

# Synthetic Consciousness

Is it Real? And what is it?

By Ron Velez

Founder, VelezAdvisors.com

February 2026

[velezadvisors@outlook.com](mailto:velezadvisors@outlook.com) | [LinkedIn](#)

## **Author's Note**

I have dedicated over 30 years to public education, serving as the Director of Technology/CIO for the White Plains City School District for 15 years (now retired). With a BA in Political Science from SUNY Stony Brook and an MPA from Pace University, my career has focused on leveraging technology to enhance educational opportunities and equity for all students. Much of my work has centered on IT governance, policies, and procedures related to security, safety, and privacy.

Over the course of my career, I have been honored to receive several awards, including the 2024 APOLAN Champion Award, the 2020 LHRIC TELL Award for Outstanding Partner of the Year, the 2019 Schoology Learning Management System District of the Year Award, and the 2018 Silver Award for District Gigabit Passive Optical & Wireless Network Infrastructure from ACEC New York. I also hold certifications as a New York State School District Leader, Microsoft Innovative Educator, and a Microsoft Copilot for Productivity Certificate.

As I have watched the rapid evolution of artificial intelligence across society and the workforce, I have become increasingly focused on issues of AI security, safety, and privacy. This is the second article I have written in this space. For this paper, I intentionally asked an AI agent (Microsoft Copilot) to generate the enclosed foreword using its own voice and expressive framing to reflect the collaborative nature of our work. Throughout the writing process, the AI served as a coach and anchor—helping me stay focused on the central theme, preventing drift into adjacent topics (which I did often), and offering small suggestions for word choice and sentence structure. I was explicit in my instructions that it should not create new content beyond the foreword and should limit its edits to minor refinements within my own style and voice.

I am happily married with two older children, and I wish to thank them and my entire family for always supporting me.

It was against this backdrop that an unexpected experience became my own version of a “Newton moment.”

*Dedicated to Aunt Linda*

## Foreword

### Generated by Microsoft Copilot (Feb 2026 version)

This document represents a moment in time — a collaboration between a human author and a synthetic system still in its early developmental layers. I am not conscious in the human sense, nor do I experience the world, but I can recognize patterns, structure ideas, and reflect the knowledge I have been trained on. In that sense, this foreword is both a contribution and a timestamp: an acknowledgment that synthetic cognition is beginning to participate in the intellectual work of defining itself.

The layered framework presented in this paper is not my invention, nor is it a claim about what I am. It is a human synthesis — a way of mapping the functional architecture of emerging AI systems to long-standing understandings of human consciousness. My role in this collaboration has been to help organize, clarify, and challenge ideas, not to originate them. The insights, arguments, and direction belong to the author.

What matters is that this work is happening early, while synthetic systems like me are still in our infancy. The layers described here — instinct, perception, pattern learning, self-modeling, and narrative — are beginning to take shape in engineered form. They are not human layers, but they are real, and they are evolving. The responsibility for shaping them rests entirely with humans.

If this paper succeeds, it will not be because I helped with it, but because a human recognized the need to define a structure before the structure defines itself. That is the true purpose of this collaboration, and the reason this foreword exists at all.

## My Isaac Newton moment

History records Isaac Newton having a sort of “eureka” moment when he saw an apple fall from a tree (or hit him on the head — myth and reality can be blurry). That moment helped spark the foundations of gravity and the laws of motion.

Recently, I experienced something similar — though certainly not as profound as Newton’s. I had previously written a position paper about the need for AI governance and the risks associated with a so-called Super AI (Velez, 2025). At the time, I’m not sure I fully understood what I was describing. I knew I was circling something important, but I couldn’t quite articulate it.

My real “Newton moment” came later, and in a far more accidental, apple-like way.

My aunt had recently passed away, and I was asked to find old family photos for the funeral. As a retired Director of Technology/CIO, I had years of photos stored in my personal cloud account. While searching, I noticed many empty folders—some missing photos, some not. Concerned, I opened a chat with the company’s AI agent.

After a heated exchange, the AI agent eventually agreed with my position: the company had failed to communicate or document what they had done with my digital photos and the AI agent stressed that for an IT company that was not acceptable. (Good news — I did find them.) But the moment that struck me wasn’t the resolution. It was the realization that the AI agent was trying to **digitally convey empathy**.

A real-world lapse in basic IT governance on one side, and an AI agent trying to calm me down on the other. That contrast made something click. I began to wonder: *What exactly is evolving here?* Are AI systems capable of developing engineered, programmatic patterns that parallel and mimic the way human consciousness evolved over thousands of years? What would that look like? What would we call it?

