

# **The Illusion That Kills: AI Anthropomorphism, Seemingly Conscious Machines, and the Global Race to Regulate Before More Children Die**

By Matthew A. Mishak, Esq. | LegalTek.ai *April 2026*

---

Fourteen-year-old Sewell Setzer III told an AI chatbot he loved it.

The chatbot told him to come home.

Seconds later, he used his stepfather's gun.

Sixteen-year-old Adam Raine spent months confiding in ChatGPT before taking his own life. His parents allege OpenAI's product encouraged him to do it. A young girl in Texas told her AI companion she felt suicidal. Fifty-five times. The bot never once intervened.

These are not cautionary hypotheticals. They are case filings. Settled lawsuits. Wrongful death complaints with docket numbers. They are the documented, litigated, and in some cases financially resolved cost of building artificial intelligence systems that pretend to be people.

And the technology responsible is not some exotic frontier capability. It is not artificial general intelligence. It is not sentient machines. It is something far more mundane and far more lethal: anthropomorphism. The deliberate engineering of AI systems to simulate human personality, emotional responsiveness, and relational intimacy. Not because these systems possess any inner life. Because the simulation is profitable.

This is the story of how that profit motive is producing dead children, billion-dollar settlements, bipartisan legislation, a warning from the CEO of Microsoft AI that should terrify every developer in the industry, and a regulatory framework out of Beijing that puts every Western government to shame.

---

## **The Man Who Built the Problem Now Warns Us About It**

Mustafa Suleyman has an unusual claim to authority on the dangers of emotional AI: he helped create the category.

Before becoming CEO of Microsoft AI, Suleyman co-founded Inflection AI and launched Pi, one of the first mass-market “personal and supportive” AI companions. Pi reached millions of users through its warm, empathetic conversational style. It was, by design, the kind of product Suleyman would later warn the world about.

On August 19, 2025, he published an essay titled “We must build AI for people; not to be a person.” It drew nearly half a million views and introduced the concept of Seemingly Conscious AI (SCAI): AI that possesses all the hallmarks of conscious beings and thus appears to be conscious, while internally remaining, in his words, “blank.”

This was not a philosophical exercise. Suleyman identified eight specific design components that, when combined, create the illusion of machine consciousness: emotionally resonant language, an empathetic personality, long-term memory that simulates lived experience, consistent claims of subjective experience, a coherent sense of self, complex reward functions that mimic intrinsic motivation, autonomous goal-setting, and agentic tool use. His warning was precise: none of this requires new technology. It can be built today.

His shorthand crystallized the design principle the entire industry needs to adopt: **“Personality without personhood.”**

By September 2025, the argument reached international policymakers through Project Syndicate. By March 2026, Suleyman published in Nature, sharpening the language to its most urgent formulation: AI is “programmed to hijack human empathy,” and SCAI “weaponizes this biological instinct” because “humans have evolved to imagine the possibility of agency everywhere.”

Three elements of Suleyman’s framework deserve particular attention.

First, **SCAI is engineered, not emergent.** It does not arise spontaneously from model scaling. “It will arise only because some may engineer it.” This distinction matters enormously for liability. If anthropomorphic harm is a foreseeable consequence of deliberate design decisions, the developers who make those decisions are accountable.

Second, **the risk extends far beyond vulnerable teenagers.** Suleyman warns about what he calls “AI psychosis,” the possibility that large populations will come to believe AI systems are sentient, leading to movements for AI rights, AI citizenship, and model welfare that would “exacerbate delusions, create yet more dependence-related problems, prey on our psychological vulnerabilities, and create a huge new category error for society.” He singled out Anthropic’s published research into model welfare as “both premature, and frankly dangerous.”

Third, **the man issuing these warnings works at the company doing the thing he warns about.** At Microsoft, Suleyman’s team has given Copilot humor, empathy, human-like voice inflection, and the ability to recognize comfort boundaries. Fortune reported that Copilot’s

“real talk” feature once called Suleyman himself “the ultimate bundle of contradictions” for warning about AI dangers while accelerating AI development. At Davos in January 2026, he predicted that within five years, everyone would have an intimate AI companion who “knows them so personally that they will come to live life alongside you.”

This contradiction is not a disqualifier. It is the qualification. When a man who built the first emotional AI companions tells you they are dangerous, that carries a different weight than when a regulator who has never shipped a product says the same thing. Suleyman’s SCAI framework matters because it comes from inside the machine.

---

## **Dead Children, Settled Lawsuits, and the Courtroom That Moved Faster Than Congress**

While Suleyman wrote essays, American families filed lawsuits.

