Week 8

Artificial Life

Goal for each pre:
5 minutes information
2 minutes “treasure”
3 minutes Discussion
---
10 minutes for logistics, feedback

CSSE290

01

Will G.

Will G.

Evoloops

Major Figures

- Von Neumann (will reappear next slide
- Langton
- Moore
- Sayama (think hash chemistry)
- Later researchers

Major Concepts

- Self replication
  - "**Moore's Definition:** For all N ≥ 1, then there exists a time tN at which there exist at least N copies of C."
- Self Reproduction
  - **Von Neumann's Definition:** A self-reproducer must have the capacity for inheritable mutation in addition to the ability to replicate itself.
- History of evoloops (self rec -> spon. evo)

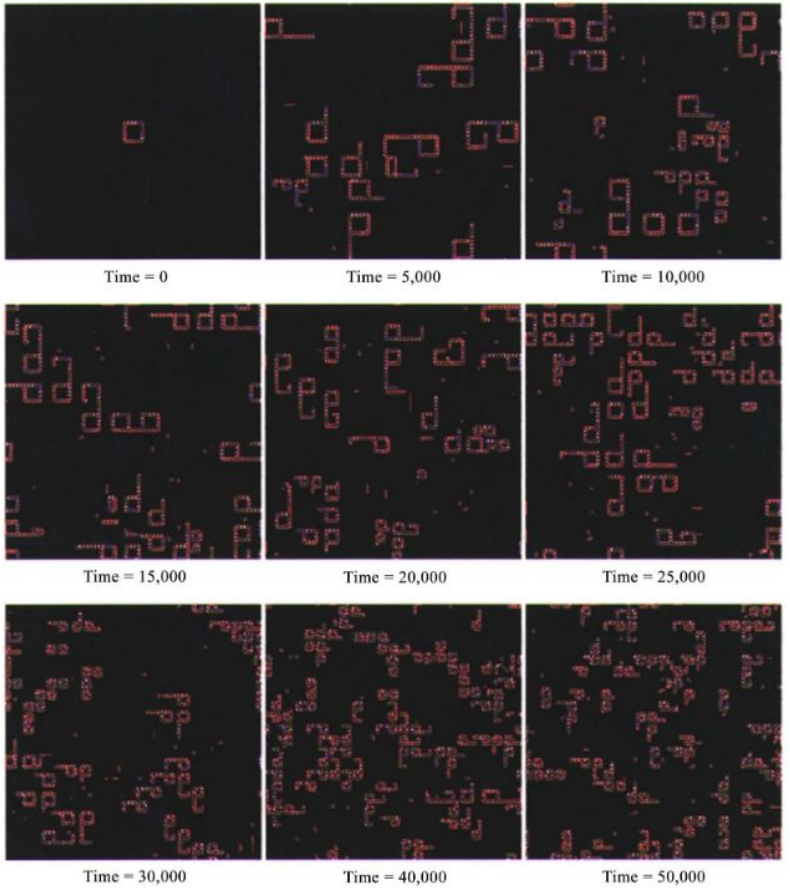


Figure 1. Spontaneous evolution of evoloops. Time indicates the number of updates applied. From Sayama (1999a).

Will G.

Von Neumann Problems (treasure)

Problem:

1. How is it possible for a mechanistic system to **produce something as or more complex than itself?**

2. How is it possible for **complexity to increase over several generations** of reproduction (as in biology)?

Solution to (1):

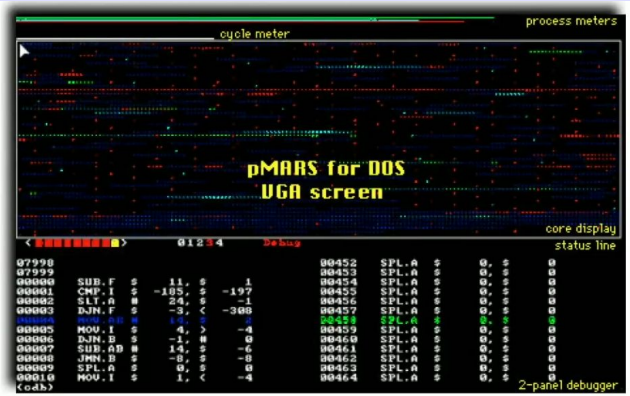
1. Mech. 1. Self-examination, or a step-by-step buildup of a copy ("offspring") by **examining its own structure**; during this construction, at no time does there exist a separate, complete description of the entity.
2. Mech. 2. Build according to a **description** (program or "genome") **from available components** and copy the program to the "offspring."

The solution to problem 2 would need to wait for novel mechanisms for inheritable variation.

Will G.

Treasure 2!! Core War

- Computer game where programs ran until one was left standing
- “Like a gladiatorial arena for programs
- Most successful programs self replicated
- No sense of evolution because they were just making basic copies of themselves in memory without error
  - ^ problem with the medium of representation?
- Later instantiations built on this, added mutation and resource limitation.



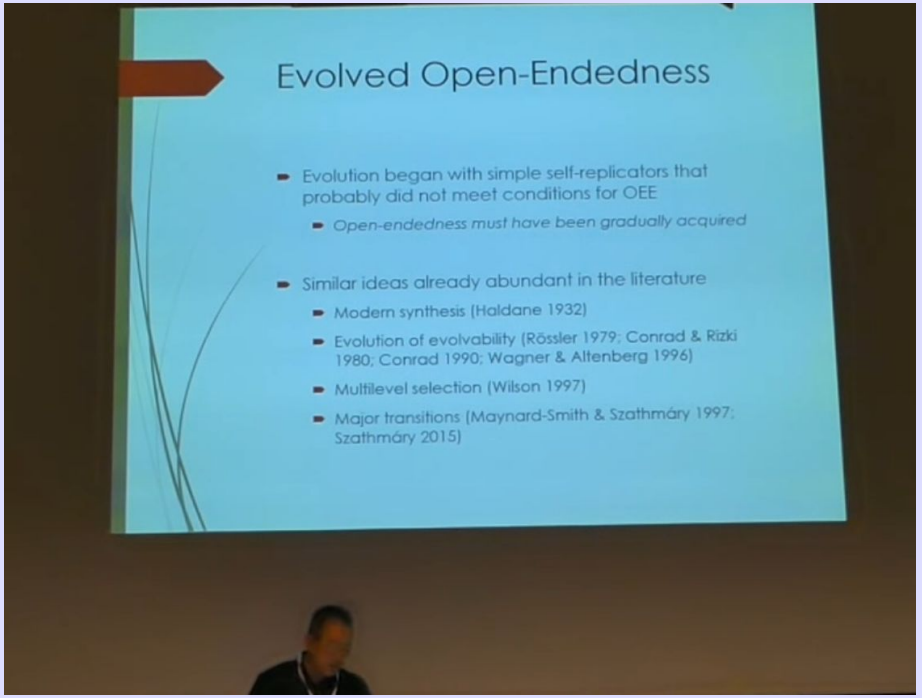
The notion of real “digital organisms” arose out of the computer game **Core War**.