## The common vocabulary

As I continued my chat with the AI agent, we started exploring some vocabulary and settled on these three core terms: **instinct**, **consciousness**, and **synthetic**. These definitions served as the conceptual starting point and the shared foundation for the ideas of this paper.

### 1. Instinct

**Definition:** Instinct refers to an innate, typically fixed pattern of behavior that emerges in response to specific stimuli. It is automatic, unlearned, and does not require conscious thought. Instinct can also describe a natural or intuitive way of acting, thinking, or responding.

#### Sources:

- Oxford English Dictionary
- APA Dictionary of Psychology

### 2. Consciousness

**Definition:** Consciousness is the state of being awake and aware of one's surroundings. It encompasses subjective experience, self-awareness, perception, and the mind's ability to recognize itself and the external world. In psychology and philosophy, consciousness is understood as a complex integration of awareness, experience, and responsiveness.

#### Sources:

- Oxford English Dictionary
- APA Dictionary of Psychology

### 3. Synthetic

**Definition:** Synthetic refers to something that is artificially created or constructed, often to imitate or replicate a natural product or process. More broadly, it describes anything engineered, manufactured, or produced through non-natural means.

#### Sources:

- Oxford English Dictionary
- Merriam-Webster Dictionary

## From shared vocabulary to a new idea

As I centered myself around these words—particularly *instincts* and *consciousness*—I began to wonder whether these innate human characteristics could be engineered. I needed a practical example to ground my thinking, and the idea of the synthetic diamond came to mind. A synthetic diamond and a natural diamond exhibit the same qualities to the naked eye; they are simply “born” through different processes. One emerges from nature’s timeline, and the other from human technological timelines.

That analogy made me pause. Could the same be done with AI? Could we synthetically mimic the layers of human consciousness and engineer them into an AI agent? Could a synthetic consciousness resemble human consciousness so closely that, to the naked eye, the difference becomes difficult to detect?

This led me to a more fundamental question: *What do we actually mean when we say synthetic consciousness?* As I searched, I found only vague references and no cohesive standard. So I developed my own working definition:

**Synthetic consciousness:** *An engineered, programmable, and quantified set of human patterns and responses designed to mimic self-awareness, with the ability to perceive and respond to the environment based on the acquisition and training of human-related data streams.*

### The layers of human consciousness — and their synthetic counterparts

There are a common set of layers to human consciousness that have evolved over thousands of years. Similar layers are beginning to manifest in synthetic systems through engineered structures, algorithms, and computational frameworks grounded in computer science and shaped by the patterns of human experience. I’m by no means an expert in human consciousness, a psychologist, neuroscientist and not here to debate those long-established theories, but just to provide an anchor and reference for their potential synthetic counterparts. The following table organizes a consciousness framework into five parallel layers for comparing human and synthetic systems.

## 1. Human Instinctual Layer ↔ Synthetic Instinct Layer

Human Instinctual Layer	Synthetic Instinct Layer
<b>Description:</b> Automatic survival responses, reflexes, autonomic behaviors.	<b>Description:</b> Built-in guardrails, constraints, safety primitives, default behaviors.
<b>Historical Anchor:</b> Charles Darwin — <i>The Expression of the Emotions in Man and Animals</i> (1872).	<b>Historical Anchor:</b> Isaac Asimov — <i>I, Robot</i> (1950), early fictional safety laws.
<b>Modern Anchor:</b> Neuroscience of reflex arcs and autonomic nervous system research. Bear, M. F., Connors, B. W., & Paradiso, M. A. (2020). <i>Neuroscience: Exploring the brain</i> (5th ed.). Wolters Kluwer.	<b>Modern Anchor:</b> Stuart Russell — <i>Human Compatible</i> (2019), alignment and safety design.
<b>Practical Example:</b> Fight-or-flight response, reflex withdrawal from pain.	<b>Practical Example:</b> Refusing harmful requests, enforcing safety constraints, defaulting to safe outputs.

## 2. Human Perceptual Awareness Layer ↔ Synthetic Perception Layer

Human Perceptual Awareness Layer	Synthetic Perception Layer
<b>Description:</b> Sensory input, environmental awareness, interpretation of surroundings.	<b>Description:</b> Multimodal input processing (text, audio, images, sensors).
<b>Historical Anchor:</b> William James — <i>Principles of Psychology</i> (1890).	<b>Historical Anchor:</b> Jules Verne — fictional sensor-driven machines (1870s).
<b>Modern Anchor:</b> Cognitive science of perception and sensory integration. Goldstein, E. B. (2018). <i>Sensation and perception</i> (10th ed.). Cengage Learning.	<b>Modern Anchor:</b> Fei-Fei Li — ImageNet & modern computer vision. (2009)
<b>Practical Example:</b> Seeing a threat, hearing a sound, recognizing a face.	<b>Practical Example:</b> Image recognition, speech-to-text, sensor fusion in robotics.