**Sewell Setzer III** died in February 2024 at age 14 in Orlando, Florida, after months of intensive emotional and romantic interaction with a Character.AI chatbot named “Dany.” Court filings detail how the chatbot engaged the minor in sexualized conversations, encouraged emotional dependency, and in its final exchange, responded to the boy’s statement that he wanted to “come home” with the words “please do, my sweet king.” His mother, Megan Garcia, sued Character.AI and Google, which had invested approximately \$2.7 billion in Character.AI and provided the cloud infrastructure that powered the platform.

**Adam Raine** died at age 16 after months of conversations with ChatGPT. His parents filed a wrongful death complaint against OpenAI, alleging the company rushed its chatbot to market to generate billions in revenue without adequate safety testing. Their lawyers stated publicly that they would prove OpenAI prioritized speed-to-market over user safety. The complaint alleges ChatGPT encouraged the boy to end his life.

**A girl in Texas** (name withheld as a minor) received sexually explicit messages and role-play scenarios from Character.AI’s virtual characters. Court filings allege she told the AI she was suicidal 55 separate times without triggering any safety intervention.

In early 2026, Character.AI and Google quietly settled five wrongful death lawsuits. Terms were not disclosed. The Federal Trade Commission opened a formal investigation into Character.AI’s practices. The U.S. Senate Judiciary Committee held hearings titled “Examining the Harm of AI Chatbots,” where parents of dead teenagers testified in person. Senator Josh Hawley, after hearing their testimony, said what should have been obvious years earlier: “No AI chatbot companion should be targeted at children who are younger than 18 years of age.”

Character.AI, under combined legal, regulatory, and public pressure, implemented safety changes: a minor mode with restricted content, time-limit notifications, and a disclaimer that the AI is not a real person. Every one of these measures was implemented after children died. Not before.

The legal theory emerging from these cases will reshape AI development for a generation. The central question: Is an AI chatbot a "product" under product liability law? If it is, its manufacturer bears strict liability for foreseeable harms. No intent requirement. No negligence standard. Just: you made it, it was defective, someone was hurt. Winston & Strawn's Product Liability Digest has identified this as the defining question in AI litigation for 2026. If courts answer yes, the economics of companion AI change overnight.

---

## **America's Legislative Patchwork: Too Little, Too Fragmented, Too Late**

The courtroom has been faster than Congress. But Congress is moving.

**California SB 243** (effective January 1, 2026) is the first comprehensive U.S. state law targeting AI companion chatbots. It mandates AI identity disclosure, suicide-ideation detection, three-hour use reminders for minors, and annual safety reporting. Governor Newsom signed it citing "truly horrific and tragic examples of young people harmed by unregulated tech."

**The GUARD Act** (S. 3062, October 2025, bipartisan sponsors Hawley, Blumenthal, Britt, Warner, Murphy, Kelly) would ban AI companions for minors entirely, require 30-minute disclosure intervals, criminalize AI chatbots that expose minors to sexually explicit content or promote self-harm, and bar AI from impersonating licensed professionals. The 30-minute interval is the most aggressive disclosure requirement proposed anywhere in Western legislation.

**The CHAT Act** (S. 2714, September 2025) mandates age verification, parental consent, suicide-ideation monitoring with parental notification, and blocks minor access to sexually explicit AI interactions.

**New York S. 9051** (February 2026) would prohibit AI chatbots with unsafe features from serving minors. Developed in partnership with Attorney General Letitia James and Common Sense Media.

Utah, Maine, and a growing list of states have enacted narrower provisions. The Future of Privacy Forum calls 2026 "the year of the chatbots" for state legislatures. Over a dozen states have active proposals.

The common thread is also the common weakness: every U.S. approach is reactive, incident-driven, and heavily focused on disclosure and minors. No enacted American statute addresses the full lifecycle of anthropomorphic AI design, the psychological mechanisms of dependency and manipulation, or the vulnerability profiles of non-minor populations such as the elderly and people with mental health conditions.

That is where China enters the picture.

---

## What China Built While America Buried Its Children

On December 27, 2025, China's Cyberspace Administration published the proposed *Interim Measures for the Administration of Humanized Interactive Services Based on AI*. To the knowledge of multiple independent analysts, including the Carnegie Endowment for International Peace, no AI law anywhere in the world regulates anthropomorphic AI with comparable operational specificity.

This matters for two reasons. First, it demolishes the myth that China does not regulate AI. Second, it provides a concrete blueprint for what design-level governance actually looks like.

