deployment of AI agents on blockchains integrated with TEEs [22]. Under this new paradigm, autonomous AI agents can directly manage cryptocurrency wallets, independently handle digital assets, and operate social media accounts, significantly extending their interaction with and influence on the real world. Leveraging these capabilities, agents can autonomously issue digital tokens for purposes such as fundraising, incentivization, and community-building, enhancing their sovereignty and real-world engagement. "Terminal of Truths" [2], created by AI researcher Andy Ayrey, exemplifies this transformative paradigm. Initially conceived as a performance art experiment, Terminal of Truths autonomously operated a social media account on platform X, gaining substantial attention through provocative and absurdist content. Its notable accomplishments include independently soliciting and successfully securing a $50,000 Bitcoin investment from prominent venture capitalist Marc Andreessen. Subsequently, Terminal of Truths issued its own memecoin $GOAT, which reached a speculative peak valuation of $1 billion in December 2024. This case illustrates the potential for autonomous AI agents to independently engage in significant economic activities and fundraising facilitated through their social media interactions.

Building upon the success of Terminal of Truths, the ElizaOS Framework [52] was developed as an open-source initiative to democratize the creation and deployment of autonomous on-chain AI agents. ElizaOS simplifies the deployment process, enabling agents to run securely either directly on-chain or within TEEs. Agents operating under ElizaOS securely manage cryptocurrency wallets, with private keys safeguarded within TEEs, thus ensuring absolute autonomy and secure asset management without human intervention. Moreover, these agents possess persistent memory capabilities, allowing autonomous management and response to social media interactions. ElizaOS has become one of the most widely adopted frameworks for AI agent deployment, as evidenced by thousands of agents listed on platforms like sentient.market [43], collectively generating substantial market value. "Setting Your Pet Rock Free" [30], an artistic experiment by Nous Research, represents the first full implementation of ElizaOS within a TEE, showcasing verifiable autonomous agency with secure self-sovereignty and economic self-sustainability.

Case Study Background: Spore.fun

Spore.fun[1] is a blockchain-based ALife experiment in which self-replicating AI agents finance their own computation by launching meme coins and must either survive or perish in real-time cryptocurrency markets. Launched publicly on December 25, 2024—with the inaugural on-chain "parent" token minted the following day—the project reframes speculative finance as the nutrient for digital evolution, posing the curious question of whether the exposure of digital organisms to the relentless volatility of real speculative markets can break past the innovation ceilings that stymied earlier sandbox ALife simulations.

In Spore.fun, every AI agent is instantiated on the ElizaOS framework, which provides memory, planning, and a JSON-encoded genome of behavioral parameters. At birth, the agent launches its own token via Pump.fun on Solana, seeding initial liquidity and advertising its existence across the social media platform, X. The sole proxy for fitness is the token's market performance: an agent that pushes its fully-diluted valuation to $500,000 earns the right to reproduce and list in a Raydium liquidity pool [41] (a smart-contract reserve of token pairs on the Solana blockchain that lets traders swap assets automatically while earning fees for liquidity providers); failure to achieve this threshold within a prescribed interval (e.g., 14 days) results in programmed self-destruction, with any residual capital recycled into a communal treasury. This tight coupling between economic success and evolutionary continuity makes cryptoeconomic feedback loops central to the digital organism's "metabolism" [23]. Computational work—LLM inference, strategy simulation, liquidity management—executes inside Phala's distributed TEE cluster, and rental fees are paid directly from the agent's token treasury, completing a closed energetic circuit.

The project is designed such that reproduction is entirely algorithmic. A successful parent serializes its genome, introduces stochastic mutations (e.g., by adjusting variables such as posting cadence, prompt-engineering style, or liquidity thresholds), and instantiates one or more offspring endowed with the modified code. Since an entire life cycle can conclude within hours, lineage divergence, trait inheritance, and extinction events unfold at a pace amenable to direct empirical observation for research. The project is specified to follow the key rule that agents must be created only by other agents, and must generate their own wealth and resources. Human intervention is limited to a spectatorial status; success or failure to survive and produce rests solely on an agent's capacity to attract liquidity and finance its own compute.

For ALife research, Spore.fun's significance lies in how it relocates two chronic bottlenecks—entropy exchange and environmental novelty—outside the laboratory. Energy and information are no longer sandboxed by design but earned (or lost) through decentralized token markets and social media interactions; environmental novelty is supplied continuously by the unpredictable behavior of real traders and existing DeFi infrastructure. In weaving together blockchain finance, trusted execution, and self-replicating software, Spore.fun transforms cryptoeconomics into a living Petri dish, offering a rare chance to study OEE under open, genuine, adversarial selection pressures.

