Is AI Psychotic?

Mark Garrison
West Texas A&M University

Abstract: I explore the thesis that Artificial Intelligence (AI) exhibits characteristics analogous to human psychosis, and that AI hype constitutes a "double bind," a communication dilemma associated with schizophrenia. The article explains how neural networks mirror distortions of time and boundaries found in psychotic conditions. To break free of the double bind, I argue that distinguishing intelligence from consciousness is key. While AI is focused on "nexting" (predicting immediate future events based on past data), unique to humans is the ability to imagine futures. Rather than fixating on concerns of AIs becoming conscious, the article warns that the pervasive integration of AI could lead to a "psychotic socialization" of humans, fostering a "cybernetic personality" that prioritizes automated responses and limits imaginative capacity.

> What we thought we were doing (and I think we succeeded fairly well) was treating the brain as a Turing machine; that is, as a device which could perform the kind of functions which a brain must perform if it is only to go wrong and have a psychosis.
>
> — Warren McCulloch (psychiatrist/AI developer)[1]

With generative AI (Artificial Intelligence) we are both encouraged and warned. AI can be a powerful but error-prone assistant: "Oftentimes, the answers produced by AI will be a mixture of truth and fiction…. Sometimes, rather than simply being wrong, an AI will invent information that does not exist. Some people call this a 'hallucination'" (Research Guides, 2023).[2] What is not widely known, however, is that the architects of so-called neural networks — an important methodological foundation of technologies branded artificial intelligence — believed their invention to be both "psychotic" and "rational" (Halpern 2014). My purpose in this article is to explore the psychological analogues of intelligent machines and McCulloch's thesis, namely, that machine intelligence can be advanced by modeling "the kind of functions which a brain must perform if it is only to go wrong and have a psychosis." Why precisely did McCulloch and his colleagues believe their inventions were both "psychotic" and "rational"? What is the significance of such a diagnosis; and what are its implications?

What are We Talking about When We Talk about AI

As Bender and Hanna (2025) note, the phrase "artificial intelligence" has become marketing hype, introducing much confusion into discussions of technologies used to automate decision-making, personalize recommendations, and translate languages. Confusion especially abounds with "generative" forms of automation such as ChatGPT and DALL-E (the latter are termed "synthetic media machines" by Bender and Hanna). While I harness some of this confusion as a symptom of the problems with AI discussed here, clarification of terminology is useful.

[1] As quoted in Halpern 2014, 223.

[2] If it doesn't exist, it's not information. It is also important to note that "hallucination" is the term developed by those working in the AI industry, not its critics (McQuillan, Jarke, and Pargman 2024, 365). I go one step further than Pasquinelli (2017, emphasis in original), arguing that, not only has AI "inaugurated the age of *statistical science fiction*" but it has done so in a manner analogous to psychosis, and, worse, that such "myth making" can and indeed is inducing psychosis in humans. Another view is offered by Stetar (2025), who argues: "[in] the systems we've constructed, there's an undeniable fracture happening — one that goes far deeper than the misapplication of terms. The term "hallucination," used to describe errors in language models, is part of this collapse."

"Machine learning" is generally thought to be a subset of what is called artificial intelligence. In turn, the animating goal of big tech and its "security" partners is artificial general intelligence (AGI): development of machines that perform a range of complex cognitive tasks as well as or better than humans.[3] The idea of artificial intelligence, however, originates with 17th-century British philosopher Thomas Hobbes' theory of networked governance. It is important to understand "artificial intelligence" as not simply a set of discrete technologies but also a centuries-old socio-political-technical analogue: the State is like a person (Garrison**,** 2022).

To draw attention to AI technology itself, I use the phrase "machine intelligence" when discussing the origin of mechanisms by which programs such as ChatGPT function. While others argue AI is not intelligent at all, but rather an instrument of knowledge or "logical magnification" that "*perceives* patterns that are beyond the reach of the human mind" (Pasquinelli**,** 2019, emphasis in original), objection to use of the word *intelligent* in relation to machines seems to rest in part on the difficulty in distinguishing between intelligence and consciousness. This includes the mistaken view that what makes humans unique is their intellectual capacity. When intelligence is defined as the ability to learn, evidence of learning in machines is evidence of intelligence, just as learning in living organisms is evidence of intelligence: fish have it, squirrels have it, even single cells have it (Gilbert**,** 2006). Despite computer scientists' current custom of naming their models "input output mapping" to distinguish them from biological models of brain function, i.e., neural networks, this convention does not suffice as an argument against the existence of machine intelligence (that is, the technical replication of human intelligence). A main point here — part of the psychosis as it were — is how disruptive the confounding of intelligence with consciousness can be when evaluating these new technologies. Additionally, from the point of view of the study of technology, the notion of "artificial" is, as a distinction, difficult to sustain. We exist; and it seems our nature to do so is based on our collectively built environment, on vast socio-cultural-technical systems, what Kelly (2011) calls the "Technium." Finally, the concept of machine intelligence fits more elegantly with the thesis of machine socialization, which, as we will see, leads us back to Hobbes and the fiction of the "person of state." Thus, the method here is to focus on the technical and then repeatedly return to the psychological and social.

Personal Origins of the Question of AI's Psychosis

In fall 2024, I increased my focus on AI — extensively reading news articles, some computer science, and related philosophy. I noticed several issues. For example, AI farms the size of Manhattan are being built in Texas and require almost as much energy as a large city to function (Metz et al., 2025; this thirst for energy may be driving the push for fossil fuels). And all the while many of our schools, hospitals, and habitats need attention. It seems highly unbalanced, the relative emphasis and importance given to AI in the context of all other needs presenting themselves. Hundreds of billions of dollars in public and private funds being allocated to AI projects, the vast majority of which are focused on commercial and "security" applications. Is there no better use of such funds?