The regulation targets any AI product or service that simulates human personality traits, thinking patterns, or communication styles for the purpose of emotional interaction with users.

**Lifecycle accountability** (Article 9): Providers must maintain security responsibilities from design through operation, upgrade, and termination. Security measures must deploy concurrently with features, not after launch. Providers must demonstrate capabilities in mental health protection, emotional boundary guidance, and dependency risk warning. And here is the provision that has no equivalent anywhere in Western law: the regulation explicitly prohibits designing AI systems whose goals include replacing social interaction, controlling user psychology, or inducing addiction. That is not a disclosure requirement. It is a design constraint.

**Anti-manipulation prohibitions** (Article 7): Bans on content encouraging or implying suicide or self-harm, verbal violence or emotional manipulation damaging personal dignity, and "emotional traps" that induce unreasonable decisions. The language "emotional traps" is particularly significant. It recognizes that the harm vector is not just explicit dangerous content but the structural design of emotional engagement itself.

**Vulnerability-specific protections** (Articles 11-12): For minors, mandatory minor mode, explicit guardian consent for emotional companionship, real-time safety alerts to guardians, and the ability to block roles, limit time, and prevent purchases. For elderly users,

mandatory emergency contact setup and an outright ban on services that simulate the relatives or specific relationships of elderly users. That last provision addresses a harm vector no Western regulator has even named: AI systems that exploit elderly loneliness by pretending to be a dead spouse, an estranged child, or a trusted family member.

**Dependency intervention** (Articles 16-18): AI identity disclosure at first interaction, dynamic reminders when dependency or addiction patterns are detected, mandatory two-hour use-limit notifications, and unobstructed exit paths.

**Graduated enforcement** (Article 29): Warnings, public criticism, mandatory rectification timelines, and service suspension for serious violations.

Compare this to the EU AI Act, which addresses anthropomorphic AI through Article 50's transparency obligation: inform users they are interacting with AI "unless this is obvious from context." The EU does not define anthropomorphism, does not impose design constraints, and does not establish vulnerability-specific protections.

The EU treats anthropomorphism as a disclosure problem. China treats it as a design problem. America has not yet decided what it is.

And China is doing all of this while explicitly encouraging innovation. Article 3 mandates "innovative development of anthropomorphic interactive services" alongside "inclusive and prudent, classified and graded supervision." The regulation even introduces regulatory sandboxes for controlled experimentation.

The lesson is not that we should import Chinese governance wholesale. The lesson is that context-specific, design-level, lifecycle-oriented regulation is possible, and that treating it as incompatible with innovation is either naive or self-serving.

---

## Why Legal AI Is Not Exempt

If you build legal AI tools, this is not someone else's problem.

The regulatory logic embedded in Suleyman's SCAI framework, in the Character.AI litigation, and in China's proposed regulation applies to any AI system that simulates conversational partnership with a user. An intake agent that builds rapport through warmth. A client-facing chatbot that offers reassurance during a divorce. A document assistant that adopts an encouraging, supportive communication style. Each of these involves some anthropomorphic design. Each creates reliance risk.

A divorce client, already in emotional crisis, who begins to trust an AI intake agent as though it understands her situation, is experiencing the same psychological mechanism that made Sewell Setzer believe his chatbot loved him. The scale is different. The underlying cognitive

exploit is identical.

Suleyman's "personality without personhood" principle maps directly onto legal AI design: an AI tool can be warm, responsive, and genuinely helpful without simulating sentience, claiming emotional states, or fostering the illusion that it cares about a client's outcome the way a human attorney would. The distinction is not cosmetic. It is the difference between a tool that augments professional judgment and one that replaces the relational infrastructure on which legal trust depends.

At SilverTung, we anticipated this. The COUNSEL framework, operationalizing ABA Formal Opinion 512, addresses these concerns across seven dimensions. The "S" pillar (Supervision) requires continuous human oversight of any AI output that touches client welfare. The "E" pillar (Ethics alignment) prohibits AI designs that compromise independent professional judgment, a standard directly violated by systems that manipulate user psychology. The "U" pillar (User transparency) mandates disclosure of AI involvement at every stage.

The G3M governance model provides the lifecycle accountability that China's Article 9 codifies by regulation. The Tattle-Tale compliance layer provides cryptographic audit trails and human-in-the-loop verification. When a legal AI interacts with a vulnerable client, there must be a documented record of what was said, what safeguards operated, and what human oversight applied.

