The False Promise of Artificial Agency

We chase machines that think like people and call it progress, but simulation is not sentience, and mistaking one for the other misdirects what AI can actually do for us.

The agency illusion

For decades, technologists have worked toward artificial general intelligence, a machine that can think, decide, and act with human-like autonomy. The pitch is seductive: give software a goal, some sensors, a way to plan, and it will behave like a person with agency. But the dream rests on a simple mistake. It confuses simulation with sentience.

An agent can mimic reasoning. It can predict, correlate, and generate. It can search a space of options and select a next step that appears sensible. None of this requires awareness. Awareness is not a property of code. It is a state of relation between body, mind, and world. Without that relation, what looks like agency is an increasingly clever imitation of it.

If we are precise about the problem, we save ourselves from false hope and poor design. We stop outsourcing responsibility to systems that cannot carry it. We set a clearer aim for AI: not artificial people, but better tools for human cognition.

What code can do

Modern systems are powerful at simulation. Feed them data and they learn patterns. Ask for an answer and they assemble likely continuations. Wrap them in a loop and they can plan, call tools, and iterate. The effect can feel agentic. It is still prediction, correlation, and generation.

This is not an insult. Simulation is useful. It can surface options we would miss, compress noise into signal, and draft first passes so humans can decide. But we should name the limit clearly: code manipulates representations; it does not have a lived stake in the world that gives those representations meaning.

> Code manipulates representations; it does not have a lived stake in the world that gives those representations meaning.

Call that stake consequence felt in a body. Call it a thread of experience that ties perception to action and back again. In machines, we have inputs, models, outputs, and feedback. In people, we also have sensation, affect, history, and a world that pushes back through lived consequence. That difference matters.

When we label prediction loops as agents, we invite a category error. We start to treat outputs as intentions. We imagine a mind where there is only a model. That move can blur accountability and inflate expectations. Clear language keeps us honest about what these systems do, and what they do not.

What awareness requires

Awareness is relational. It is a state of consciousness grounded in an entity's connection to its own internal states and to a world it inhabits. On this view, awareness is not something you can upload into code like a feature flag. It depends on embodiment, a body that moves, senses, and is changed by its environment.

Embodiment is not a romantic add-on; it is the ground of awareness. A living system learns because it must. It coordinates perception and action to stay alive. Its experience is shaped by friction, limits, and consequence. That is a very different substrate than symbols manipulated by rules or weights adjusted by gradients.

There are counterpoints worth acknowledging:

- Some argue that if a system replicates the functional organization of a mind, it would share its mental states, regardless of the material it runs on.
- Others suggest that complex models might display a non-biological form of proto-consciousness.
- It is also true that definitions of consciousness and awareness are not settled, so strong impossibility claims can overreach.
- And insisting on embodiment can be read as biological chauvinism, prematurely closing the door on non-biological minds.

These positions highlight open questions. Still, the core claim here stays modest and clear: simulating the outputs of awareness is not the same as being aware. Until we can show that a system has a lived, relational stake, body, mind, and world entwined, calling it conscious is like mistaking a reflection for the person standing before the glass.

Rethinking the goal toward cognitive extension

Once we stop promising artificial agency, a better purpose comes into view: extend human cognition. Build systems that help us think, not pretend to think for us. That shift answers to both the power and the limit of code.

Practical principles follow:

- Treat AI as an instrument. Use it to surface patterns, draft possibilities, and test options. Keep meaning-making with people.
- Design for clarity. Make systems show their steps, assumptions, and limits so humans can judge where they are reliable.
- Keep responsibility human. Let machines recommend; let people decide and remain accountable.
- Favor augmentation over autonomy. Aim for tools that improve attention, memory, planning, and craft, not proxies for judgment.
- Respect the boundary. Do not project intention onto outputs. Evaluate them as simulations that may be helpful, not as statements from a mind.

This is not a retreat. It is a strategy that aligns power with purpose. Simulation scales insight. Humans bring awareness. Together, they form a workable loop: models accelerate options; people apply context, values, and consequences felt in the real world. That pairing is where these systems can do their best work.

The mirror and the face

To imagine a machine as conscious is to mistake a mirror for a face. A mirror can show a faithful image. It can even surprise you with angles you had not noticed. But it never looks back. It does not know you are there.

> AI is a mirror of our data and our prompts. It returns patterns that resemble reasoning.

AI is a mirror of our data and our prompts. It returns patterns that resemble reasoning.

When wrapped in tools and loops, it can act in ways that look agentic. None of this makes it aware. The face, the felt relation of body, mind, and world, remains human.

We chase artificial agency when we should build cognitive extension. Code that predicts and generates cannot be aware because awareness requires embodiment, a lived stake in the world that gives meaning to representations. When we stop confusing simulation with sentience, we stop outsourcing responsibility to systems that cannot carry it. We build tools that help us think better rather than pretend to think for us.

To translate this into action, here's a prompt you can run with an AI assistant or in your own journal.

Try this...

Before calling an AI system an agent, ask: does it have a lived stake in the world, or does it only manipulate representations of one?