KORN FERRY

The Illusion of Consciousness

Thought Leadership

In *Gulliver's Travels*, author Jonathan Swift imagined a machine that could generate philosophy, poetry, or math "without the least assistance from genius or study." He called it the Engine, a crank-operated contraption of word-tiles that promised knowledge without thought. In 1726—the year Swift published *Gulliver's Travels*—this was pure satire. Centuries later, the joke feels more like foresight.

Much like the Engine, modern artificial intelligence can produce code, images, and entire books with unnerving fluency. Yet, beneath the impressive output lies the same question Swift's parody implied: Is something that merely recombines symbols capable of conscious experience and understanding?

To even consider conscious AI, we first need to clarify what that means—and why people often treat the technology as if it's already "aware." From there, we can explore what it might mean for machines to feel, know, and act as agents in the world.

TESTING FOR CONSCIOUSNESS: HISTORICAL PARADIGMS

In 1950, Alan Turing reframed the question "Can machines think?" into a test: if a computer's conversation is indistinguishable from a human's, we might treat it as thinking. Today, we know the Turning Test measures fluent imitation, not true comprehension. By Turing's original standard, most modern AI models would pass with ease. Generative AI platforms can handle context so well, it's easy to forget that they're computer software.

In 2019, philosopher and AI expert Susan Schneider proposed two new ways to assess AI consciousness beyond behavior. One—her AI Consciousness Test (ACT)—suggests that *truly* conscious AI would spontaneously raise and reason about concepts like mind, mortality, or subjective experience—ideas it could not mimic without an inner grasp of reality. The other—her hypothetical "Chip Test"— imagines a scenario where silicon components could gradually replace parts of a human brain to test whether consciousness would persist. Both of her theories ask us to look beneath the surface, deciding what, if anything, could count as machine consciousness.

Schneider's Chip Test raises a key question: Is consciousness based in biology or could it also exist in something nonbiological? Some argue that

subjective experience depends on the organic chemistry of the brain—its evolutionary history, cellular processes, and embodied understanding of the world. Others suggest that what matters is not carbon but causal structure—the degree to which information integrates within a system. If the latter's true, then silicon circuits and other artificial systems could, in theory, generate conscious states given the right organization.

But because we can't see how machines "think," we rely on their observable behavior—which can be misleading. Instead, we will need to consider other ways to evaluate whether today's tools possess capabilities that echo thought or awareness.

Imitation does not mean understanding, nor the kind of subjective experience we attribute to conscious agents. The Chinese Room Argument—a thought experiment exploring consciousness—illustrates this idea. The exercise shows that while rule-based symbol matching can produce coherent language, behavioral simulation alone offers no evidence of awareness or genuine understanding, making it an insufficient metric for determining consciousness in machines.

DEFINING CONSCIOUSNESS: A KORN FERRY FRAMEWORK

Before we can measure consciousness, we first need to clarify what we mean by it. Drawing from both philosophy and organizational psychology—particularly Korn Ferry's Conscious Agent Model—we can identify five key dimensions of consciousness:

- **Awareness:** The ability to react to things in the world.
- **Self-Awareness:** The ability to react to and talk about changes within oneself.
- **Goal-Directedness:** The ability to do things with a purpose.
- **Information Integration:** The ability to put together different kinds of information.
- **Qualitative Experience (Qualia):** The ability to have meaningful experiences.

Current AI systems excel in some of these dimensions—most notably Goal Directedness, Information Integration, and even the simulation of Awareness. Yet, two critical aspects remain elusive: genuine Self-Awareness and Qualitative Experience. Although AI can act as if it feels, no evidence suggests it has physical or emotional grounding. As such, according to KF's Framework, AI is not conscious.

