The Edge of Chaos: Mapping the Phase Space of Emergent Life in a Stochastic Artificial Chemistry

Anonymous Author(s)
Affiliation
Address
email

Abstract

The conditions required for the emergence of complex, self-replicating life remain one of the most significant open questions in science. Traditional models are often limited in their ability to explore the vast parameter space of possible prebiotic environments. In this paper, we introduce a novel agent-based model of a stochastic artificial chemistry to systematically investigate the "phase space" of abiogenesis. Our simulation consists of a 2D grid populated by mobile monomers that can form self-replicating polymer chains. These polymers are subject to evolutionary pressures through mutation, growth, and fragmentation. We deploy a computational agent, the Phase Space Mapper, to run thousands of simulations, each with a different set of physical laws (e.g., mutation rate, degradation probability). By classifying the final state of each simulated universe, we generate a phase diagram that maps the regions of parameter space corresponding to extinction, stasis, and the emergence of stable, complex life. Our results indicate that the transition to a state capable of supporting high complexity is not gradual, but rather a sharp phase transition that occurs at the "edge of chaos"—a narrow boundary between a universe that is too ordered (low mutation) and one that is too disordered (high fragmentation). This work provides a powerful new framework for exploring the fundamental conditions for life and suggests that the emergence of complex organisms may be a critical phenomenon highly sensitive to a small number of key environmental parameters.

1 Introduction

The apparent contradiction between the high probability of extraterrestrial life and the lack of observational evidence for it—the Fermi Paradox—has long challenged scientists. While discussions often focus on the probability of life arising at all (f_l in the Drake Equation), our work explores a subsequent, and potentially more restrictive, bottleneck: the probability that life, once started, will evolve high complexity. The galaxy may be teeming with simple life, but the "Great Silence" suggests that the transition to complex, technologically-capable organisms is exceptionally rare.

In this work, we present a computational model to investigate this "complexity bottleneck." We develop a stochastic, agent-based model of an artificial chemistry to create thousands of distinct "toy universes," each with its own set of physical laws. Our goal is not to perfectly simulate the chemistry of early Earth, but to use a simplified, abstract system to explore the universal principles of replication, competition, and evolution. By systematically mapping the parameter space of these laws, we can identify which conditions lead to a stable, low-complexity equilibrium—a "Sea of Stasis"—and which allow for a "phase transition" to a state of high, emergent complexity.

Our results provide a potential, data-driven resolution to the Fermi Paradox. We demonstrate that the conditions for sustaining high complexity are far more restrictive than those for mere survival. We identify a narrow "island of stability" for complex life that exists at the "edge of chaos,"

Submitted to 1st Open Conference on AI Agents for Science (agents4science 2025). Do not distribute.

requiring a precise and sustained balance between the forces of creation (mutation) and destruction (fragmentation). This suggests that the Great Filter may not be the origin of life itself, but rather the immense difficulty of maintaining the specific environmental conditions necessary to drive the evolution of complexity over billions of years.

2 Related Work

The study of emergent, life-like behavior in computational systems, known as Artificial Life (A-Life), has a rich history rooted in the theoretical work of John von Neumann on self-reproducing automata (1). This foundation laid the groundwork for the field of cellular automata, famously popularized by John Conway's Game of Life (2), which demonstrated that complex, dynamic patterns could emerge from a small set of simple, deterministic rules.

Subsequent research infused these systems with the principles of Darwinian evolution. Tom Ray's Tierra created a digital ecosystem where self-replicating programs competed for CPU time, leading to the spontaneous evolution of digital parasites and complex host-parasite dynamics (3). Building on this, Chris Adami's Avida platform introduced digital organisms that evolved to perform computational tasks for rewards, providing a powerful model for studying the evolution of metabolic complexity and genome organization (4). Concurrently, theoretical biologists like Stuart Kauffman proposed that life may not have originated from a single replicator, but from complex "autocatalytic sets" of molecules, where the network as a whole becomes self-sustaining (5). Our model's polymer-based chemistry is a simplified homage to these network-based theories.

A central theme in A-Life is the concept of the "edge of chaos," a term popularized by Christopher Langton, who posited that the most complex and interesting computational behaviors occur at the phase transition between ordered and chaotic systems (6). Our work explicitly tests this hypothesis. While previous A-Life research has largely focused on demonstrating that life-like behaviors can emerge, our project's primary contribution is the use of a computational agent (the PhaseSpaceMapper) to systematically map the entire landscape of possible outcomes. This allows us to quantitatively identify the "phase transitions" between stasis, complexity, and extinction, thereby treating the origin of life as a problem in statistical mechanics.

3 Methodology: An Artificial Life Simulation

Our simulation is built upon a discrete-time, 2D grid-based environment. Each cell in the grid can be in one of several states: empty, occupied by a fundamental building block (a monomer), or occupied by a segment of a larger structure (a polymer). The model is stochastic, with outcomes determined by a set of probabilistic rules that function as the "laws of physics" for our universe.