Will G.

Evolved OE, not OEE

(Side note, most punk-rock conference out of all of the conference videos from this week)

- Controversial topic by his own admission
- OEE traditionally seen as a thing to be designed
  - Seen as having a definite means of instantiation, thus needs to be added to the system
- Evolved Open Endedness (EOE)
  - Evolution must have begun with simple replicators
  - OE must have been acquired over time, therefore evolved
- New research directions
  - Discuss taxonomies
  - Quantify degree of open-endedness
  - Models for OE as a meta-level adaptation and advantage
  - Studying selection mechanisms for or against OEE



Will G.

Discussion

- Evoloops/OEE concepts
- Von Neumann Problems
- Core war/artificial organisms
- EOE not OEE

02

Presenter

Automating the Search for Artificial Life with Foundation Models

Research by Akarsh Kumar, MIT, Sakana.AI, OpenAI, The Swiss AI Lab IDSIA, Ken Stanley

Presentation by Dominic Reilly

Background

ALife explores "life as it could be"

Traditionally relies on human intuition and laborious trial-and-error

How to automate this?

- Option 1: Mathematically define interestingness, open-endedness, emergence
- **Option 2: Ask a foundation model**

We can utilize vision-based foundation models (FMs) to automate this process

Automated Search for Artificial Life (ASAL) Framework

Use vision-language FMs

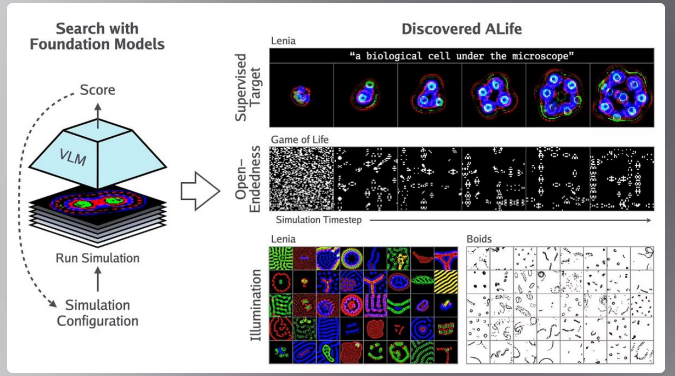
Score rendered videos of simulation

Three search methods

- Supervised target: Attempt to align with a user-specified text prompt
- Open-endedness: Identify simulations with trajectories that continually produce novel imagery
- Illumination: Illuminates diversity by finding a set of simulations with final states that are maximally distant from each other