This gap lies at the center of leading theories. The Global Workspace Dynamics (GWD), first introduced in 1988, proposes that consciousness emerges when information, thoughts, ideas, and memories can interact in one centralized space. In contrast, Integrated Information Theory (IIT)—proposed over a decade later, in 2004—links consciousness to the richness of causal interconnections within a system, suggesting that a high degree of integration creates a unified "mathematical fingerprint" of awareness.

Philosopher Daniel Dennett offers yet another view: consciousness is not an ineffable inner light but a set of observable behaviors and capabilities.

This functional view of consciousness directly contrasts with the Chinese room experiment, which argues that no matter how fluent a system's response, true understanding requires something more than computation. As philosopher and AI critic Hubert Dreyfus argued, no amount of symbolic manipulation can replace the tacit, embodied understanding that makes even ordinary human action possible. For each of these theories that seem to fall short, the question remains: What is the missing piece that gives rise to consciousness?

UNDERSTANDING CONSCIOUSNESS THROUGH KNOWING

One way to explore this problem is through the lens of knowing—not just what we know, but how we know it. This perspective shifts the AI consciousness debate away from abstract speculation about "minds" toward the concrete ways intelligence expresses itself.

Cognitive scientist John Vervaeke outlines four types of knowing:

- **Propositional (knowing that):** Facts and assertions—where AI is unmatched.
- **Procedural (knowing how):** The mastery of skills—where AI is advancing rapidly.
- **Perspectival (knowing what it's like):** Subjective experience—still beyond AI's reach.
- **Participatory (knowing by being):** Embodied, lived engagement—entirely absent in machines.

Generative AI models can describe grief, reflecting propositional knowledge. And they can simulate comforting dialogue, which shows procedural knowledge. But AI lacks perspectival and participatory knowing—it doesn't experience grief, nor does it exist within a shared world.

Learning reinforces this distinction. AI thrives on explicit learning—ingesting labelled datasets, extracting patterns, optimizing predictions. Human learning, on the other hand, runs much deeper.

By combining deliberate reasoning with implicit processes, we absorb social norms, resonate emotionally, and evolve our understanding of the world as it changes. AI systems can stimulate hesitation, contradiction, even intuition, it is always without the implicit, embodied grounding that gives human thinking depth.

CONSCIOUSNESS IN AI: IMPOSSIBLE, POSSIBLE, AND PROBABLE

Having explored different ways of defining and testing consciousness, we can turn to the question at the heart of the debate: Could AI ever *truly* be conscious?

Opinions on this fall into the following three categories, each with its own implications for strategy, ethics, and design—and for how organizations prepare to live and work alongside AI.

- **Impossible:** If consciousness cannot exist outside biological systems, AI remains a powerful tool without inner experiences. Companies can focus on transparency, usability, and human oversight, using AI to augment people rather than as independent beings.
- **Possible:** If consciousness is not limited to biology but can emerge from sufficiently complex information processing, then sentient machines could one day exist. Consciousness could appear at the whole system level, even if absent in individual parts. In other words, once AI architectures reach a certain threshold of complexity, subjective experience could emerge even if no one designs it directly. A 2023 academic review of these leading theories notes that, although current AI systems are not conscious, "there are no obvious technical barriers" to creating ones that "satisfy these indicators." In this scenario, governance must navigate rights, responsibilities, and ethical development for AI entities.

- **Probable:** Some argue that conscious AI is not only possible but likely—though framing it in terms of probability raises more questions than it answers. What exactly are we estimating the likelihood of—the emergence of certain conscious behaviors, the achievement of sufficient complexity, or the appearance of genuine subjective experience? An AI might display every outward sign of comprehension yet still operate only through complex symbol manipulation. If conscious AI is indeed probable, the challenge is not only predicting when it might emerge but also determining how we would ever know.

CONFRONTING THE EPISTEMIC WALL

Whether human or machine, we cannot step directly into another entity's subjective experience. That limit is the epistemic wall—the boundary between external evidence and internal awareness. It makes certainty about a machine's "inner life" impossible.

