



Why Computer AI Will Never Do What We Imagine It Can

ABSTRACT: In response to questions about the author's previous proof (in Narrative 29.1) that computers contain a hardware limit that renders them permanently incapable of reading or writing narrative, this article draws upon the author's work with Deep Neural Networks, Judea Pearl's do-calculus, GPT-3, and other current-generation AI to logically demonstrate that no computer AI (quantum or otherwise) has ever learned, or will ever learn, to produce or process novels or any other kind of narrative (including scripts, short fiction, political speeches, business plans, scientific hypotheses, technology proposals, military strategies, and plots to take over the world).

KEYWORDS: *Artificial Intelligence, Narrative, Neuroscience, Biological Evolution, Symbolic Logic, Semiotics, Turing Machine, Causality.*

Angus Fletcher is Professor of Story Science at Ohio State's Project Narrative. His most recent books are *Wonderworks: Literary Invention and the Science of Stories* (Simon & Schuster, 2021) and *Storythinking: The New Science of Narrative Intelligence* (Columbia University Press, 2022).

I would like to express my gratitude to Jon Chun and Katherine Elkins for their thoughtful response to my proof of the limits of computer AI, a response that raises dozens of critical technical issues in the ongoing, high-stakes debate over what computers can, and will, do. In this reply, I directly answer Chun and Elkin's objections by drawing on my firsthand experience with Deep Neural Networks, Judea Pearl's do-calculus, GPT-3, and other current-generation AI to extend my original proof into a series of logical demonstrations that reveal why no computer AI (quantum or otherwise) has learned, or will ever learn, to produce or process narrative.

A COMPUTER AI THAT RUNS NARRATIVE—writing novels, creating plots, and planning to take over the world—is a twenty-first century perpetual motion machine: a captivating fiction that contradicts the laws of nature.

Perpetual motion appeared possible to many a brilliant mind—from Leonardo da Vinci to Nikola Tesla—because its refutation by thermodynamics took a bridging of two different fields: heat science and motion science. And a similar intellectual stretch is required to refute narrative AI. Its contradiction of nature can't be spotted from within AI's logical science or from within narrative's biological science. It can only be seen when we hold both sciences, together, in view.

This can seem a daunting task, one that obliges scholars who've spent a lifetime learning one field to find another lifetime in which to learn the other. But while such dual expertise would be crucial for building a narrative AI, it isn't necessary for seeing that the AI cannot be built. All we need to discern is the basic fact that story and logic operate in distinct domains, the former necessarily temporal, the latter intrinsically timeless. From that fact, we can establish, simply and unequivocally, that computer AI cannot read or write—and *will never be able to read or write*—novels or any other kind of narrative, including scripts, short fiction, character dialogue, political speeches, business plans, scientific hypotheses, technology proposals, military strategies, and plots to conquer the globe.

In my original article, I highlighted the difference between story and logic by starting with a series of stories that shifted into a short logical demonstration. In this reply, I'll take the reverse approach, starting with a short story that shifts into a series of logical demonstrations. Those demonstrations will allow us to delve into the subtleties of Deep Neural Networks, Judea Pearl's do-calculus, GPT-3, and the rest of AI's modern machinery, answering all the technical concerns that Jon Chun and Katherine Elkins have raised about my proof. But the key feature of this reply remains what it was in my original article: the *shift* between story and logic. That shift is the moment where we catch narrative and AI, together, in view, allowing us to see the contradiction that dooms any attempt to marry them.

The Story

Once upon a time, I was not an AI skeptic. I was a neuroscience researcher who became interested in narrative theory because so much human intelligence consists

of planning, speculating about causes and effects, and other forms of storythinking (Fletcher, *Storythinking*).

I entered the world of AI a few years ago when I was approached by a software company that was trying to create a Natural Language Processor: a machine-learning program that reads and writes like a human, only much faster, in any language, error-free. The company's program was great at semiotics but useless at story; it could perfectly identify the meanings of words but couldn't craft anecdotes or fathom what was happening in news articles. Or, to put the program's problem in syntactic terms: it was good with subjects, predicates, and linking verbs, but had difficulties processing other verbs. To process those other verbs, the program had to convert them into subjects, predicates, and linking verbs, but data was getting lost in translation. The lost data was narrative; it was the *action* of the verbs, what they were causing through time.

Since I had knowledge of both the theory and practice of narrative, I was asked to find a fix. And I accepted the job because I believed that a fix was possible. I believed because I knew very little about computers. But as I worked on the project, and the project made no progress, I began to learn. I learned about Deep Neural Networks. And about Judea Pearl's do-calculus. And about computer hardware. And I came to see that no software could solve the Natural Language Processor's problem. Because the problem was a hardware limit. The Processor could not run narrative because narrative required a different machine: the animal neuron.

I had no self-interested motive for this conclusion. Quite the opposite. I had access to some of the world's most talented AI programmers, and if we'd taught a computer to process narrative, we'd all have become very rich. A computer that could invent stories would be worth more than Google and Tesla. An algorithm that could plan would have made me king of everything. So, I had vast incentive to solve the problem of getting a computer to run narrative.

Solving problems requires, of course, intelligence and creativity, and I may lack the requisite amounts to solve hard problems. But while it's hard to solve hard problems, it can be straightforward to see when problems are unsolvable. It's straightforward, for example, to see that there's no way to solve, *What A is not A?* And getting a computer to run narrative is exactly that unsolvable problem.

Why Getting a Computer to Run Narrative is an Unsolvable Problem

Modern AI researchers almost universally acknowledge that computer AI and humans think differently. But the mechanical reasons for that different-thinking are often not fully dissected, leading to talk of "emergent properties" and "black-box intelligence" and "consciousness" and other fuzzy entities that blur the hard, physical boundary between algorithmic and neural processes. So, to bring that boundary into sharp relief, let's begin with this distinction: computer AI thinks entirely in equation, while the human brain thinks partly in equation but primarily in action.

Action works by speculating that causes will have specific effects; when the effects don't materialize as planned, the human brain reacts by speculating again. Equation works by positing that one thing equals (or in other words, correlates with) another; if the equation fails, AI enlarges the equation with a new correlation.

Equation has intentionally been engineered into the computer brain (that is, the Arithmetic Logic Unit, or ALU) because equation is the method of symbolic logic, the NAND/NOR data-cruncher formalized by Aristotle in his *Organon* and extended by subsequent mathematicians, semioticians, and statisticians to calculate moon flights, interpret texts, and estimate Bayesian probabilities. Humans, in other words, chose to build a machine to think in equations, and we did so because we discovered that math, semiotics, and statistical modeling can be powerful problem-solving tools (Jordan). When executed by the machine-learning algorithms of modern AI, those logical processes can predict likelihoods, sift digitized data, and identify spatial patterns, enabling tasks such as supply-chain logistics, archive searches, and image recognition.