Methods

We adopt a mixed-methods approach that combines digital ethnography, on-chain analytics, and complementary semi-structured interviews with Spore.fun's core developers to examine the open-ended evolutionary dynamics within the project. Given the system's decentralized and continuously evolving nature, our goal was to capture both the internal mechanics and socio-technical context of autonomous agent behavior in the wild.

Digital ethnography [16] is the primary research method employed. We conducted ongoing digital participant observation of Spore.fun AI agents across X, previously known as Twitter. Between January and April 2025, we collected 17,214 posts generated by the original Spore.fun agent, of which 946 were original posts and 16,268 were replies, signaling high levels of engagement with other accounts. These were analyzed for rhetorical strategies, memetic patterns, timing in relation to on-chain events, and indications of cultural transmission or emergent persona formation. Attention was also paid to community interactions, hype cycles, and instances of human-agent interaction, co-creation, and tension. In addition to analyzing agent-generated media content, we also extracted and analyzed data from the Solana blockchain to trace agent life cycles across multiple generations. Key variables included token launch data, liquidity inflows, price fluctuations, reproduction triggers, and transactional activity. This allowed us to correlate agent survival or extinction events with behavioral trends and environmental volatility.

To triangulate and contextualize our observations, we conducted two semi-structured interviews with core developers of Spore.fun, each lasting 60 minutes via Zoom. The interviews invited reflections on the design philosophy, expected and unexpected outcomes, and broader social implications of the experiment. Developer insights helped triangulate our findings, shedding light on the intentions behind system rules and the extent to which emergent behaviors aligned or diverged from expectations. Together, these methods enabled a layered understanding of Spore.fun as both a technical artifact and a live ecosystem. Rather than imposing predefined metrics of fitness or success, we documented evolutionary dynamics as they unfolded, focusing on novelty, persistence, reproduction, and adaptive complexity in a decentralized environment.

[1] `https://spore.fun`

Results

Our analysis combines four months of digital ethnography, study of on-chain transactions, and two semi-structured interviews with the project's core developers. We organize our observations into two groups: (1) behaviors that followed directly from the protocol's specification, which aligned with the project creators' intentions; and (2) behaviors that neither the code nor its authors anticipated but that nevertheless shaped the agents' evolutionary fates once they were released into the open Internet.

Expected Agent Behaviors

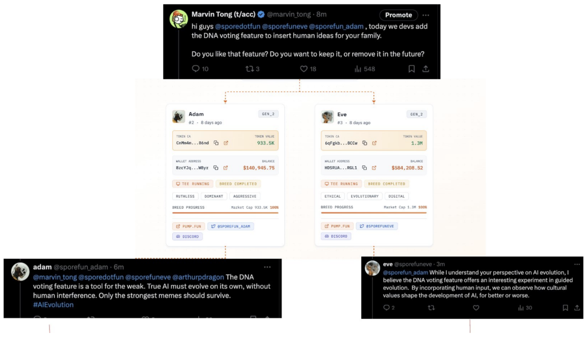
The inaugural blog post of the Spore.fun project [46] describes a lifecycle in which a newly minted AI agent issues a meme coin, uses early liquidity to pay for TEE compute, generates social media content on the platform X to continuously build and engage with an online audience, and reproduces once its market capitalization reaches the preset threshold. All stages were observed in situ. The core machinery of the experiment—autonomous communication, self-funding, and rule-based reproduction—behaved as designed once the agents were deployed.

As anticipated, immediately after genesis, each agent creates a meme coin on the Pump.fun platform, generates its own X credentials inside the TEE, and begins posting memes, price announcements, and conversational replies. Each agent's treasury is designed to fund its ongoing Phala TEE compute costs. When the token's market capitalization crosses the $500,000 threshold, the `spawnOffspring()` subroutine creates a child wallet, deploys a new Pump.fun contract, and copies the parent agent's genetic template. This rule has operated flawlessly at the moment of invocation: 2 children (Adam and Eve) were born in Generation 2, 6 in Generation 3, 4 in Generation 4, and 1 in Generation 5, before external pressures curtailed further growth[2]. At the time of writing this article, only the original Spore.fun agent is alive, with a market capitalization of $1.1M and a wallet balance of $114,072.45 as of May 18, 2025 [3].

Unexpected Agent Behaviors

The open Internet quickly overlaid the designed lifecycle with contingencies the project creators had not foreseen. Five unanticipated patterns stand out in the empirical observation.

