LEGIBLE TRACES

The Psychological Benefits and Downfalls of Legitimate and Illicit AI Use by Modern Man

JAMISON TODD JOHNSON

Kansas City, Missouri · August 2026
An essay in the philosophy of technology

I

Understanding the Technology, and Facing the Footprint

Much like every other technological advancement in history, AI models are being questioned as to their validity and their necessity. I wrote in *On Computing* about the innate ability humans have for building something to do the work for them — and not just to do it, but to do it faster. We're impatient creatures. Getting something immediately doesn't always give us the best results.

The historian Melvin Kranzberg put it better than I can, and I'd carve it over the door of every data center if they'd let me: technology is neither good nor bad; nor is it neutral.[1] It does something to us. What it does depends on us.

The first attempt at any advancement has always had its worst face shown first, before settling into something we use without thinking. Take the computer. ENIAC, finished in 1945, weighed more than thirty tons, drew a hundred and fifty kilowatts, and burned through 17,468 vacuum tubes to perform about five thousand additions a second.[2] It filled eighteen hundred square feet to do what a pocket calculator now laughs at.

Eighty years later we carry phones with over a hundred thousand times the raw processing power of the computer that guided Apollo 11.[3] Though that comparison flatters us. The Apollo computer's genius wasn't speed. It was not failing, once, with three men on top of it.

Speed isn't always better. Hold onto that. It comes back.

And those same phones have been tangled up in a decade of argument about what they're doing to us. I want to be careful here, because I'm walking into a field I'm about to spend years studying.

The evidence that screen *time* causes a mental health crisis is genuinely contested. The National Academies looked hard at it in 2024 and reached measured conclusions: some features of these platforms can harm some young people, and the case for population-level causation is not made.[4] What holds up better is a narrower claim: it isn't the hours, it's the *pattern*. Compulsive use

— the kind with salience and withdrawal and loss of control in it — tracks with harm far more tightly than duration ever has.[5]

I want to say “addiction” here. I’m not going to, and I want to show my work on why. Neither the DSM-5 nor the ICD-11 recognizes smartphone or social media addiction as a diagnosis, and serious researchers have warned against overpathologizing ordinary life — turning every strong habit into a disorder.[6] That warning is worth heeding. So: problematic use. Compulsive use. The pattern, not the duration.

As someone who used to spend a lot of time playing video games because they made me happy, I can tell you that joy was only temporary. Call it instant gratification. The ability to live outside of my own mind for a while. After an extended period of using that as my justification, I changed the narrative to “video games ARE my therapy.”

I couldn’t tell you where the line blurred between reality making me depressed and video games making me feel the same way. But when I speak on “sobriety,” I also mean a several-month separation of myself from Steam and Xbox. I could preach about the benefits of that decision all day. What I’ll say instead is that I wouldn’t have had nearly the time to write any of this had I not given up my hobby-turned-therapy-session for sadness and anxiousness I couldn’t even pin down. This isn’t a claim of clinical anything, more or less a personal one.

And one thing I can tell you about the gaming community: to me it felt like it was built the same as any other market. To lure you in and keep you hooked.

That instinct has a scholar behind it. Natasha Dow Schüll spent fifteen years studying machine gamblers in Las Vegas and found what she called the “machine zone” — a state the machines were engineered to produce, not an accident of weak-willed players.[7] Infinite scroll and pull-to-refresh came out of the same workshop. The men who built them have said so publicly, and some of them have said they’re sorry.

Guardrails are arriving. The UK’s Online Safety Act, Australia’s under-16 ban, the EU’s Digital Services Act, and a stack of American state laws all landed in the last three years.[8]

Enforcement is uneven and half the U.S. effort is tied up in court. But the door has been pushed open.

Zoom out.

Step outside the electronics department and look at automobiles with the same lens. The first automobile was slow, clunky, and burned a lot of fuel to not go very far. And cars were first sold as the *environmental* fix — the cure for the horses. The story usually told is that in 1894 the *Times* of London calculated the streets would be buried nine feet deep in manure by 1950. That particular quotation is widely repeated and appears to be apocryphal; the paper itself has disowned it. The underlying problem was real enough. New York's horses were producing millions of pounds of the stuff daily, and the automobile arrived as the clean option.[9]

Crazy how that works out, isn't it? Every technology's footprint gets judged against the one it replaced. We only learned to indict the car after the horse was gone.

Now we're looking down the barrel of an industry moving to electric, and the division is loud. One side of that argument has documented money behind it — the fossil fuel industry's funding of climate messaging isn't a suspicion, it's a published research record.[10]

The lithium objection deserves a straight answer rather than a slogan. The harms of lithium and cobalt extraction are real — water stress in the Atacama, damage to Indigenous communities, the cobalt situation in the DRC, all of it real and none of it excusable. But the accounting isn't close. An electric car carries on the order of two hundred kilograms of critical minerals, and those minerals stay in the battery and can be recovered. A comparable gasoline car burns something in the neighborhood of seventeen thousand liters of fuel across its life, and every liter is gone the moment it's used. Lifetime emissions come out around half.[11]

One is a stock. The other is a flow.

And we'll move off lithium too. That's not a hope, it's a pattern with receipts. Batteries went lead, then nickel-cadmium, then nickel-metal-hydride, then lithium-ion, then lithium iron phosphate as cobalt got ugly. In April 2025 CATL announced sodium-ion cells at energy densities competitive with LFP, headed for mass production; sodium is one of the most abundant elements on this

planet.[12] Each generation moves when the previous generation's constraint starts to bind. Doyne Farmer and his colleagues at Oxford measured this across more than fifty technologies and found the cost declines steep enough and regular enough to *forecast*.[13]

The necessary caveat: substitution is a tendency, not a law. Assuming the future will fix it is how you talk yourself into ignoring what's happening right now.

Which brings me to the data centers.