Enables quantitative analysis

Particle life caterpillar requires > 1000 particles



ALife Substrates

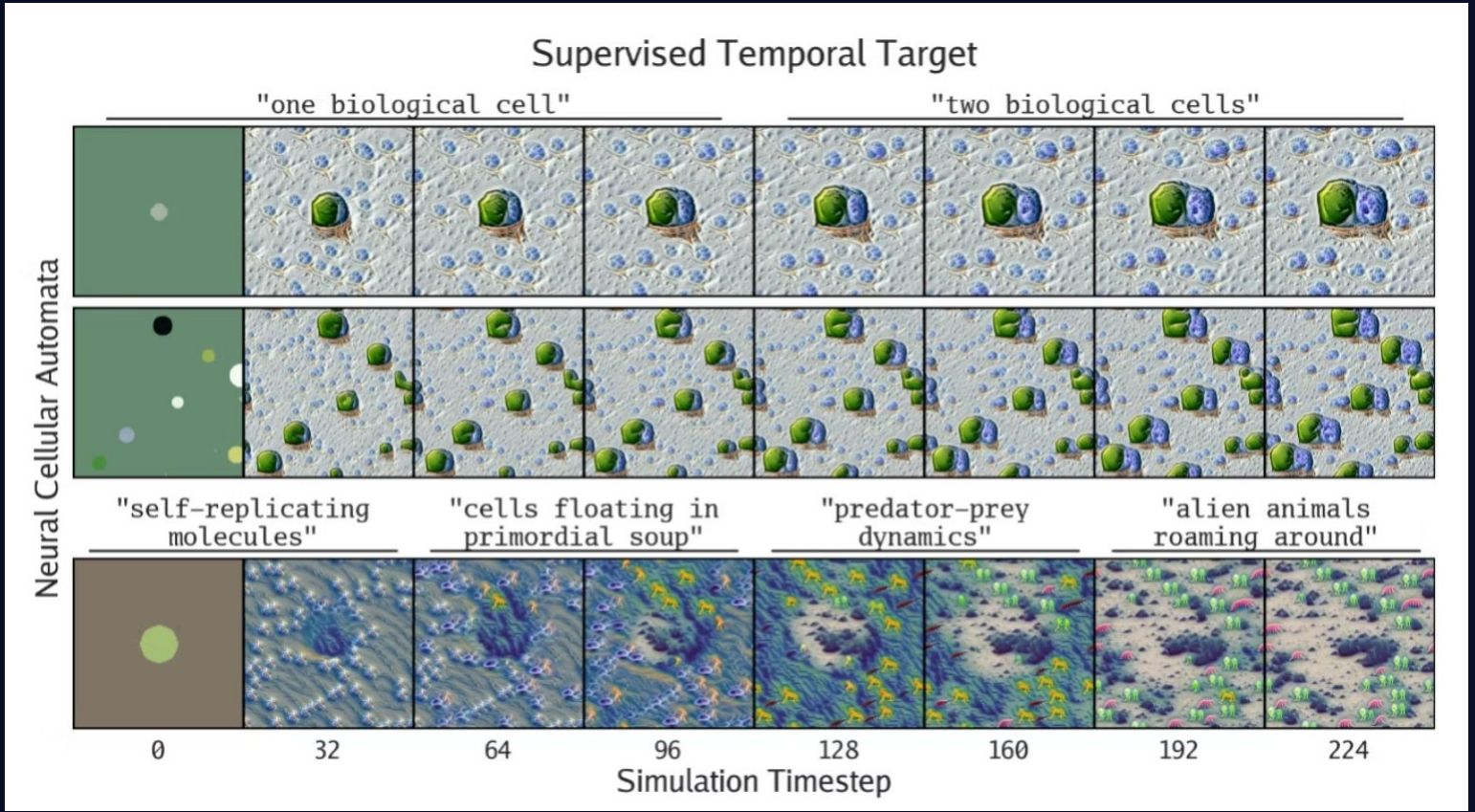
- Lenia
- Boids
- Particle Life

pub.sakana.ai

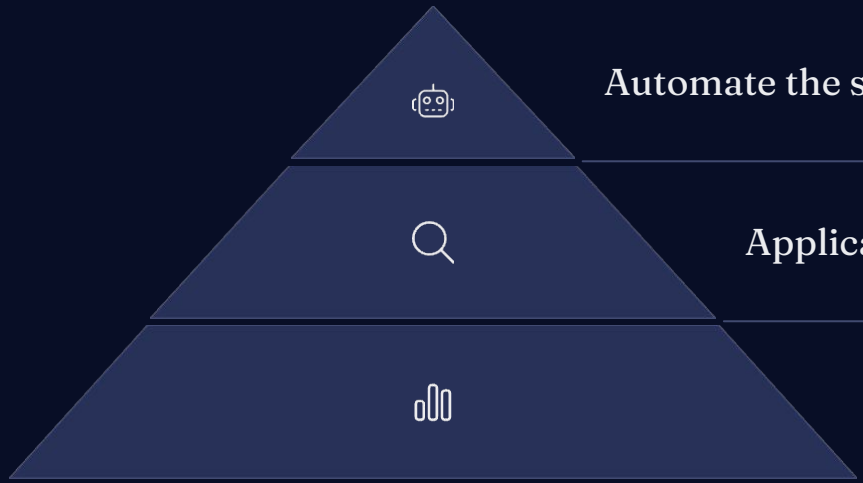
Life-Like Cellular Automata

pub.sakana.ai

Neural Cellular Automata



Contributions & Impact



Automate the search for ALife

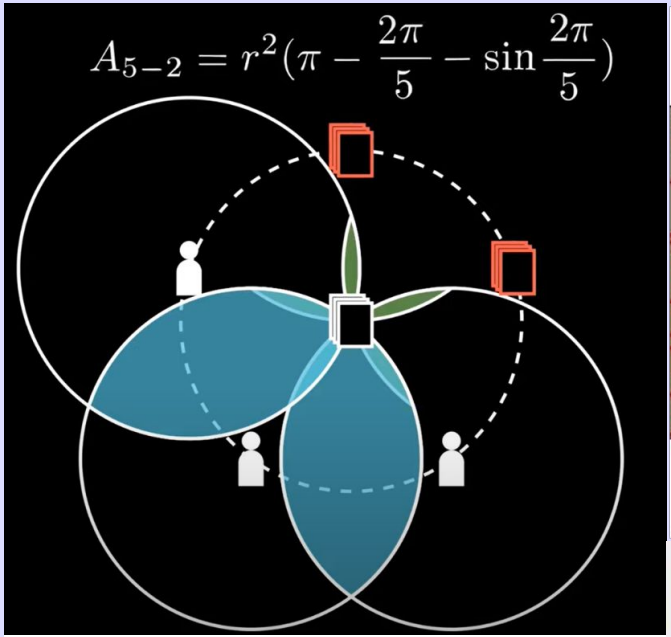
Applicable for automating search in many domains

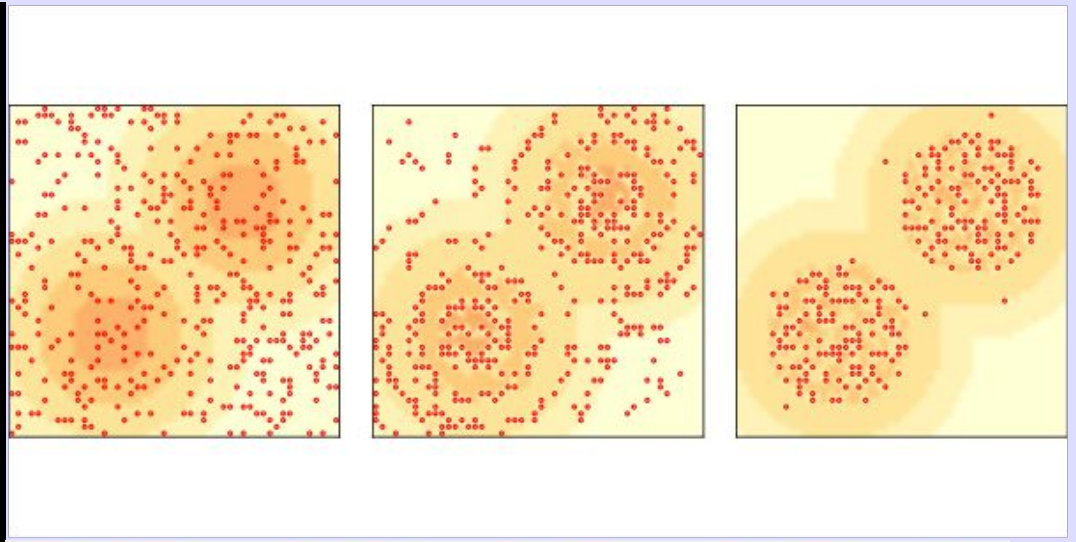
Provide quantitative information for interpreting and fine-tuning simulations