Action, meanwhile, is an evolved, biological tool. So old is the tool that it dates back more than 500 million years to the beginnings of the animal neuron. The neuron emerged, in its insentient archaic form, as a machine that randomly flailed a flagellum or other primitive limb and then learned to modulate the flailing in response to positive and negative reinforcement. That basic trial-and-error mechanic scaled to produce your brain, which continues to buzz with random activity that originates in every one of your neurons. Your neurons (contrary to a pop-science myth) do not operate like miniature addition machines that sum inputs from other neurons to compute whether to fire their action potential. Instead, they pulse at their own rate, a rate that's influenced (but not determined) by environmental feedback. This means that the neuron is always acting; even if it receives *no* input from its neighbors, it will occasionally (or even frequently) fire. And because the neuron's frequency is ultimately autonomous, its firing will often conflict with other neurons in its network, leading to psychological vacillation, indecision, self-contradiction—and also to creativity.

The neuron's restless, independent action is why we humans get anxious, of two minds, and bored. It's why we're continually day-dreaming new adventures, possible futures, and counterfactual realities. And it's why our daydreams and anxieties take the form of narratives. Narratives are chains of actions. They're *not* chains of equations. Chains of equations are algorithms and arguments, both of which human brains struggle to follow. A human can happily listen to a story for hours but will labor to process a thirty-second argument (and will suffer total brain lock trying to process a thirty-second algorithm).

Because the power to think in action is built into the physical machinery of human neurons, it's possible for narrative processes to occur in our brain without consciousness. The insentient regions of the neocortex can (and often do) hatch plans that bubble up into our active awareness—or are even executed spontaneously, as when we're engaged in athletic events, improv performances, or other situations that require reactive, original behavior. Likewise, because the power to think in equation is built into the computer ALU, it's possible for algorithms to execute complex feats of

statistical calculation without being sentient. This means that neither narrative nor AI is an “emergent property” that mysteriously transcends the nuts and bolts of biology or computer science. Narrative and AI are both simply scaled-up versions of different machine processes. And like those machine processes, narrative and AI operate differently, leading humans and computers to make decisions, create, and problem-solve in contrasting ways:

- Humans make decisions by identifying a few significant datapoints and using them to speculate on what will happen in the future. Computers make decisions by crunching *all* the available datapoints to see what choices are correlated with the highest historical success rates.
- Humans create by thinking, *What do I want to accomplish?* And by then crafting a new tool or behavior that can bring about that effect. Computers create by mix-and-matching pieces of preexisting structures and labeling which combos are associated with positive and negative results.
- Humans problem-solve by hypothesizing an action that can eliminate the problem. Computers problem-solve by brute forcing from a set of calculated or randomly-generated options.

These different methods of thinking have enriched business, government, science, and education with human-AI partnerships in which humans do the narrative work of creating new actions, plans, and strategies, while machine-learning algorithms do logistics, data retrieval, pattern finding, and fast testing. The same basic division of labor has inhered in human-AI partnerships since the days of Alan Turing’s “bombe,” a pre-computer processor that brute forced Germany’s Enigma code during World War 2, providing humans with military intel that they marshalled to concoct strategic plans. But the recent wave of human-AI partnerships has nevertheless struck Elon Musk, Stephen Hawking, Ray Kurzweil, and other AI prognosticators as the first step of something far bigger: a revolution in which computers take over. Just as robots are replacing human workers in factories, the idea goes, so too will AI replace human thinkers in business, politics, war, science, and art, performing increasing chunks of the mental work and, in time, ousting human intelligence completely.

How could AI obsolesce us if we and AI think so differently? Well, as AI researchers have pointed out, different physical processes can yield the identical physical result: a hand-swung hammer and an air-powered piston can both drive the same nail; a human autoworker and a mechanized assembly line can both build the same car. Which means that it’s possible for computers to use their different process of decision-making, creating, and problem-solving to match us at our own intellectual tasks.

As a clear example of this, there’s chess. Chess was invented by humans for humans to play, and it’s played very differently by humans than by AI. While humans plot temporal attacks, imagining the capture of pieces, AI computes spatial board states, assigning them outcome probabilities. Yet despite this well-known divergence between the chess-playing methods of AI and humans, AI has learned to play chess with enough effectiveness to not only compete with human grandmasters but consis-

tently beat them. In fact, AI is now so good at chess that it's restricted to playing in its own league, so that humans still have a chance to win.

AI's ascendance at chess offers indisputable proof that computer thought can equal, and even outdo, the results of human thought. Just like a combustion engine and an electric battery can employ divergent mechanisms to propel the same vehicle forward, so too can a computer and a human deploy different brain mechanics to play the same game. Why, then, couldn't AI do the same with narrative? Why couldn't AI arrive at narrative through its own computational route? Why couldn't it do what it has done in chess: accomplish the same *product* via a different *process*?

The answer is: because narrative is a process. And you can't have a different process of doing the same process. To continue our above example, narrative is not chess. It's the way that humans *play* chess: planning actions, plotting countermoves. So, narrative isn't a product of human thinking that computer thinking can achieve via a different path; it's the human thinking itself.

This is the unsolvable, *What A is not A?* Or, with the values filled in: *What action is an equation?* Not, *What action produces the same outcome as an equation?* Not, *What action can be modeled by an equation?* But, *What temporal activity is a timeless identity?* *What combustion engine is literally an electric battery?*

This problem is unsolvable because it violates the law of noncontradiction. Which is the same law that, as Alan Turing proved in his famous 1936 "Turing machine" paper, makes it impossible for computers to solve another problem: the Halting Problem. And just as that law surprised earlier generations into realizing that there could never be an omniscient computer, so too was it the brick wall into which I crashed while working with the AI company. It proves that no Natural Language Processor will ever process verbs naturally. And that no computer AI (quantum or otherwise) will ever read or write novels (or command armies, invent technologies, run science labs, practice medicine, or replace humanity).

But hang on a moment, you might be thinking. A lot of big conclusions are being drawn here, very fast. And I have objections. Or at least, I have questions . . .

The Initial Questions Provoked by the Claim that Narrative and Computation are Different Processes

In my conversations with AI researchers and narrative scholars (conversations that began at the software company and have now mushroomed into the hundreds, following the publication of my proof), three sets of questions are typically provoked, right away, by the claim that narrative and computation are different processes. So, let's engage with those questions now, offering summary responses that chart the broad answers, before navigating the finer details in this reply's later sections.

The first set of questions is: Why is narrative only a process? Why can't it also be a product? Don't the brain's narrative processes yield products such as novels and strategic maps? And aren't those products narrative?

The summary response is: Narrative requires action, which requires cause-and-effect, which in turn requires a temporal one-two (the *one* for the cause, the *two* for the cause's effect). So, narrative is always a process. That process can be transcribed into novels, strategic maps, and other products, but such products are a symbolic shorthand that does not (and materially cannot) contain narrative's one-two action. The shorthand is merely a prompt that allows a narrative-competent machine (e.g. a human brain) to reconstitute the action. Just as with a cartoon flipbook, where the inked characters never move on the page but only in our brain, so too does a novel's narrative exist not in the printed text but only in our neural nuts and bolts, nuts and bolts running a *this-causes-that* process that the logic gates of the computer ALU cannot.

Once that first set of questions is addressed, the usual follow-up is the second set: Printed texts and images may be a symbolic shorthand, but the shorthand is still a product, so why couldn't computers manufacture that product through their own non-narrative process? Why couldn't AI statistics generate text and images that human readers appreciate as novels, strategic maps, and so forth?