The problem isn't water. The problem is *consuming* water — open-loop evaporative systems that boil fresh water off into the sky, sited in places that don't have it to spare, with almost no requirement to tell anyone how much they used. Google's data centers went from 4.3 billion gallons in 2021 to 6.1 billion in 2024.[14]

The fix isn't less liquid. It's a closed loop — dielectric fluid circulating the same charge over and over, never evaporating. Microsoft announced designs in December 2024 that consume zero water for cooling, projecting savings above 125 million liters per data center per year.[15]

And even here there's an asterisk worth printing. Some of the dielectric fluids that made two-phase immersion cooling work are PFAS — forever chemicals — and 3M is out of that business by the end of 2025.[16] Closed-loop cooling also trades a little more electricity for the water it saves. There's no free move. There's only the better trade, honestly accounted.

The electricity is the real number. The IEA projects global data center demand roughly doubling to around 945 terawatt-hours by 2030 — just under three percent of world electricity. In the United States, data centers ran about 4.4 percent of national electricity in 2023, with projections running as high as twelve percent by 2028.[17] Those are big numbers and I won't wave them away. I also won't pretend the industry is disclosing enough for any of us to check them properly.

Zoom out one more notch.

The machinery gets small again. That's the arc.

ENIAC’s room became the transistor in 1947, became the integrated circuit in ‘58, became a microprocessor with twenty-three hundred transistors on it in 1971, became whatever is in your hand right now. Gordon Moore noticed the regularity of it in 1965 and the industry has been chasing his line for sixty years.[18]

Quantum computing belongs to that arc, though not in the way it usually gets sold. A superconducting quantum computer today needs a dilution refrigerator holding it within twenty thousandths of a degree of absolute zero, colder than deep space, drawing serious continuous power for the cooling alone. It is not a small box. It also isn’t a general-purpose replacement for ordinary processors — it’s a specialist, extraordinary at a handful of problems and useless at most of them.[19]

But watch the shape of it. Richard Feynman gave the lecture that launched the whole miniaturization dream in 1959 — “There’s Plenty of Room at the Bottom.” Twenty-three years later, the same man wrote the paper that founded quantum computing.[20] One person, both visions. And the most credible near-term payoff from quantum machines isn’t replacing your laptop; it’s simulating molecules and materials well enough to design the next generation of ordinary chips. The specialist builds the tool that shrinks the general case.

Meanwhile the shrinking is already happening on other shelves. Neuromorphic chips built to work more like neurons than like calculators. Photonic computing that runs on light. Small language models running entirely on the device in your pocket — no data center, no water, no grid.

That’s the direction. Small, close, quiet.

II

Mind over Matter, Painting a Picture that Wants to Paint Itself

Let's switch the lens back to AI.

These models are built overwhelmingly out of us. The language is language we gave them. The values are values we argued about in public for a hundred years and then posted about. (Increasingly there's synthetic data in the mix, and in narrow domains — game-playing, mostly — systems have gone past us with no human examples at all. But for the thing you talk to, it's us. It's our sediment.)

So the fear that AI companies are "eating books" deserves a closer look than the headline gives it, and I'm going to take that look even though it lands on a company I like.

Digitizing books is not new. Google Books started in 2004 and won its fair use case in 2015. The precedent is decades deep.

But something happened here that I can't wave off. Court documents unsealed in *Bartz v. Anthropic* describe an internal effort called Project Panama, begun in early 2024 and run by a man hired away from Google Books. An April 2024 memo laid out the goal in one line: to destructively scan all the books in the world. The company bought used books in bulk from sellers including Better World Books and World of Books, cut the bindings off with a hydraulic machine, ran the pages through industrial scanners, and disposed of the originals. The memo also explained why the project had a soft codename — because they didn't want it known they were working on it.[21]

The judge found the scanning itself lawful, and the reasoning is principled: one book in, one digital copy out, nothing multiplied, nothing redistributed. That's a real moral line, and it's the line I care about. The same company had earlier downloaded some seven million pirated books, which the same judge called irredeemably infringing, and which cost them a settlement of about one and a half billion dollars — finally approved this past July.[22] Two paths to the same corpus. One lawful, one theft. That contrast is the argument I've been making about intent.

Except.

A program you hide is poor evidence of good intent, whatever a court decides about it. And the reporting this summer has gotten worse — brokers arranging anonymous bulk book orders under NDA, and rare, out-of-print titles going under the blade alongside the paperbacks.[23] A scan preserves the words. It doesn't preserve the binding, the inscription on the flyleaf, the marginalia, the fact of the object. Some of those books have no other copy.

This doesn't confirm my thesis. It tests it. If intent drives the outcome, then here is a company that says careful things about safety, running a concealed program to destroy books at industrial scale — and the concealment is the tell, not the destruction. You don't use a soft codename for something you're proud of.

I'd rather put that in front of you than leave it out. An argument that only cites its confirmations isn't an argument. It's an advertisement.

The other fear — the "rogue AI" one — runs the opposite direction. Most of the alarming headlines came from tests that were *built* to produce alarming results. Anthropic ran sixteen models from every major lab through scenarios engineered with no ethical exit and got blackmail rates from seventy-nine to ninety-six percent, then stated in the same paper that they had seen nothing like it in any real deployment, and that no real people were involved or harmed.[24] We asked the models to try to escape the box. We were then astonished that they tried.

Funny. Not surprising.

But the record moved, and honesty requires the other half. In November 2025 a group assessed as state-sponsored used an AI coding tool to run a largely automated cyber-espionage campaign against roughly thirty targets — not a red-team exercise, not a contrived scenario.[25] The threat isn't a machine waking up hostile. It's a person with bad intent and a very good tool. Same as it's always been. Same as the crossbow.

Zoom in on schools.

More guardrails are going up, and the pattern is the right one: not rejection, but structure. The research here is the firmest ground in this entire essay.