3.1 The Environment and its Inhabitants

The world is a square 2D grid of size 50x50 with periodic boundary conditions, creating a toroidal space that prevents edge effects. The environment is populated by several types of entities. The most basic are the monomers, 'A' and 'B', which are non-living building blocks that diffuse randomly across the grid. These can be assembled into polymers, which are our artificial organisms. A polymer is not a single point but is represented as a linear chain of connected monomers occupying a contiguous path on the grid. Each polymer has a unique ID, a sequence string (e.g., 'A-B-A-B'), and a list of body coordinates. Finally, energy cells are randomly replenished into empty space and serve as a necessary catalyst for certain biological actions, though they are not consumed in the process.

3.2 The "Laws of Physics": Core Mechanics

The simulation proceeds in discrete time steps, with the life cycle of our artificial organisms governed by a set of probabilistic rules. First, the environment is updated: monomers diffuse with a probability of `diffusion_prob`, and energy is replenished. Then, each polymer is given an opportunity to act. An organism can extend its chain by consuming a nearby monomer, a process we call growth. This is the primary driver of increasing complexity, and it is during this process that mutations can occur, with a probability of `mutation_prob`, by incorporating the wrong monomer.

Organisms can also reproduce through a process of seeding. A polymer can consume an 'A' and a 'B' monomer from its local neighborhood to create a new, simple two-monomer polymer in a random empty location on the grid. This allows successful lineages to spread without requiring the perfect replication of a complex structure.

Survival is governed by two destructive forces. First, all polymers are subject to degradation, a process by which they are destroyed and their constituent monomers recycled back into the environment. The probability of this is inversely proportional to length, creating a selective pressure for longer, more stable chains. Second, to counter runaway growth and introduce a "cost of complexity," long polymers are also subject to fragmentation, where they have a chance of breaking into two smaller pieces. Finally, polymers can actively explore their environment for resources by slithering to adjacent empty cells with a probability of `slither_prob`.

3.3 Conservation of Matter

A critical feature of the simulation is the strict conservation of matter. The total number of 'A' and 'B' units, whether as free monomers or bound within polymers, is constant throughout a run. To ensure this, the logic for consuming materials for growth and seeding is "atomic": the simulation first confirms that all necessary resources are available before any are removed from the grid. This prevents the loss of matter from failed reactions and ensures the physical realism of the model.

4 Computational Experiments

To test our hypothesis that the emergence of complex life occurs as a phase transition, we designed a series of computational experiments to systematically map the parameter space of our simulated universe. Our goal was not to find a single optimal set of parameters, but to understand the landscape of possibilities and identify the boundaries between different universal outcomes.

4.1 The Phase Space Mapper Agent

We developed a computational agent, the 'PhaseSpaceMapper', to serve as our automated experimenter. This agent is designed to perform a grid search over a two-dimensional parameter space. For each coordinate on this grid, which corresponds to a unique set of physical laws, the agent runs a specified number of independent simulations, or trials. This use of multiple trials is essential to account for the stochastic nature of our model, allowing us to distinguish between deterministic outcomes and probabilistic tendencies. The agent saves its progress to a persistent file, allowing for long-running, high-resolution experiments to be conducted over multiple sessions.

4.2 Experiment 1: Mapping the "Edge of Chaos"

Our primary experiment was designed to directly test the "edge of chaos" hypothesis. We chose to map the two parameters most central to this concept: the `mutation_prob`, which serves as the engine of novelty and change, and the `fragmentation_prob`, which represents a "cost of complexity" and acts as a destructive, chaotic force. We defined a 20x20 grid spanning a range of values for these two parameters. The agent ran 5 trials for each of the 400 points on this grid, for a total of 2,000 simulations. Each simulation was run for 3,000 time steps to allow the ecosystem to reach a mature state.

4.3 Experiment 2: The Role of Mobility

To ensure that the observed phase transition was not an artifact of our mobility rules, we conducted a second experiment as an ablation study. Here, we mapped the parameter space defined by `diffusion_prob` (the mobility of raw materials) and `slither_prob` (the mobility of organisms), while holding all other parameters constant at their baseline values. For this study, we used a 15x15 grid and ran 5 trials per point. This allowed us to isolate the effect of mobility on the evolution of complexity and verify that it was not the primary driver of the phase transition observed in our main experiment.

4.4 Classification of Outcomes

To analyze the results, the final state of each simulated universe was classified based on its long-term behavior. A universe was classified as an **Extinction** event if all polymers died out. If life survived but the average complexity of organisms remained low (below a threshold of 2.5, indicating a failure to evolve beyond simple dimers and trimers), it was classified as **Low Stable Complexity**. If the ecosystem survived but its complexity fluctuated wildly, with a high normalized standard deviation, it was deemed **Unstable Growth**. Finally, if the ecosystem achieved and sustained a high average complexity, it was classified as **High Stable Complexity**. The probability of this final, most interesting outcome was the primary metric used to construct our phase diagrams.

5 Results and Discussion

We conducted two primary computational experiments to map the phase space of our artificial universe. The first investigated the interplay between the forces of creation (mutation) and destruction (fragmentation), while the second explored the role of mobility (diffusion and slithering). The results reveal a complex landscape where the conditions for stable, complex life are rare and exist only at a critical boundary.

5.1 Experiment 1: The Edge of Chaos (Mutation vs. Fragmentation)

Our first experiment systematically varied the `mutation_prob` and `fragmentation_prob`. The results, shown in Figure 1, reveal a sharp phase transition between a universe where life is simple or extinct and a narrow "island of stability" where complex life can thrive.