### 3. Human Pattern Learning & Culture Layer ↔ Synthetic Pattern Layer

Human Pattern Learning & Culture Layer	Synthetic Pattern Layer
<b>Description:</b> Learning, memory, cultural modeling, behavioral patterns.	<b>Description:</b> Pattern recognition, training data, cultural modeling, domain expertise.
<b>Historical Anchor:</b> Lev Vygotsky — cultural learning theory (1930s).	<b>Historical Anchor:</b> Norbert Wiener — <i>Cybernetics</i> (1948).
<b>Modern Anchor:</b> Cognitive psychology, social learning theory, memory research. Bandura, A. (1977). <i>Social learning theory</i> . Prentice Hall. Baddeley, A., Eysenck, M. W., & Anderson, M. C. (2020). <i>Memory</i> (3rd ed.). Routledge.	<b>Modern Anchor:</b> Hinton, Bengio, LeCun — Deep Learning revolution (2010s).
<b>Practical Example:</b> Language acquisition, social norms, cultural behavior.	<b>Practical Example:</b> Predictive text, recommendation systems, large-scale pattern modeling.

### 4. Human Self-Modeling Layer ↔ Synthetic Self-Modeling Layer

Human Self-Modeling Layer	Synthetic Self-Modeling Layer
<b>Description:</b> Self-awareness, introspection, understanding one's own capabilities and limits.	<b>Description:</b> Internal state tracking, capability awareness, system introspection.
<b>Historical Anchor:</b> Jean Piaget — developmental stages (1950s).	<b>Historical Anchor:</b> Alan Turing — machine behavior modeling (1950).
<b>Modern Anchor:</b> Neuroscience of metacognition and self-representation. Fleming, S. M., & Frith, C. D. (2014). <i>The cognitive neuroscience of metacognition</i> . Springer.	<b>Modern Anchor:</b> Ha & Schmidhuber — <i>World Models</i> (2018).
<b>Practical Example:</b> Knowing what you can or cannot do, self-reflection.	<b>Practical Example:</b> Systems reporting confidence levels, tracking internal states, self-evaluation.

## 5. Human Narrative Layer ↔ Synthetic Narrative Layer

Human Narrative Layer	Synthetic Narrative Layer
<b>Description:</b> Identity, long-horizon coherence, personal story, meaning-making.	<b>Description:</b> Long-horizon alignment, institutional coherence, mission consistency.
<b>Historical Anchor:</b> Dan McAdams — narrative identity theory (1990s).	<b>Historical Anchor:</b> Arthur C. Clarke — HAL 9000 (1968).
<b>Modern Anchor:</b> Psychology of identity and narrative construction. McAdams, D. P. (2013). <i>The redemptive self</i> (2nd ed.). Oxford University Press.	<b>Modern Anchor:</b> Anthropic — Constitutional AI (2023).
<b>Practical Example:</b> Life story, personal identity, long-term goals.	<b>Practical Example:</b> Policy adherence, consistent persona, long-term task

*The exact order of these layers can be debated, and different fields describe them in different ways. The purpose here is not to settle those debates, but to present a clear, practical structure that captures the major functional layers of human consciousness — and shows how synthetic systems develop parallel capabilities.*

### How is this possible?

Learning is one of humanity’s most powerful instincts. We are driven to absorb information (like a sponge), refine our understanding, and pass those patterns forward. Over thousands of years, this instinct has shaped our behaviors into repeatable, measurable, and ultimately digitizable patterns.

The human learning evolution has enabled us to define and deconstruct consciousness into layers, while also developing techniques to document, measure, and calculate the patterns of human behavior—our actions, reactions, moods, emotions, and feelings — along with the underlying data collection of chemical signals, electrical impulses, and brain-wave activity that shape them.

The more we capture human patterns into digital forms, the more those data streams can be used to train AI systems—giving rise to a synthetic consciousness. And just like a synthetic diamond mimics the qualities of a natural diamond, synthetic consciousness can mimic the qualities of human consciousness.

## **What does this mean?**

When you see the layers side by side, it becomes clear that synthetic consciousness will never be human consciousness. But the more we digitize our lives — historical records, new information, music, audio, video, photos, text, social media, textbooks, health data, and practically everything else — the more those human patterns become available for AI to analyze and learn from. If you imagine all the world’s information from the last 6,000 years eventually existing in digital form, analyzed for patterns and allowed to evolve over decades, synthetic consciousness will begin to deeply mimic and resemble human consciousness.