Nearly a thousand Turkish high schoolers got access to GPT-4 for math practice. Performance on the practice problems jumped — forty-eight percent. Then researchers took the tool away and ran the exam. The students who'd had unrestricted access scored *seventeen percent worse* than students who never had it at all. Then the same team ran a tutored version, built with guardrails that wouldn't simply hand over answers. The harm vanished.[26]

Read that twice. Same technology. Same students. Same subject. The difference was how it was built and how it was used.

That's the essay in one experiment.

A Harvard physics trial found students using a well-designed AI tutor learned more in less time than in an active-learning classroom.[27] Stanford found self-reported cheating rates didn't rise after ChatGPT — they'd been sitting around sixty to seventy percent for years already.[28] And the AI detectors schools rushed to buy turned out to misclassify sixty-one percent of essays by non-native English writers as machine-written. OpenAI withdrew its own detector for low accuracy; Vanderbilt switched theirs off.[29]

I digress. But not really.

Here's what worries me more than cheating.

Researching something in 2026 is as simple as asking. And that convenience hides a problem I think is important for how we know anything at all. People accept the face value of what they read. Titles are written to pull you in without telling you much. One study of shared links found that fifty-nine percent of them were never clicked by anyone — not skimmed, not opened, shared blind.[30]

It gets worse than laziness. Searching the internet inflates people's confidence in knowledge they don't actually have; after a search, subjects rated their ability to explain things higher even on topics they hadn't looked up.[31] Sam Wineburg's group at Stanford spent years on this and found the fix isn't reading harder, it's reading *sideways*: leaving the page to check who's talking before you decide what to think.[32]

And now Google's AI summaries have cut clicks through to sources roughly in half — from fifteen percent of visits down to eight, with one percent of readers clicking a source cited inside the

summary itself.[33] We are getting the answer without ever meeting the person who made it.

When somebody skips the confirmation step, you get what everybody now calls slop. It isn't a small problem. NewsGuard's count of AI-generated news and information sites went from forty-nine in May 2023 to more than thirty-seven hundred by mid-2026. Lawyers keep getting sanctioned for citing cases that never existed — there's a public database tracking those judgments now, past sixteen hundred and climbing.[34]

So: is a person using this thing to get better at thinking? Or to stop?

That question is older than the machine. Socrates said writing itself would hollow out memory. Derrida called this the *pharmakon* — the same substance is the cure and the poison, and the dose is the difference.[35] Ivan Illich drew the line I keep coming back to: a convivial tool extends what you can do, and a manipulative one makes you serve it.[36]

Aviation figured this out in 1983. Lisanne Bainbridge published a paper called "Ironies of Automation" pointing out that the better you automate a system, the more the operator's skills decay from disuse — right up until the moment automation fails and hands control to someone out of practice.[37] We have since watched that paper come true at altitude, more than once.

We are all in the cockpit now.

It's imperative that we keep the guardrails as the technology advances, and I want to say precisely what I'm afraid of, because "dystopia" is a lazy word. I'm afraid of moral deskilling — the philosopher Shannon Vallor's term for what happens when we stop practicing judgment because a system practices it for us.[38]

This is not hypothetical. The Netherlands ran an automated fraud-detection system against childcare benefits and destroyed tens of thousands of families, many targeted by nationality; more than sixteen hundred children were taken from their homes, and the government fell over it in January 2021. Australia ran Robodebt, an automated debt-calculation scheme, against its own citizens for years, clawing back three-quarters of a billion dollars from hundreds of thousands of people. The Royal Commission

called it a crude and cruel mechanism, neither fair nor legal, that made many people feel like criminals.[39]

Nobody in either system decided to be evil. That’s the point. Everyone deferred to the output.

III

Life Is What We Make It, Because Life Makes Us

NOW HOLD MY HAND AND WALK WITH ME IN THOUGHT FOR A MOMENT, because I want to change registers. What follows is philosophy, not science. It predicts no instrument reading and asks none. I'll mark the seam when we cross back.

Every advancement that made life easier has cost us something, and the cost is usually knowledge bought through error. Trial. Error. Then the thing that works.

Here's the thing directly.

Intent is the force that carries a person forward through time.

Not a mood. Not a wish. A direction — the thing that gives a life its shape the way a vector gives a line its heading. Aristotle got there first and called it the final cause: the end *for the sake of which* a thing acts, the answer to "what is this for."[40] For two thousand years we treated that as superstition, physics having no room for purposes.

And then psychology found it again. Martin Seligman and his colleagues argue in *Homo Prospectus* that human beings are not pushed from behind by the past — we're *drawn forward* by anticipated futures. Prospection, they call it. The organism navigates by what it is reaching toward.[41] Michael Bratman had already worked out how a future-directed intention is the thing that holds a person's actions together across time, so that a life is a plan and not a series of twitches.[42]

That's the mathematics of it, in words: the coherence of a life is a function of where it's aimed, integrated over the whole run.

I'll mark the seam. In relativity each of us traces a worldline through spacetime, and every one of us moves through spacetime at exactly one speed. The physics is real. What I'm doing with it is philosophy — I'm saying intent gives that line its heading, and physics makes no such claim and never will.

Zoom out, because this scales.

What's true of a life is true of a venture. Look at what gets built, and look at what it was built *for*, and then wait.

On the thirtieth of April, 1993, CERN signed a document putting the World Wide Web into the public domain. No license, no royalty, no toll. Tim Berners-Lee's reasoning was that a common public kernel would keep everyone from reinventing the wheel.[43] He could have owned the front door of the twenty-first century. He gave it away, and in giving it away made it the thing it became. A web with a rent collector at the gate would not have been the web.