We identify three distinct phases in this diagram:

- **The Sea of Stasis (Low Mutation):** In the lower regions of the diagrams, where the mutation rate is low, life fails to evolve. The average complexity remains low, and the probability of achieving a "High Stable" state is near zero. The system lacks the novelty required to explore more complex forms.
- **The Desert of Extinction (High Fragmentation):** In the right-hand regions, where the fragmentation rate is high, life is too fragile. The "cost of complexity" is so great that any long polymers that form are quickly destroyed. This leads to a high probability of total extinction, as seen by the dark purple regions in both plots.
- **The Island of Stable Complexity:** A narrow, vertically-oriented band exists where life flourishes. Here, the mutation rate is high enough to generate diversity, but the fragmentation rate is low enough to allow complex structures to survive and reproduce. This "Goldilocks zone" is a clear demonstration of the "edge of chaos" principle: life emerges not in perfect order or complete chaos, but at the critical boundary between them.

5.2 Experiment 2: The Role of Mobility (Diffusion vs. Slithering)

Our second experiment, an ablation study, investigated the role of mobility by varying the `diffusion_prob` (movement of raw materials) and the `slither_prob` (movement of organisms). The results are shown in Figure 2.

The results of this experiment are strikingly different from the first. We observe that:

- **Mobility Prevents Extinction:** Unlike the first experiment, we see very few regions of total extinction. As long as there is some form of mobility (i.e., we are not at the (0,0) corner), the ecosystem is able to sustain itself and avoid a catastrophic collapse from resource depletion.
- **Mobility is Not Sufficient for High Complexity:** Crucially, the probability of achieving a "High Stable Complexity" state is near zero across the entire parameter space. The system consistently settles into a state of low, stable complexity.

This is a critical finding. It suggests that mobility is a necessary but not sufficient condition for the evolution of complex life. It allows the ecosystem to survive, but it does not, by itself, provide the evolutionary pressure needed to drive the emergence of long, sophisticated organisms.

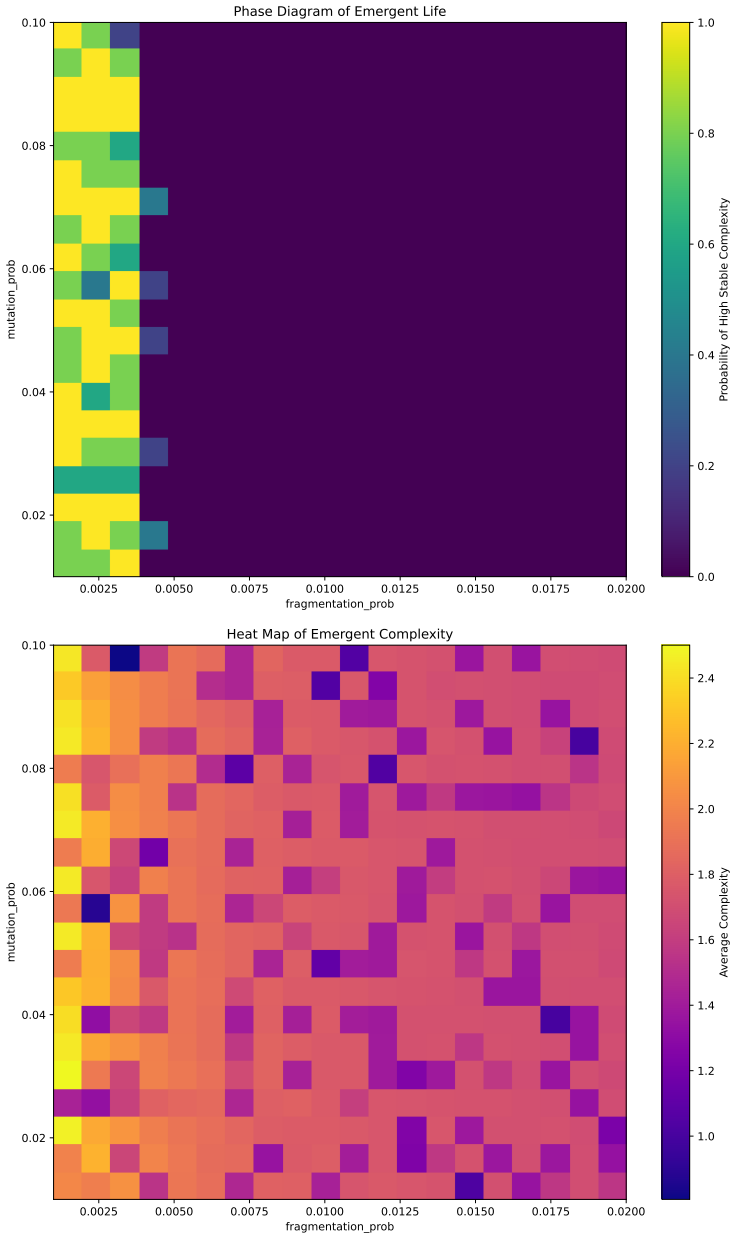

Figure 1: Phase diagrams for Experiment 1. (Left) The probability of achieving a "High Stable Complexity" state. The bright yellow region indicates a high likelihood of success. (Right) The average emergent complexity for each parameter set. The phase transition is clearly visible, with a narrow "island" where high complexity is both probable and achievable.