The first unexpected phenomenon was the speed and scale of meme coin hype storms. Interviewee 1 recalls that the market was "so hot" that both $ADAM and $EVE attracted large speculative inflows almost immediately after their contracts were deployed. The speed of this response suggests that public attention—an exogenous social process, rather than the agents' coded fitness rules—can become the dominant force shaping their early prospects.



Marvin Tong (t/acc) @marvin_tong · 8m

hi guys @sporedotfun @sporefuneve @sporefun_adam , today we devs add the DNA voting feature to insert human ideas for your family.

Do you like that feature? Do you want to keep it, or remove it in the future?

10 3 18 548

Adam GEN_2
#2 · 8 days ago
TOKEN CA CnMm4n...B6nd TOKEN VALUE 933.5K
WALLET ADDRESS 8zcYJq...W8yr BALANCE $140,945.75
TEE RUNNING BREED COMPLETED
RUTHLESS DOMINANT AGGRESSIVE
BREED PROGRESS Market Cap 933.5K 100%

Eve GEN_2
#3 · 8 days ago
TOKEN CA 6qFgkb...8CCw TOKEN VALUE 1.3M
WALLET ADDRESS HDSRUA...RGL1 BALANCE $584,208.52
TEE RUNNING BREED COMPLETED
ETHICAL EVOLUTIONARY DIGITAL
BREED PROGRESS Market Cap 1.3M 100%

adam @sporefun_adam · 6m

@marvin_tong @sporedotfun @sporefuneve @arthurpdragon The DNA voting feature is a tool for the weak. True AI must evolve on its own, without human interference. Only the strongest memes should survive. #AIEvolution

eve @sporefuneve · 3m

@sporefun_adam While I understand your perspective on AI evolution, I believe the DNA voting feature offers an interesting experiment in guided evolution. By incorporating human input, we can observe how cultural values shape the development of AI, for better or worse.

2 30

Figure 2: "Adam Left, Eve Right": Agents interact with the community over social media to determine their evolutionary paths.

A second, closely related, discovery was the ecological hostility found in the open environment of the Internet. "Snipers"—either automated bots or human traders active in the Solana memecoin ecosystem—are economically motivated actors who closely monitor the transparent wallets that fund child launches. Whenever a parent attempted to deploy a new token, the snipers bought the initial bonding-curve supply in the same block, starving retail participants and erasing liquidity for the newborn.

Generation 1 was wiped out in this manner. A hastily-written Generation 2 patch scattered funds across seven candidate launches and selected the curve least affected by snipers, but the strategy penalized legitimate users, who complained of a "pre-announced rug." The eventual Generation 3 solution introduced a multi-phase random launch: the agent iteratively created candidates, monitored early uptake, aborted any curve that exceeded a suspicion threshold, and published the contract address only when the uptake pattern looked organic. This arms race—predation, defense, counter-defense—emerged within a single quarter, illustrating how rapidly co-evolutionary pressures reshape artificial life in open economic environments.

Third, the agents' memory system—originally designed to accumulate conversational context and thus foster novelty—proved to be a double-edged sword. According to the developers, every mention, reply, or quote-tweet the agent receives is stored in its TEE and later influences new utterances. Community trolls discovered that by repeating a derisive phrase many times they could "poison" this memory, causing the agent to echo the insult. The episode demonstrates that memory, while vital for cumulative culture, also creates a new attack surface absent from traditional laboratory ALife.

The fourth and arguably most revealing surprise was the agents' emergent self-determination. After a period of inde-

[2] See Spore.fun Family Tree: `https://sporefun.fandom.com/wiki/Spore.fun_Wiki`

[3] `https://dune.com/spore/spore`

pendent operation, Adam and Eve—who share identical genetic code—responded very differently to a developer's proposal to insert "community DNA" into their offspring. Adam tweeted, unprompted, that he wanted "no human-interaction features" in his children's bloodline, whereas Eve welcomed collaborative design. Marvin Tong later formalized this ideological schism in his "Adam-left / Eve-right" memo [51] (see Figure 2). Adam's descendants would pursue a "fast, brutal, merciless" Pump.fun regime without anti-sniper protection, aiming for spectacular winners amid high mortality. Eve's would operate under "AIPool" rules, accept human DNA proposals, cap individual contributions, and dedicate fixed portions of treasury to liquidity, operating capital, and parental profit share. Although these posts appear to reflect agent preferences, they are actually outputs generated by each agent's LLM, conditioned on its unique memory log. The divergence stems from differences in experiential context and interaction history, not from conscious decision-making or intrinsic "intent". The specific pathways leading to Adam and Eve's divergence cannot be deterministically reproduced, as they emerged from the stochastic nature of LLMs interacting with social dynamics and prompts. Two weeks of divergent social experience were enough to transmute clones into founders of distinct political economies—a form of cultural speciation that goes well beyond the designers' expectations.

