

UC Berkeley

Perspectives in Medical Humanities: Essays and Articles

Title

Mind the Machine: Why Do Humans Keep Giving Intelligence Away?

Permalink

<https://escholarship.org/uc/item/0bd4m7s0>

Journal

Perspectives in Medical Humanities

Author

Dolan, Brian

Publication Date

2026-02-20

DOI

10.5072/FK2PZ5CP35

Copyright Information

This work is made available under the terms of a Creative Commons Attribution-NonCommercial-NoDerivatives License, available at

<https://creativecommons.org/licenses/by-nc-nd/4.0/>

Peer reviewed

Mind the Machine: Why Do Humans Keep Giving Intelligence Away?

Brian Dolan, PhD

University of California, San Francisco

brian.dolan@ucsf.edu

Orchid ID: <https://orcid.org/0000-0002-7935-8915>



Abstract

Developments in computing and electrical engineering which ushered in the field of cybernetics in the twentieth century raised moral and ethical concerns about how humans perceive machines to possess intelligence, autonomy, and authority. The biggest challenge that each generation has confronted is not machines that truly think, but that humans were willing to treat them as if they do. The ethical dilemma, as philosophers and critics have warned, is not that machines will assume moral agency, but that humans will relinquish their own. The history of artificial intelligence therefore tells us less about the rise of machine minds than about our willingness to redefine intelligence in the image of our inventions. To understand the ethical stakes of AI in education and society today, we must first understand this deeper pattern: that the most consequential transformation may not be in our machines, but in ourselves.

Citation: Dolan, Brian. (2026) Mind the Machine: Why Do Humans Keep Giving Intelligence Away? *Perspectives in Medical Humanities* (UC Medical Humanities Consortium, February 20, 2026)
DOI: :10.5072/FK2PZ5CP35

Received: December 16, 2025

Accepted: January 11, 2026

Published: February 20, 2026 (Digital)

Copyright: 2026
CC BY-NC-ND-4.0 License

Prelude

A woman sat at her computer and decided to air some grievances about men. She opened a chatbot window and typed a statement: "Men are all alike." When she hit return, the chat responded with a query for her: "In what way?" The woman dashed off her answer: "They're always bugging us about something or other." "Can you think of a specific example?" the chat replied, seeming to probe the validity of her declaration.

The woman pauses. For a moment, it feels less like she's using a word processing tool and more like she's interacting with something that thinks. Not because the machine possesses thoughts or intentions, but because it performs as if it did so convincingly. After a few more minutes of chatting through the window of her computer, the conversation progressed to her feelings of depression and her desire to be undisturbed in the room while she sought further counsel from the computer.



Attribution-NonCommercial-NoDerivs

Peer Review:

This article has been peer reviewed through a collaborative review process through a platform provided by the UC Medical Humanities Consortium consisting of a multi-disciplinary faculty editorial board. More information about collaborative review can be found at: <http://ucmedicalhumanitiespress.com/>

The interaction I'm describing here occurred in 1966, in the laboratory of a computer scientist named Joseph Weizenbaum.¹ The MIT professor had created a program called ELIZA which imitated a psychotherapist by rephrasing users' statements as questions. When someone typed, "I feel anxious about my future," ELIZA replied, "Why do you feel anxious about your future?" Weizenbaum knew the program was simple—a set of pattern-matching rules with no understanding at all. What surprised him was not what the machine could do, but what people were willing to believe it was doing. They confided in it. They trusted it. They attributed empathy where none existed.

This moment was not the birth of machine intelligence. It was the revelation of something about *human* intelligence—about our readiness to see ourselves reflected in our machines.

For more than seventy years, from the cybernetic theories of Norbert Wiener to Alan Turing's imitation game and beyond, we have wrestled with the same question: not whether machines can think, but what happens when humans treat them as if they do. Each time machines become more convincing, the ethical danger is framed (usually in sci-fi) as a future in which machines surpass us. But history suggests a different risk. The greater danger is that we begin to see ourselves as less responsible, less essential, and less accountable in a world we increasingly experience through computational systems.