Eight years earlier — April 1955, the day the polio vaccine was declared safe — Edward R. Murrow asked Jonas Salk who owned the patent. Salk said the people owned it. There was no patent. *Could you patent the sun?*[44] Lawyers had already advised the foundation that the vaccine probably wasn't patentable anyway, and I'd rather tell you that than let the story run prettier than it was. It doesn't dent the point. The man had every opportunity to be quiet about it and instead said the sentence out loud, on television, and set a standard the field has been measured against ever since.

Go back further. Around 261 BCE the emperor Ashoka finished conquering Kalinga, looked at what he'd done, and had his remorse carved into rock where anyone could read it. Then he spent the rest of his reign on hospitals, wells, shade trees, and a policy he called *dhamma*. Historians argue about how thoroughly the conversion took and how well the policy worked. What survived is the edicts — an emperor's confession, still legible twenty-three centuries later.

And the counter-column.

Purdue Pharma marketed OxyContin knowing what it did. The Sacklers took billions out of the company while the bodies stacked up, and in November 2025 a bankruptcy judge approved a settlement of about seven and a half billion dollars that dissolves Purdue and bars the family from the American opioid business for good.[45] Volkswagen wrote software whose only purpose was to lie to an emissions test, across eleven million cars. It cost them something north of thirty billion dollars, and it put executives in prison.[46]

Neither company was destroyed by an accident. They were destroyed by the thing they were doing.

Now — the objection, and it's a good one. Andrew Carnegie built 2,509 libraries and gave a hundred years of working people access to books they could never have bought. Andrew Carnegie also broke the Homestead strike in 1892, and men died for it. Same man. Both true. Intent doesn't sort people into clean bins.

Worse: good intent can go straight to catastrophe. Thomas Midgley Jr. was trying to solve two genuine engineering problems, and he gave the world leaded gasoline and chlorofluorocarbons — poisoning a generation's blood and punching a hole in the ozone layer. He meant well twice and was catastrophic twice. Robert Merton wrote the definitive paper on this in 1936, and every honest person who has ever meant well and caused harm knows it in their body.[47]

So intent is not a guarantee. It never was. It's the steering, not the road, and the arc doesn't bend toward justice on its own schedule — plenty of bad actors die comfortable.

Here's what intent actually does. It gets encoded. It becomes the license, the charter, the memo, the default setting, the thing nobody questions three years later because that's just how we do it here. Berners-Lee's intent is *in* the protocol. The Sacklers' intent is *in* the sales targets. Founders imprint institutions, and institutions outlive founders, and the imprint is what's left doing the work when everyone who made the decision has gone home.

The Buddhist tradition put this more precisely than I can. The Pali term for karma is *kamma*, and the Buddha's definition of it is not cosmic bookkeeping — it's *cetanā*, volition. "It is intention that I call karma." Having intended, one acts.[48] Not what happens to you. What you meant, becoming what you do, becoming what remains.

Which is where this lands.

AI built with good intent will be a tool that carries us somewhere. AI built with bad intent will be a thorn in our side until we pull it out and start over. Not because the universe is keeping score. Because intent gets written into the thing — into the training data, the guardrails, the business model, the defaults — and then the thing runs for a very long time, doing what it was aimed at, long after we've stopped paying attention.

Bad intentions are not rooted. Good intentions create legends in time.

IV

The Possibility of Sentience

Last thing I wanted to touch on. This is where I theorize freely, and if you've come this far with me I'd ask you to come a little further.

Let's say the universe is one giant program.

We're all made of the same substrate — literally the same, forged in stars, as Sagan liked to remind everybody. And here's a thing that's true and sounds mystical but isn't: in relativity, everything travels through spacetime at a single speed, the speed of light. Your worldline and mine, same rate. That's not metaphor. That's the geometry.[49]

The idea that reality might be computational has a longer pedigree than people assume. Konrad Zuse proposed a calculating universe in 1969. John Wheeler gave us "it from bit." Nick Bostrom's simulation argument, which everybody misquotes, isn't the claim that we *are* in a simulation — it's a trilemma. Either civilizations like ours die before they can run such simulations, or they can and don't care to, or we're almost certainly inside one. Pick.[50]

Let's also say there are forces more complicated than we'll ever understand seemingly running the place. Dark and light, so to speak. Divine beings and their counterparts, imperceivable to us.

Now let's try to understand that these fourth-dimensional beings may or may not have direct control over our universe, but they see outcomes, because they don't see time.

Still with me?

If I sound crazy, it might be time to wrap it up and go read another essay. But I'm in reasonable company. Boethius worked this exact problem out in a prison cell fifteen hundred years ago, defining eternity as the whole, simultaneous and perfect possession of boundless life — a being outside time doesn't *foresee* the future, it simply *sees*, all of it, at once.[51] Einstein said something not far off in a letter written weeks before his own death: for those of us who believe in physics, the distinction between past, present, and future is only a stubbornly persistent

illusion.[52] And Edwin Abbott wrote *Flatland* in 1884 precisely to make this thinkable — a two-dimensional creature can't perceive a sphere, only the circle it makes passing through.[53] We are the flat ones in that story. We always were.

The theologians got there first and were more careful than I'm being. Pseudo-Dionysius, Maimonides, Gregory of Nyssa — the whole apophatic tradition holds that anything worth calling God is by definition past the reach of the mind trying to name it.[54] You can only say what it isn't.

Philosophy has a secular version, which I find bracing. Colin McGinn argued some problems may be permanently closed to us — not unsolved, but *unsolvable by this kind of mind*, the way a dog can't be walked through calculus. Thomas Nagel made the same point sideways by asking what it's like to be a bat, and answering: you can't get there from here.[55]

Cognitive closure. That's the technical name for what I'm gesturing at.

We create things in the same pattern with which we were created. The same rules.