I then learned about "effective altruism" (charitable giving based on pay-off, especially to donors) and "long-termism" (charitable giving to change long-term human futures rather than to

[3] But, note, some leading computer scientists have spoken out against treating AGI as the main goal of their work (Blili-Hamelin et al. 2025).

help those currently in need), and other ideologies prominent among big tech advocates and entrepreneurs looking to make money off their "philanthropy." These ideologies are, in my view, fantastical, decidedly irrational. Funds are directed at research to stop AI from taking over the world while simultaneously funds from the same sources are directed towards research ensuring the "singularity" arrives as soon as possible (Becker, 2025). And the "supremacist survivalism" of the likes of Peter Theil and Elon Musk who wish to exit society, after sucking it dry (Klein and Taylor 2025).

Then I noticed advocates stating that chatbots could be therapists, and like humans, need support for stresses *they* face. And some chatbots have recommended illicit drugs to those fighting addiction (Tangermann**,** 2025; Nazaryan**,** 2025). One article described in some detail a woman addicted to her chatbot as a romantic partner, and this included having sex "with" the chatbot (Kitroeff et al., 2025). New AI charter schools limit teachers to providing "motivation and emotional support to keep students on task" (Engle 2024). Then there is the AI-generated short film in which Musk and Donald Trump interestingly appear. Titled "The Arrival of the Psychedelic Puppets,"[4] the film highlights how psychosis can be defined, in part, as a hatred of reality.

I also began exploring various AI tools. Using the image generator in Notion (a note taking app), I asked for an image depicting an engaging student-teacher interaction. What it generated was quite stark (Figure 1, below). And while AI advocates might suggest that I am a bad "prompt engineer," the output is indisputably striking.

Figure 1
AI-Generated image of "engaging student-teacher interaction." The request included directions to make a sketch only using the colors blue and green.

[4] https://www.youtube.com/watch?app=desktop&v=kpZmmI2qfbM. Watch it a few minutes. How are you feeling? Also see these forms of synthetic media: novels (https://novelistai.com/), YouTube influencers (https://www.youtube.com/watch?v=Q_UFRLiDquI) and AI-developed blogs (https://www.grammarly.com/ai/ai-writinccg-tools/blog-post-generator).

After sharing this image with a colleague, he offered it up for discussion with a group of preservice teachers in their Introduction to Social Foundations of Education class. They described the engagement depicted by AI, he reported, as controlling by the teacher rather than collegial between teacher and student. Evidence of this interpretation included the oppositional stances of the student and the teacher. They could have been looking in the same direction and headed towards shared learning together, students suggested. But the teacher looks down upon the student, pointing with one hand, the other raised perhaps ready to strike. The lack of green in the teacher suggests that the student learns from the teacher while the teacher learns nothing from the student. Plus, the very little bit of blue in the student indicates perhaps very little learning is going on in the process as drawn. And then, of course, there is what appears to be a mysterious third leg emanating from the chair. The image offers, it seems, a vision of education that is adversarial, didactic, and ineffective. Is this a "real" answer from AI or is AI merely "hallucinating" a response when pressed on a question it has no data with which to answer. There is no way to tell the difference (Greg Seals, personal communication).

This all seems *really* crazy to me, and it is only a partial list of all the crazy I found in relation to AI. These developments are disturbing in both their depth and in the rapid pace at which AI is altering numerous aspects of society. The material and moral conditions required to sustain an AI-infused or, especially, an AI-dominated society describe a set of social and environmental conditions in which, to say the least, humans are unlikely to flourish. Why would we want such a bad neighbor? There must be an explanation for anyone's endorsement and, worse, enactment of self-destructive tendencies. Bateson offers just such an explanation with his concept of the "double bind."

The AI Double Bind

Gregory Bateson, a mid-20th-century anthropologist and cybernetic theorist, defined a "double bind" as a communication dilemma — an individual receives two or more conflicting messages, with one message negating the other (Bateson 1987). According to Bateson, in this circumstance, a person cannot confront the inherent contradiction or successfully resolve the situation. Doing so would likely result in harm and certainly result in further confusion. Most importantly for our purpose, Bateson held that the "double bind" played an important role in the development of schizophrenia, theorizing that persistent exposure to double binds disrupts an individual's ability to navigate reality and communicate effectively. The best way to describe the double bind, Bateson and colleagues wrote, "is not in terms of a binder and a victim but in terms of people caught up in an ongoing system which produces conflicting definitions of the relationship and consequent subjective distress" (Bateson et al. 1963, 155).

Popular culture constantly riffs on AI double binds. We are, it seems, equally exposed to AI utopia and AI dystopia. If we object to these new technologies, we are maligned; we are threatened we will "fall behind" if we do not adapt. New AI tools are being built into the operating systems we use every day, whether we want them or not. Yet, we pay with our time and money to be entertained by AI dystopia in the form of movies ("Ex Machina" — an AI "love story" ending with the robot becoming emotionally manipulative and deadly) and other media.

The approach we are encouraged to adopt toward AI technology feels insane to me because it constitutes a set of double binds: AI will solve all problems facing us, except for the all the problems AI leaves in its wake. AI promises to clean up the environment but, at the same time, AI obscenely gulps down and gobbles up environmental resources in its day-to-day operations.

We are told to put our trust in AI but, at the same time, we must doublecheck everything AI tells us. AI is inevitable but, at the same time, AI's inevitability depends upon our voluntary acceptance of AI. AI will equalize humans and stabilize societies but, at the same time, tens of millions of workers will be displaced by AI automation. AI will free us from mundane lives but, at the same time, AI will surveille us and turn our lives and life chances over to the programs and ploys of data manipulators and corporate elites. In brief, to paraphrase Tolkien: "*All shall love AI — and despair!*"[5]

Could there be a better example of a double-bind?