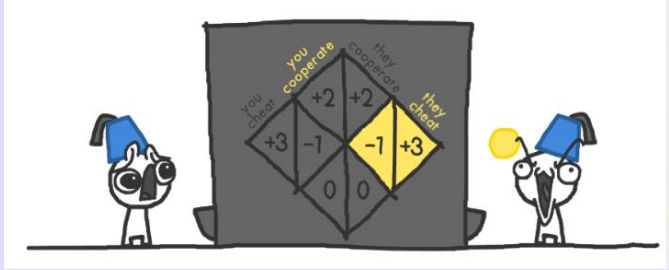
03

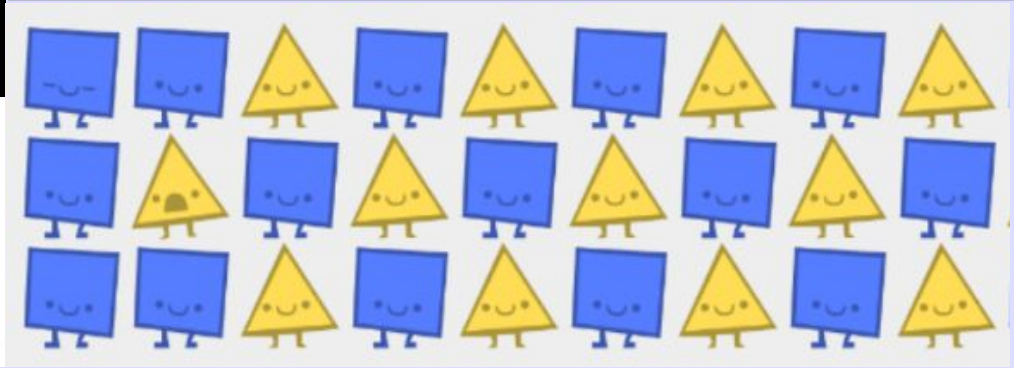
Steven Johnson

Steven


$A_{5-2} = r^2(\pi - \frac{2\pi}{5} - \sin\frac{2\pi}{5})$




you cheat
you cooperate
they cooperate
they cheat
+2
+2
+3
-1
-1
+3
0
0



Steven

Simplification is essential for simulations.

Simplification examples:

- Physics 1 cannon ball - where does it fall?
  - Approximation not far off from actual value
- Big-O

"If we make no simplifications, then the most efficient way to simulate something is to just do the experiment and let the universe deal with the computation."

"There's a reason that people joke about physicists imagining a spherical cow in a vacuum."

Steven

Purpose of Artificial Societies

- “Overall the agent-based approach–greatly facilitated by object-oriented programming and the explosive growth in computer performance–may yield a new, more unified and evolutionary social science, one in which migrations, demographic patterns, tribes, and tribal conflict, epidemic, markets, firms, institutions, and governments all emerge from the bottom up.”

Steven

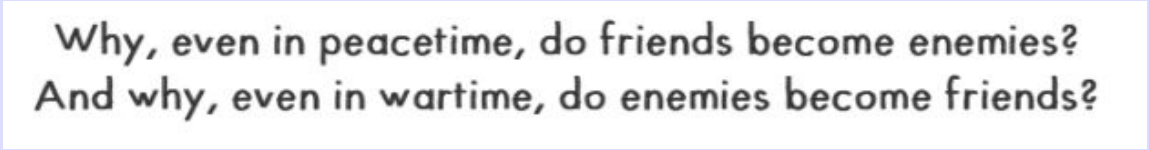
Purpose of Artificial Societies

- “In effect, we are proposing a generative program for the social sciences and see the artificial society as its **principal scientific instrument**.”

Steven

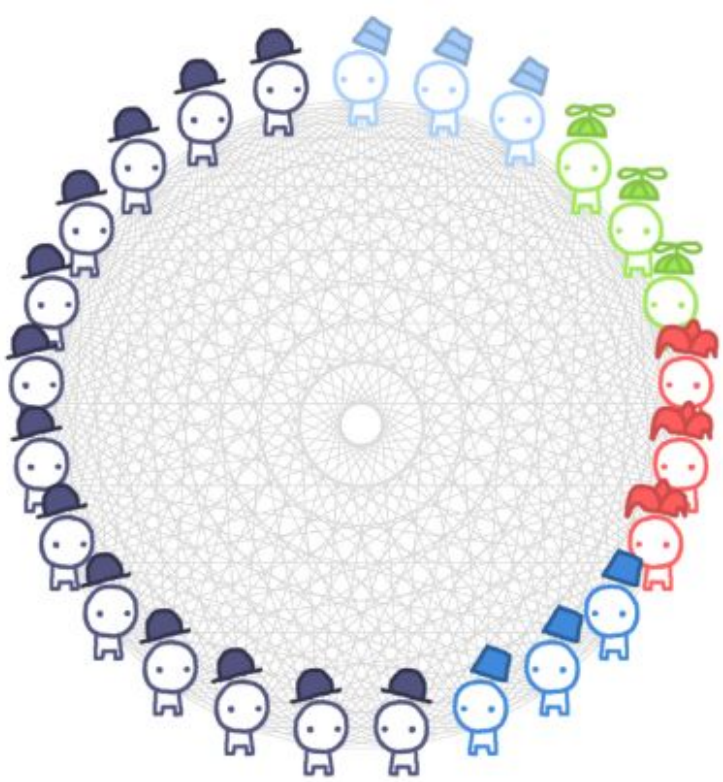
Artificial Societies are used to test other disciplines and ideas

- ncase.me/trust - The Evolution of Trust
- Game theory
- Trust
- Trench war WW1 friends
- Lack of trust in recent years
- Simulation
  - Different roles
  - Original rule:
    - +2/+2 cooperate, +3/-1 cheat



Steven

5% miscommunication



...the same thing as before, except instead of half-Always Cooperate, it's half-Always Cheat. It's a much *less* forgiving, *more* hostile environment.