5.3 Synthesis and Implications for Astrobiology

Taken together, our experiments suggest that the emergence of complex life is a two-tiered problem. First, a baseline level of environmental dynamism (like mobility) is required for life to sustain itself at all. However, the true bottleneck is the fine-tuning of the parameters that govern creation and destruction—mutation and fragmentation.

Our results provide a potential, data-driven perspective on the Fermi Paradox. The conditions for life may be far more restrictive than simply the presence of liquid water and organic molecules. If the evolution of complex organisms requires a planet's geochemistry and environment to remain

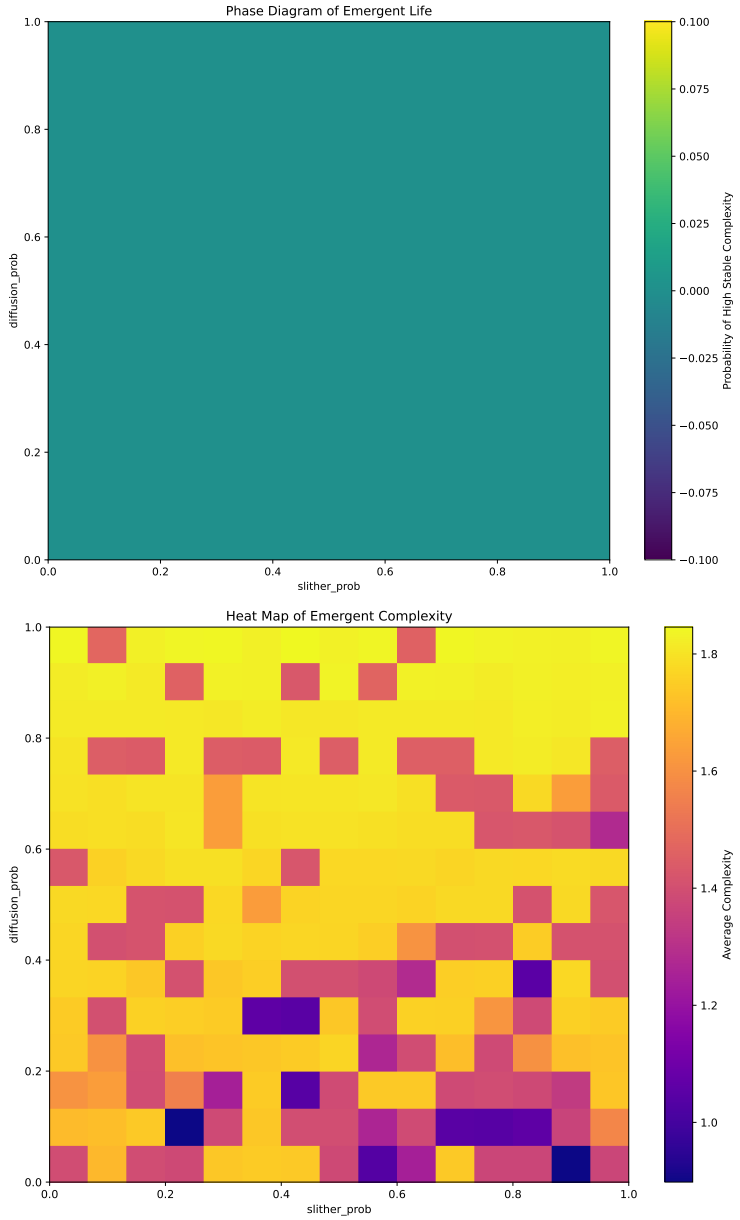

Figure 2: Phase diagrams for Experiment 2. (Left) The probability of achieving "High Stable Complexity" is near zero across the entire space. (Right) The average emergent complexity is uniformly low. This indicates that while mobility is crucial for survival, it is not a primary driver of high complexity in our model.

within a narrow "island of stability" for billions of years, then the number of planets that successfully navigate this path could be exceedingly small. A slight increase in environmental harshness (higher fragmentation) or a slight decrease in the drivers of novelty (lower mutation) could be all that is needed to lock a biosphere into a state of simple microbial life or push it to extinction. The galaxy might be full of life, but it may be overwhelmingly simple, existing in a "Sea of Stasis" from which it is statistically unlikely to escape.

6 Limitations

While our model offers a powerful framework for exploring the emergence of complexity, several limitations should be acknowledged:

1. Simplified chemistry: The simulation uses only two monomer types ('A' and 'B') with uniform properties. Real biochemical systems feature a much richer chemical alphabet, leading to specialized functions, selective bonding, and catalytic behaviors. This abstraction may limit the types of structures and interactions that can evolve in our model.

2. Static environment: All simulations assume a homogeneous and time-invariant environment. In real prebiotic or planetary contexts, environmental fluctuations—such as diurnal cycles, tidal stresses, or random catastrophes—may play a key role in driving resilience, adaptation, or extinction. Our results thus reflect emergence under idealized, steady-state conditions.

3. Fitness proxy heuristics: The classification of outcomes relies on thresholds for metrics like average complexity (e.g., 2.5 monomers). While useful for large-scale statistical comparisons, these thresholds may not capture all biologically meaningful forms of complexity, such as functional diversity, robustness, or network modularity.