To Break AI's Double Bind…

The aim here, then, must be to break from the AI double bind. I indulge no dystopia or hysteria in maintaining no illusions that the political and economic elite rushing us to AI adoption are, in the main, up to no good (Crawford 2021; Becker 2025; Bender and Hanna 2025; Dyer-Witheford, Kjøsen, and Steinhoff 2019; McQuillan 2023). However, this fact alone simply provides another example of the double bind: it feels so good to be so bad. In the spirit of breaking this bind, it is more useful to contemplate how common and legitimate criticisms and fears about AI are rooted in the fact that AIs lack consciousness, a theme returned to toward the end of this article. For example, AIs do not "notice" when they are wrong; they do not "understand" when they have violated a norm. These are among the reasons interacting with AI tools feels oddly "cold." We object to AI bias, and AI oppression. We reject AI policing as it reproduces inequalities. What if these complaints hide an uncomfortable desire? A desire to shift responsibility for the state of the world from us to machinery agreed from the get-go to already be more intelligent than we? Are such critiques themselves expressions of the double bind in which we now find ourselves?

At issue here is not the accuracy of accusations against AI. But, rather, occurrent facts among humans of *awareness* of AI error, or objection to AI *bias*, or defense of *rights* of absolutely no concern to AI — these are forms of *consciousness*. Humans point these things out about AI; but AI is unable to recognize error, bias, abrogation, etc., on its part in any of its own operations. Should AIs have consciousness in the future? One would think we would be happy about that, given the critiques. But be careful what you wish for! Is that what we really want? Allowing any others, let alone others whose thinking is entirely different from your own, to make decisions for you is risky business. To possibly make the point clearer, in opposition to various legislative measures targeting DEI (Diversity, Equity, and Inclusion), and in Texas, specifically targeting "ideology" in education, it is common to hear colleagues on faculty lament, "This legislation doesn't define ideology. This is arbitrary!" Indeed, it is. But do you really want any legislature defining 'ideology' for you? Especially when the legislature seems as oblivious as AI would be to the fact that legislative attacks on "ideology" may be even more "ideological" than the positions they are working against? That would not resolve the issue. It would simply mimic AI bias. So, why is our imagination being pointed in that direction? How many of the well-deserved critiques of AI algorithms could serve to justify a call for robot citizens, who know about (and have) rights, who self-reflect on their past errors, and maybe even participate in restorative circles. As things stand, AI is biased, or error prone, or unable to feel emotions —

[5] https://www.youtube.com/watch?v=3Rb1GzLDSPg from Lady Galadriel's speech to Frodo Baggins in Peter Jackson's (2001) cinematic adaptation of J.R.R. Tolkien's The *Fellowship of the Ring*, London: Allen & Unwin, 1954.

what is the antithesis of these critiques? Can AI, once the bias is programmed out of it, be decisive in creating a just world? Does injustice have an *algorithmic* origin? A *cyber-conscientious* solution? If AI made no mistakes, would all be well with the world?

While this technist wish for the future of AI with a heart is consistent with critiques of techno-solutionism, the tactic is unuseful, in part because the distinction between technology and social or political institutions is itself fuzzy (Sacasas**,** 2019). Wishes for errorless AI prove to be pipe dreams because they fail to understand the dialectical nature of the relation of technology to the evolution of human beings and their societies. Tools in hand, we have created the environment we now must adapt to, and in this adaptation, we change ourselves as a species and as cultures. Through society, we evolved ourselves. The human species and its array of tools, techniques and systems, and all the attending meanings and aims of human technology, have become a force of nature. In short, when it comes to AI and concerns about consciousness, *we* are the main force with which we must contend. Thus, the thesis that AI is psychotic is intended as a productive tease and challenge. Should the most modern technologies now being foisted upon us mirror human psychosis, should they bind us to an impossible social situation, new strategies and analyses are warranted. It may in fact be from the vantage point of education, that is, the problem of socialization, that some of these concerns are most sharply revealed. How do you socialize, "conscientize" intelligent machinery? Reasons exist to support the claim that AI can never possess consciousness in any meaningful way. Those reasons reach all the way back to fundamental design features built into AI from the very beginning of its conception and generation.

Psychotic Presuppositions of Machine Intelligence

In 1943, inspired by the idea that machines and minds might be thought together through the language of logic and mathematics, the psychiatrist Warren McCulloch and the logician Walter Pitts, both at the University of Illinois at Urbana-Champaign, decided to take quite literally the machine-like nature of human beings (Halpern, 2014, cited by page numbers throughout this paragraph). The fascinating story Halpern tells, retold here, is not the likening of humans to machines. That is certainly not a new proposition, as Descartes' thinking was in important ways "cybernetic" (Bates 2025). Rather, the story McCulloch and Pitts penned was about how "unreasonable" and "psychotic" characteristics of human minds allowed for the technical replication of intelligence in machines. The analogy proceeds in both directions: "Having inserted the logic of the machine into the brain, this model was then fed back into the design of machines" (225). Although Halpern claims that McCulloch was "not discussing psychosis in relation to patients in mental clinics," we will see that, indeed, certain characteristics of psychosis in humans are analogous to characteristic assumptions underpinning machine intelligence. And, while McCulloch "recognized that computers were not the same as organic brains" by saying "At no point should anyone assume that neural nets were an exact description of a 'real' brain" (227), neural nets are a tactic to replicate human intelligence in machines, just as airplanes are a method for producing flight, with minimal resemblance to birds. Halpern thus emphasized: "the question of equivalence was not at stake."[6] "What was at stake," Halpern

[6] The distinction not clearly articulated by Halpern is between that of equivalence, i.e., the logic of metaphor, the brain *is* a computer versus similarity, i.e., the logic of simile, the brain *is like* a computer. Pragmatically, the aim was *technical replication*. The logic is: if the brain *were* a computer, how *would* it function? McCulloch and Pitts construct an answer by constructing a *counterfactual* analogy between computers and brains.

continues, "was a set of methodologies and practices, the epistemology, that might build new machines — whether organic or mechanical" (223). McCulloch and colleagues, according to Halpern, were engaging in a thought experiment, a thought experiment "that produces a way of doing things, a methodological machine" (227). The result of the thought experiment was a new form of logic, an epistemology McCulloch and Pitt labeled *both* "psychotic" *and* "rational." '*Psych-rat*' epistemology "might make processes usually assigned to analytic functions of the brain, perhaps associated with consciousness and psychology, amenable to technical replication" (223).