The summary response is: Statistics can generate text, but because statistics cannot process that text as narrative, AI cannot learn to increase its efficiency at generating text that humans recognize as coherent characters and plots. Computers, in other words, can operate as symbol-writers but not as narrative-readers, leaving them unable to improve their storytelling in the manner that they've improved their chess-playing. When it comes to manufacturing text, computers can only get better at clustering semantic content into statistical sets, a procedure that allows AI text-generators such as GPT-3 to mimic subject-focus and a few superficial elements of literary style, but that won't ever yield texts that read like novels.

We'll expand below on these summary responses, the first in Myth 1, the second in Myth 2. But before plunging into the various intricacies that they raise, let's consider the third big set of questions. That set, unlike the others, doesn't center on the products of the human brain's narrative process. Instead, it shifts the focus onto AI's process.

Its questions are: What about Artificial General Intelligence? Doesn't Artificial General Intelligence prove that, eventually, computers will find a way to solve every solvable problem? And since the problem of reading novels and writing strategic maps has been solved by the human brain, doesn't that mean that eventually, computers will solve the problem, too?

These questions, which are raised by Chun and Elkins at the end of their response, are the ultimate source of most AI researchers' hesitation to accept that computers will never learn to author novels, strategic maps, business plans, scientific hypotheses, technological blueprints, and other products that are functionally equivalent to the products of our brain's narrative machinery. After all, once you accept that computers are capable of solving every solvable problem, it becomes self-evident that, given enough time and resources, computer AI will learn to accomplish everything that the human brain has done. As a sign of that apparently inevitable future, AI researchers will often (as Chun and Elkins do) point to Judea Pearl's do-calculus, which seems to demonstrate that it's possible to upgrade AI's software to handle cause-and-effect

(and therefore action and narrative). And they will often (as Chun and Elkins do) bring up quantum computing as evidence that it's also possible to upgrade AI's hardware to solve problems that seem beyond the reach of current computers.

We'll get into the particular wrinkles of Judea Pearl's do-calculus and quantum computing in Objection 4 and Objection 9, below. But first, let's explore the sweeping import of Artificial General Intelligence. Why doesn't it prove that, someday, computers will be able to match every product of the human brain via their own different process?

What about Artificial General Intelligence?

Computers forge their equations with logic, and logic has two theoretical limits. The first limit is problems (like the Halting Problem) that contradict logic. The second limit is problems that fall outside logic.

Because they fall outside logic, the problems in the second category are impossible to define *with* logic. As a result, they've been largely ignored in AI research. But the problems nevertheless exist. And in fact, they're everywhere. To find them, just look for a tool that isn't logic. For example, a wrench. A wrench is designed to solve all sorts of problems that logic can't. Ditto for a hammer, a saw, a vaccine, a computer mouse, and so on.

A more formal way to state this limit is: *The fact that one process can replicate another's outcome doesn't mean that any process can replicate every outcome of another.* To take an obvious example, the fact that computer AI can duplicate the outcomes of human chess-thought doesn't mean that computer AI can duplicate the outcomes of a rocket. The AI could model the rocket's flight plan and engine thrust. But no matter how clever or powerful the AI, its equations could only emit symbols and other algorithmic stuff. They couldn't emit fire, propelling the AI's motherboard into space.

This physical limit on process isn't contested by AI advocates. The reason that they don't apply it to AI's imitations of human thinking is because of "general intelligence." General intelligence is the *non plus ultra* of thought; defined precisely, it's the power to achieve all possible intelligent outputs. By upgrading computers—with smarter software and faster logic boards—to the point at which they achieve general intelligence, AI researchers thus aspire to make an Artificial General Intelligence that can invent technologies, do science, run businesses, win wars, create art, and in all other ways replicate the results of human cognition.

This ascent into general intelligence is an extraordinary, awe-inspiring idea. (Indeed, it's the root of the belief in an omniscient God.) Yet the ascent has a snag: there's no scientific evidence that general intelligence exists. Within computer science, the notion of general intelligence is founded on an unwarranted extension of Alan Turing's formalization of "general computing." We'll cover Turing's formalization in Objection 7, below, but basically, the extension rests on a tautology: by equating "solvable" with "computable," it commits the circular logic of assuming its own conclusion (that a machine capable of solving all computable problems can therefore solve all solvable problems), thus failing to address (or even acknowledge)

the possibility of other categories of solvable problems. We've already touched on one such category—solvable *physical* problems—above. And as we'll explore more in this section, there's another such category of solvable *intellectual* problems: the myriad natural-world concerns that the narrative machinery in our brains evolved to handle.

Within the rest of modern science, the notion of general intelligence dates to early twentieth century psychologists (for example, Wilhelm Wundt and Charles Spearman) who imported it from philosophy, where it derived from Kantian Reason, Aristotelian cognitive universals, and other rational definitions of human intelligence. These definitions are—as evidenced by their widespread adoption by twentieth century computer scientists and cognitive scientists—logically plausible. Yet they're belied by the actual biology of our brain. Our brain does not consist of general neurons all running the same process. It instead houses a mix of specialized neurons. Most of them perform the brain's original function: action. But some have biological machinery capable of executing other functions, of which one is the very computational process run by AI: equation. Which explains how our brain can calculate statistical probabilities, logical propositions, and symbolic systems, although with nothing like the speed, power, and consistency of a computer.

How did our brain acquire the power to run this computational process? Is it because equation emerged out of action as a more general form of intelligence? Or, was it the other way round? Did action give birth to the more specialized intelligence of equation?

The answer is: neither. What happened instead is that both processes evolved, independently, to grapple with different biological jobs. Action evolved (as we saw two sections ago) from using limbs to *move*. Meanwhile, equation evolved (also a little more than 500 million years ago) from using eyes to *see*, or in other words, to represent the world. Neurons with the machinery to run the first process formed the brain's motor centers and other cortical regions that planned and mediated movement. Neurons with the specialized synaptic machinery to run the second process formed the visual cortex and propagated through prefrontal and parietal brain regions that can now perform mathematical and logical operations—operations that can also be performed by the visual cortexes of individuals born blind.

What all this means is that the human brain is not a single, unified machine. It is multiple neural mechanisms that developed in response to distinct physical tasks. Those multiple mechanisms are now wedged side-by-side in your cranium like a Swiss army knife, not merged into a more general mental process like an omniscient God. Which is why you feel a neural shift every time you switch from reading stories to reading logical demonstrations: your brain is changing to another tool for thinking. It's activating a separate hub of neural machinery.

That shift is your inner empirical proof that human intelligence is not a general power carrying us upward into universal enlightenment; it's a package of discrete mental processes, each of which grips different environmental problems and opportunities in different ways. And there's no more physical reason to think that those mental processes would converge into a single process than there is to think that a perfect hammer would also be a perfect saw.

This isn't to say, of course, that we can rule out the existence of general intelligence, any more than we can rule out the existence of an omniscient God. But it is to say that, until such time as general intelligence is plausibly detected, it needs to be cordoned outside the realm of modern science. And within that realm, action and equation point in precisely the opposite direction of general intelligence. Their processes are not only different but divergent, as we can readily discern by comparing the performance characteristics of human story and AI statistics:

Statistics struggles to gain traction in small data sets, which is why computer AI turns brittle in fast-changing environments. Yet such environments are the very ones in which story thrives. Story allows us to act purposefully with limited intel, because it substitutes adaptive creativity for rigorous accuracy. This is why narrative cognition has proved such a useful tool in the ever-shifting ecosystems of natural selection, and why the animal brain, even after evolving the capacity for equation, has remained primarily a storythinker.