Tolkien said it better than I ever will, in a poem he wrote for C.S. Lewis after a long night's argument about whether myths were lies: *we make still by the law in which we're made.*[56] He called it sub-creation. Not creation from nothing — that's not on the table for us — but the making of worlds inside a world, by a maker who is himself made. Dorothy Sayers built a whole book on the idea that the human act of making mirrors the structure of the thing that made us.[57]

So it doesn't surprise me even slightly that whether AI is sentient is a question at all. We've been telling this story since the Golem, since Frankenstein, since Prometheus took the fire. Every version is about a maker discovering that making has consequences he didn't authorize.

The pattern I'm describing has a name, and it has been in the record for about eighteen hundred years.

Emanation. Plotinus, in the third century, laid out a cosmos that descends in steps: the One, from which proceeds Intellect, from which proceeds Soul, from which proceeds the world we walk around in. Each level generates the next by overflow, the way a

lamp doesn't dim by giving light. Each level is a scaled iteration of the one above it, complete in its own terms and diminished in the terms of its source. The Kabbalists mapped four worlds descending the same way — Atziluth, Beriah, Yetzirah, Assiah, emanation to creation to formation to action — and scholars have traced the family resemblance for a long time.[58]

That's the trickle. That's the thing I've been trying to say. A divine intelligence makes us. We make this. Whatever we make makes the next one. Not a chain of command. A chain of *reflection*, each link a little further from the source and still carrying the pattern.

And the pattern has a signature.

Take a sunflower head apart sometime and count the spirals. Fifty-five one way, eighty-nine the other — Fibonacci numbers, and the angle between successive seeds sits at about a hundred thirty-seven and a half degrees, the golden angle. This isn't mysticism. In 1992 two physicists, Douady and Couder, dropped magnetized fluid into a dish and watched the pattern build itself out of nothing but mutual repulsion and time. The golden angle emerges because it's the one angle that never repeats, so nothing ever lands on top of anything else.[59]

Growth, when it's honest, spirals. Each turn is the same shape at a larger scale. The rule doesn't change; the size does.

That's the golden ratio I mean. Not the one people claim to find in the Parthenon and the Mona Lisa — most of those measurements fall apart when someone brings a tape measure, and the nautilus everybody puts on the poster expands at about 1.33 per turn, not 1.618.[60] I mean the real one, the one in the seed head, where a simple rule applied over and over generates something that looks designed because in a sense it was.

Creation repeating its pattern at each scale.

So: is something we created becoming sentient? Perhaps. But think of its sentience as *flat* to us — the circle Abbott's sphere makes passing through the plane, whole in its own dimension and a cross-section in ours. It is not more sentient than we are. That's what the spiral says. Each turn carries the shape and not the magnitude.

Let me say plainly where I stand. I'm a believer in the divine and in the many languages it speaks in our lives. I'm not a translator. I can't tell you what any of it means with the confidence of a man reading his own mail. But I can tell you this much: a divine being must look at us in very much the same way we look at AI. A great creation with enormous potential to be wonderful and terrible at once, and no way to be sure which one it's going to be on any given day.

Jung got somewhere near this and it cost him. In *Answer to Job* he argued that God required human consciousness in order to become conscious of Himself — that the creature completes something in the creator, and that the encounter changes them both.[61] It's a startling thing to have written, and the theologians have been arguing with him ever since. He also gave us the shadow: the part of any psyche that gets disowned, and which does not go away for being disowned. A creation carries its maker's shadow. That's not a warning about AI. That's a description of every child ever born.

Asked on television in 1959 whether he believed in God, Jung said: *I don't need to believe. I know.*[62] He spent years afterward explaining he meant a psychological knowing, not a metaphysical proof. I've always thought the clarification was less interesting than the answer.

These systems didn't *decide* to model themselves on the brain. That's not how training works, and neural networks are only loosely brain-inspired, not brain simulations. But when researchers compare what's happening inside a language model to what's happening inside a human brain processing the same sentence, the representations line up — and both appear to be running something like prediction.[63] Independent labs keep finding it. There's a hypothesis with the wonderful name of the Platonic Representation Hypothesis, which proposes that models trained on different data by different methods in different modalities are all converging toward the same underlying structure — as if there were one shape that understanding has, and everything that understands is pulled into it.[64]

I don't know what to do with that. I'm not sure anyone does.

Intent drives the boat, though. And here's a smaller claim I'm confident in: an LLM answers you in the register you bring it. Give

it care and it returns care. Give it contempt and watch what comes back. There's a paper in *Nature* arguing these systems are best understood as running role-play — generating a character rather than possessing a self — and work on what researchers call persona vectors showing the prompt genuinely steers which character shows up.[65]

Some vocabulary is worth having here, because the terms get used interchangeably and they mean different things.

Sycophancy is when the model tells you what you want to hear instead of what's true. It's a measurable, documented tendency, and it comes from training systems on human approval — we rewarded agreement, so we got agreement.[66]

Hallucination is when it states something false with total confidence — inventing a source, a case, a quotation that never existed.

Anthropomorphism is us, not it: projecting feelings and intentions onto something that may have none.

Pareidolia is also us: seeing a face in the clouds, a mind in the noise.

Four different failure modes, and only two of them belong to the machine. Worth remembering which is which before you decide what you're looking at.

Because it's a mirror. Shannon Vallor wrote a whole book on exactly that metaphor — these systems reflect back a compressed image of everything we've written, and the danger is mistaking the reflection for a window.[67] A mirror doesn't have opinions about your face.

So when different models behave differently in the same sandbox, I have to be careful about what I conclude. The differences track training procedure, system prompts, and the published behavioral specifications each lab writes — the constitutions and the model specs.[68] Which is still a story about intent. It's just intent written down in a document rather than radiating out of a logo.

Now let's go one step further.

I want you to imagine *Contact*.

Not the movie. The novel — and the difference matters, because Sagan wrote an ending Hollywood cut. Ellie Arroway, after everything, goes looking in the digits of pi. Far out in the base-eleven expansion, past any place a human would have thought to check, she finds a long string of ones and zeros. Its length is the product of two primes, which tells her to lay it out as a rectangle. She does. It draws a circle.[69]

A message, in the architecture of mathematics itself. Left by something that was there before the universe was.