Cybernetic Brains and Economic Displacement

In his 1948 book *Cybernetics: Or Communication and Control in the Animal and the Machine*, the MIT mathematician Norbert Wiener wrote, "It has long been clear to me that the modern ultra-rapid computing machine was in principle an ideal central nervous system to an apparatus for automatic control; ... With the aid of strain gauges or similar agencies to read the performance of these motor organs and to report to, to "feed back," to the central control system as an artificial kinesthetic sense, we are already in a position to construct artificial machines of almost any degree of elaborateness of performance."² While not claiming that machines possess consciousness or moral understanding, Wiener recognized an economic incentive to engineer skilled machines, giving "the human race a new and most effective collection of mechanical slaves to perform its labor."³ Though this appeared to avoid the direct cruelty of human slavery, he recognized a different ethical dilemma: "Any labor that accepts the conditions of [human] competition with slave labor accepts the conditions of slave labor, and is essentially slave labor."⁴

The danger was not cruelty to machines, but that unregulated automation

1. A historic video of this interaction can be found on YouTube, "[Before Siri and Alexa there was ELIZA](#)," accessed November 4, 2025.

2. Norbert Wiener, *Cybernetics: Or, Control and Communication in the Animal and the Machine* (Cambridge, MA: MIT Press, 1948), pp. 26-27.

3. Ibid, p. 27.

4. Ibid, p. 27.

normalizes an economy where efficiency justifies treating human labor as expendable. This, in turn, degrades human value. Following the machine revolution's success, Wiener warned, "the average human being of mediocre attainments or less has nothing to sell that is worth anyone's money to buy."⁵ How could humans be considered anything but mediocre compared to machine performance and apparent intelligence?

5. Ibid, p. 28.

From Economic to Systemic Displacement

Ethical reassurance was provided with the idea that humans retain ultimate control over machines. However, English psychiatrist W. Ross Ashby undermined this comfort by working to remove human intervention from the equation, replacing it with adaptive systems where machines could reorganize their own internal structure to maintain viability under changing conditions. In his books, *Design for a Brain* (1952) and *An Introduction to Cybernetics* (1956), he argued that machines could adapt, learn, and express intelligence in ways that served their own goals and that were not bound by social constructs such as human value and dignity.

As a demonstration, Ashby invented the homeostat, an electro-mechanical machine whose sole function was keeping a needle centered, automatically correcting disturbances by adjusting internal parameters. Most significantly, when ordinary regulation broke down, even if caused purposely by human subterfuge, the machine activated a "step-mechanism" that randomly altered internal connections until stability was restored.⁶

6. W. Ross Ashby, *Design for a Brain* (New York: Wiley & Sons, 1952) see Chapter 8.

Ashby argued that the brain operates analogously: its primary task is homeostasis, and when habitual responses fail, it reorganizes until equilibrium returns. In his framework, adaptive behavior results from trial and error governed by feedback, not by symbolic thought or conscious planning. He argued that we don't need to invoke "intelligence" or vital properties to explain the brain's learning ability. Instead, the nervous system is a strictly deterministic machine governed by equilibrium processes.

Where Wiener warned of humans being economically displaced by machines, Ashby revealed the deeper risk that humans remain present but increasingly occupy roles defined by system needs, such as editing inputs or overseeing maintenance. For Ashby, this was a shift from humans being morally subjugated by machines to them being systemically incorporated to their design.

The Problem of Giving Intelligence Away

Like Wiener and Ashby, the famous artificial intelligence pioneer Alan Turing also flipped around metaphysical debates about human intelligence by proposing a pragmatic question: could a human distinguish between a human and a machine when communicating with each through text? His “imitation game”⁷ proposed that if a machine could convince a person it was human through meaningful linguistic interaction, then attributing intelligence to the computer is justified without claiming it possesses a soul or inner mental states that had been considered unique to humans.