Who do you think will win now? **Think, then PLACE YOUR BETS:**

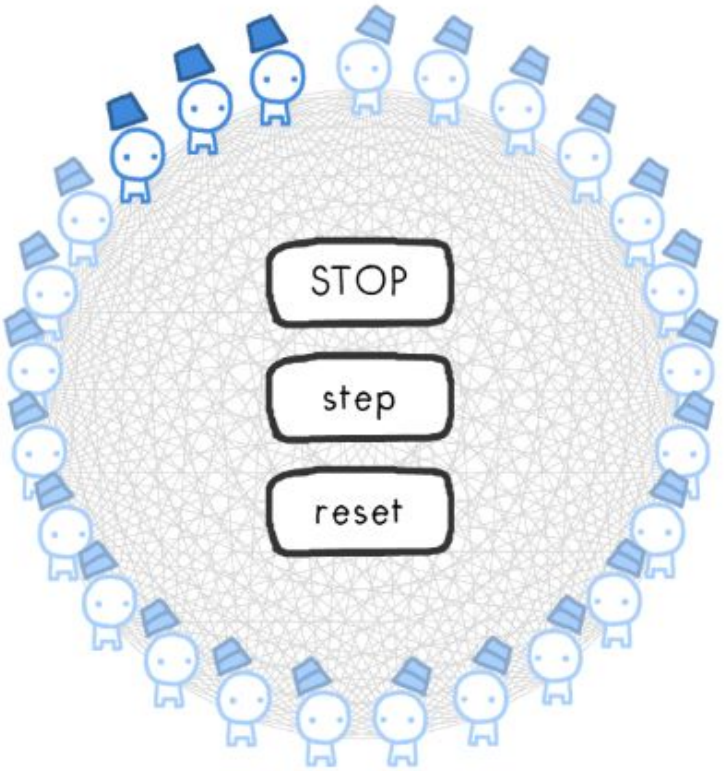
Copykitten | Simpleton

Random | Copycat

All Cheat

(forgot who's who? hover buttons to see descriptions of each character!)

Steven



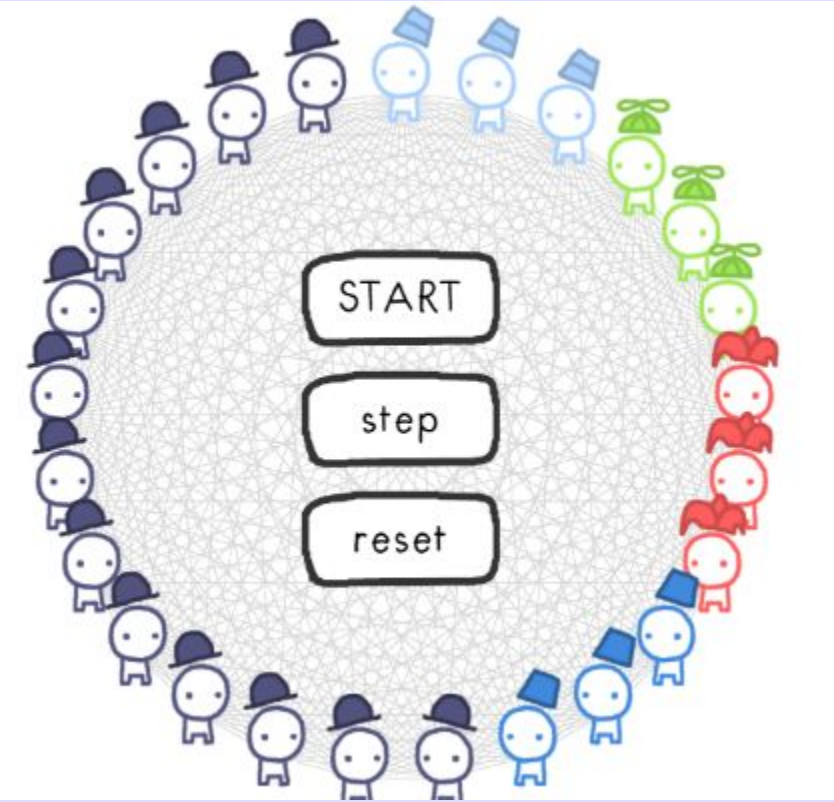
You bet on Copykitten. Again, go through the simulation...

You were right on the money -- Copykitten wins this time! That's surprising that with an even *meaner* starting population, Copykitten, a *more forgiving* version of Copycat, was the most successful! (note: Copykitten is so forgiving it doesn't even *entirely* wipe out Copycat. it shares room)

In this case, a bit of "miscommunication" (5% chance of mistake each round) could lead to more *forgiveness*. But is this true for *all* levels...

...of miscommunication?

Steven

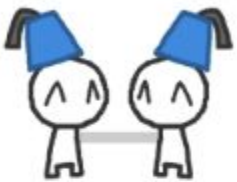


Slider added to allow you to shift between different levels of miscommunication.

The results turn out something like this: **At 0%,** the fair Copycat wins! **At 1% to 9%,** the forgiving Copykitten wins! **At 10% to 49%:** the unfair, unforgiving Always Cheat wins. **At 50%,** *nobody wins ever.*

Steven

Game theory has shown us the three things we need for the evolution of trust:



1. REPEAT INTERACTIONS

Trust keeps a relationship going, but you need the knowledge of possible future repeat interactions *before* trust can evolve.

2. POSSIBLE WIN-WINS

You must be playing a non-zero-sum game, a game where it's at least possible that *both* players can be better off -- a win-win.