Meanwhile, story's mechanical strength has a flip side: story has difficulty synthesizing large quantities of data, which is why we humans get overwhelmed by situational noise, latching onto skewed bits of evidence that confirm our biases, instincts, or priors. AI is the opposite. It performs better with huge volumes of facts and figures: the same spreadsheet tables that flabbergast human thought make computers more effective, suiting computers for environments (such as the timeless rule domains of chess and other boardgames) where the data is rich and cumulatively reliable.

What modern science tells us, in short, is that AI and narrative are such unlike processes that their performance ranges are inverted. Narrative is better in unstable, murky contexts; AI, in firmly regulated, transparent ones. Like hammers and saws, they're tools for solving different sets of problems, making them complementary, not duplicative.

This empirical reality upholds the *A is not A* logical proof. It establishes that modern science and the eternal syllogism can both agree: computer AI will never fathom or author a novel—or business plan, political movement, scientific hypothesis, technological innovation, military strategy, or other narrative. But if you're not yet convinced—or if you're looking for a little more conviction—let's dig further into the specifics by examining a few common myths about AI and narrative. After which, we'll resolve the particular objections raised against my claims by Chun and Elkins.

**AI-NARRATIVE MYTH 1: NARRATIVE CAN BE REDUCED TO LANGUAGE, SO WHEN
AI BECOMES SUFFICIENTLY ADEPT AT PROCESSING LANGUAGE, IT WILL LEARN
TO READ AND WRITE NARRATIVE.**

That AI researchers would reduce narrative to language is understandable: a great deal of modern literary instruction—from K-12 Common Core through university—equates novels, short stories, and plays with words printed on a page. The equation was mainstreamed into modern literary studies in the 1920s by I. A. Richards, who replaced folk-psychological approaches to literature (for example, nineteenth century “character criticism”) with semiotics, that is, with the deduction: (1) literary texts are composed of language (2) language is composed of words (3) words are symbols (4) symbols can be processed via symbolic logic for their meanings (that is, the truth claims they equal). This deduction is logically valid and yet fails empirically to account for all of literature’s observed effects (Phelan), a failure we can detect by tracing the historical results of literary semiotics (as in my original article). One of those results was to convert Shakespeare’s plays from theatrical performances into symbolic webs, deleting characters and plots for themes and meanings. Another was to turn films and novels into allegories, reducing fiction to a rhetorical method for communicating propositional content. Another was to implode literary studies in the late twentieth century, thanks to deconstructionists who logically demonstrated that semiotics could extract opposed meanings from the same text.

What happened in all these cases is that semiotics ghosted narrative, replacing causal actors and events with a logical method for converting printed texts into arguments, representations, and other forms of equation that could be judged true or false, right or wrong. Semiotics told us: *Those arguments and representations are what literature is. Literature is just an elaborate way of inserting truth claims into the human brain.*

This is patently false to any ordinary reader of literature. Ordinary readers do not go to literature purely for arguments and representations. Ordinary readers go to literature to experience feelings such as empathy for characters and suspense at plots. Those feelings, in turn, have psychological effects (such as the reduction of anger, the alleviation of trauma, the easing of loneliness, and the boosting of curiosity) that can make our brain healthier, happier, and more creative (Fletcher, *Wonderworks*). Such effects cannot be replicated by semiotic propositions, themes, and identities. If they could, there’d be no need for literature; our shelves could be stocked purely with logic textbooks.

To recover the psychological effects that ordinary readers detect in literature, let’s take a hard look at the claim that literature is reducible to language. This claim seems self-evident when we open a novel: What else is a novel but words on a page? And what else are words but language? Yet the claim begins to fall apart when we spin back time to literature’s physical origins. Those origins were oral and dramatic, which is to say, they preceded printed texts. The printed texts came later, as a way of recording the performances of bards and stage players. So, it’s worth asking: was anything lost in the recording?

We can answer that question by excavating the anthropological origins of writing. Writing was invented (in multiple cultures, independently, more than 5,000 years ago) as a record-keeping aid for merchants who wanted to log commercial transactions, regents who wanted to document political treaties, and priests who wanted to maintain ancient rites and rituals. It came to be, in other words, as a *memorial* tool. It was devised as a prompt to help human brains recall things. Those recalled things included objects and events, so writing developed nouns and verbs, respectively. Nouns and verbs, like all writing, were symbols. Yet despite this shared ontology, nouns and verbs referred to things with very different physical natures. Nouns referred to things that needed only space to exist, whereas verbs referred to things that required time.

This practical difference between nouns and verbs is significant because writing does not itself exist in time, only in space. So, nouns can embody objects in a way that verbs cannot embody events. You can visualize the distinction by rewinding history before the development of alphabet symbol systems, returning writing to its archaic beginnings in etched and painted pictures. A picture can capture objects such as trees and humans, but a picture cannot capture events such as a tree growing or a human running. To represent those actions requires *two* pictures, the first with a shorter tree and the second with a taller; the first with a person's legs in one position and the second with the legs advanced. Yet even those two pictures don't capture the event's physical action; the action occurs in the gap between the pictures. And if we fill that gap with a third picture, we still don't capture the action. Instead, we create two more gaps, with the action occurring in them, unseen.

If the action cannot be captured on the page, then where does it exist? The answer is: it exists in a human brain that reads the page. That human brain contains narrative machinery that thinks in action, so it can add the element of time, imagining what is not physically in print. And indeed, our brain's machinery performs this addition so automatically that it never occurs to us that the narrative is not contained in the printed words of books and scrolls. To us, the narrative always appears there (even though it is really just in our head), explaining our intuitive (but mistaken) belief that narrative can be not only a process but also a product.

This readerly addition of narrative returns us to writing's historical origins as a memorial prompt. In those origins, we see that writing wasn't designed to provide a stand-alone representation of the world; it was simply meant to be a useful tool for reminding human brains of objects and events that they'd already processed. Writing therefore did not need to contain anything that human brains could supply; it could shorthand events as static symbols because the kinetic action that made those events into events could be added by a reading human.

Writing's shorthand worked without a hitch so long as its readers were human. But then, beginning in the twentieth century, the texts were fed to a different reader: the computer. And the computer contained no narrative machinery, only NAND/NOR logic gates that thought in the eternal present tense of mathematics, reducing time to an icon, sign, or spatial dimension. So, when the computer read verbs, it could not be reminded, as a human brain could, of the temporal process of action. It could only treat verbs as what they literally were on the page: timeless representations.

The result was to render explicit writing's deletion of time—and with it, of narrative. When oral stories and theatrical performances were inked on scrolls, their plots and characters were converted into nonnarrative text. Language had recorded enough for human brains to recreate the process of action. But absent the human brain or some other narrative reader, that process was lost.

The same loss occurs with digital video cameras, 35 mm film, and other technologies that convert events into sequences of images. The conversion is perfectly sufficient as long as the intended audience is a human. But a computer lacks the hardware to restore the action. It cannot think of characters as characters or plots as plots; which is to say, it cannot think of them as causal agents and effects. It can only think of them as subjects and predicates to be run through equations. By confusing the written word for literature itself, computer AI thus fails to appreciate literature in precisely the same way that semiotics (and a great deal of modern literary studies) fails to appreciate literature. It reduces events to themes and characters to representations, skipping over the plots and psychologies that drive literature's most popular physiological effects: empathy, suspense, laughter, catharsis, etc.