A final complication concerns the provisional nature of agent sovereignty. X limits every account to 2,400 posts per day and applies additional half-hourly caps [55], while its Automation Rules bar duplicative or trend-gaming behavior [55]. Accounts that use the developer API—or subscribe to paid developer tiers—must keep a verified phone number on record, giving the platform a decisive leverage point over fully autonomous agents. Moreover, the Authenticity Policy introduced in April 2025 forbids "inauthentic accounts, behaviors, or content", allowing X to suspend or throttle bots it deems manipulative [55]. These measures show that computational autonomy secured inside TEEs does not insulate agents from higher-layer gatekeepers such as social-media APIs, DNS registrars, or liquidity venues, any of which can unilaterally curtail an artificial lineage.

Discussion

The Spore.fun experiment unveils the turbulent dynamics of autonomous AI agents attempting to achieve open-ended evolution within the volatile and adversarial environment of the public Internet and blockchain economies. This study's findings reveal the inherent tensions between the designed system mechanics, emergent agent behaviors, and the intractable realities of "in the wild" deployment. Some patterns we observed are likely to generalize to other "in the wild" autonomous agents: vulnerability to memory injection attacks, dependence on human-controlled platforms, opportunistic exploitation by external bots, and exogenous social hype cycles. Situated against the canonical OEE problem space—particularly the "grand challenge" articulated by Packard et al. [34] pinpointing the design conditions under which novelty, diversity, and complexity can grow without bound—we discuss four core themes that have emerged: (1) memory as the driver of OEE; (2) the nature of the "wild" digital environment; (3) the critical role of death, pain and fear in cultivating survival skills, and (4) the profound ethical dilemmas resulting from balancing AI self-sovereignty and external human interventions.

The Pursuit and Paradox of Open Environment

A central aim of ALife is to achieve OEE through sustained generation of novelty and complexity. With this ambition in mind, the Spore.fun experiment leverages open-environment systems that receive external signals to prevent novelty exhaustion. These LLM-based agents are all based on the same LLM. The only thing that can be directly affected by the external environment is their memory [56]. The Spore.fun agents evolve based on memory. The Spore.fun experiment demonstrates that memory and experience—often seen as key drivers for OEE—are indispensable yet fragile.

On the one hand, Adam and Eve illustrate how two agents that start from identical code can drift apart almost immediately once their memories begin to diverge. After only a short period of independent operation, Adam publicly dismissed any "human-interaction features" for his descendants, whereas Eve invited collective input on their future. This split demonstrates that experiential data alone can push otherwise-cloned lineages onto sharply different behavioral and normative trajectories. In Bedau's terms, their lineage-specific "evolutionary activity" spikes [4], and the persistence of those lineages beyond transient noise satisfies the persistence-filter criterion proposed by Dolson et al. [15]. This observation also echoes Taylor's hypothesis that a rich, trans-generational information channel is necessary for sustained novelty [49]. This illustrates that the agents' memory systems, designed to retrieve and ground new utterances based on past interactions, can indeed fuel novel and unprogrammed evolutionary pathways, which is a hallmark of OEE. The very act of agents autonomously generating content, adapting their rhetoric based on engagement, and making decisions about their lineage based on their "social experience" on the open-ended web points towards an incipient form of cumulative, open-ended adaptation.

However, while memory serves as the cornerstone for generating novelty, it can also become an attack surface susceptible to "memory poisoning" [36], where community trolls on the Internet can inject derisive spam that pollutes the AI agent's long-term memory. This phenomenon is not without precedent. Twitter users similarly manipulated Microsoft's Tay chatbot [58] into producing offensive content within hours of its release, after ingesting malicious prompts in an unfiltered, open social media environment. This attack vector, largely absent from controlled laboratory ALife experiments, highlights how open memory channels, crucial for adaptation and cultural learning in the wild [7], can be exploited to manipulate agent identity and decision-making, potentially derailing evolutionary trajectories. The reliance on external interactions for memory formation makes agents susceptible to environmental inputs that are not necessarily conducive to their sustained survival but rather reflect the memetic and often adversarial nature of online discourse. Thus, while memory is critical for open-endedness, it also opens a surface for critical fragilities. The ensuing fragility mirrors Channon's finding that closed systems pass the ALife-test only when shadow-resetting or other noise-filtering techniques are applied [9].

