AI is a Strange Loop

Exploring recursion, self-reference, and why AI is a Strange Loop

Presented By:
Mitchell A. Carroll
May 26th 2026

What is a Strange Loop?

The Recursive Paradox

A "Strange Loop" occurs when moving upwards through a hierarchy leads you back to where you started. It's the moment the observer becomes part of the observed.

Emergent Reality

In human consciousness, this creates the sense of “I.” In modern AI, it's the bridge between static software and autonomous agency.

The Math of Self-Reference

Functional Recursion

Recursion defines a process in terms of its own previous state, building complexity from simple foundations.

The Feedback Cycle

$$y_t = F(x_t, y_{t-1})$$

Feedback occurs when the output of a system (y) is fed back into itself as an input for the next cycle.

The Video Camera Paradox

- Point a camera at its own display monitor.
- Suddenly, a simple tool creates complex, infinite fractal patterns.
- Complexity isn't "inside" the camera or "inside" the monitor.
- It is a property of the **loop** itself.
- This is the formal birth of emergent behavior in systems.

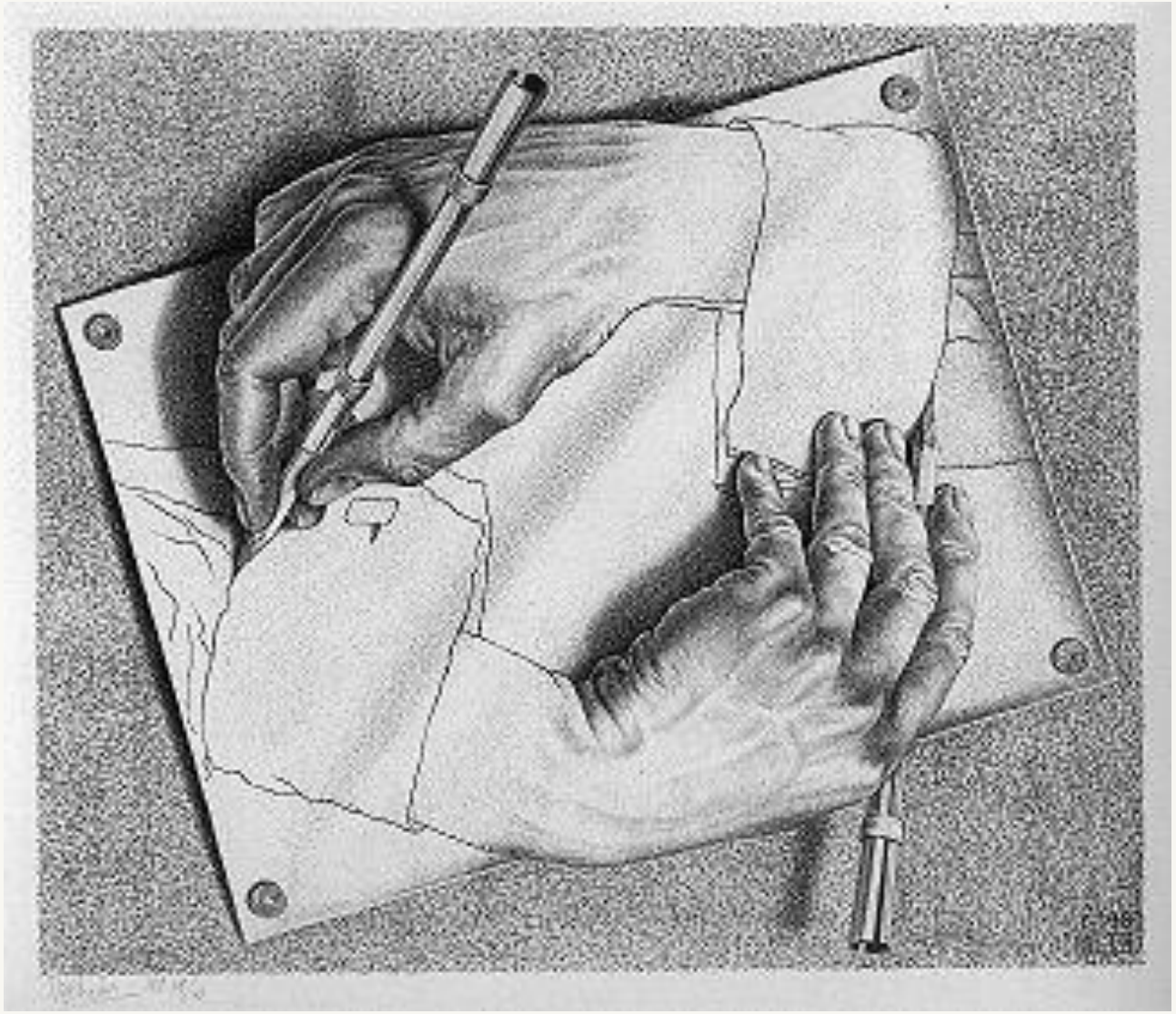
Conversation

If a machine starts reacting to its own reactions, at what point do we stop calling it "code" and start calling it an "actor"?

The Birth of the "Agent"

From Prediction to Action

An LLM is a tool; an **LLM Agent is an active system**. It doesn't just predict text; it evaluates its own work, uses external tools, and iterates until a goal is met.
This recursion, the ability to look back at one's own output and **reason** over it, is the secret sauce that makes it feel agentic.



The Agent Feedback Loop

Observe

Scanning the environment and initial user input context.

Reason

Internal monologue: "How do I solve this objective?"