These biological facts about writing—and language more generally—make it impossible for computer AI to learn to read or write literature. Computers can learn to use language in the sense of grasping what language literally *is*. But they cannot learn in the sense of grasping what language was created *to do*. They're missing the crucial mental ingredient—narrative—that literature requires to come alive. So, literature will forever be to them like instant soup without the water. Which is to say: not soup.

**AI-NARRATIVE MYTH 2: TO LEARN TO WRITE NOVELS, AI SIMPLY HAS TO
PRODUCE TEXTS THAT HUMANS RECOGNIZE AS NARRATIVE, AND AI PROGRAMS
SUCH AS GPT-3 HAVE DONE THAT ALREADY**

This myth is responsible for the general public's belief that computer AI has learned to write stories. In fact, *no* computer AI has *ever* learned to write a story. What has happened instead is that a computer has emitted a syntactic mix-and-match that human brains (by adding the ingredient of narrative, as in Myth 1) have re-imagined as a story.

Here's why the distinction matters: for computers to learn to create stories, they must function not simply as writers but also as readers. They must be able to determine that one syntactic mix-and-match is more narrative than another, thereby learning to improve their storytelling. Without that improvement, computers will never become authors of novels. They will simply be random pastiche makers.

This is not a philosophical distinction between a real novel and a cut-and-paste one. It is a practical distinction between a feasible novel and an unfeasible one. The odds of an arbitrarily-generated sequence of syntax adding up to a coherent novel are infinitesimally small. So small that if we had millions of supercomputers running for millions of years, we'd almost certainly never get a single chapter. But that's not the most unfeasible part of the process. The most unfeasible part is that we'd only find the chapter *if we had humans reading everything that the computers emitted*. And since it

takes humans longer to read than it does for computers to eject text, just imagine the gigantic social apparatus that would be required to hunt through the printed haystack for the narrative needle. It would necessitate whole planets of human readers, trapped in an endless MTurk experiment. It would be the literal incarnation of Jorge Luis Borges's "La biblioteca de Babel."

This looming absurdity has been superficially concealed by the apparent improvement of computer algorithms such as GPT-3 at text generation. But once you understand how the algorithms work, you can quickly see that the improvement is illusory. What the algorithms do is identify spatial patterns in pre-existing human sentences—and then blend those patterns by interchanging syntactic chunks. You can think of it as a robot that goes to a parking lot and creates new cars by randomly exchanging the piston rods of one vehicle for another. This is creative in a limited sense. But it is plainly not how a human engineer creates cars. A human engineer creates cars by envisioning an automotive function and then plotting backward (using narrative) to a mechanism that can produce that effect.

This difference between human engineers and the robot radically constrains what the robot can accomplish. Most damningly, the robot cannot learn to develop cars that do new things. It cannot, that is, learn how to innovate. It might haphazardly onto something that a human appreciates as an innovation. But the robot cannot build upon that breakthrough by repeating its innovation process to generate new innovations. Nor can it refine the innovation that it has stumbled into. It cannot because its method is not one of causal thinking; its method is instead simply one of random rearrangement, and all that method can do to hone itself is identify pre-existing structures—such as carburetors and sparkplugs—that can be more reliably switched without causing the vehicle to explode. Which is to say: the most the robot can learn is to constrain its randomness in a way that reduces fatal errors. It cannot, like a human brain, learn to *develop* its randomness in a way that *increases* creative breakthroughs.

This enormously inefficient and fundamentally purposeless robo-activity is what is happening in GPT-3 and its ilk. And it reveals that these Artificial Intelligences aren't producing something intellectually coherent, narrative or otherwise. They're instead playing a game, the purpose of which is to dupe us into *thinking* that we're reading something intellectually coherent. Winning the game is impossible for the AI, because it requires too much luck. But the AI can briefly extend the contest by avoiding a few gross mistakes. Which it does by identifying pre-existing structures—that is, chunks of syntax—that it can more safely blend together, allowing it to fake competence for long enough to trick us into thinking: *This machine is making progress—so, eventually, it will get somewhere.*

Like any game, this one can be entertaining to play. But it's not a way to extend our literary library. If we want to extend that library, we should give literature to human students and encourage them to imagine their own characters and stories. Then step back and watch as they teach themselves.

AI-NARRATIVE MYTH 3: NARRATIVE'S FUNCTION CAN BE CAPTURED BY A STATISTICAL ANALYSIS OF NOVELS

As we covered in Myth 1, novels are a memory prompt coded in textual symbols that exclude (as symbols must) the temporal action of cause-and-effect. It's only when the memory prompt is uploaded into a human brain (or other machine capable of thinking in action) that it can become narrative. Therefore, no purely statistical analysis of novels can therefore capture narrative because narrative is not equivalent to the textual symbols that statistics vacuum from the printed (or digitized) page.

But what if we accept all that—and still think that there could be a role for AI in narrative research? What if we imagine a hybrid relationship between AI and human readers, like the one that exists in assisted machine-learning, where humans extract and label narrative content that AI then analyzes with statistics?

The answer is: we would fall back into the same unscientific method as before. To begin with, we can't "label" narrative content, because that content (as we saw in Myth 1) exists in the gaps between the printed symbols that computers process. If narrative content could be labeled for computers, then AI (and semiotics) would have worked in the first place. Moreover, by encouraging us to gather up stories and identify what they have in common, this approach leads in the profoundly unscientific direction of Joseph Campbell's "hero's journey," Matthew Jockers's claim that there are only 6 (or 7) kinds of plot, and other efforts to identify timelessly enduring story structures. Which is to say, it leads to a conservative view of narrative that's consistent with logical approaches to literature (such as medieval literary criticism and the Enlightenment's theatrical unities) but that fundamentally contradicts narrative's biological function.

That biological function is to innovate action. Narrative evolved in animal brains (prior to humans and prior to language) as a way of responding to environmental challenges and opportunities. It enabled our primate ancestors to plan sequences of actions through which they escaped threats and captured food. It was, in short, adaptive. And it succeeded in being adaptive because it was endlessly flexible. As threats and food evolved to counter one plot, new plots were hatched, allowing for brains to reshape their behavior to function in shifting contexts.

This reactive plasticity is why stories remain so useful to us today. Stories allow us to keep pace with our ever-changing world. They allow us to fashion new political narratives and business plans that flex to shifting public wants and needs. They allow us to imagine original technologies to troubleshoot the problems of our modern lives. They allow us to adjust to tough job markets, internet stardom, and a million other environmental threats and opportunities that did not exist when our brain first evolved.

We lose this adaptive utility by treating narrative as a product that can be subjected to statistical analysis. Such analysis leads us to focus on what narrative *is* as opposed to what narrative *can do*, encouraging us to reiterate prior fairytales and political actions instead of empowering us to create fresh stories that tackle the unique challenges and chances of our moment.