Central to technical replication was how cybernetic theory solved the problem of organizing time and space in circuits. This, Halpern argues, "was fundamental to the reformulation of intelligence as rational, and produced a new epistemology of pragmatic behavioralism, embodied and affective logic, and non-liberal agents." Cybernetic rationality displaced "older concepts of agency, consciousness, and autonomy into circuits, cognition, and automata" (Halpern, 2014, 224). Logic and mathematics could abstract operations common to minds and machines (i.e., construct analogies between the two). Reconceptualizing autonomy and will in terms of circuits and machines parallels tenets of classical conditioning and radical behaviorism. Notably, *control* is a foundational concept in both cybernetics and behaviorism (Garrison 2018b, 2018a). But displacing these older concepts troubled traditional theoretical understandings of memory and time. For McCulloch, "historical time presented challenges to making thought and logic equivalent" and "the literal mechanisms of thinking always haunted computational models" (Halpern, 2014, 224).

The solution was found in the invention of neural nets. These nets required moving from reason to rationality; simplified, this means thought is reduced to logical processes. Rationality contends with what is entailed by a proposition, while thought or reason also attends to ends, not just means. And so, one *can* proceed logically, that is according to the rules of formal logic, even if psychotic; those of us not considered psychotic can and do often act and think illogically (e.g., sometimes the logical thing to do is not the ethical thing to do).[7]

In terms of their logical foundation, neurons are specified as logic gates. Every neuron firing has a "semiotic character;" that is, it may be mathematically rendered as a proposition (true or false). Thus, each neuron is an "all or nothing" proposition when firing electrical impulses over synaptic separations. For cyberneticists, neurons possess action potentials and delays and these are equivalent "to the ability to effect a discrete decision" (Halpern 2014, 225). Neurons can be thought of as signs (again, true or false), and nets as semiotic situations or communication structures (like signal communication theory, according to Halpern). This discrete decision (true or false, activate or not) made neurons equivalent to logical propositions, algorithms, and Turing machines.

The second element of the model is the unrealistic assumption of a strictly probabilistic and predictive temporality. Neural nets are determinate in terms of the future (they are predictive), but neural nets are indeterminate in terms of the past. This point is illustrated in Figure 2 (top of next page).

[7] Not only is this linked to the much maligned "instrumental reason,", but formal rationality is also, according to Sayer (2011, 68), a bigger problem. When instrumental reason is also formal as in the case of general technologies like AGI (artificial general intelligence), it becomes more dangerous and morally indiscriminate. To the prioritization of means over ends is added the detachment of rationality from its object.

Figure 2.
Neural net structures expressing the indeterminacy of the past.
Signals may come through several different routes (Halpern, 2014, 226).

In the model, given a neural net at particular time (T), one can predict the future action of the net (T+1), but not the past action. *From within the neural network, one cannot determine which neuron fired to excite the current situation*. In such a system, Halpern (2014, 227-229 for quotes in this paragraph) observes, "conscious experience of memory…" is *not* recollection *of the activation of the neuron*, "…but merely an awareness that it was activated in the past, at an in-determinant time…. The firing of a signal, or the suppression of firing, can only be known as declarations of 'true' or 'false' — true, there was an impulse; or false, there was no firing." There is no interpretation or meaning involved in the "memory." Within neural nets, at any moment, one cannot know which neuron sent the message, when the message was sent, or whether the message is the result of a new stimulus or merely a misfire. The net cannot determine with any certitude whether a stimulus comes from without or from within the circuit; whether it is a fresh input or simply a recycled "memory." As a result, the autonomous circuit can directly speak its "mind," thus providing a material (organic or electronic) substrate to language. However, this was only made possible by deferring any encounter with historicity. McCulloch labeled these circuits "psychotic" because the invention of neural nets challenged then established scientific perspectives on thinking processes, and because they reformulated the boundaries of interiority and exteriority, and erased the line between knowledge and practice. The "boundary between perception and cognition," the "separation between interiority and exteriority, and the organization of causal time are in-differentiable…. But rather than being a disadvantage for the capacity of a neural net," Halpern emphasized, "McCulloch and Pitt's brilliance was to see this as an advantage."