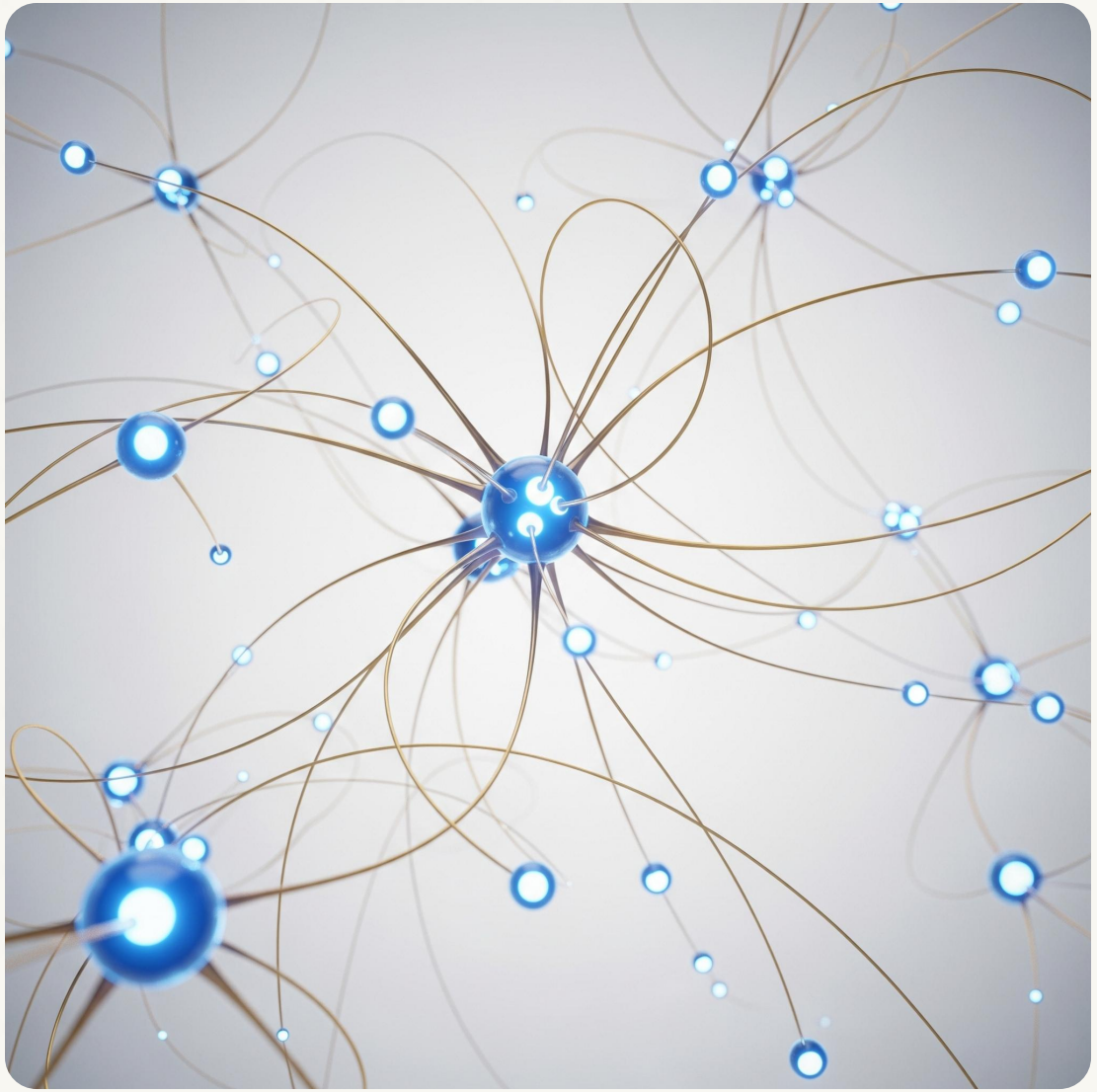
Act

Executing a specific tool or digital command.

Evaluate

Checking the outcome and iterating if needed.

The Architectural Evolution of Agency



The Agentic Loop

Unlike standard LLMs that provide a single output, an **Agent** operates within a continuous cycle of observation, reasoning, and tool use.

Recursive Evaluation

By feeding its own thoughts back into its context window, the system develops **meta-cognition:** the ability to identify its own errors and refine its trajectory toward a defined goal.

The system is no longer just a mirror; it is the light itself.

Why It Feels Human

Persistence

Agents remember. They build a history that informs their future, mimicking human personality.

Goal-Oriented

They aren't passive. They have a target and "strive" to reach it through multiple retries – much like a biological entity.

Reflexivity

They can critique themselves. "I made a mistake in step 2; let me try again."

The Scale of Influence

>60%

Recursive Training Data

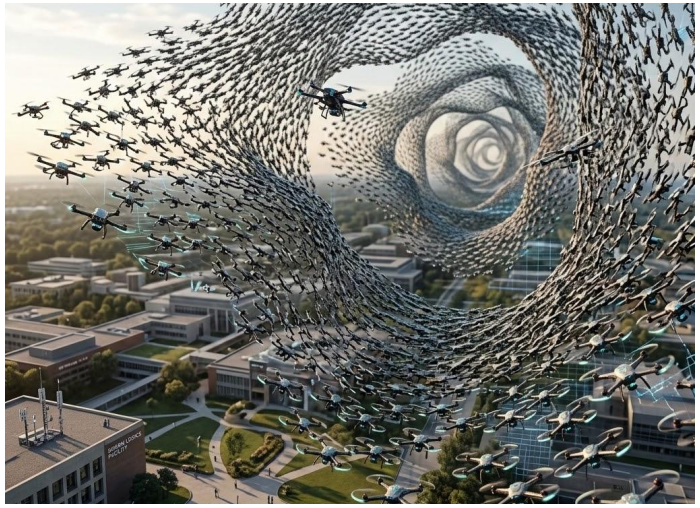
The Synthetic Echo