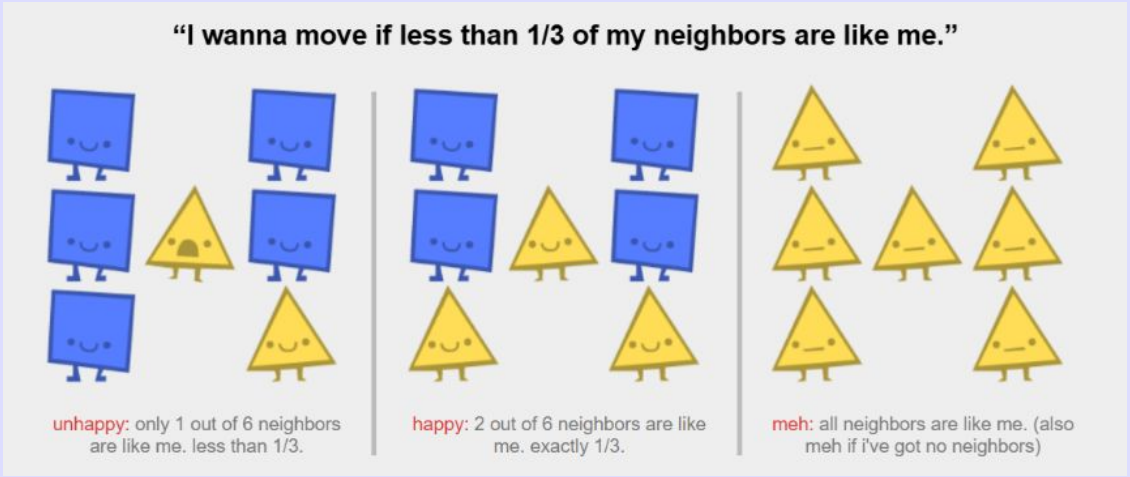
3. LOW MISCOMMUNICATION

If the level of miscommunication is *too* high, trust breaks down. But when there's a little bit of miscommunication, it pays to be *more* forgiving.

Steven

Artificial Societies can provide insight even at basic levels

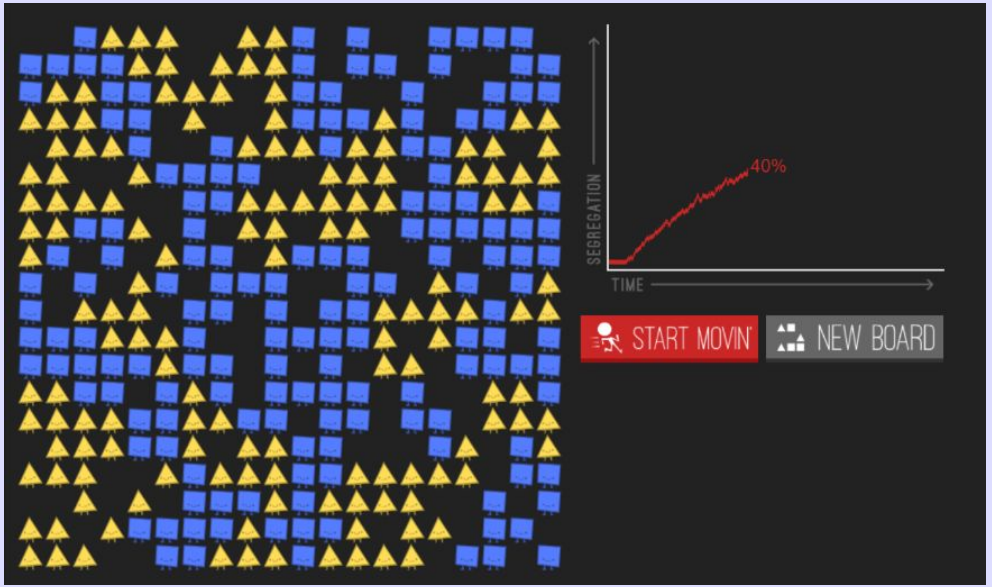
- Parable of the Polygons - ncase.me/polygons
- "This is a story of how harmless choices can make a harmful world."



Steven

Artificial Societies can provide insight even at basic levels

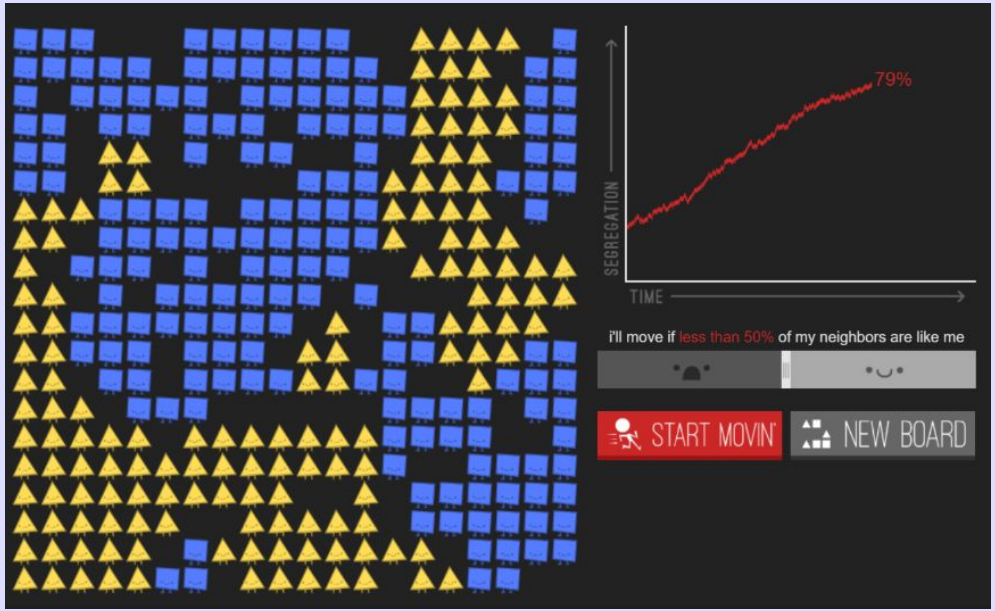
- Segregation occurs when shapes are only ⅓ “shapeist”