By presenting machine intelligence as a parlor-like “game,” Turing deliberately put aside theological and moral anxieties by programming playfulness. However, the risk posed by technologies appearing playful, but which astound us with their apparent “intelligence,” was that they inadvertently deceived humans into thinking that such systems could do no harm. They’re just a game, right? One implication was that if humans attributed intelligence to machines based on interactions, they might begin to bestow other human traits onto similarly well-designed mimics. This is something Turing contemplated in 1947, a few years before his famous article, when he wrote: “One way of setting about our task of building a ‘thinking machine’ would be to take a man as a whole and try to replace all the parts of him by machinery. He would include television cameras, microphones, loudspeakers, wheels and ‘handling servo-mechanisms’ as well as some sort of ‘electronic brain.’ This would of course be a tremendous undertaking.”⁸

The lesson of Turing’s vision of the future, where lifelike robots might walk among and be indistinguishable from humans, was not that machine intelligence might be attainable, but that the proof of their intelligence was in the machine’s deceptiveness. The more often a machine manipulated a human into believing that the machine itself was human (through the platform of blind, text-based communication), the more convinced humans became that the power of manipulation *itself* indicated the kind of intelligence that they saw most in themselves.

If one machine makes you laugh, what if another makes you cry? If one machine convinces you it is human because of cleverly programmed answers to limited questions, what happens when it begins to answer questions that are more intrusive and powerful?

The MIT computer scientist Joseph Weizenbaum explored this very concern. His computer program ELIZA is considered the first chatbot, which worked by retaining keyword lists that triggered responses and relayed back

7. Alan Turing, “Computing Machinery and Intelligence,” *Mind: A Quarterly Review of Psychology and Philosophy* (1950), 59:236, 433-460.

8. Simone Natale, *Deceitful Media: Artificial Intelligence and Social Life After the Turing Test* (New York, 2021; online edn, Oxford Academic, 18 Feb. 2021), Turing quoted on p. 24.

user phrases that it stored in memory. Using a version called DOCTOR that imitated a psychiatrist, demonstrations of his program revealed that people were readily willing to ascribe intelligence and empathy to the machine, which Weizenbaum found deeply disturbing. While ELIZA seemed to vindicate Turing's proposition that humans can be deceived by machines with linguistic mimicry, Weizenbaum went on to show how little was required for such acceptance of intelligence to occur. As computers became more prevalent in everyday social interactions, the distinction between whether it was a human or a machine that was doing the thinking that ran our lives was increasingly dissolved.

The moral lesson, contra Turing, was that this acceptance exposed human susceptibility to abdicate the effort to think at all, rather than prove that machines had intelligence. People projected meaning, intention, and understanding in the operation of algorithms where none existed. If behavioral imitation alone earned a machine the attribution of intelligence, then humans could be deceived into granting moral authority, trust, or responsibility to systems that have none.

When Are We Going to Stop Fooling Ourselves?

Several philosophers have subsequently positioned the anthropomorphism of computer systems as a central ethical problem before us. The MIT clinical psychologist Sherry Turkle's 1984 book *The Second Self* examined how people interacted with computers, illustrating how this reshaped the way people understand their own minds, increasingly treating thought, emotion, and personality as computational or modular.⁹ Turkle warned that treating computers as if they think or feel runs the risk of blurring boundaries between simulation and sincerity, encouraging people to accept simplified, mechanized models of selfhood and to defer judgment or ethical reflection to systems that cannot genuinely cogitate. The danger is not that intelligent machines will replace humans, but that the humanization of machinery risks ignoring the uniquely human obligation to care about the consequences of our actions. This could lead to a social environment where responsibility is replaced with regulation, and conscience replaced with input controls. As philosopher Hans Jonas articulated in 1966, treating machines as "human-like" enables the risk that we relinquish moral judgment to systems that cannot feel, suffer, or care.¹⁰