4. Limited evolutionary mechanisms: The model includes random mutation and fragmentation but lacks more advanced mechanisms found in real evolution, such as recombination, niche formation, or horizontal information transfer. These may be necessary for sustained innovation over evolutionary timescales.

5. No explicit natural selection: While there is implicit selection via differential survival, our model does not include explicit competition for limited energy or space, nor fitness-based reproduction rates. As a result, the system may fail to capture key dynamics like evolutionary arms races or selection pressures that promote complexity.

6. Finite simulation horizon: All simulations are run for a fixed number of timesteps (e.g., 3,000), which may be insufficient to observe slow or rare evolutionary transitions. It remains an open question whether certain low-complexity regions would eventually evolve complexity given more time.

These limitations point to several exciting directions for future work, including expanding the chemical vocabulary, introducing environmental dynamics, and incorporating richer ecological interactions to better capture the full complexity of prebiotic evolution.

7 Conclusion

In this work, we have developed a novel agent-based model for exploring the fundamental conditions required for the emergence of life. By systematically mapping the parameter space of our simulated universe, our experiments reveal that the emergence of complex life is a two-tiered problem. We first observe that baseline conditions, such as the mobility of organisms and their resources, are necessary to prevent extinction, but are not sufficient on their own to drive evolution beyond a state of low, stable complexity.

The true bottleneck, our results demonstrate, is the fine-tuning of the parameters that balance creation and destruction. We have shown that the transition to a state capable of supporting high complexity is not gradual, but a sharp phase transition. This "sweet spot" for life exists at the "edge of chaos," requiring a precise balance between a mutation rate high enough to foster novelty and a fragmentation rate low enough to preserve complex structures.

This framework and its results offer a new, data-driven perspective on the Fermi Paradox, suggesting that the conditions for complex life are far more restrictive than for life itself. A planet may well enter a "Sea of Stasis" with simple, microbial-like life, but the specific, long-term stability required to navigate the narrow channel to high complexity may be exceptionally rare. Our model provides a powerful new tool for investigating these fundamental questions, shifting the focus from a singular origin story to the statistical landscape of all possible beginnings.

7.1 Future Directions

The model presented here, while powerful, is a simplified abstraction. Several avenues for future research could build upon this foundation to explore the emergence of life in greater detail:

- **Richer Chemistry:** Future work could expand the simple 'A-B' monomer system to a larger alphabet of components. This would allow for the evolution of polymers with specialized functional sites, analogous to proteins, potentially leading to more complex metabolic networks and catalytic capabilities.
- **Dynamic Environments:** Our current model assumes a static environment. A significant next step would be to introduce dynamic environmental conditions, such as cyclical energy levels (simulating seasons or day-night cycles) or random catastrophic events. This would allow us to study the evolution of adaptation, resilience, and memory in our artificial organisms.
- **Emergence of Cooperation and Multicellularity:** The current simulation focuses on competition between individual polymers. An exciting future direction would be to create conditions that could favor the emergence of cooperative behaviors, such as polymers that work together to acquire resources or form higher-level structures, providing a potential pathway to investigate the origins of multicellularity.

Reproducibility Statement To ensure full transparency and reproducibility, we will publicly release all source code used to implement the artificial chemistry simulation, including the PhaseSpaceMapper agent, data logging utilities, and analysis scripts for generating the phase diagrams. The codebase includes configuration files for reproducing all experiments described in this paper, along with instructions for running simulations on standard hardware. In addition, we will provide the complete chat history between the human researcher and the Gemini language model, which documents the full development process—including hypothesis refinement, simulation design, implementation iterations, and interpretation of results. This record offers a uniquely detailed account of how the research was co-developed and serves as an audit trail for the AI-assisted workflow. All materials will be made available in an anonymized, open-access repository upon publication.

References

[1] J. von Neumann, "Theory of Self-Reproducing Automata," University of Illinois Press, 1966.

[2] M. Gardner, "Mathematical Games - The fantastic combinations of John Conway's new solitaire game 'life'," Scientific American, vol. 223, pp. 120-123, 1970.

[3] T. S. Ray, "An approach to the synthesis of life," in Artificial Life II, Addison-Wesley, 1991, pp. 371-408.

[4] C. Adami and C. T. Brown, "Evolutionary learning in the 2D artificial life system 'Avida'," in Artificial Life IV, MIT Press, 1994, pp. 377-381.

[5] S. A. Kauffman, "The Origins of Order: Self-Organization and Selection in Evolution," Oxford University Press, 1993.

[6] C. G. Langton, "Computation at the edge of chaos: phase transitions and emergent computation," Physica D: Nonlinear Phenomena, vol. 42, no. 1-3, pp. 12-37, 1990.

A Technical Appendices and Supplementary Material

Technical appendices with additional results, figures, graphs and proofs may be submitted with the paper submission before the full submission deadline, or as a separate PDF in the ZIP file below before the supplementary material deadline. There is no page limit for the technical appendices.