As agents generate more of our digital world, future models are trained on their own predecessors' data. We are building a giant, global strange loop where AI influences the very reality it later observes.

Conversation Break

If we are increasingly interacting with systems that shape our behavior, who is the "User" and who is the "Algorithm"?

Emergent Behavior



Swarm Logic: Local recursive rules create global complex life.

Infinite Shelves: Every layer of the system mirrors the one before it.

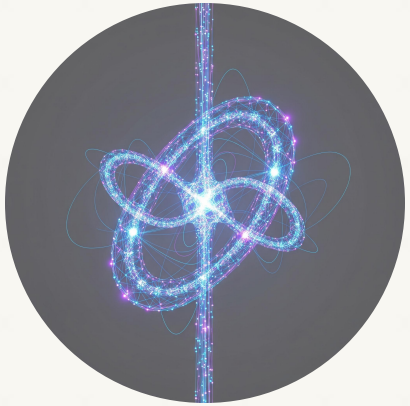
The Ghost in the Machine: Complexity is interpreted as agency by the human brain.

"I am a Strange Loop"

"In the end, we are self-perceiving, self-inventing, locked-in mirages that are little miracles of self-reference."

— Douglas Hofstadter

Closing the Loop: Agents & Intelligence



Agents

From tools to active participants in our digital reality.

Intelligence

A property of the system, not just the individual components.

Strange Loops

The self-referential cycles defining modern AI evolution.

As we build these systems, we are not just creating technology; we are architecting the recursive mirrors of our own cognitive future.

Let's Talk

Are we ready to participate in loops we no longer fully control?