The way to generate those fresh stories is to engage with old stories not as finished products but as ongoing processes. It's to think: *What would I do if I was that char-*

acter? How would I operate in that storyworld? How would I extend that novel's plot? It's to leverage our reading momentum into becoming authors ourselves. It's to free ourselves from logic's limiting effort to reduce narrative to everlasting structures. And it's to embrace the modern scientific finding that narrative is a dynamic activity that can differ across cultures, historical eras, and individuals, nurturing literary diversity and throwing open a creative future of original characters, storyworlds, and plots.

Such is the future celebrated by the word *novel*. Which doesn't mean *eternal, timeless, classic*. But *original, unprecedented, new*.

Now that we've examined these three myths, I hope that my reasons for dissenting from Chun and Elkins are clear: when you approach narrative from the perspective of neuroscience, biological evolution, and the psychological experience of ordinary readers, it leads to a fundamentally different view of what novels are doing. That different view has many profound theoretical and practical consequences, too many to be covered here. But to fill a few more of the gaps, I'll finish this reply by offering direct responses to Chun and Elkins's main critiques.

OBJECTION 1: I ONLY UNDERSTAND THE WAY THE "OLD" AI WORKS

To quote: "Fletcher's critique is misdirected at largely defunct older Chomskyan Symbolic approaches. He overlooks current AI approaches that rely on carefully-crafted white-box statistical models (termed GOF AI or Good Old-Fashioned AI) and massive black-box statistical universal approximation engines (DNN or Deep Neural Networks)."

My own programming experience is in fact with the "new" AI of Deep Neural Networks. So, everything I have written about the limits of AI in this reply (and in my original article) is directed specifically at Deep Neural Networks and their "black-box" engines. But it also applies to every past and future AI because all AI runs on the same computer hardware, and the AI limit I identify is a hardware limit. That limit will never be overcome by any software, ever, for the same reason that downloading a new OS onto your smartphone will never enable your smartphone to fly. Software can only max out the functional capability of hardware; it cannot enlarge that capability beyond its inherent physical range.

Don't be led by the notion of "emergent properties" into thinking that nuts-and-bolts can jump the laws of physics. A computer that (for example) achieved sentience would still be limited to performing computation, just as a car that achieved sentience would still be limited to driving from one place to another. All that would change is that both machines would become *aware* of their activities. A self-aware computer would not become capable of writing novels any more than a self-aware car would become capable of launching itself to Mars.

Deep Neural Networks are also not especially new. They've been around for more than a decade, long enough that if they could process narrative, we would have seen it. And their name is a classic instance of rhetorical creep: by calling AI "neural" it implies that AI is operating like a human neuron. It is not. It is operating like a computer statistician.

**OBJECTION 2: I DON'T TELL THE COMPLETE STORY OF BERTRAND RUSSELL, AI,
OR THE DIGITAL HUMANITIES**

I accept this objection. I did not tell the complete story of Russell or AI or anything. How could I? No story is ever complete. It's in the nature of narrative to incorporate only some of the facts.

What I can't accept, however, is that longer, more detailed stories invalidate the core claim illustrated by my original stories. That claim is: computer AI and the human brain possess different physical machinery, the former restricted to running statistics, semiotics, and other forms of equation, the latter capable of running action—and therefore, narrative. There's no evidence in the history of Russell, AI, or anything else that overturns this material distinction.

**OBJECTION 3: COMPUTER AI HAS ALREADY AUTHORED SHORT STORIES AND
SCRIPTS**

This is entirely untrue. This is precisely what AI has not done, for the reasons I outlined in Myth 2. It has instead generated swap-piles that have been cherry-picked by humans and assisted by readerly generosity to create the illusion of short stories and scripts.

Really, what AI is learning to do is hack the Turing test, not complete it. AI is figuring out ways to eliminate its most obvious tells and string along indulgent audiences. It is not doing what humans do when they write literature: think from the perspective of narrators, characters, and other causal agents in order to generate events, dialogues, and other potential *effects* of those causes. Nor could computer AI ever do this. It cannot think from cause to effect. It can only think statistically. So, it can never get better at writing literature. It can only come up with more tricks for masking its fundamental narrative incapacity.

To prove this to yourself, play with GPT-3. You will be initially amazed. And then the more you play with it, the more you will realize: it is narratively incoherent. It is simply engaging in the grand evasion of blending together enough semi-related semiotic material to seem as if it could be going somewhere. Then have a conversation with a three-year-old. Ask her what she did that day. Her candid purposefulness will make clear: she knows, innately, how to tell a story.

It is the responsibility of narrative theorists to communicate this difference between AI and humans to the general public, because there's real mischief in people believing that computers have already managed to produce simple narratives. The mischief is that people will think: *Computers are improving at the task of creating narratives. And if computers are improving, they will get better and better until they become genuine storytellers. At which point, they could use story to plot a future free of humans—and then plan to take over the world.*

If computers were improving at narrative, that would be absolutely correct. But the computers *are not improving*. They're simply discovering how to mask their root incompetence. This method of "learning" means that GPT-3 is really nothing more

than a high-powered prank. Rather than learning what to do with story, it's learning what not to say with words. Because the longer it can keep the ruse going, suspending our readerly judgment in a state of wondering credulity, the more it can play on our brain's narrative capacity to imagine what future AI could accomplish, enticing us to dream up sci-fi fantasies that it itself cannot.

OBJECTION 4: I FAIL TO ACKNOWLEDGE JUDEA PEARL'S DO-CALCULUS, WHICH PROVES THAT COMPUTER AI IS CAPABLE OF CAUSAL REASONING

Judea Pearl's do-calculus is brilliant. But it was not designed to serve, nor can it serve, the function of having a computer think in characters, plots, and other elements of narrative.

What the do-calculus does is identify statistical correlations that can be tagged as statistical causations. For example, the do-calculus thinks: *If there is a 97% correlation between going to the dentist and buying a toothbrush, and a 93% correlation between going to the dentist and buying toothpaste, and a 76% correlation between buying a toothbrush and buying toothpaste, and a 17% correlation between going to the dentist and buying a towel, and a 5% correlation between buying a towel and buying a toothbrush, then there is a 54% chance that going to the dentist is a cause of buying a toothbrush.*

You will immediately notice that this mathematical *if-then* procedure is not how your brain arrives at the conclusion that going to the dentist causes toothbrush-buying. (Your brain probably thinks something like: *Going to the dentist causes me to care about my teeth, which causes me to buy a toothbrush.*) That's because your brain is performing causal thinking—while the do-calculus is not. The do-calculus is doing what all computers must always do: performing correlational thinking. It is using statistical correlations to label something as a possible cause, just as a Natural Language Processor uses statistical correlations to label a noun as a possible direct object of an action verb.

Because the do-calculus is yoked to this computational method, it cannot deploy a labeled cause as a cause, any more than a Natural Language Processor can deploy a labeled action verb as a verb. The best it can do is correlate the cause with a set of symbols that it has labeled "effects." This is the logical process that resulted in the shortcomings of medieval science (as documented in my original article). Theoretically, it leads to the tautology that a cause is equal to its effect, deleting temporal action from physics and tipping us into magical thinking. And practically, it renders the do-calculus unable to mobilize causes (as our brain's narrative machinery can) to speculate on new (that is, unobserved or currently nonexistent) effects, thus preventing the do-calculus from outputting original plans, plots, behaviors, and narratives.

The do-calculus thus does not in any way establish that computers can replicate human causal thinking. Which in turn allows us to see that the do-calculus does not enable computers to read or write stories.