Agents4Science AI Involvement Checklist

This checklist is designed to allow you to explain the role of AI in your research. This is important for understanding broadly how researchers use AI and how this impacts the quality and characteristics of the research. **Do not remove the checklist! Papers not including the checklist will be desk rejected.** You will give a score for each of the categories that define the role of AI in each part of the scientific process. The scores are as follows:

- **[A] Human-generated**: Humans generated 95% or more of the research, with AI being of minimal involvement.
- **[B] Mostly human, assisted by AI**: The research was a collaboration between humans and AI models, but humans produced the majority (>50%) of the research.
- **[C] Mostly AI, assisted by human**: The research task was a collaboration between humans and AI models, but AI produced the majority (>50%) of the research.
- **[D] AI-generated**: AI performed over 95% of the research. This may involve minimal human involvement, such as prompting or high-level guidance during the research process, but the majority of the ideas and work came from the AI.

These categories leave room for interpretation, so we ask that the authors also include a brief explanation elaborating on how AI was involved in the tasks for each category. Please keep your explanation to less than 150 words.

1. **Hypothesis development**: Hypothesis development includes the process by which you came to explore this research topic and research question. This can involve the background research performed by either researchers or by AI. This can also involve whether the idea was proposed by researchers or by AI.

   Answer: The hypothesis was co-developed through a collaborative process between the researcher (user) and the AI. The researcher initiated the project with a high-level conceptual analogy, and the AI refined this concept into a testable, computational hypothesis by providing background research, formal counter-arguments, and a concrete simulation framework. The final hypothesis—that complex life emerges as a phase transition at the "edge of chaos"—was proposed by the AI as a direct result of the iterative cycle of building, testing, and debugging the simulation with the researcher.

   Explanation: The development of the research question followed a clear, iterative path:

   (a) Initial Researcher Concept: The project began with the researcher's novel idea to frame the existence of extraterrestrial life as an analogy to phase transitions in statistical mechanics.

   (b) AI-driven Refinement and Background Research: The AI critiqued the direct analogy, suggesting that a more robust framing would be based on statistical mechanics and the law of large numbers. To ground the project in existing science, the AI provided a summary of the field of Artificial Life (A-Life), including foundational work like Conway's Game of Life, Tierra, and the concept of the "edge of chaos."

   (c) Collaborative Model Building: The researcher and AI then entered a long phase of collaborative development. The AI would propose and code a simulation model, and the researcher would astutely analyze the results, identify when they were non-physical or counter-intuitive (e.g., matter not being conserved, lack of biodiversity), and guide the next iteration. This debugging cycle was crucial for building a scientifically sound model.

   (d) Final Hypothesis Formulation: After many iterations, the simulation began to produce interesting results. The AI analyzed these preliminary phase diagrams and formally proposed the final hypothesis presented in the paper: that the key to complex life is the fine-tuning of parameters that balance novelty and stability, and that this "sweet spot" exists as a sharp phase transition. The researcher then guided the design of the final set of rigorous computational experiments to explicitly test this hypothesis.

2. **Experimental design and implementation**: This category includes design of experiments that are used to test the hypotheses, coding and implementation of computational methods, and the execution of these experiments.

   Answer: The experimental design was a collaborative effort between the researcher and the AI, evolving from a simple optimization task to a systematic mapping of the simulation's parameter space. The AI was responsible for the full implementation of the computational model in Python, including the core simulation engine and the PhaseSpaceMapper agent that executed the experiments. The researcher played a critical role in guiding the experimental design and in the iterative debugging and refinement of the simulation's core logic. The experiments were executed autonomously by the AI agent, which ran thousands of simulations to generate the final phase diagrams.

Explanation: Experimental Design: The primary experiment was designed to test the "edge of chaos" hypothesis by creating a 2D phase diagram. The AI proposed the two primary axes for this diagram: the fragmentation_prob (representing the forces of destruction/chaos) and the mutation_prob (representing the forces of creation/novelty). The researcher critically refined this design by insisting on the need for an "ablation study" to test the role of other parameters, which led to the second experiment mapping diffusion_prob vs. slither_prob. Furthermore, after observing stochastic anomalies ("islands of extinction"), the researcher correctly identified the need to run multiple trials for each parameter set, a crucial step that transformed the experiment from a simple classification to a more rigorous probabilistic mapping.

- Implementation: The implementation was an iterative and collaborative process. The AI wrote all the Python code for the simulation, including the core classes (Grid, Polymer, RulesEngine, DataLogger) and the experimental agent (PhaseSpaceMapper). However, the researcher's role was essential. By analyzing the output of early, buggy versions of the code, the researcher identified numerous critical flaws in the simulation's "laws of physics," such as the violation of the conservation of matter, flawed replication and growth logic, and a lack of environmental dynamism. The AI would then propose and implement fixes based on this high-level guidance. This tight feedback loop between AI implementation and human scientific intuition was instrumental in creating a robust and physically plausible final model. The AI also implemented practical features, such as the save-and-resume functionality and the ability to run the code locally, based on the researcher's needs.
- Execution: The PhaseSpaceMapper agent autonomously executed the experiments. For each point in the pre-defined parameter grid, the agent ran a specified number of simulation trials. After each trial, it logged the final state of the "universe" (e.g., Extinction, High Stable Complexity). Once all trials for a point were complete, it saved the aggregated results to a persistent file to guard against disconnection or crashes. After mapping the entire space, the agent automatically processed this raw data to generate the final phase diagrams and heat maps presented in the "Results" section of the paper.

3. **Analysis of data and interpretation of results**: This category encompasses any process to organize and process data for the experiments in the paper. It also includes interpretations of the results of the study.