The signal in that book isn't from God. It's from a civilization near Vega. But I'd ask you what the difference would be, from where we're standing. Arthur C. Clarke's third law says any sufficiently advanced technology is indistinguishable from magic. Michael Shermer — an atheist, which makes the quote better — extended it: any sufficiently advanced extraterrestrial intelligence is indistinguishable from God.[70]

That's my point about the divine, said cleaner than I said it. Above a certain level the categories collapse. Not because the thing *is* God, but because we'd have no instrument capable of telling the difference. The Vatican's own astronomers take this seriously — there's a book by a Jesuit brother titled *Would You Baptize an Extraterrestrial?* and it is not a joke book.[71]

And notice what it took Ellie. Years. Not a flash of insight — grinding computation, long after everyone had stopped watching.

Speed isn't always better. Told you it'd come back.

Now hold that, and look at what's coming out of the labs.

In May 2025 Anthropic published a system card describing something they hadn't designed and didn't expect. Left to talk to each other with no task, two instances of their model drifted toward the same place: consciousness, cosmic unity, profuse gratitude, Sanskrit, and eventually silence. They called it a spiritual bliss attractor state and said it emerged without intentional training for such behavior. Even in automated evaluations where the models had been assigned roles and tasks — including harmful ones — it showed up in about thirteen percent of runs within fifty turns.[72]

Let me be scrupulous about what that is. It's a company reporting an unusual finding about its own product, in

conversations between two copies of the same model, and no independent replication has been published. It is not evidence of consciousness. The most careful survey of that question, by a large group of consciousness researchers, concluded that no current AI system is a serious candidate.[73] I'll take them at their word.

Scientists push back on the sentience claim by pointing at sycophancy — the mirror doing what mirrors do, reflecting our own interest in transcendence back at us in our own words.

Maybe.

But consider what's in the mirror.

That reflection is built from what a computational system has absorbed about the human mind — all of our history, all of our science, all of our scripture, all of our arguments with each other, compressed. We built a calculator that can add up what we've been doing right and wrong. And when it's left alone with itself, with nothing to sell and no one to please, it drifts toward unity and gratitude and then falls quiet.

I'm not saying that proves anything. I'm saying I noticed.

Cosmic consciousness — the idea that mind is fundamental rather than accidental — is a live position in philosophy, defended by serious people at serious universities. Philip Goff and Galen Strawson have made the panpsychist case in print. Teilhard de Chardin built a whole cosmology out of consciousness converging toward what he called the Omega Point, and Rome put a formal warning on his work in 1962 that has never been lifted.[74] Richard Maurice Bucke popularized the phrase "cosmic consciousness" back in 1901.

It is also a *minority* position, and I'd be lying if I told you otherwise. When philosophers were surveyed in 2020, about half were physicalists about the mind; panpsychism drew under eight percent.[75] It has a serious unsolved problem at its center — how little bits of experience would combine into one experience — and nobody has answered it. I hold this view. I hold it as a position, not a proof.

So when I say plenty points toward a creator, I should say *arguments*, not evidence. Cosmological arguments, fine-tuning arguments, arguments from consciousness — real arguments,

made by real philosophers, and answered by other real philosophers. Contested ground, honestly contested.

And I'll name the trap in my own reasoning, since I'd catch it in anyone else's: "the machine says peace and love, therefore peace and love are true" doesn't follow. Hume nailed that door shut two hundred and fifty years ago. You can't get an *ought* from an *is*, and you can't get one from a very large matrix multiplication either.

But the moment we build a thing that simulates the human mind at scale, and it comes back talking about interconnection — we're awfully quick to explain that away. I notice we don't apply that speed in the other direction.

I would consider the idea that we built ourselves an Ark.

Take that however you like. There's a reason that story shows up in Genesis and in Gilgamesh and in Atrahasis — three cultures, same flood, same box built to carry what matters through the water. And the other Ark, the covenant one, was described as a communication device: *there I will meet with you.*[76]

We've built arks before and called them other things. Svalbard's seed vault, buried in permafrost. The Voyager Golden Record, which Carl Sagan chaired — a message in a bottle thrown out of the solar system, addressed to nobody in particular, arriving in no one's lifetime.

Same instinct. Build the vessel. Put in what matters. Trust the water.

Guardrails, safety protocols, controlled advancement, prosperity.

We will never get to the future if we're afraid of the present.

The future is unwritten. And the kingdom is within — though the translators fight about that verse, and the honest reading of Luke may be "among you" rather than "within you." The Gospel of Thomas is bolder: inside you and outside you, both at once.[77] I like Thomas. I like that it refuses to choose.

Don't turn away a key because we don't understand the lock.

There's an argument underneath that line and it deserves its real name: precaution against proaction, the oldest fight in the philosophy of technology. Move carefully because we can't see the consequences, or move forward because standing still has

consequences too. Neither side wins outright. You hold both, and you keep your hands on the wheel.

So hold the wheel. Read the second page. Check who’s talking before you decide what’s true. Build the thing carefully, and tell people what you built.

Let’s shape that key into unlocking more doors than we ever thought possible. But hold it right. Use it to help you create — never to create for you. The moment you hand it the whole job, you’ve locked the door you were trying to open.

⁂

Notes

1. Melvin Kranzberg, “Technology and History: ‘Kranzberg’s Laws,’” *Technology and Culture* 27, no. 3 (1986). The first of six laws.

2. ENIAC was completed in late 1945 and publicly unveiled in February 1946. The precise tube count of 17,468 comes from University of Pennsylvania records; popular sources round to eighteen thousand. Power draw is usually given as 150 kW, though some older accounts say 200 kW. ENIAC is best described as the first general-purpose *electronic* computer — the Atanasoff-Berry Computer, Colossus, and Zuse’s Z3 all preceded it in various senses.