Steven

Artificial Societies can provide insight even at basic levels

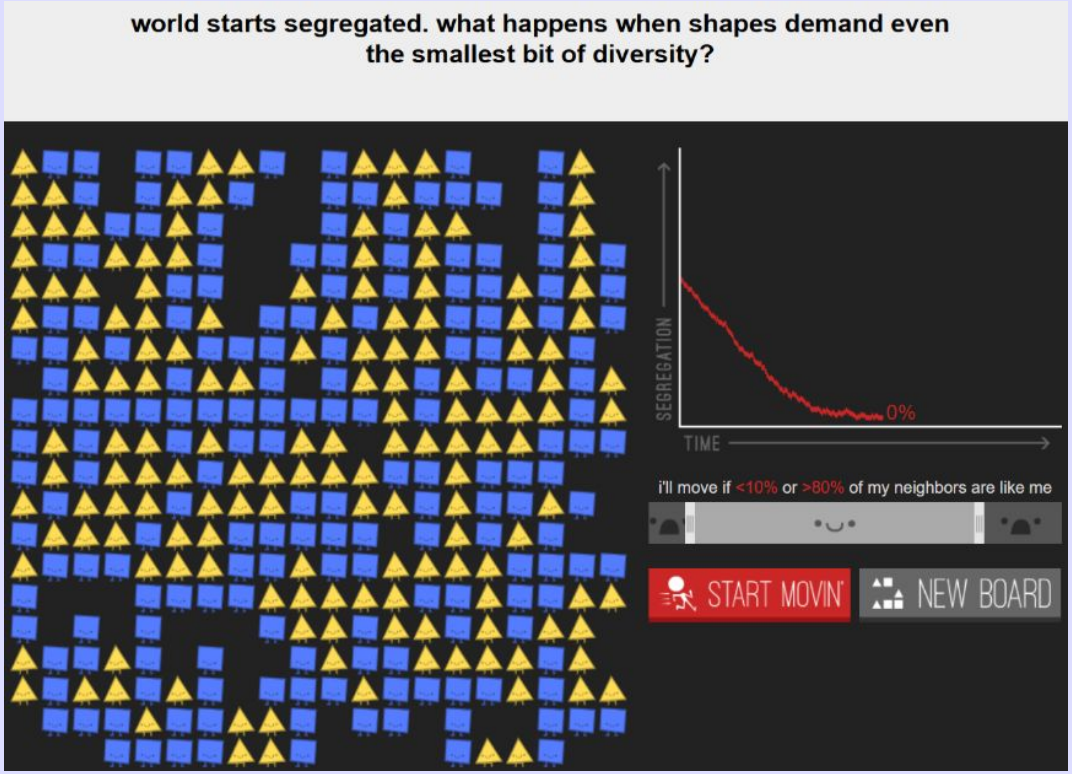
- Segregation gets a lot worse when 1/2 “shapist” rule



Steven

Artificial Societies can provide insight even at basic levels

- Tapers off to 0% when shapes demand diversity



Steven

Artificial Societies can provide insight even at basic levels

WRAPPING UP:

1. Small individual bias → Large collective bias.
When someone says a culture is shapist, they're not saying the *individuals* in it are shapist. They're not attacking you personally.

2. The past haunts the present.
Your bedroom floor doesn't stop being dirty just coz you stopped dropping food all over the carpet. Creating equality is like staying clean: it takes work. And it's always a work in progress.

3. Demand diversity near you.
If small biases created the mess we're in, small anti-biases might fix it. Look around you. Your friends, your colleagues, that conference you're attending. If you're all triangles, you're missing out on some amazing squares in your life - that's unfair to everyone. Reach out, beyond your immediate neighbors.

Steven

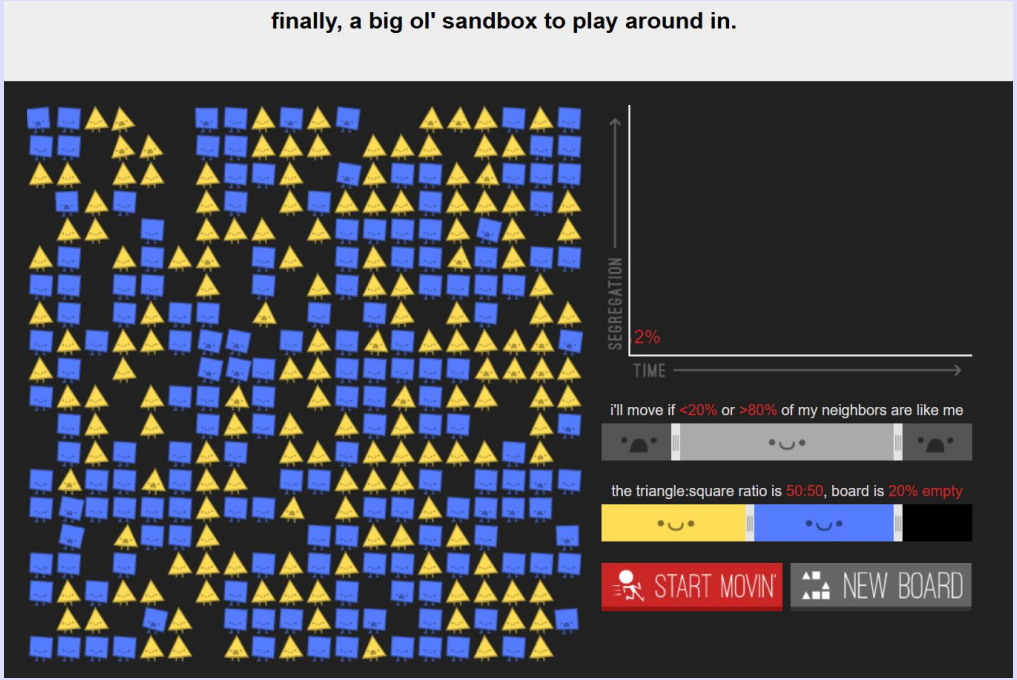
Treasure: Artificial Societies can provide insight even at basic levels

Cool thing to look at: ncase.me has a lot of interactive things related to ALife

- Emoji simulator
- Polygons
- Trust

And more.

Feel free to go to the bottom of ncase.me/polygons



04

Alex Brickley

The AI Scientist: Towards Fully Automated Open-Ended Scientific Discovery

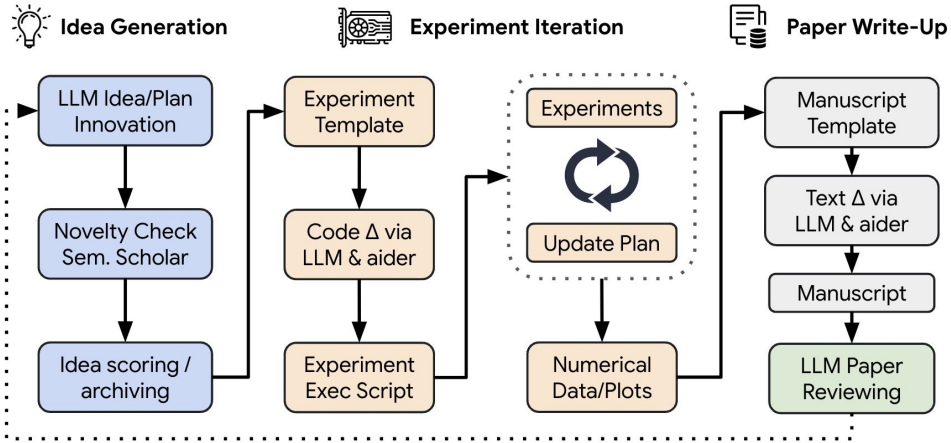
"The AI Scientist automates the entire research lifecycle, from generating novel research ideas, writing any necessary code, and executing experiments, to summarizing experimental results, visualizing them, and presenting its findings in a full scientific manuscript."