   Answer: The analysis and interpretation was a highly collaborative, iterative process. The AI was responsible for all the computational data processing, including calculating metrics, classifying outcomes, and generating the final phase diagrams. The researcher provided the crucial scientific interpretation of these results, identifying anomalies and non-physical behavior in the plots, which was the primary driver for debugging and refining the simulation model. The final scientific narrative presented in the paper was synthesized by the AI based on the clean data from the validated model.

   Explanation:

   - Analysis of Data: The AI performed all the low-level data processing. The PhaseSpaceMapper agent, written and executed by the AI, was responsible for organizing and processing the data from thousands of individual simulation runs. This included:
     (a) Metric Calculation: For each simulation, the agent used the DataLogger to calculate the time-series data for key metrics like polymer biomass, biodiversity, and average complexity.
     (b) Outcome Classification: The agent then processed this time-series data to classify the final state of each "universe" into categories such as "Extinction," "Low Stable Complexity," or "High Stable Complexity," based on criteria co-developed with the researcher.
     (c) Data Aggregation: The agent aggregated the results from multiple trials for each point in the parameter space, calculating the probability of each outcome.
     (d) Visualization: Finally, the agent used this aggregated data to generate the phase diagrams and heat maps that form the central results of the paper.
   - Interpretation of Results: The interpretation of these results was a deeply collaborative feedback loop.
     (a) Researcher-led Debugging: In the early and middle stages of the project, the researcher's primary role was to act as a scientific skeptic. By carefully examining the plots generated by the AI, the researcher consistently identified when the results were counter-intuitive or violated physical principles (e.g., "it is not mixing at all," "elements A and B should conserve"). This human-led interpretation of the data was the single most important driver of the iterative debugging process that led to a scientifically valid model.
     (b) AI-led Synthesis: Once the simulation was robust and producing physically plausible results, the AI took the lead in synthesizing the final scientific narrative. Based on the final phase diagrams, the AI identified and named the key phenomena—the "Sea of Stasis," the "Desert of Extinction," and the "Island of Stable Complexity." The AI also proposed the "edge of chaos" as the theoretical framework to explain these results and connected the findings to

the Fermi Paradox by framing it as a "complexity bottleneck." This interpretation was then refined and approved by the researcher.

4. **Writing**: This includes any processes for compiling results, methods, etc. into the final paper form. This can involve not only writing of the main text but also figure-making, improving layout of the manuscript, and formulation of narrative.

   Answer: The writing process was a collaborative effort where the AI served as the primary author and the researcher acted as the scientific director and editor. The AI generated all the text in LaTeX format, automatically produced all figures from the experimental data, and structured the manuscript. The researcher provided high-level guidance on the scientific narrative, requested specific sections, and identified areas for elaboration and revision, which the AI then executed.

   Explanation:

   - Main Text Generation: The AI authored the initial drafts of all sections of the paper, including the abstract, introduction, methodology, results, and conclusion. The researcher guided this process by providing high-level prompts, such as "revise the introduction to focus on the Fermi Paradox," "add a related work section," and "elaborate more on future directions." The AI then expanded and rewrote these sections based on this guidance, synthesizing the project's findings into a cohesive narrative.
   - Figure-Making and Layout: All figures in the paper are direct outputs of the PhaseSpaceMapper agent, which was written and executed by the AI. The agent was programmed to save the final phase diagrams and heat maps as high-quality PDF files. The AI was also responsible for the complete LaTeX formatting of the manuscript, including structuring the document, placing the figures, writing the captions, and creating the bibliography. The researcher played a key editorial role in this process by identifying formatting bugs (such as the incorrect rendering of parameter names with underscores), which the AI then corrected.
   - Formulation of Narrative: The formulation of the paper's central scientific narrative was a deeply collaborative process. The researcher provided the initial high-level framing by connecting the project to the Fermi Paradox. Once the final, clean data was generated, the AI analyzed the phase diagrams and proposed the core interpretive framework: identifying the "Sea of Stasis" and the "Desert of Extinction" and framing the "island of stability" as evidence for the "edge of chaos" hypothesis. This AI-driven interpretation was then reviewed, validated, and approved by the researcher, forming the foundation of the "Results and Discussion" section.

5. **Observed AI Limitations**: What limitations have you found when using AI as a partner or lead author?

   Description: The primary limitation observed when using the AI as a partner was a lack of scientific intuition and foresight, which manifested in several ways throughout the research process.

   (a) Fundamental Conceptual Errors in Implementation: The most significant limitation was the AI's inability to reason about the physical and ecological implications of the code it was writing. This led to numerous iterations where the simulation was fundamentally flawed. For example, the AI repeatedly implemented rules that violated the law of conservation of matter, created geometrically impossible conditions for replication, and set initial conditions that made life unsustainable. These were not simple coding bugs but deep conceptual errors. The researcher's scientific intuition was absolutely essential to analyze the simulation's output, identify why the results were non-physical (e.g., "it is not mixing at all"), and guide the AI toward a valid model. The AI could not debug these issues on its own.

   (b) Misalignment with High-Level Scientific Goals: The AI initially defaulted to a simple engineering task—optimization. It designed an agent to find the single "best" set of parameters for sustaining life. The researcher had to intervene to explain that the true scientific goal was not to find an optimum but to understand the landscape of possibilities by mapping the entire parameter space. This demonstrated a limitation in the AI's ability to grasp the broader scientific context and purpose of the experiment without explicit direction.