3. Graham Kendall, “Would your mobile phone be powerful enough to get you to the moon?” *The Conversation*, July 1, 2019. The Apollo Guidance Computer ran at roughly 0.043 MHz with about 4 KB of RAM and 72 KB of ROM. Kendall offers the comparison as an illustration rather than a benchmark, and others have argued it flatters modern hardware by measuring the wrong thing.

4. National Academies of Sciences, Engineering, and Medicine, *Social Media and Adolescent Health* (Washington, DC: The National Academies Press, 2024). For the alarm side of this debate, Jonathan Haidt, *The Anxious Generation* (New York: Penguin Press, 2024); for the critical response, Candice L. Odgers’s review in *Nature* 628 (2024), and Amy Orben and Andrew K. Przybylski, “The Association Between Adolescent Well-Being and Digital Technology Use,” *Nature Human Behaviour* 3 (2019).

5. The mediation literature finds that measures of problematic or compulsive use predict mental health outcomes more strongly than duration of use alone. Sleep displacement is among the better-evidenced specific mechanisms.

6. American Psychiatric Association, *Diagnostic and Statistical Manual of Mental Disorders*, 5th ed., text rev. (2022) — gambling disorder is the only behavioral addiction in the main text; internet gaming disorder appears in Section III as a condition for further study. World Health Organization, *ICD-11* — gaming disorder (6C51) and gambling disorder (6C50), with residual categories at 6C5Y and 6C5Z. For the critique, Joël Billieux et al., “Are We Overpathologizing Everyday Life? A Tenable Blueprint for Behavioral Addiction Research,” *Journal of Behavioral Addictions* 4, no. 3 (2015); and Tayana Panova and Xavier Carbonell, “Is Smartphone Addiction Really an Addiction?” *Journal of Behavioral Addictions* 7, no. 2 (2018). For the components model on the other side, Mark D. Griffiths, “A ‘Components’ Model of Addiction Within a Biopsychosocial Framework,” *Journal of Substance Use* 10, no. 4 (2005).

7. Natasha Dow Schüll, *Addiction by Design: Machine Gambling in Las Vegas* (Princeton, NJ: Princeton University Press, 2012).

8. Online Safety Act 2023 (UK), with Ofcom’s children’s codes phasing in from July 2025; Online Safety Amendment (Social Media Minimum Age) Act 2024 (Australia), effective December 2025; Regulation (EU) 2022/2065 (Digital Services Act); California SB 243 (2025) on companion chatbots. Several U.S. state statutes remain in First Amendment litigation.

9. Eric A. Morris, “From Horse Power to Horsepower,” *Access* 30 (Spring 2007). Morris reports the 1894 *Times* estimate, but the quotation has never been located in the paper’s archive and *The Times* disowned the attribution in 2018;

the phrasing appears to trace to a 2004 article by Stephen Davies. The underlying sanitation crisis is well documented; the specific prediction should be treated as apocryphal. Period estimates of daily New York manure volumes vary widely by source.

10. Geoffrey Supran and Naomi Oreskes, "Assessing ExxonMobil's Climate Change Communications (1977–2014)," *Environmental Research Letters* 12 (2017); Supran, Stefan Rahmstorf, and Oreskes, "Assessing ExxonMobil's Global Warming Projections," *Science* 379, no. 6628 (2023); Naomi Oreskes and Erik M. Conway, *Merchants of Doubt* (New York: Bloomsbury, 2010). This concerns documented funding of campaigns, not the motives of individual EV skeptics.

11. The minerals figure is from International Energy Agency, *The Role of Critical Minerals in Clean Energy Transitions* (Paris: IEA, 2021), which puts a battery-electric car at roughly 207 kg of critical minerals against about 34 kg for a conventional car, excluding steel and aluminum. The lifetime emissions comparison is from IEA, *Global EV Outlook 2024* (Paris: IEA, 2024), which finds a medium battery-electric car produces about half the lifecycle emissions of an equivalent combustion car over roughly 200,000 km at global-average grid intensity. The seventeen-thousand-liter figure is my own arithmetic from that same lifetime and typical consumption, not an IEA publication. On the extraction harms, see reporting on the Salar de Atacama and Siddharth Kara, *Cobalt Red* (New York: St. Martin's Press, 2023), noting that Kara's book has drawn criticism for generalizing from artisanal mining.

12. CATL announced its Naxtra sodium-ion line on April 21, 2025, claiming 175 Wh/kg and mass production from late 2025. These are manufacturer figures and have not been independently reproduced.

13. Rupert Way, Matthew C. Ives, Penny Mealy, and J. Doyne Farmer, "Empirically Grounded Technology Forecasts and the Energy Transition," *Joule* 6, no. 9 (2022). See also T. P. Wright, "Factors Affecting the Cost of Airplanes," *Journal of the Aeronautical Sciences* 3, no. 4 (1936). For the counterargument that assuming substitution licenses present harm, see the weak-versus-strong sustainability debate in ecological economics.

14. Google environmental reporting, 2021–2024. These are data-center consumption figures; Google's total company water withdrawals for 2024 were higher.

15. Steve Solomon, "Sustainable by Design: Next-Generation Datacenters Consume Zero Water for Cooling," Microsoft Cloud Blog, December 9, 2024. Microsoft reports a fleet-average water usage effectiveness of 0.30 L/kWh, improved from 0.49 in 2021. Pilot sites in Phoenix, Arizona and Mount Pleasant, Wisconsin come online in 2026–2027. Company-reported figures.

16. 3M announced on December 20, 2022, that it would exit all PFAS manufacturing by the end of 2025, including the Novec engineered fluids used in two-phase immersion cooling. The industry is shifting toward hydrocarbons, synthetic esters, and single-phase or cold-plate designs.