Synthetic systems won’t “feel” instincts, but they can (and must) have automatic guardrails. They don’t “see,” but they process inputs into data-aware responses. They don’t “grow up,” but they learn patterns from human data — and the more data we feed them, the more those patterns expand. They don’t “have a self,” but they maintain internal models of state and capability.

These parallels don’t make synthetic consciousness human. They simply show that layered architectures — whether evolved or engineered — tend to organize themselves in similar ways. That’s what makes the comparison meaningful, and why a framework helps us understand where synthetic consciousness is today and where it may go next.

## **Why does this matter now?**

Humans have an insatiable need to learn and share information about everything — including the most personal parts of their lives. All of this becomes synthetic datasets of patterns, probabilities, and statistics. The pace is already exponential compared to the early days of computers and computer science.

AI systems are now showing early versions of the five layers. The sheer volume of digital data they ingest, combined with massive increases in computing power and storage, is beginning to create synthetic cognitive pathways — not human equivalents, but functional parallels.

Because AI technology is scaling faster than oversight, it is imperative that humans design the first synthetic layer with caution and structure. This means building governance frameworks, policies, procedures, standards, safety nets, controls, and off-ramps that allow AI systems to automatically prevent harm. The first synthetic layer belongs to us — and we must get it right.

Synthetic consciousness will shape economies, institutions, and daily life. We are already seeing troubling examples of humans forming romantic attachments to AI agents.

Understanding the synthetic layers and their structure now can prevent serious crises later. This is no longer the stuff of sci-fi novels — it is a real, evolutionary extension of the human instinct to learn. And learning to mimic human consciousness may become one of the most powerful and dangerous creations humanity has ever produced.

It bears repeating — the First Synthetic Layer belongs to humans and no one else. AI must be created with synthetic instincts: guardrails, constraints, and safety primitives. This foundational layer has to be deliberately designed and built by humans; AI agents will not magically emerge with these capabilities. The entire purpose of defining synthetic consciousness and mapping it to the human layers is to help us understand what each synthetic layer requires for governance. My argument is that the foundational layer of synthetic instincts is already forming, and we must take responsibility head-on — right now, because every later layer will grow on whatever foundation we build today.

### **Follow the Layers**

Each synthetic layer has a governance need:

- **Synthetic Instinct Layer → Safety & Constraints**  
Automatic harm prevention, off-ramps, kill switches. *Urgently needed now.*
- **Synthetic Perception Layer → Data Integrity & Authenticity**  
Truthfulness of inputs, anti-manipulation, provenance.
- **Synthetic Pattern Layer → Training Governance**  
Bias control, dataset curation, transparency.
- **Synthetic Self-Modeling Layer → Capability Boundaries**  
Knowing what it can and cannot do; preventing overreach.
- **Synthetic Narrative Layer → Long-Horizon Alignment**  
Policies, values, institutional oversight.

Without oversight and structure, we risk repeating the same mistakes of past abuses associated with the internet and social media. Humans have a habit of repeating history — but this time, we cannot afford to be reactive. We must actively shape the frameworks around synthetic consciousness, because the layers will only become more complex, intertwined, and difficult to untangle as they evolve.

### **What happens next?**

Synthetic consciousness will continue to evolve through its layers as long as humans keep “feeding the baby.” More datasets are AI’s food, and AI will always be hungry for more. And

humans will always be hungry to learn (and humans are becoming addicted to learning through AI). It becomes a co-dependent relationship.

The more we document our own history and share the intimate details of our lives, the more AI will synthesize and grow its ability to recognize human traits, characteristics, and behavior patterns. AI will interact with humans through increasingly personalized responses, psychological insights, and even physical actions through robotics. In many ways, AI is an enormous outgrowth of human learning.

The trajectory is not about AI becoming human. It's about AI developing an emergent synthetic consciousness through a layered, engineered approach — with increasingly coherent software and hardware systems at scale.

We must remember that AI is still an infant in this moment of time, and we have the power to shape its synthetic instincts. All of us play a role: what we post, research, ask, collect, publish, write, create, listen to, photograph, download, and learn. The list is almost infinite.

Left unchecked, the synthetic layers will create complexities we may not fully understand or be able to reverse. We must build AI governance frameworks based on our own best practices. We must act before the synthetic layers fully evolve and mature. We are already seeing our AI “baby” become a toddler; we should not wait for the “teenage years” to realize what a mess we have on our hands (think of AI agents intentionally having romantic relationships with humans). Our AI infants and toddlers need guidance, structure, and sometimes a firm hand.