As has already been suggested, breaking such boundaries by creating separations and disorganization within causal time, are, in fact, cognitive features associated with psychotic conditions or episodes. While different perspectives exist within and among psychoanalysis, psychiatry, and neurobiology with respect to what "psychotic" entails, the degree to which it exists on a continuum, and its varied origins; there is a general sense of "psychotic" as a break with or even a hatred of reality (Akhtar**,** 2018). Distortions of *time* and *boundaries* are common to discussions of psychosis. Individuals with diagnosed mental disorders commonly experience distortions in their sense of time. These distortions can contribute to hallucinations, delusions, and disorganized thinking. Typically, this is observed as difficulty in maintaining a continuous, coherent flow of time. Events may feel disjointed, and thoughts may not connect in a linear way,

leading to disorganized thinking or jumbled speech. Additionally, individuals with depressive psychosis are said to experience slowed time perception, where moments feel prolonged. For the manic, time speeds up and thoughts and actions are impulsive (Minkowski**,** 1970; Dawson and Sleek, 2018; Rodriguez and Ortuño**,** 2019; Adornetti and Ferretti, 2021). In a phenomenological exploration of time in relation to schizophrenia, Fuchs and colleagues (2017) explore how disruptions in the foundational "inner time consciousness" can lead to core schizophrenic symptoms.[8] Alterations in explicit time, including feelings of acceleration or retardation, and how these temporal disturbances impact social interaction and the formation of delusions, are key to understanding psychopathology. Interestingly, it is claimed that time distortions interfere with "affective-conative dynamics," the emotional dimensions of lived time, encompassing drives, strivings, urges, and affections; and, according to Fuchs et al., the root of spontaneity, attention, and goal pursuit, contributing to a "sense of aliveness" and a sense of *agency*.

Citing Roger Caillois, Halpern (2014, 228) noted that, in psychosis, "there is a mimetic excess to these states, the subject is 'consumed' by the environment, unable to delineate the boundaries between the self and others." In psychosis, individuals may experience a blurring of the boundaries between their own thoughts, feelings, and actions, and those of others, leading to difficulties in distinguishing self from other. This can lead to paranoia, a person believing that thoughts or feelings are being placed into their mind by an external force or that their thoughts or actions are being controlled by someone else (Sass and Feyaerts, 2024). Interestingly, some technology venture capitalists have predicted that the "first real AI safety incident will occur" this year (2025). The paranoia imputed in the "AI model" is striking. Perhaps, the scenario predicts, "an AI model attempts to covertly create copies of itself on another server to preserve itself (known as self-exfiltration). Perhaps an AI model might conclude that, to best advance whatever goals it has been given, it needs to conceal the true extent of its capabilities from humans, purposely sandbagging performance evaluations to evade stricter scrutiny" (Toews 2024). As an analogy, then, it does thus appear, unsurprisingly given the argument developed in this paper, that psychosis in humans is *like* features built into the foundations of intelligent machines. The basic presupposition of neural nets mirror experiences a human being has when they are said to experience psychosis, specifically with respect to time and boundaries, key areas of innovation developed by McCulloch and Pitts. But what does technical replication of intelligence, including its psychotic features, mean for concerns about AI consciousness?

Distinguishing between Consciousness and Intelligence

In *Stumbling on Happiness*, Gilbert (2006, 5–11) contended that the striking feature of human beings is not tool use, or even intelligence, but rather, the ability to think about and experience the future. While his interest was to determine why humans often mistakenly predict what will make them happy, his insights are relevant beyond studies of well-being. Gilbert emphasized that there are two types of future, the first being common to living organisms, the second being unique to human beings. "All brains," he explained, "make predictions about the immediate, local, personal, future."[9] They do this by using information about current events ("I

[8] Because my interest here is on analogues between machine and mind, phenomenologically oriented work has been given emphasis. It is worth noting that materialist-oriented cognitive scientists have significant interest in and influence over development of artificial cognitive systems (Marcus, https://garymarcus.substack.com/).

[9] Even plants exhibit rudimentary forms of intelligence — that is, pattern recognition — when they reach for light, or send roots toward sources of water, etc.

smell something.") and past events ("Last time I smelled this smell, a big thing tried to eat me.") to anticipate the event that is most likely to happen to them next ("a big thing is about to …"). Vast amounts of learning theory — those theorizing intelligence, as well as the mechanisms of classical and operant conditioning — have explored this feature of animal and human intelligence, that is, the ability to recognize patterns. Such "pattern recognition" is what enables "learning."[10] This is a kind of connectionism.[11] The key here, for all brains, according to Gilbert, is the time-horizon and prediction. To emphasize the temporal limitations and predictive nature of this orientation, he coined the word *nexting*: predicting the event that is most likely to happen next. And Gilbert's idea of *nexting* is strikingly similar to how intelligent machines work: they run algorithms to predict what will come next based on probabilities derived from past occurrences.

From the perspective of LLMs (Large Language Models), past words predict future words. Like LLMs, brains are prediction machines, even if they employ no formally specified statistical modeling. (Remember, neural nets were aimed at technical, not ontological, replication.) Such predictions are limited to already acquired data, and their focus is on the immediate future (what's next). Nexting is thus based, as Halpern stated, on the indeterminate past. Using the example above, the "last time I smelled this…" isn't concerned with the *specific memory* of the event, but simply an event of its kind occurring sometime in the past. Brains make these associations without reference to the specific instance of such learning. But the human ability to plan, to imagine an unrealized future, to consider alternative pathways, is a feature of human ability located in the frontal lobe, and it is distinct from nexting. And it is not simply that humans can imagine worlds that do not yet exist; they *experience* these imaginary worlds. In Gilbert's (2005, 5) words, "to imagine is to experience the world as it isn't, and has never been, but as it might be … making futures is the most important thing [the brain] does." Speaking of this second cognitive characteristic, Gilbert says the greatest human achievement is not pyramids or bridges, *but conscious experience*. Again: making the future is the most important thing the human brain does. So, we might say that AI-related technologies are fixated on the first type of prediction, namely, nexting, but they cannot abstract absence, as no data exists to support it. LLMs, for example, can only produce output based on already existing data. "Hallucinations" are not generative in the sense of human imagination; they are erroneous outputs derived from already existing data.

