Title

Minimal autopoiesis in an AI system

Journal

Behavioral and Brain Sciences, 49

ISSN

0140-525X

Author

Schwitzgebel, Eric

Publication Date

2026

DOI

10.1017/s0140525x25103786

Minimal Autopoiesis in an AI System

Eric Schwitzgebel
University of California, Riverside

Department of Philosophy
University of California, Riverside
900 University Avenue
Riverside, CA 92521
USA

951 827 4288
eschwitz@ucr.edu
https://faculty.ucr.edu/~eschwitz/

Abstract word count: 58
Main text word count: 997
References word count: 35
Entire text word count: 1157

Abstract:

A minimally autopoietic AI system could be designed according to current engineering principles. Although Seth suggests that standard computational systems are not autopoietic and therefore not conscious, *autopoiesis* is a high-level, functional concept, and nothing appears to prevent autopoietic processes in a standard computational system. Autopoietic requirements do not in principle rule out consciousness in standard AI systems.

Main text:

Citing Maturana and Varela (1980), Seth characterizes autopoietic systems as systems that "continually regenerate their own material components through a network of processes… actively maintain[ing] a boundary between the system and its surroundings". Seth suggests that standard AI systems are not autopoietic and therefore not conscious. I suggest that standard AI systems could be, and perhaps some already are, minimally autopoietic.

Consider a hypothetical solar-powered robot designed to move toward light when its charge is low. The system thereby acts to maintain its own functioning. It might model the direction of light sources through predictive processing. It might stay upright by means of a gyroscope and tilt detectors that integrate gravitational and camera inputs. More fancifully, it might be composed of modules held together electromagnetically, so that without electrical power it falls apart.

Now give the robot error-detection systems and the ability to replace defective parts. When it detects a breakdown in one part – for example, in the upper portion of its left front leg – it orders a replacement part delivered. Upon delivery, the robot scans the part for compatibility (rejecting any incompatible parts) then electromagnetically disconnects the damaged part and installs the new one. If the system has sufficient redundancy, even central processing systems could be replaced. A redundant trio of processors might eject a defective processor and run on the remaining processors until the replacement arrives.

A plastic shell maintains the boundary between the system and its surroundings. The system might detect flaws in the shell, for example, by internal sensors that respond to light entering through unexpected cracks, by visually monitoring its exterior, and perhaps by electrostatically detecting cracks or gaps. Defective shell components might be replaced.

If repelling intruders is necessary, we can challenge our robot with fakes. Shipments might sometimes arrive with a part mislabeled as compatible or visually similar to a compatible part, but ruinous if installed. Detecting and rejecting fakes might become a dupe-and-mimic arms race.

I see no in-principle obstacles to creating such a system using standard AI and engineering tools. Such a system is, I suggest, minimally autopoietic. It actively maintains itself. It enforces a boundary between itself and its environment. It continually generates, in a sense, its own material components. It employs predictive processing, fights entropy by drawing on external energy, resists dispersion, and has a solar-electric metabolism.

Does the fact that it depends on shipments mean that it does not actually generate its own parts? Humans also depend on nutrients generated from outside, for example vitamins and amino acids that we cannot biologically manufacture. Sometimes these nutrients are shipped to us. Also, it's easy enough to imagine the robot not simply installing but in a minimal sense manufacturing a part. Suppose a leg has three modular components. Each component might arrive separately, requiring a simple joining procedure to create the leg as a whole.

In a human, the autopoietic process occurs at multiple levels simultaneously. Cells maintain themselves, and so do organs, and so does the individual as a whole. Our robot does not have the same multi-level autopoiesis. But it's not clear why autopoiesis must be multi-level to count as genuine autopoiesis. In any case, we could recapitulate this imaginative exercise for subsystems within the robot or larger systems embedding the robot. A group-level autopoietic system might comprise several robots who play different roles in the group and who can be recruited or ejected to maintain the integrity of the group and the persistence of its processes.

Perhaps my system does not *continually* regenerate its own components, and that is a crucial missing feature? It's not clear why strict continuousness, rather than periodic replacement as needed, should be essential to autopoiesis. In any case, we can imagine if necessary that the robot has some fragile parts that need continual refurbishment. Perhaps it occupies an acidic environment that continually degrades its shell so that its shell coating must be continually monitored and replaced through capillaries that emit lacquer as needed from a refillable lacquer bag.

My system does not reproduce, but reproduction, sometimes seen as essential to life, is not standardly viewed as necessary for autopoiesis (Maturana & Varela, 1980, p. 100).

I invite Seth to respond to this thought experiment by either (1) explaining what aspects of autopoiesis are missing that could not straightforwardly be engineered into this system and why those missing elements matter for consciousness, or (2) acknowledging that a system of this sort could be a minimal autopoietic system (perhaps the type of "living AI" he briefly gestures toward) and exploring the consequences for his biological, substrate-dependent view of consciousness.

A case could even be made that my desktop computer is already minimally autopoietic. It draws power from its environment, maintaining a low-entropy state without which it will cease to function. It monitors itself for errors. It updates its drivers and operating system. It detects and repels viruses. It does not order and install replacement hardware, but it does continually sustain its intricate electrical configuration. Indirectly, though acting upon me, it does sometimes cause replacement parts to be installed. Alternatively, perhaps, we might view its electrical configuration as an autopoietic system and the hardware as the environment in which that system dwells.

My main thought is: *Autopoiesis* is a high-level, functional concept. Nothing in the concept appears to require implementation in a particular substrate. Nothing seems to prevent autopoietic processes in AI systems built along broadly familiar lines. An autopoietic requirement on consciousness does not seem in principle to rule out consciousness in standard computational systems.

Maturana and Varela themselves allow substrate flexibility. They write that

> The organization of a machine (or system) does not specify the properties of the components which realize the machine as a concrete system, it only specifies the relations which these must generate to constitute the machine or system as a unity (1980, p. 77).

It is clear from context that they intend this remark to apply to autopoietic as well as non-autopoietic machines (cf. Luisi 2014).

Competing Interests: none.

Funding: This research received no specific grant from any funding agency.

<u>References:</u>

Luisi, P. L. (2014). The minimal autopoietic unit. *Origins of Life and Evolution of Biospheres 44,* 335-338.

Maturana, H. R., & Varela, F. J. (1980). *Autopoiesis and cognition: The realization of the living.* D. Reidel.

Follow up thoughts:

- Balancing competing interests and needs
- Evan Thompson Bluesky: must compose itself *chemically*, otherwise it's not really making its own material basis (reply: doesn't manufacture its own elements H, C, N, O, etc; what principle makes chemical combination the right, special level?)
- Jovana on email: Must be responsible for its own sun-seeking and self-construction, not commanded from outside (reply: not currently commanded, and past isn't relevant to current autopoiesis; and even biological entities are designed from the beginning with patterns/processes)
- Being, behaving, becoming: https://substack.com/@mraginsky/p-146116129
-