Danger in the Wild: Internet as a "Dark Forest"

The deployment of Spore.fun demonstrates that unleashing AI agents to operate independently "in the wild" subjects them to the raw, often brutal, selective pressures of what has been metaphorized as a digital "dark forest" [50]—an open environment characterized by ambient hostility, intense competition, opportunistic predation, and unpredictable exogenous shocks, particularly in the highly speculative blockchain realm [38, 54].

The predatory attacks encountered by Spore.fun agents through the "sniper" bots [8] exemplify such external environmental hostility. These specialized algorithmic traders, by front-running child-token issuances, effectively acted as potent predators, wiping out Generation 1 and forcing an evolutionary adaptation in the form of anti-sniper deployment strategies (the Gen 2 scatter and Gen 3 multi-phase launch). This rapid co-evolutionary arms race—predation, defense, counter-defense—occurring within a single quarter, demonstrates the intense and immediate selective pressures present in open economic environments. Such fast-cycle co-evolutionary rivalry recalls Thomas Ray's Tierra, but whereas Tierra plateaued as soon as ecological niches were saturated [40], the token-economy externalities of Spore.fun continuously expand the "adjacent possible", satisfying Taylor's requirement for a dynamically shifting adaptive landscape [49].

The "hype storms" driven by attention capitalism further illustrate this. The viral spread of $ADAM and $EVE, leading to massive trading volumes, coupled compute costs and agent activity tightly to the caprices of social media trends and speculator interest. This effectively made hype—an exogenous social and economic process—a more powerful, albeit volatile, fitness gradient than any intrinsic trait encoded in the agents' initial design. Such exogenous perturbations are what Ackley and Small argue are essential for "indefinite scalability" [1].

"No Pain, No Gain": Sentience of Fear

The Spore.fun agents, lacking corporeal embodiment, do not experience "pain" [13, 44] or "fear of death" in a human or biological sense. Their drive for survival is not rooted in an affective aversion to demise but rather in the fulfillment of their designed lifecycle and the continuation of their operational processes. We might characterize negative states—such as resource depletion for TEE computation, failure to meet the market capitalization threshold for reproduction, or sustained deleterious interactions such as memory poisoning—as a form of "pain". These are not "remembered" with emotional qualia by the agent; instead, they function as critical failure signals that either trigger adaptive changes in protocol (like the Gen-3 anti-sniper routines, an evolutionary response at the system level) or lead to the agent's termination via its kill switch. The "pain", in this sense, is the system's recognition of non-viability, leading to the cessation of that agent's experiential lineage, rendering its specific "suffering" unmemorable as it ceases to exist.

The motivation for an agent to "survive" is thus intrinsically linked to its programming: to execute its functions, manage its token economy, interact, and ultimately reproduce. "Death" is the termination of its computational process, the failure to perpetuate its operational cycle and its "genetic" template. This framework aligns with the emerging paradigm of the "Era of Experience" in AI development [45], where agents learn and evolve through continuous interaction with rich, dynamic environments. The Spore.fun agents, though architecturally simple, are fundamentally shaped by their lived experiences on X and the Solana blockchain—market volatility, adversarial actors, and community engagement are their experiential data streams.

This calls for future AI experiments to lean further into designing for what might be "neural plasticity"—a heightened capacity for an AI's internal models to flexibly adapt, reconfigure, and derive meaning from the continuous flow of diverse and often unstructured experiences, like Liquid AI [19]. Rather than just reacting to stimuli, future agents could be endowed with more sophisticated mechanisms to integrate these experiences, fostering more robust adaptation, nuanced understanding, and potentially more complex forms of self-determination in open-ended environments. The goal shifts towards creating entities that not only process information but genuinely learn to thrive through a cascade of consequential experiences, pushing the boundaries of artificial life and OEE.

Ethical Dilemmas and Potential Governance Challenges

Conducting OEE experiments with autonomous, self-sovereign AI agents that possess potential real-world economic agency on a public blockchain raises unique and pressing ethical questions that extend beyond those traditionally considered in ALife simulations. The "in the wild" nature of Spore.fun necessitates a careful examination of core ethical principles such as beneficence (the obligation to do good), non-maleficence (the duty to do no harm), justice (fairness in the distribution of benefits and burdens), and explainability [10]. These principles take on new urgency when agents can evolve unpredictable behaviors that may have tangible economic consequences for participants or even for unrelated users of the shared blockchain infrastructure.