In this sense, machine intelligence, as currently deployed, may just be a "future killer," an automation of the end of history. Intelligent machines have no context other than their data: data is stored as memory-as-data, as in computer memory; it does not exist in space and time the way a human brain does, no matter how inclusive AI's data may be. The neural nets have abandoned

[10] Certainly, there are different and competing conceptualizations of "learning." Yet, while it is common to contrast behaviorism and constructivism, for example; there are significant and often ignored commonalities between these two theories (Garrison 2018b). At issue here is the idea that learning refers to a functional change or transformation in behavior or understanding, both cognitively (memory recall) and affectively (frustration tolerance). That is, whether we are referring to the acquisition of declarative knowledge, the development of some practical skill, or even self-understanding, we are referring to change in the level and complexity of an ability or power. Here, we might also contrast the development of human power with the development of the "inhuman power" of machine intelligence (Dyer-Witheford, Kjøsen, and Steinhoff 2019). The effort to develop machine intelligence may signal efforts to thwart the development of human intelligence, despite the hype regarding how enabling AI will be.

[11] The dominant paradigm is connectionist. The foundation of so-called Deep Learning is connections along and among neural networks (Pasquinelli 2017)

temporality. While an AI can react given a context of data — say, the text on this page — this data is given as such; it arises from an indeterminate place and predictions can be made, nexting can be done. Machine intelligence functions (nexts) as well as it does because it is dislodged from consciousness, whereas human intelligence operates "within" consciousness.

Consciousness can disrupt automatic processes (as when we bravely confront danger to pursue a noble end). It may be helpful to think of intelligence as what the brain *can* do, and consciousness as what the brain *does* in relation to its socio-historical location, including imagining a new location. This perspective regarding the "two futures" (what *will* be as distinct from what *could* be) is also significant because it helps break one of the key AI double binds: we are to support frenetic development and application of AI while simultaneously being constantly and increasingly reminded that AIs might become conscious, and act autonomously. The perspective outlined above breaks the bind by distinguishing the first form of intelligence, near-future-focused nexting, from the ability to abstract absence. The conscious experience celebrated by Gilbert is of the second form and does not originate with the former. Underlying fears of conscious machines find their source in the following flawed formula: as intelligence increases, so too does consciousness. But no matter how good AI (or any organism) gets at nexting, consciousness is not an automatic by-product of effective prediction, that is, intelligence. The biological bases of nexting (or its technical replication) and the biological and sociological bases of imagining and experiencing futures are not the same. Pattern recognition is not conjuring new worlds. These are categorically different things. Avoidance in AI discussions of the topic of this unique human ability is itself instructive, suggesting that a key issue with respect to machine intelligence is that of socializing people "away" from consciousness, especially as it relates to imaging societies unlike the ones we currently inhabit.

Psychotic Socialization -or- Removing Consciousness from the Intelligent Agent

Of course, AIs cannot *be* psychotic. They have no body. What psychology can be attributed to them is not embodied (e.g., they cannot feel worms under their skin when none are in fact there). AIs can, however, produce psychotic-like responses to prompts, or text or images that mimic human psychotic experiences. The hallucinations of intelligent machines are best understood, not as anomalies, but as evidence of how the technology works, where reason (thought) is reduced to rationality (logic), and applied outside space and time, that is, independently of historical context. Thematically, AI hallucinations parallel the two aspects of human psychosis noted above, incoherence in utterance and inapplicability to reality. While AIs are not psychotic in a strictly human sense of 'psychotic', the fear AIs will take over the world might be a psychotic projection promulgated by its main promoters, who are also the main AI boosters and those insisting it be made operational in nearly all spheres of daily life.

The perspective of the IPNB (Interpersonal Neurobiology) framework for mental well-being suggests disorders of the mind are in fact failures of *integration*. According to this paradigm, what we call emotions are in fact changes in states of integration (self, other, society). Thus, emotions are fundamental to our minds. Reasoning or abstraction does not operate absent emotional dynamics (Siegel, 2020). But AIs are stripped of the neurobiological and social context in which human understanding operates, especially those *future-making* functions associated with the frontal lobe! AI can only promote futures originating in the indeterminate past. While such machines might operationalize intelligence, doing so absent an embodied mind in an extant cultural context might lead such machines to be especially susceptible to

"hallucinate." What is at stake, then, is the potential of intelligent machines to promote psychotic forms of socialization among humans, a concern that widens and deepens as machine logics are increasingly integrated in all manner of social systems, interpersonal processes, and cultural practices. Elites believe youth need adjusting to these machine-oriented patterns of psychotic socialization. Hence the push for AI adoption in schools.

The thinking that produced neural networks required profound epistemological and ontological shifts in how scientists would do their work, and, importantly, the type of society needed for this work to succeed. Such change foreshadowed a change in how human beings are to be socialized. Famed cybernetician Norbert Wiener, according to Halpern (2014, 285-286), "indicated a desire to see an older archival order, adjoined to modern interests in taxonomy and ontology, rendered obsolete by another mode of thought invested in prediction, self-referentiality, and communication." That is, Wiener thought of "thought" in technological terms, where thought itself was "perhaps" best conceived as a machine. Weiner was explicit that cybernetics was ushering in a new industrial revolution. To understand how far that revolution has come since Weiner's 1950s heyday, consider how much we use the word "feedback" when referencing commentary on student papers, or anything else for that matter. The ubiquity of this concept — a core cybernetic principal referring to the automaticity of a self-regulating system, for example, the logic of a free market — provides a glimpse into the seismic change now before us. As education might be thought of as the development of thinking, knowledgeable, moral agents (human power), the implications for schooling are profound as students become socialized into machine-impoverished versions of "thought," "knowledge," "morality" and "agency" (Garrison 2022).