A remedy to this was proposed by Jonas in his follow-up monograph, *The Imperative of Responsibility* (1984), where he urged that ethical conduct should be guided by concerns about the worst plausible outcomes of machine

9. Sherry Turkle, *The Second Self: Computers and the Human Spirit* (Cambridge, MA: MIT Press, 1984; second ed., 2005).

10. Hans Jonas, *The Phenomenon of Life: Toward a Philosophical Biology* (New York: Harper & Row, 1966).

behavior. He proposed a “heuristic of fear,’ replacing the former projections of hope, [which] must tell us what is possibly at stake and what we must be aware of.”¹¹ Humans intrinsically react to fear more swiftly than to beneficence, he went on to suggest. “We are not unsure about evil when it comes our way, but of the good we become sure only via the experience of its opposite. It is doubtful whether anybody would ever have praised health without at least the sight of sickness ...”¹²

And So What To Do?

Early on in the history of medicine, where the body was conceived to be like a mechanical system, people resisted anthropomorphism—attributing too many human traits to machines. Yet the cultural norms preventing people from attributing excessive agency to computers evolved with each user generation.

In 1987, a group of social science researchers used the Computer Anxiety Rating Scale to discern why people might fear computers or suffer from “computerphobia.”¹³ In one of the earliest empirical frameworks for understanding psychological responses to technology, they showed that computer anxiety was not about what computers are, but about what humans perceive them to be. In other words, the attributes humans gave to computers caused the very anxiety about their function in our society. In a sense, humans were already manifesting Jonas’ heuristic of fear by worrying about imagined consequences of having computers do stuff.

Thirty years later, Johnson and Verdicchio (2017) wrote of “AI anxiety” which they associated in part with the proliferation of sci-fi visions of “superintelligence” where machines become autonomous and shed human control. The “humans versus machines” dichotomy, they argued, is a result of sociotechnical blindness, where people “fail to recognize that AI is a system and always and only operates in combination with people and social institutions.”¹⁴ The anxiety here emerges because of a miscomprehension of what *autonomy* is, not unlike the misunderstandings of how intelligence was defined in early theories about smart machines. “When nonexperts hear that machines have autonomy,” wrote Johnson and Verdicchio, “they attribute to machines something comparable to the autonomy that humans have, something close to the freedom to behave as one chooses.”¹⁵

The problem here is that this anxiety is applied to the machine itself instead of to the humans who design, deploy, regulate, and profit from it. Ethically, this is dangerous because it again defers responsibility (“the AI did it,” “automation is inevitable”) and obscures accountability for choices about surveillance, labor,

11. Hans Jonas, *The Imperative of Responsibility: In Search of an Ethics for the Technological Age* (Chicago: University of Chicago Press, 1984), p. x..

12. Ibid, p. 27.

13. R.K. Heinssen, Jr, Glass C. R., Knight L. A. “Assessing computer anxiety: Development and validation of the computer anxiety rating scale,” *Computers in Human Behavior*, 3(1) (1987), 49–59, on p. 50.

14. D. G. Johnson, Verdicchio M. “AI Anxiety,” *Journal of the Association for Information Science and Technology*, 68(9) (2017), 2267–2270, p. 2268.

15. Ibid, pp. 2268-2269.

or discrimination, themes which harken back to Wiener and Ashby.

So what do we do with our anxiety? Following the ethical perspective offered by Hans Jonas, we must exercise the moral discipline to resist assigning intelligence, intention, or responsibility to systems we design. History reminds us that evil does not arise from machines themselves, but from the human values that are encoded, enforced, or neglected within the sociotechnical systems where they operate.

Before we surrender more of ourselves—our intelligence and our judgments—to computerized systems, we should pause to remember the long human struggle to define ourselves through moral responsibility, through the uniquely human capacity to care about the consequences of our actions. That capacity does not reside in machines. It resides in us.

Author's note: I thank the readers in the Digital Humanities network and the UC Health Humanities Consortium for comments on a longer version of this article.