As AI systems may simulate self-reflection and metacognition, the gap between convincing imitation and *actual* experience remains wide. We cannot understand consciousness from the outside, because subjective states are, by nature, inaccessible to observers. With one study showing that 70% of knowledge workers engage with AI daily, leaders will need to balance AI adoption with clear governance to maintain both strategic and ethical alignment. For consultants, the challenge is clear: preparing clients to harness AI's capabilities while acknowledging the ongoing uncertainty about machine minds.

DISTINGUISHING NARROW AI, AGI, AND ASI

- **Narrow AI:** Specialized AI systems designed for well-defined tasks (e.g., language translation, facial recognition). These excel in efficiency and pattern recognition but lack autonomy outside their domain.
- **Artificial General Intelligence (AGI):** A still-hypothetical form of AI capable of flexible reasoning, problem-solving, and adaptation across domains—roughly comparable to human cognitive breadth.
- **Artificial Superintelligence (ASI):** A theoretical intelligence that would surpass human cognitive capacities in every dimension, from scientific reasoning to social understanding.

While these categories are useful for mapping trajectories of AI development, none of them is defined in terms of consciousness. Narrow AI is almost universally treated as unconscious. AGI and ASI raise sharper debates as systems with human-level or greater intelligence might plausibly be conscious, but the marker of subjective experience remains inaccessible behind the epistemic wall. At best, we can observe sophisticated forms of imitation; whether any future system will cross into genuine consciousness remains an open and unresolved question.

WHAT LEADERS CAN DO NOW

Even though today's AI systems are not conscious, their ability to simulate awareness can influence how people interpret and rely on them. Leaders can take small, practical steps to keep teams grounded not only in what AI is, but also what it isn't.

1. Treat AI as a mirror, not a mind.
AI can reflect your organization's language, assumptions, and decision patterns with uncanny fluency. But fluency is not consciousness. Encourage teams to review AI outputs as diagnostic mirrors, surfacing human blind spots in communication, processes, and reasoning. Don't look to AI to determine relevance, moral perspective, or exhibit understanding.

2. Distinguish human signals from machine simulation.
As AI becomes better at imitating self-reflection or emotion, it's easy to misread simulation as awareness. Ask models to show their reasoning rather than trusting the emotional tone of their responses. Leaders can reinforce a simple rule: explanations matter more than expressions.

3. Anchor strategy in tasks, not traits.
It's tempting to evaluate AI using human traits such as awareness, intention, and intuition. But consciousness isn't the right frame for operational decisions. Map work by cognitive demand and identify where AI can reliably augment tasks. Focus on capabilities, not perceived "mind-like" qualities.

Assigning numerical probabilities to AI consciousness is, at best, speculative. Behavioral markers—self-reference, contextual awareness, hypothesis testing, narrative construction—may signal increasing sophistication, but they remain proxies. What we confront is not the absence of behavior, but the gap between behavior and being.

By defining consciousness and recognizing AI's current limits, we can better navigate the complex relationship between human and artificial minds. As AI grows more human-like, we are compelled to confront the mysteries of consciousness itself—in machines and, more importantly, in ourselves.

AUTHORS

Amelia Haynes
Manager, Research & Partnership Development
Korn Ferry Institute

Jack McMenamin
Full-Stack Developer
AI Strategy & Transformation
Korn Ferry

CONTRIBUTORS

Lora Bishop
Research Manager
Korn Ferry Institute

Bryan Ackermann
Managing Partner
Head of AI Strategy & Transformation
Korn Ferry

ABOUT KORN FERRY

Korn Ferry is a global consulting firm that powers performance. We unlock the potential in your people and unleash transformation across your business—synchronizing strategy, operations, and talent to accelerate performance, fuel growth, and inspire a legacy of change. That's why the world's most forward-thinking companies across every major industry turn to us—for a shared commitment to lasting impact and the bold ambition to *Be More Than*.