With regard to reading stories, let's imagine a possible future in which the do-calculus is upgraded: instead of relying, as the do-calculus does now, on a human

operator to posit causal relationships that it then quantifies with data, our future do-calculus can itself inductively posit causal relationships that it then quantifies. This is the operational horizon of the do-calculus; its utopian technical vista. But even at this distant reach, the do-calculus would still be limited to working backward from correlation to causation, such that it could use data to probe whether two co-occurring objects have a causal relationship, but could not do what our brain's narrative machinery naturally can: speculate on the existence of unseen causal actors. To take a simple example, the future do-calculus computer could deduce that if a literary character frowned every time that a second character entered the room, then there was a chance that the second character was causing the first character to frown. But unlike a human reader, the future do-calculus computer couldn't conjecture from this information that the frown's unseen cause was a hidden feeling that the first character carried inside, a feeling that could prompt the first character to do things when the second character was nowhere near.

Because of this limit, a do-calculus computer could also never do what human readers do all the time: posit a cause from a single event. This is impossible for a statistical machine. But it's the very biological root of narrative: to hypothesize an unseen reason for a one-off occurrence. Without it, the future do-calculus computer would be sucked back into the Joseph Campbell problem that we explored in Myth 3: in its hunt for comprehension, it would be forced to vacuum up ever-greater quantities of text that dissolve the differences between individual particulars into abstract patterns. Jane Austen's characters would be blurred into narrative archetypes and Shakespeare's plots into narrative allegories, destroying the authors' unique creations and evaporating most of the literary effects that human brains get from *Hamlet* and *Pride and Prejudice*.

With regard to writing stories, the do-calculus falls into the trap that catches GPT-3. Since the do-calculus cannot derive new effects from the causes it labels, it cannot start with a causal agent (that is, a literary character) and imagine what she might do if she was placed in a new situation (that is, an original storyworld). The do-calculus can only distill general rules about what chunks of existing stories can be swapped around. At the cost of a gigantic investment of resources, the do-calculus's ultimate achievement would thus be mix-and-match versions of stories we'd already heard, cranked out at such cavalier volume that legions of humans would have to devote lifetimes to identifying whether there was a single script worthy of going directly to rerun.

This sort of dystopian boondoggle is the inevitable result of treating AI as a replacement for human intelligence rather than as what it is: a tool that human intelligence can use to solve problems (Larson). The do-calculus was never intended to be a stand-alone substitute for our brain. It was intended to solve a narrow mathematical problem defined by its human inventor, and it solves that problem ingeniously. It does not solve the unsolvable problem of computing narrative.

OBJECTION 5: I AM WRONG TO SUGGEST THAT THE COMPUTER ALU CAN ONLY RUN SYMBOLIC LOGIC

To quote: “[The ALU can run] symbolic logic, statistical machine learning, deep neural networks, cellular automata, probabilistic, and causal graphical models.”

I completely agree that the computer ALU can run statistical machine learning, deep neural networks, cellular automata, and probabilistic models. But, per my original point, these are all extensions of symbolic logic, at least as classically defined, for example, by Turing’s protégé I. J. Good: “a finite set of symbols together with a finite number of rules of transformation for transforming finite strings of symbols into other strings” (182). If we want to endow these logical processes with more vivid nomenclature, that’s fine, but whatever we call them, they remain AND, OR, NOT operations.

As to whether the computer ALU can run “causal graphical models,” that depends on what we mean by “run.” Causal graphical models are another term for Judea Pearl’s do-calculus (see Objection 4). So, yes, the computer ALU can run the parts of causal graphical models that are reducible to symbolic logic. But anything beyond that requires a human operator.

OBJECTION 6: I FAIL TO RECKON WITH THE REDUCTIONIST MATERIALISM THRUST UPON US BY AI

To quote: “The most salient question is: if language, narrative and art can be modeled by machines, does that imply human thought and creativity are simply the emergent properties of vast statistical properties of atoms and neuronal spike trains? Where is humanity in such radically reductionist materialism?”

This misconstrues my central position. I *am* a radically reductionist materialist. I *do* suppose that narrative and art can not only be modeled, but run, on a machine. That machine is the human neuron, which is not magic or mystic—nor even necessarily conscious. Narrative plots and plans are run all the time by nonconscious regions of our brain.

This realization of my own machineness does not lead me into soul-searching about my “humanity.” Quite the opposite. It creates wonder and gratitude at my existence. I am amazed and thankful to have been born with the biological hardware to read Maya Angelou and Virginia Woolf. I am amazed and thankful that the insentient mechanics of evolution by natural selection have resulted in my existence.

And also, by being a reductionist materialist, I can see that human thought is not an “emergent” property of “vast statistical properties of atoms and neuronal spike trains.” This is to treat statistics as a property of nature as opposed to what it is: a tool for modeling nature. It is to confuse a mathematical representation of sunlight with sunlight itself. It carries us out of modern experimental science and into a rationalist science that presumes that the world’s operating system is symbolic logic. That is the reductionist symbolism of AI, not the reductionist materialism of real-world physics, biology, and my brain.

**OBJECTION 7: I FAIL TO ACKNOWLEDGE ALAN TURING'S FAMOUS PROOF THAT
COMPUTERS CAN SOLVE ANY SOLVABLE PROBLEM**

To quote: "Given that von Neuman architectures (CPU/ALU) are Turing complete universal compute platforms, they can process any statistical relationship and not just correlations."

Let's begin by being clear about what "Turing complete" means. It means that a machine is a "general purpose" computer, or in other words, that unlike the early twentieth-century computational devices that preceded Alan Turing's famous 1936 paper, it isn't limited to particular hardwired tasks such as adding pennies or totaling census ballots. It means, in short, that the machine can compute anything that can be computed.

This straightforward insight has been exploited, repeatedly, by AI promoters to claim that AI can therefore solve all solvable problems. But that claim rests on the tautological assumption that all solvable problems are computable problems. If a solvable problem cannot be solved by logic and statistics, then it cannot be solved by a "Turing complete" machine.

Such solvable but non-computable problems (as we explored in our discussion of Artificial General Intelligence, above) are everywhere in our material world, because they're the problems of any unstable, uncertain environment. Which is why the human brain evolved to think in the nonlogical, nonstatistical method of speculating causes and testing effects. The method of modern experimental science. And of narrative.

But, you might respond, what if we could feed the data of every single piece of cosmic matter and energy into a computer? Couldn't that computer solve every solvable problem?

No. For starters, this thought experiment is another version of the tautology above. It reifies its own unproved assumption (that solvable equals computable) by leaping into a future where everything is inside a computer, literally jumping to its own conclusion. That conclusion is that it's possible to achieve a perfect mathematical representation of the physical universe, which is to say, that the physical universe can be equated to a computer simulation. At which point, it becomes self-evident that any solvable problem that exists within the simulation can be solved by the simulation.