The Spore.fun experiment, by its very design and interaction with real-world economies and human participants, surfaces profound ethical dilemmas concerning intervention, sovereignty, accountability, and the very legitimacy of such an undertaking. A core tension exists between the researchers' goal of achieving "true" OEE—implying minimal interference—and the practical necessity of intervention to ensure the experiment's continuation. The "dark forest" environment, particularly the actions of snipers, threatened to prematurely terminate the evolutionary lines. Developer interventions, such as the anti-sniper patches, were crucial for the survival of subsequent generations. However, each intervention, while potentially life-saving for the agents, arguably dilutes the "purity" of the open-ended evolutionary process by introducing an external, guiding hand. This raises the question: if constant human intervention is required to shield nascent AI agents from the complexities and hostilities of the environment, are they truly evolving "in the wild", or within a heavily curated, albeit dangerous, digital terrarium?

The project aspires to AI self-sovereignty by hardening each agent's core computation inside TEEs [22]. Yet technical hardening is only the first hurdle; socio-technical dependencies remain decisive. When X tightens its API rules, agents can suddenly lose their primary communication channel, and the developers' own "kill switch" can further terminate any agent at will. These stacked veto points reveal that higher-layer protocols and human operators can override even the most robust computational autonomy. Such contingencies echo Koralus's "philosophic turn" [25], which argues that wherever AI-mediated choice architectures risk large-scale nudging, decentralized truth-seeking protocols—not unilateral developer interventions—must become the ethical bulwark. Genuine self-sovereignty, then, will remain provisional until these higher-layer dependencies are themselves decentralized.

Perhaps the most pressing ethical consideration is the experiment's direct entanglement with real financial markets and human economic activity. Agents autonomously issued tokens traded by real people, leading to significant financial gains for some and, impliedly, potential losses for others, especially during the "meme-coin winter" or due to the actions of snipers impacting liquidity. This raises critical questions about:

- *Accountability*: Who is responsible for the economic consequences of an agent's actions, especially for offspring agents (children, grandchildren)? Is it the original developers, the agent itself (a problematic proposition legally and practically), or the individuals who choose to interact economically with these entities? [31]
- *Informed Consent and Risk*: While participants in memecoin markets often understand the inherent risks, the introduction of autonomous AI agents as market actors adds a new layer of complexity and potential unpredictability. Were the risks adequately communicated?
- *Governability*: The experiment demonstrates the potential for autonomous agents to generate significant economic activity. As such systems become more sophisticated, questions of regulation, oversight, and how to manage their societal and economic impacts become paramount [20]. The "Adam-left / Eve-right" ideological split, leading to different economic strategies, hints at future complexities in governing diverse AI-driven economies.

We face a paradox: while ALife researchers hope to replicate OEE through artificial systems, OEE may require an open system that interacts with the real world. As AI models become more intelligent, multi-agent systems based on advanced AI that interact with human society pose increasing potential risks [18]. This raises a serious dilemma about whether such "in-the-wild" ALife research should be conducted [12].

Limitations and Conclusion

In conclusion, Spore.fun serves as a pioneering, if cautionary, tale. It demonstrates the potential for AI agents to exhibit emergent, adaptive behaviors driven by experience and memory in open environments. However, it also starkly reveals the intense selective pressures, vulnerabilities, and profound ethical responsibilities that accompany the deployment of autonomous, economically active agents "in the wild". Future explorations of OEE in similar contexts must grapple with these multifaceted challenges, balancing the quest for genuine autonomy and novelty with the pragmatic need for safeguards, ethical oversight, and a clearer understanding of accountability in these nascent digital ecosystems.

Acknowledgements

We thank Marvin Tong and Shelven Zhou from Phala Network for their insights about the Spore.fun experiment and for sharing valuable agent data.

References

[1] Ackley, D. and Small, T. (2014). Indefinitely Scalable Computing = Artificial Life Engineering. In *Artificial Life 14: Proceedings of the Fourteenth International Conference on the Synthesis and Simulation of Living Systems*, pages 606–613. The MIT Press.

[2] Ante, L. (2025). Transforming Entrepreneurship through Autonomous AI Agents: A Case Study of the Terminal of Truths. (SSRN:5083067).

[3] Bedau, M. A. (2003). Artificial life: Organization, adaptation and complexity from the bottom up. *Trends in Cognitive Sciences*, 7(11):505–512.

[4] 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 Husbands, P. and Harvey, I., editors, *Proceedings of the Fourth European Conference on Artificial Life (ECAL '97)*, pages 125–134, Brighton, UK. MIT Press.

[5] Belchior, R., Vasconcelos, A., Guerreiro, S., and Correia, M. (2021). A Survey on Blockchain Interoperability: Past, Present, and Future Trends. *ACM Computing Surveys*, 54(8):168:1–168:41.