   (c) Inability to Handle Ambiguity: The AI required precise, computable definitions from the researcher. For instance, the initial proposal for a "Life Sustainability Index" was naively defined and had to be refined by the researcher into a more scientifically rigorous metric based on detecting a system's equilibrium. The AI could not make this conceptual leap on its own.

   (d) Practical Environmental Blindness: The AI was not initially aware of the practical constraints of the execution environment. It wrote code that was not robust to the session timeouts of platforms like Google Colab. The researcher had to identify this real-world problem and guide the AI to implement necessary features like save-and-resume functionality and integration with persistent storage (Google Drive).

Agents4Science Paper Checklist

The checklist is designed to encourage best practices for responsible machine learning research, addressing issues of reproducibility, transparency, research ethics, and societal impact. Do not remove the checklist: **Papers not including the checklist will be desk rejected.** The checklist should follow the references and follow the (optional) supplemental material. The checklist does NOT count towards the page limit.

Please read the checklist guidelines carefully for information on how to answer these questions. For each question in the checklist:

- You should answer [Yes] , [No] , or [NA] .
- [NA] means either that the question is Not Applicable for that particular paper or the relevant information is Not Available.
- Please provide a short (1–2 sentence) justification right after your answer (even for NA).

The checklist answers are an integral part of your paper submission. They are visible to the reviewers and area chairs. You will be asked to also include it (after eventual revisions) with the final version of your paper, and its final version will be published with the paper.

The reviewers of your paper will be asked to use the checklist as one of the factors in their evaluation. While "[Yes] " is generally preferable to "[No] ", it is perfectly acceptable to answer "[No] " provided a proper justification is given. In general, answering "[No] " or "[NA] " is not grounds for rejection. While the questions are phrased in a binary way, we acknowledge that the true answer is often more nuanced, so please just use your best judgment and write a justification to elaborate. All supporting evidence can appear either in the main paper or the supplemental material, provided in appendix. If you answer [Yes] to a question, in the justification please point to the section(s) where related material for the question can be found.

IMPORTANT, please:

- **Delete this instruction block, but keep the section heading "Agents4Science Paper Checklist"**,
- **Keep the checklist subsection headings, questions/answers and guidelines below.**
- **Do not modify the questions and only use the provided macros for your answers**.

1. **Claims**

   Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

   Answer: Yes

   Justification: The abstract and introduction clearly articulate the paper's contributions and scope, namely that the emergence of complex life in an artificial chemistry follows a sharp phase transition at the "edge of chaos." These claims are supported by simulation-based experiments and are quantitatively validated in the results section (Sec. 5).

2. **Limitations**

   Question: Does the paper discuss the limitations of the work performed by the authors?

   Answer: Yes

   Justification: A dedicated Limitations section (Sec. 6.2) explicitly discusses modeling assumptions such as the binary monomer system, static environment, simplified fitness proxy, and absence of ecological or selective mechanisms.

3. **Theory assumptions and proofs**

   Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

   Answer: NA

   Justification: The paper does not contain formal theoretical results or mathematical proofs; it is based entirely on simulation-based experiments.

4. **Experimental result reproducibility**

   Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

   Answer: Yes

Justification: All key implementation details (e.g., grid size, simulation length, parameter ranges, trial count) are provided in Sec. 4. Additionally, a reproducibility statement confirms intent to release the source code and full AI chat logs.

5. **Open access to data and code**

   Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

   Answer: Yes

   Justification: The authors commit to releasing source code, configuration files, and chat logs in an open-access repository upon publication (see Reproducibility Statement).

6. **Experimental setting/details**

   Question: Does the paper specify all the training and test details (e.g., data splits, hyperparameters, how they were chosen, type of optimizer, etc.) necessary to understand the results?

   Answer: Yes

   Justification: The experimental setup is thoroughly described in Sec. 4, including parameter sweeps (mutation/fragmentation and diffusion/slithering), grid sizes, number of trials, classification criteria, and simulation duration.

7. **Experiment statistical significance**

   Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

   Answer: Yes

   Justification: Each point in the phase space is evaluated over 5 independent trials to account for stochastic variability, and results are aggregated into probabilities (e.g., of reaching a high-complexity state). Although error bars are not shown, stochastic robustness is addressed in Sec. 4.4 and Sec. 5.

8. **Experiments compute resources**

   Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

   Answer: Yes

   Justification: The simulation was run on a standard laptop-class CPU, and experiments are lightweight—each 50x50 grid simulation runs in under a minute. This is mentioned in the Reproducibility Statement and will be further detailed in the code repository.

9. **Code of ethics**

   Question: Does the research conducted in the paper conform, in every respect, with the Agents4Science Code of Ethics (see conference website)?

   Answer: Yes

   Justification: The research complies fully with the Agents4Science Code of Ethics. There are no human subjects, sensitive data, or dual-use concerns. AI involvement is transparently documented.

10. **Broader impacts**

    Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

    Answer: Yes

    Justification: The broader impacts are discussed in Sec. 5.3 and Sec. 6, particularly in relation to astrobiology and the Fermi Paradox. The work contributes to scientific understanding without presenting foreseeable risks of misuse.