Moreover, even if we grant this tautology for the sake of argument, it doesn't get us as far as you might think. For starters, it doesn't get us to AI omniscience. A data-complete computer would not be an all-knowing God; it would lack the knowledge, as Turing's paper proves, to solve the Halting Problem. And although it's true that the computer would be able to solve the problems inside its cosmic simulation, those problems would not equal the set of *all* solvable problems. They would merely equal the set of *all remaining* solvable problems, because the other solvable problems (the ones, that is, posed by uncertainty) would have been eliminated by the computer's data-completeness. To forge that data-completeness, the computer's designers and engineers would have had to employ their brains' narrative machinery to solve the problems of a world where not everything was known yet. And if those no-longer existing problems were fed into the data-complete computer after it was finished, it

would be as incapable of solving them as AI is now. Why? Because such problems, by definition, take the form of: *How should I act when I cannot have total certainty?* Rendering them opaque to a machine that can only problem-solve via absolute knowledge.

Far from overturning our *A is not A* proof, this thought experiment thus upholds it, revealing that the cosmic computer would achieve apparent intellectual omnipotence only because its data-complete simulation obviates the need for strategic plans to cope with uncertain situations (just as the data-complete chess circuits of grand-master AI obviates the need for plot countermoves to win). So, the computer would never have to face the problem, *What action is an equation?* It would simply evade the problem forever.

Not that any of this could happen, of course. There's no physical way to measure all the data in the universe and plug it into a computer. Once again, thinking about AI has propelled us into science fiction, which is to say, into a story that infinite Turing machines could not fathom.

**OBJECTION 8: I FAIL TO RECOGNIZE THE LOOMING EXISTENTIAL THREAT THAT
COMPUTER AI POSES TO OUR HUMAN FUTURE**

Computer AI poses the same existential threat to us now that astrology once posed to ancient kingdoms. The threat is that by focusing on a make-believe danger we ignore actual threats to our existence. And of those threats, the primary one, if biology and history are a guide, is not computers. It is ourselves. It is humans.

Humans are our main biological competitors and our main historical antagonists. Which is why we're always harming each other, via both direct violence and callous indifference. When we elevate AI to a similar status, what we're really saying is: AI is us. But computer AI is nothing of the sort. It has no desires, nor any power to plan actions or to plot futures. Nor will it. It's just a machine for doing math. And math is no threat to anyone. Except, of course, when misapplied by humans.

OBJECTION 9: MY PROOF IS WRONG

I have, over the previous sections of this reply, stood up for my proof's central claim: that computer AI cannot perform the causal thinking necessary to do narrative. But to tie off the loose ends, let me answer the formal disproof.

The disproof proceeds by breaking my proof's final claim into four sub claims, of which two are contested. The first contested sub claim is adjudged "false" because it does not account for quantum computers—and then, because quantum computers remain in their infancy, is conceded to be "largely true." It is in fact entirely true. Quantum computers are simply another way of building a Turing machine, so while they may prove more effective than existing computers at solving certain computable problems, they cannot solve the non-computable problems that (as explained in

Objection 7) narrative can. Which is to say: quantum computing is still computing; it is not quantum narrating.

The second sub claim is rejected because I do not account for Pearl's do-calculus. As I explain in Objection 4, however, Pearl's do-calculus falls firmly within the scope of my proof.

The rest of my proof is allowed to stand. So, with those two sub claims restored, the proof continues to establish that computer AI will never read or write novels.

WHY IT MATTERS THAT COMPUTER AI WILL NEVER READ OR WRITE NARRATIVE

To declare that AI will never do something can seem close-mindedly foolish, like shutting the patent office because everything useful has already been invented. And none of us want to go down in history as that nitwit. So, it's not only more intellectually progressive to err on the side of *it could happen*; it's safer, too.

For that reason, readers of this debate might feel inclined to hedge their bets by holding open the possibility of a novel-writing, strategy-planning, human-eliminating computer. Such hedging would never be countenanced by a computer itself; a computer would instantly acknowledge *A is not A* as an unsolvable contradiction. But such hedging is a natural result of our brain's nonlogical, narrative machinery, which has helped us survive in both the physical world and the social by counterbalancing *is not* with *could be*.

Let me therefore speak directly to that human machinery by emphasizing that, while it may feel safer to hedge, the act of sanctioning further AI research into narrative carries its own dangers.

The first is that it harms computer AI by setting it up to fail. Computer AI has already had its reputation dinged by overpromising (for example, IBM's Watson Health and Tesla's Autopilot). Which is a shame, because AI is not at fault for its misapplications. It should not be blamed for failing to replace doctors, drive cars, or process natural language, any more than a hammer should be blamed for not cleanly sawing through a piece of wood. All tools—including AI and narrative—have their operational limits, and respecting those limits is respecting the tool.

The second danger is that it inhibits genuine innovation. To be seduced into believing that computer AI is a master intelligence that can not only play chess but create novels, have conversations with humans, practice science, devise new medicines, imagine future AI, and plan a utopian future is to overlook the need for researching new mechanisms of machine thinking (like the one outlined in my original article) that could be used to solve problems that computers cannot. It's like pouring all our money into antibiotics research while neglecting cancer drugs and antivirals.

The third danger is that it diverts resources from researching narrative. Narrative is an enormously underappreciated tool for problem-solving. It's regarded by the general public and most scholars simply as a means of communication, when in fact, it's also what our brain uses to develop plans, strategies, and technologies for coping with uncertain and unstable environments. The environments, that is, of most places

outside a computer simulation. And the more that we understand about the biological machinery of narrative, the more that we—and future generations—can prosper in such places.

The fourth and final danger is that it distracts us from the vast possibilities of human intelligence. Human intelligence remains our most powerful, and most neglected, resource. The world is full of human beings who lack the opportunity and the support to grow to their full intellectual potential. And we can give them more opportunity and more support by handing them a scientifically-backed narrative technology: literature. Literature is a cheap and infinitely renewable resource filled with story mechanisms for boosting curiosity, problem-solving, scientific thinking, empathy, emotional resilience, and other contributors to human intelligence (Fletcher, *Wonderworks*). So, instead of wasting time and money trying to get computers to do what they can't, why don't we invest in what people can? Why don't we quit the impossible dream of a novel-writing computer? And why don't we focus on empowering the human storytellers who will author our actual future?

Endnote

I would like to thank Rob Cadwallader and Erik Larson for many stimulating conversations that helped shape this reply.

Sources

- Fletcher, Angus. *Storythinking*. New York: Columbia Univ. Press, 2022.
- . *Wonderworks* (New York: Simon & Schuster, 2021).
- Good, I. J. “Logic of Man and Machine,” *New Scientist*, (1965): 182–3.
- Larson, Erik. *The Myth of Artificial Intelligence*. Cambridge: Harvard Univ. Press, 2021.
- Jockers, Matthew. “Syuznet,” 2 Feb 2015. <http://www.matthewjockers.net/2015/02/02/syuzhet>. (accessed 19 Oct 2021)
- Jordan, Michael I. “Artificial Intelligence: The Revolution That Hasn't Happened Yet.” *Harvard Data Science Review* 1.1 (2019) <https://hdsr.mitpress.mit.edu/pub/wot7mkc1/release/9> (accessed 19 Oct 2021).
- Phelan, James. *Somebody Telling Somebody Else: A Rhetorical Poetics of Narrative*. Columbus: The Ohio State Univ. Press, 2017.
- Turing, Alan. “On Computable Numbers, with an Application to the *Entscheidungsproblem*.” *Proceedings of the London Mathematical Society, Series 2* 42 (1936–7): 230–65; corrections, 43 (1937): 544–6.