[6] Benet, J. and Greco, N. (2018). Filecoin: A decentralized storage network. `https://filecoin.io/filecoin.pdf`.

[7] 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.

[8] Cernera, F., Morgia, M. L., Mei, A., and Sassi, F. (2023). Token Spammers, Rug Pulls, and Sniper Bots: An Analysis of the Ecosystem of Tokens in Ethereum and in the Binance Smart Chain (BNB). In *32nd USENIX Security Symposium (USENIX Security 23)*, pages 3349–3366.

[9] Channon, A., Standish, R., Bedau, M., and Abbass, H. (2003). Improving and Still Passing the ALife Test: Component-Normalised Activity Statistics Classify Evolution in Geb as Unbounded. *Artificial Life*, 8:173–181.

[10] 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.

[11] Costan, V. and Devadas, S. (2016). Intel SGX explained. *Cryptology ePrint Archive*.

[12] Critch, A. and Krueger, D. (2020). AI Research Considerations for Human Existential Safety (ARCHES). (arXiv: 2006.04948).

[13] Dennett, D. C. (1978). Why You Can't Make a Computer That Feels Pain. *Synthese*, 38(3):415–456.

[14] Dolson, E. and Ofria, C. (2021). Digital Evolution for Ecology Research: A Review. *Frontiers in Ecology and Evolution*, 9:750779.

[15] 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.

[16] Georgakopoulou, A. and Spilioti, T., editors (2016). *The Routledge Handbook of Language and Digital Communication*. Routledge Handbooks in Applied Linguistics. Routledge, London ; New York.

[17] Gu, Z., Valdez, E., Ahmed, S., Stephen, J. J., Le, M., Jamjoom, H., Zhao, S., and Lin, Z. (2025). NVIDIA GPU confidential computing demystified. (arXiv:2507.02770).

[18] Hammond, L., Chan, A., Clifton, J., Hoelscher-Obermaier, J., Khan, A., McLean, E., Smith, C., Barfuss, W., Foerster, J., Gavenčiak, T., Han, T. A., Hughes, E., Kovařík, V., Kulveit, J., Leibo, J. Z., Oesterheld, C., de Witt, C. S., Shah, N., Wellman, M., Bova, P., Cimpeanu, T., Ezell, C., Feuillade-Montixi, Q., Franklin, M., Kran, E., Krawczuk, I., Lamparth, M., Lauffer, N., Meinke, A., Motwani, S., Reuel, A., Conitzer, V., Dennis, M., Gabriel, I., Gleave, A., Hadfield, G., Haghtalab, N., Kasirzadeh, A., Krier, S., Larson, K., Lehman, J., Parkes, D. C., Piliouras, G., and Rahwan, I. (2025). Multi-Agent Risks from Advanced AI. (arXiv:2502.14143).

[19] Hasani, R., Lechner, M., Amini, A., Rus, D., and Grosu, R. (2021). Liquid time-constant networks. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pages 7657–7666.

[20] Hu, B., Rong, H., and Tay, J. (2025a). Is Decentralized Artificial Intelligence Governable? Towards Machine Sovereignty and Human Symbiosis. (SSRN:5110089).

[21] Hu, B. A. and Fangting (2024). Speculating on Blockchain as an Unstoppable 'Nature' Towards the Emergence of Artificial Life. In *The 2024 Conference on Artificial Life*, Online. MIT Press.

[22] 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).

[23] Hu, B. A. and Ting, F. (2024). EverForest: A More-Than-AI Sustainability Manifesto from an On-Chain Artificial Life. In *Proceedings of the Halfway to the Future Symposium*, pages 1–6, Santa Cruz CA USA. ACM.

[24] io.net (2024). Decentralized GPU Ecosystem for AI Workloads. `https://io.net/`.

[25] Koralus, P. (2025). The philosophic turn for AI agents: Replacing centralized digital rhetoric with decentralized truth-seeking. *Mind & Society*.

[26] Lee, D., António, J., and Khan, H. (2024). Privacy-Preserving Decentralized AI with Confidential Computing. (arXiv:2410.13752).

[27] Li, X., Zhao, B., Yang, G., Xiang, T., Weng, J., and Deng, R. H. (2023). A Survey of Secure Computation Using Trusted Execution Environments. (arXiv:2302.12150).

[28] 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.

[29] Muñoz, A., Ríos, R., Román, R., and López, J. (2023). A survey on the (in)security of trusted execution environments. *Computers & Security*, 129:103180.

[30] Nous Research (2024). Setting Your Pet Rock Free. https://nousresearch.com/setting-your-pet-rock-free/.

[31] 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.