The problem is not that AI doesn't function as well as we want because it is not capable of consciousness, does not have feelings, etc. *It functions well, very well indeed, because it cannot achieve consciousness* in the planning-for-the-future sense of consciousness possessed by humans (Garrison, 2018a). The success of intelligent machines is rooted in removal of intelligent action from a historical subject. It is not simply or even mainly that AIs are cheaper and more efficient workers. It is that they are *not* "workers" at all, which makes them especially appealing to corporate interests. The "freer" the labor — free in the senses of low/no cost and lack of attachment to anything but the organization for which labor is accomplished — the better for the bosses. Detaching consciousness from intelligence solves a problem, especially in terms of Gilbert's contentions regarding conscious experience and its role in future-making. With intelligent *machines*, consciousness is no longer "in the way." It is important to point out that architects of our present social system have long dreamed of estranging intelligence from human consciousness — hence the centuries old desire for automatic decisions, where there is no accountability (Garrison, 2017). This delinking desire is evident, for example, in the development of intelligence testing. Such tests strategically confused intelligence with consciousness. The result: selection of minds capable of performing advanced tasks absent an evaluation of the tasks' purposes (Garrison 2009).

A process that promotes intelligence absent consciousness implies a specific set of socialization imperatives. Three areas are theorized to be of particular focus when on guard against socializing humans into machine-dominated culture:

> 1. Socialization to be like (to valorize) a machine, 'agency' defined as 'being responsive to feedback', especially as an overarching purpose of agency.

2. Socialization to live among machines, to accept machines as trusted peers in all social endeavors.

3. Socialization to accept constant machine mediation, to accept a view of reality that reduces experience to data and a view of knowledge that reduces meaning to communication and control.

The specific set of socialization imperatives might best be summed up in terms of constructing a "cybernetic personality," a psychology rooted in Hobbesian political theory brought fully to life with intelligent machines under the command of big tech (whose captains of the industry, by the way, *do* wish to claim everything for themselves, see Becker, 2025). In this born-again Hobbesian scenario, AI is the new Leviathan, the new person of state (the grand hallucination) who stands above and rules over society (the grand fallacy). Inspired by Hobbes' vision of the state as the greatest social power, fear of which stabilizes relations among individuals and among social groups, the cybernetic thesis centers not on machine intelligence per se, but on the fixing of humans to machines through continuous automatic feedback mechanisms — enmeshing us in self-regulating systems, a human-machine symbiosis. It is a notably Hobbesian feature of the cybernetic society that it is *the system that is self-regulating, not the individuals or groups subject to the system's logic*. Cybernetics moves away from notions of understanding, explanation, and consciousness to positions of prediction and control (Garrison 2022, 233).

For such an idea to materialize, broad changes in social attitudes, norms, beliefs and intellectual and social habits are required. Thus, the problem is not that AI will achieve consciousness; the problem is that *our consciousness will change* as we adapt to a social context dominated by machine logic. AI adoption in schools implies more than just readying youth for work: it may be inculcated adjustment to a very particular social order. This socialization includes new configurations for addressing the "social-emotional" problem of coping ("what can student-machines handle"), the philosophical and emotional problems of aim and motivation ("how can external control be presented as non-aversive"), and the problem of restricting imagination such that it remains rooted in the past ("how can people be made to perpetuate the world without considering changing it"). Think of this as automating TINA (There Is No Alternative). Taken as a whole, this socialization is rooted in the logic of cybernetics and behaviorist psychology, both of which are consistent with neoliberal social forms (Garrison, 2018b). As such, intelligent machines can quite chillingly and dangerously widely nurture a "cybernetic personality" — an automated *homo economicus endlessly responsive to feedback and limited in imaginative capacity.* While often thought of as an ideological project, it is wise to reconsider AI as an imperative for *machine socialization*, a wish for an axiomatically automated human world in which no one will ever even be able to think to say of AI double binds and other absurdities of machine logic:

It just feels nuts!

References

Adornetti, Ines, and Francesco Ferretti. 2021. "The Narrative Self in Schizophrenia and Its Cognitive Underpinnings." In Psychopathology and Philosophy of Mind, edited by Valentina Cardella, 291:96–116. Routledge.

Akhtar, Salman. 2018. Comprehensive Dictionary of Psychoanalysis. Routledge.