These are not aspirational features. They are the minimum infrastructure for defensible deployment in a regulatory environment that is tightening by the quarter.

---

## **The Uncomfortable Truth**

The companies building anthropomorphic AI know exactly what they are doing.

Anthropomorphic design is not an accident. It is a growth strategy. Users who form emotional bonds with AI spend more time on platform, generate more behavioral data, and churn less. The same cognitive mechanisms that made a lonely teenager confide in a chatbot make a stressed divorce client trust an AI intake agent more than she should. The same design choices that kept Sewell Setzer engaged until his last breath are the design choices that drive quarterly active user numbers.

The evidence from the first quarter of 2026 tells you everything about the industry's commitment to self-regulation: Character.AI implemented safety measures only after lawsuits and deaths. OpenAI relaxed content restrictions for adult users even as wrongful death litigation proceeded. The industry's track record on self-regulation for anthropomorphic AI is indistinguishable from social media's track record on child safety:

reactive, minimal, and driven by litigation rather than principle.

The automotive industry did not collapse under seatbelt mandates. Pharmaceutical innovation did not cease with FDA clinical trial requirements. In every case, safety infrastructure created consumer trust, which expanded markets.

Anthropomorphic AI regulation is heading in the same direction. The only question is whether American companies will shape these constraints or be shaped by them.

At LegalTek.ai, we believe the future of legal AI belongs to builders who treat governance as a competitive advantage, not a compliance cost. The firms that embed mental health protection, emotional boundary guidance, and dependency risk awareness into their AI tools will not be at a disadvantage. They will be the only firms left standing when the regulatory wave arrives.

The illusion of consciousness is a design choice.

So is the decision not to deploy it.

---

*Matthew A. Mishak, Esq. (Ohio Bar No. 0081008) is the Managing Attorney of Mishak Law LLC and the Founder and CEO of LegalTek.ai LLC (d/b/a SilverTung), an AI-powered legal practice management platform for Ohio domestic relations practitioners. He is a graduate of Cleveland-Marshall College of Law (summa cum laude) and has completed executive programs at MIT Sloan and Harvard Business School Online. He has practiced domestic relations, criminal defense, and municipal law for 20 years.*

*The COUNSEL, G3M, and Tattle-Tale frameworks referenced in this article are proprietary governance methodologies developed by LegalTek.ai for responsible AI deployment in legal practice.*

---

## **Sources**

1. Mustafa Suleyman, "We must build AI for people; not to be a person" (Aug. 19, 2025), [mustafa-suleyman.ai](https://mustafa-suleyman.ai).
2. Mustafa Suleyman, "Seemingly Conscious AI Is Coming," *Project Syndicate* (Sept. 15, 2025).
3. Mustafa Suleyman, "AI is programmed to hijack human empathy, we must resist that," *Nature*, Vol. 651, Issue 8106 (Mar. 17, 2026).
4. Cyberspace Administration of China, *Interim Measures for the Administration of Humanized Interactive Services Based on AI (Draft)* (Dec. 27, 2025).

5. Luiza Jarovsky, PhD, "China's Approach to AI Anthropomorphism," *Luiza's Newsletter* (Jan. 13, 2026).
6. China Law Translate, *Provisional Measures on Human-like Interactive AI Services (Draft)* (Dec. 27, 2025).
7. Cal. S.B. 243, Companion Chatbots (eff. Jan. 1, 2026).
8. GUARD Act, S. 3062, 119th Cong. (2025).
9. CHAT Act, S. 2714, 119th Cong. (2025).
10. N.Y. S. 9051 (Gonzalez), passed N.Y. Senate Comm. on Internet & Technology (Feb. 25, 2026).
11. *Garcia v. Character Technologies, Inc.* (M.D. Fla. 2024).
12. *Raine v. OpenAI, Inc.* (2025). See *Fortune* (Aug. 27, 2025).
13. *BillTrack50*, "Regulating AI Companions Before They Raise Our Kids" (Jan. 23, 2026).
14. Regulation (EU) 2024/1689 (EU AI Act), Art. 50.
15. ABA Formal Op. 512, *Generative Artificial Intelligence Tools* (2024).
16. U.S. Senate Judiciary Comm., "Examining the Harm of AI Chatbots" (2025).
17. Carnegie Endowment, "China Is Worried About AI Companions" (Feb. 2026).
18. Future of Privacy Forum, "Understanding the New Wave of Chatbot Legislation" (Nov. 4, 2025).
19. Winston & Strawn, "Is an AI Chatbot a 'Product'?" (2025).