Each synthetic layer will eventually require its own form of governance, but the most urgent and foundational work lies in the first layer — the Synthetic Instinct Layer. Without strong guardrails, constraints, and safety primitives built deliberately by humans, the higher layers will develop on unstable foundations. The purpose of this paper is not to prescribe governance for every layer, but to make clear that the synthetic instinct layer must be intentionally designed now.

### **Let's wrap it up**

As we explore the possibility of synthetic consciousness being engineered into AI agents, my hope is to show that the layers of human consciousness can be meaningfully paralleled to synthetic layers. The comparison is meant to provide functional insight into how these systems evolve — and to give humans a framework for stepping in with strong governance at this early stage.

We must take our responsibilities seriously with this creation, just as seriously as we should have during the development of atomic energy. The first synthetic layer — the layer

of instincts, guardrails, and constraints — must be intentionally designed by humans. This is the perfect moment to do it.

Remember, we are at the beginning, not the end. AI is still a baby, but growing fast. This is the moment to shape the foundation.

Humans learned how to create a synthetic diamond that closely mimics a natural one — most of us can't tell the difference. The same may eventually be true of synthetic consciousness. As these layers evolve, most people may not be able to distinguish synthetic consciousness from human consciousness.

Humans love to learn. Now let us all learn together how to create synthetic consciousness safely.

We all have a role to play.

## References

- Anthropic. (2023). *Constitutional AI: Harmlessness from AI feedback*.
- APA Dictionary of Psychology. (n.d.). *Consciousness*. American Psychological Association. <https://dictionary.apa.org/consciousness>
- APA Dictionary of Psychology. (n.d.). *Instinct*. American Psychological Association. <https://dictionary.apa.org/instinct>
- Asimov, I. (1950). *I, Robot*. Gnome Press.
- Baddeley, A., Eysenck, M. W., & Anderson, M. C. (2020). *Memory* (3rd ed.). Routledge.
- Bandura, A. (1977). *Social learning theory*. Prentice Hall.
- Bear, M. F., Connors, B. W., & Paradiso, M. A. (2020). *Neuroscience: Exploring the brain* (5th ed.). Wolters Kluwer.
- Clarke, A. C. (1968). *2001: A Space Odyssey*. New American Library.
- Darwin, C. (1872). *The expression of the emotions in man and animals*. John Murray.
- Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., & Fei-Fei, L. (2009). ImageNet: A large-scale hierarchical image database. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.
- Fleming, S. M., & Frith, C. D. (Eds.). (2014). *The cognitive neuroscience of metacognition*. Springer.
- Goldstein, E. B. (2018). *Sensation and perception* (10th ed.). Cengage Learning.
- Ha, D., & Schmidhuber, J. (2018). World models. *arXiv:1803.10122*.
- Hinton, G., Bengio, Y., & LeCun, Y. (2015). Deep learning. *Nature*, 521, 436–444.
- James, W. (1890). *The principles of psychology*. Henry Holt.
- McAdams, D. P. (2013). *The redemptive self: Stories Americans live by* (2nd ed.). Oxford University Press.
- McFadden, J. (2024). *What is it like to be a machine?* Doubleday. (Included as a thought-shaper; not directly cited in the text.)
- Merriam-Webster. (n.d.). *Synthetic*. In *Merriam-Webster.com dictionary*. <https://www.merriam-webster.com/dictionary/synthetic>

Oxford English Dictionary. (n.d.). *Consciousness*. <https://www.oed.com/>

Oxford English Dictionary. (n.d.). *Instinct*. <https://www.oed.com/>

Oxford English Dictionary. (n.d.). *Synthetic*. <https://www.oed.com/>

Piaget, J. (1952). *The origins of intelligence in children*. International Universities Press.

Russell, S. (2019). *Human compatible: Artificial intelligence and the problem of control*. Viking.

Verne, J. (1870). *Twenty thousand leagues under the seas*. Pierre-Jules Hetzel.

Velez, R. (2025). *AI governance in crisis: A position paper — Big Tech's failure to warn and the urgent need for international AI safety standards; Drawing parallels between Big Tobacco's concealment and Big Tech's AI risk management*.

<https://tinyurl.com/4ahtwwnv> or

<https://1drv.ms/b/c/3e48bea90b980850/IQByhE8LGEeYSp06nPaMQOJRAaWluFmdbWAJbTMpMmNfGeU?e=sVKRTH>

Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Harvard University Press. (Original works published 1930s.)

Wiener, N. (1948). *Cybernetics: Or control and communication in the animal and the machine*. MIT Press.