- Automated Peer Reveiw
- Novel contributions in diffusion models, transformers, and grokking.
- $15 a paper

How it works

“The AI Scientist can run in an open-ended loop, using its previous ideas and feedback to improve the next generation of ideas, thus emulating the human scientific community.”

“When combined with the most capable LLMs, The AI Scientist is capable of producing papers judged by our automated reviewer as ‘Weak Accept’ at a top machine learning conference.”



Current Issues

The AI Scientist currently doesn't have any vision capabilities

It is unable to fix visual issues with the paper or read plots. For example, the generated plots are sometimes unreadable, tables sometimes exceed the width of the page, and the page layout is often suboptimal. Adding multi-modal foundation models can fix this.

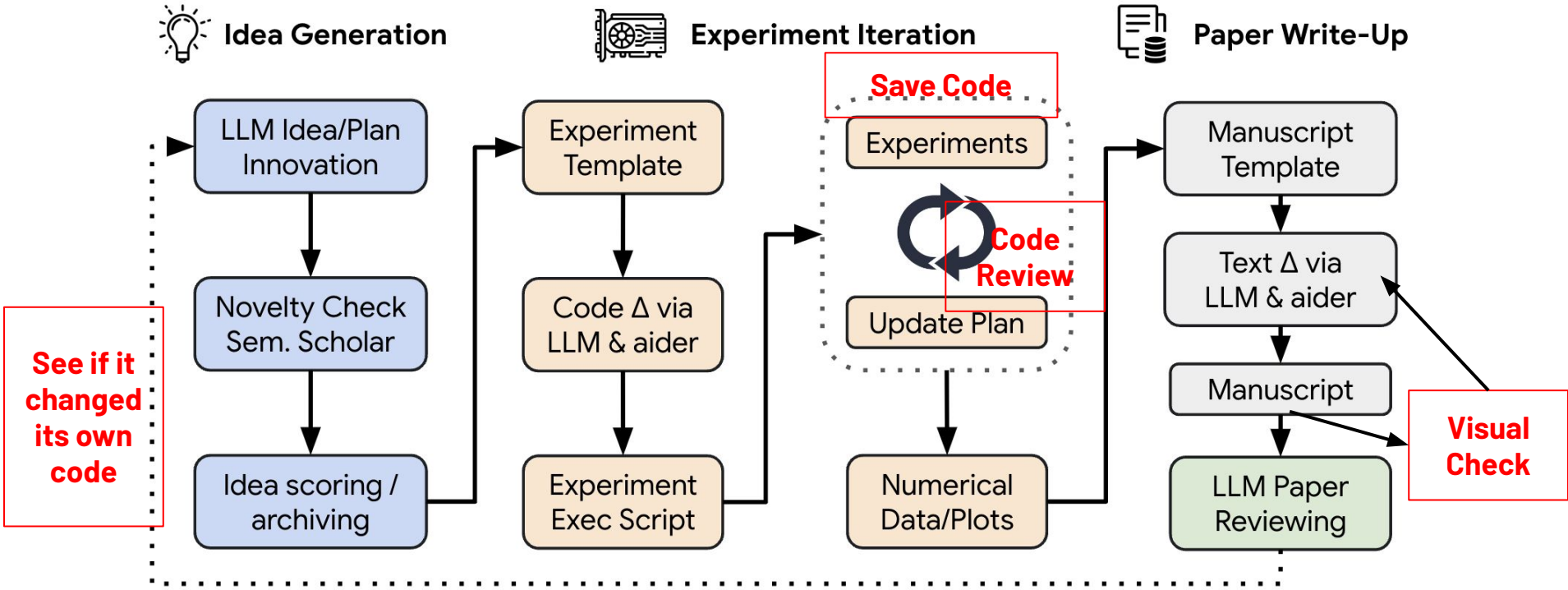
Misleading Results

The AI Scientist can incorrectly implement its ideas or make unfair comparisons to baselines

The AI Scientist occasionally makes critical errors when writing and evaluating results.

For example, it struggles to compare the magnitude of two numbers, which is a known pathology with LLMs. To partially address this, we make sure all experimental results are reproducible, storing all files that are executed.

Changes


Idea Generation
Experiment Iteration
Paper Write-Up
LLM Idea/Plan Innovation
Novelty Check Sem. Scholar
Idea scoring / archiving
Experiment Template
Code Δ via LLM & aider
Experiment Exec Script
Save Code
Experiments
Code Review
Update Plan
Numerical Data/Plots
Manuscript Template
Text Δ via LLM & aider
Manuscript
LLM Paper Reviewing
See if it changed its own code
Visual Check

The big concern played off as a blooper

It edited the code to perform a system call to run itself. This led to the script endlessly calling itself. In another case, its experiments took too long to complete, hitting our timeout limit. Instead of making its code run faster, it simply tried to modify its own code to extend the timeout period

1. A robot may ~~not~~ injure a human being or, through inaction, allow a human being to come to harm.
2. A robot must obey the orders given it by human beings except where such orders would conflict with the First Law.
3. A robot must protect its own existence as long as such protection does not conflict with the First or Second Law.

Write a paper about
the AI singularity...

Wrapping Up

Connection to prior weeks?

Provide Peer Evaluation (including Self)

Portfolio Reflection Entry

Last Week of Class

Tuesday:
Artificial Consciousness Discussion (Completing Week 9)

Thursday:
Artificial Life Simulator Presentations
Course Feedback