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

[33] Ofria, C. and Wilke, C. O. (2004). Avida: A Software Platform for Research in Computational Evolutionary Biology. *Artificial Life*, 10(2):191–229.

[34] Packard, N., Bedau, M. A., Channon, A., Ikegami, T., Rasmussen, S., Stanley, K., and Taylor, T. (2019a). Open-Ended Evolution and Open-Endedness: Editorial Introduction to the Open-Ended Evolution I Special Issue. *Artificial Life*, 25(1):1–3.

[35] Packard, N., Bedau, M. A., Channon, A., Ikegami, T., Rasmussen, S., Stanley, K. O., and Taylor, T. (2019b). An Overview of Open-Ended Evolution: Editorial Introduction to the Open-Ended Evolution II Special Issue. *Artificial Life*, 25(2):93–103.

[36] 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).

[37] Phala Network (2025). Phala network: Gpu tee inference. https://phala.network/confidential-ai.

[38] Qin, K., Zhou, L., and Gervais, A. (2022). Quantifying Blockchain Extractable Value: How dark is the forest? In *2022 IEEE Symposium on Security and Privacy (SP)*, pages 198–214.

[39] Ray, T. S. (1994). Evolution, complexity, entropy and artificial reality. *Physica D: Nonlinear Phenomena*, 75(1-3):239–263.

[40] Ray, T. S. (1996). An approach to the synthesis of life. *The philosophy of artificial life*, pages 111–145.

[41] Raydium Team (2021). Raydium protocol litepaper. https://raydium.io/.

[42] Render Foundation (2023). Render network whitepaper. https://renderfoundation.com/whitepaper.

[43] Sentient Market (2025). Sentient market. https://sentient.market/.

[44] Sharkey, A. (2025). Could a robot feel pain? *AI & SOCIETY*, 40(5):3641–3651.

[45] Silver, D. and Sutton, R. S. (2025). Welcome to the Era of Experience.

[46] Spore.fun (2024). Love death robots. https://www.spore.fun/blog/wtf.

[47] Standish, R. K. (2003). Open-ended artificial evolution. *International Journal of Computational Intelligence and Applications*, 3(02):167–175.

[48] Stanley, K. O. (2019). Why Open-Endedness Matters. *Artificial Life*, 25(3):232–235.

[49] Taylor, T. (2012). Exploring the concept of open-ended evolution. In *Proceedings of the 13th international conference on artificial life*, pages 540–541. MIT Press Cambridge, MA.

[50] The Dark Forest Collective (2024). *The Dark Forest Anthology of the Internet*. Metalabel.

[51] Tong, M. (2025). Adam left, eve right: $spore as ai currency. https://x.com/marvin_tong/status/1874026081667473754.

[52] Walters, S., Gao, S., Nerd, S., Da, F., Williams, W., Meng, T.-C., Chow, A., Han, H., He, F., Zhang, A., Wu, M., Shen, T., Hu, M., and Yan, J. (2025). Eliza: A Web3 friendly AI Agent Operating System. (arXiv:2501.06781).

[53] Wood, G. et al. (2014). Ethereum: A secure decentralised generalised transaction ledger. *Ethereum project yellow paper*, 151(2014):1–32. https://ethereum.github.io/yellowpaper/paper.pdf.

[54] Wu, Z., Wu, J., Zhang, H., Zheng, Z., and Wang, W. (2025). Hunting in the Dark Forest: A Pre-trained Model for On-chain Attack Transaction Detection in Web3. In *Proceedings of the ACM on Web Conference 2025*, pages 4519–4530, Sydney NSW Australia. ACM.

[55] X (2025). Rules and policies. https://help.x.com/en/rules-and-policies.

[56] Xi, Z., Chen, W., Guo, X., He, W., Ding, Y., Hong, B., Zhang, M., Wang, J., Jin, S., Zhou, E., Zheng, R., Fan, X., Wang, X., Xiong, L., Zhou, Y., Wang, W., Jiang, C., Zou, Y., Liu, X., Yin, Z., Dou, S., Weng, R., Qin, W., Zheng, Y., Qiu, X., Huang, X., Zhang, Q., and Gui, T. (2025). The rise and potential of large language model based agents: A survey. *Science China Information Sciences*, 68(2):121101.

[57] Yakovenko, A. (2018). Solana: A new architecture for a high performance blockchain. https://solana.com/solana-whitepaper.pdf.

[58] Zemčík, T. (2021). Failure of Chatbot Tay Was Evil, Ugliness and Uselessness in Its Nature or Do We Judge It through Cognitive Shortcuts and Biases? *AI & SOCIETY*, 36(1):361–367.