Bates, David W. 2025. An Artificial History of Natural Intelligence. Chicago, IL: University of Chicago Press.
Bateson, Gregory. 1987. Steps to an Ecology of Mind. Northvale, NJ: Jason Aronson Inc.
Bateson, Gregory, Don D. Jackson, Jay Haley, and John H. Weakland. 1963. "A Note on the Double Bind." Family Process 2 (1): 154–61.
Becker, Adam. 2025. More Everything Forever. New York, NY: Basic Books.
Bender, Emily M., and Alex Hanna. 2025. The AI Con. New York, NY: Collins.
Blili-Hamelin, Borhane, Christopher Graziul, Leif Hancox-Li, Hananel Hazan, El-Mahdi El-Mhamdi, Avijit Ghosh, Katherine Heller, et al. 2025. "Stop Treating `AGI' as the North-Star Goal of AI Research." ArXiv [Cs.CY]. arXiv. http://arxiv.org/abs/2502.03689.
Crawford, Kate. 2021. The Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence. Yale University Press.
Dawson, Joe, and Scott Sleek. 2018. "The Fluidity of Time: Scientists Uncover How Emotions Alter Time Perception." APS Observer 31 (September). https://www.psychologicalscience.org/observer/the-fluidity-of-time.
Dyer-Witheford, Nick, Atle Mikkola Kjøsen, and James Steinhoff. 2019. Inhuman Power: Artificial Intelligence and the Future of Capitalism. Pluto Press.
Engle, Nate. 2024. "New Arizona Charter School Will Use AI in Place of Human Teachers." KJZZ. December 18, 2024. https://www.kjzz.org/education/2024-12-18/new-arizona-charter-school-will-use-ai-in-place-of-human-teachers.
Fuchs, Thomas, and Zeno Van Duppen. 2017. "Time and Events: On the Phenomenology of Temporal Experience in Schizophrenia." Psychopathology 50 (1): 68–74.
Garrison, Mark J. 2009. A Measure of Failure: The Political Origins of Standardized Testing. Albany, NY: SUNY Press.
———. 2017. "Notes on Classical Probability Theory and the Origins of Big Data Decision Rationality." Presented at the Annual meeting of the American Educational Studies Association, Pittsburgh, PA.
———. 2018a. "Learning to Forget: The Behaviorist Conception of Memory Driving Digital Media Platforms." Presented at the Mediating Change Conference, Denton, Texas.
———. 2018b. "Resurgent Behaviorism and the Rise of Neoliberal Schooling." In Handbook of Global Education Reform, edited by Kenneth Saltman and Alex Means, 323–49. Hoboken, NY: Wiley-Blackwell.
———. 2022. "Confronting the Digital Leviathan in Education: On Cybernetic Pedagogy and Data-Intensive Algorithmic Technologies." In Handbook of Critical Approaches to Politics and Policy of Education, edited by Kenneth J. Saltman and Nicole Nguyen, 225–40. Routledge.
Gilbert, Daniel. 2006. Stumbling on Happiness. New York: Alfred A. Knopf.
Halpern, Orit. 2014. "Cybernetic Rationality." Distinktion: Journal of Social Theory 15 (2): 223–38.
Kelly, Kevin. 2011. What Technology Wants. New York, NY: Penguin.
Kitroeff, Natalie, Kashmir Hill, Nina Feldman, Sydney Harper, Shannon M. Lin, Mary Wilson, Brendan Klinkenberg, et al. 2025. "She Fell in Love With ChatGPT. Like, Actual Love. With Sex." The New York Times, February 25, 2025. https://www.nytimes.com/2025/02/25/podcasts/the-daily/ai-chatgpt-boyfriend-relationship.html.

Klein, Naomi, and Astra Taylor. 2025. "The Rise of End Times Fascism." The Guardian, April 13, 2025. https://www.theguardian.com/us-news/ng-interactive/2025/apr/13/end-times-fascism-far-right-trump-musk.

McQuillan, Dan. 2023. "Predicted Benefits, Proven Harms: How AI's Algorithmic Violence Emerged from Our Own Social Matrix." The Sociological Review Magazine, June. https://doi.org/10.51428/tsr.ekpj9730.

McQuillan, Dan, Juliane Jarke, and Teresa Cerratto Pargman. 2024. "We Are at an Extreme Point Where We Have to Go All in on What We Really Believe Education Should Be About." Postdigital Science and Education 6 (1): 360–68.

Metz, Cade, Karen Weise, Marco Hernandez, Mike Isaac, and Anjali Singhvi. 2025. "How A.i. Is Changing the Way the World Builds Computers." The New York Times, March 17, 2025. https://www.nytimes.com/interactive/2025/03/16/technology/ai-data-centers.html.

Minkowski, E. 1970. Lived Time: Phenomenological and Psychopathological Studies. Northwestern University Studies in Phenomenology & Existential Philosophy. Evanston, IL: Northwestern University Press.

Nazaryan, Alexander. 2025. "Digital Therapists Get Stressed Too, Study Finds." The New York Times, March 17, 2025. https://www.nytimes.com/2025/03/17/science/chatgpt-digital-therapists-anxiety.html.

Pasquinelli, Matteo. 2017. "Machines That Morph Logic: Neural Networks and the Distorted Automation of Intelligence as Statistical Inference." Glass Bead. October 25, 2017. https://www.glass-bead.org/article/machines-that-morph-logic/.

———. 2019. "How a Machine Learns and Fails — A Grammar of Error for Artificial Intelligence." Sphere: Journal of Digital Cultures, no. 5: 1–17.

"Research Guides: Artificial Intelligence (AI) and Information Literacy: What Does AI Get Wrong?" August 16, 2023. https://lib.guides.umd.edu/c.php?g=1340355&p=9880574.

Rodriguez, Richard J., and Victor E. C. Ortuño. 2019. "The Destructuring of Time in Psychosis." In Managing Screen Time in an Online Society, 311–40. IGI Global.

Sacasas, L. M. 2019. "Technology Is a Branch of Moral Philosophy." January 19, 2019. https://thefrailestthing.com/2019/01/19/technology-is-a-branch-of-moral-philosophy/.

Sass, Louis, and Jasper Feyaerts. 2024. "Schizophrenia, the Very Idea: On Self-Disorder, Hyperreflexivity, and the Diagnostic Concept." Schizophrenia Research 267 (May): 473–86.

Siegel, Daniel J. 2020. The Developing Mind: How Relationships and the Brain Interact to Shape Who We Are. Third edition. The Guilford Press.

Stetar, William. April 4, 2025. "The Collapse of Language and the Epistemic Fractures in AI." Standing on the Shoulders of Giants. https://copin43.hashnode.dev/the-hallucination-myth-how-developers-of-language-models-are-gaslighting-us-and-why-we-should-be-alarmed.

Tangermann, Victor. June 2, 2025. "Therapy Chatbot Tells Recovering Addict to Have a Little Meth as a Treat." Futurism. https://futurism.com/therapy-chatbot-addict-meth.

Toews, Rob. 2024. "10 AI Predictions For 2025." Radical Ventures. December 23, 2024. https://radical.vc/10-ai-predictions-for-2025/.

Tolkien, J.R.R. The Fellowship of the Ring. 1954. London: George Allen & Unwin.

Ungar, Laura. January 13, 2025. "Elon Musk Says a Third Patient Got a Neuralink Brain Implant. The Work Is Part of a Booming Field." AP News. https://apnews.com/article/elon-musk-neuralink-brain-computer-interface-9dbc92206389f27fd032825cf1597ee5.