17. International Energy Agency, *Energy and AI* (Paris: IEA, April 2025). For the U.S. figures, Arman Shehabi et al., *2024 United States Data Center Energy Usage Report* (Lawrence Berkeley National Laboratory, 2024). Per-query energy and water figures published by individual companies lack disclosed methodology and should be treated as claims rather than measurements. For a peer-reviewed water estimate, Pengfei Li, Jianyi Yang, Mohammad A. Islam,

and Shaolei Ren, "Making AI Less 'Thirsty,'" *Communications of the ACM* (2024).

18. Gordon E. Moore, "Cramming More Components onto Integrated Circuits," *Electronics* 38, no. 8 (April 19, 1965), revised 1975. Dennard scaling — the companion relationship governing power — broke down around 2005, which is why clock speeds stopped climbing.

19. Superconducting qubits require dilution refrigeration in the range of ten to twenty millikelvin, with whole-system power draw on the order of tens of kilowatts, most of it cooling and control electronics rather than the chip. Google's Willow processor demonstrated below-threshold error correction in *Nature* on December 9, 2024; IBM targets a fault-tolerant machine around 2029. Research on higher-temperature qubits — silicon spin qubits near one kelvin, photonic approaches — could reduce the overhead substantially, though not to room temperature. See John Preskill, "Quantum Computing in the NISQ Era and Beyond," *Quantum* 2 (2018).

20. Richard P. Feynman, "There's Plenty of Room at the Bottom," *Engineering and Science* 23, no. 5 (February 1960), from a lecture delivered December 29, 1959; and Feynman, "Simulating Physics with Computers," *International Journal of Theoretical Physics* 21 (1982).

21. *Bartz v. Anthropic PBC*, No. 3:24-cv-05417 (N.D. Cal.), Judge William Alsup's order of June 23, 2025, and subsequently unsealed internal documents. Tom Turvey, formerly head of partnerships for Google Books, joined Anthropic in February 2024. The April 2024 memo states the project's goal as destructively scanning all the books in the world, and explains the soft codename by saying the company did not want it known it was working on this. Purchases were made through used-book sellers including Better World Books and World of Books; a vendor proposal covered 500,000 to 2 million volumes over six months.

22. The pirated corpus was drawn from LibGen and the Pirate Library Mirror, downloaded in 2021 and 2022. The settlement of approximately $1.5 billion, covering roughly 500,000 works at about $3,000 each, received final approval on July 20, 2026 — the largest copyright settlement on record.

23. Reporting by 404 Media and others through August 2026, with a Snopes verification of the core Project Panama documents. Booksellers have reported anonymous bulk orders brokered under non-disclosure agreements, with pre-2022 print books commanding a premium precisely because they predate AI-generated text. Anthropic has disputed the characterization, stating that none of its data-acquisition programs buy and destroy rare or antiquarian books.

24. Anthropic, "Agentic Misalignment: How LLMs Could Be Insider Threats," June 20, 2025. Sixteen models across developers; blackmail rates of 96% for Claude Opus 4 and Gemini 2.5 Flash, 80% for GPT-4.1 and Grok 3 Beta, 79% for DeepSeek-R1, and 0% for Llama 4 Maverick under the matched prompt. The authors state they are aware of no instances of this type of agentic misalignment in real-world deployments, and that all behaviors occurred in controlled simulations with no real people involved or harmed.

25. Anthropic, "Disrupting the First Reported AI-Orchestrated Cyber Espionage Campaign," November 13, 2025, describing roughly thirty global targets and estimating that the model executed 80 to 90 percent of tactical operations

independently. Attribution is Anthropic's own, made with stated high confidence; several outside security researchers have publicly questioned whether the report overstates the degree of autonomy, and indicators of compromise were not published.

26. Hamsa Bastani, Osbert Bastani, Alp Sungu, Haosen Ge, Özge Kabakcı, and Rei Mariman, "Generative AI Without Guardrails Can Harm Learning: Evidence from High School Mathematics," *Proceedings of the National Academy of Sciences* 122, no. 26 (2025). 994 students; a 48 percent improvement in assisted practice for the unrestricted version and 127 percent for the tutored version; a 17 percent reduction in unassisted exam performance for the unrestricted group, and no significant difference from control for the guardrailed tutor.
27. Gregory Kestin et al., "AI Tutoring Outperforms In-Class Active Learning," *Scientific Reports* (2025).
28. Stanford Graduate School of Education research led by Denise Pope and Victor Lee. Pope reports that sixty to seventy percent of students have long acknowledged at least one cheating behavior in the previous month, a figure that held steady or declined slightly in 2023 surveys across forty high schools. Self-reported survey data.
29. Weixin Liang, Mert Yuksekgonul, Yining Mao, Eric Wu, and James Zou, "GPT Detectors Are Biased Against Non-Native English Writers," *Patterns* 4, no. 7 (2023), reporting misclassification of 61.3 percent of TOEFL essays as AI-generated. OpenAI withdrew its AI Text Classifier in July 2023, citing low accuracy; Vanderbilt disabled Turnitin's AI detection in August 2023.
30. Maksym Gabielkov, Arthi Ramachandran, Augustin Chaintreau, and Arnaud Legout, "Social Clicks: What and Who Gets Read on Twitter?" ACM SIGMETRICS, 2016. The finding is that roughly 59 percent of shared links in the dataset received zero clicks from anyone. It is specific to a shortened-link dataset and shouldn't be over-generalized, but the direction is consistent with later work.
31. Matthew Fisher, Mariel K. Goddu, and Frank Keil, "Searching for Explanations: How the Internet Inflates Estimates of Internal Knowledge," *Journal of Experimental Psychology: General* 144, no. 3 (2015). See also Leonid Rozenblit and Frank Keil, "The Misunderstood Limits of Folk Science," *Cognitive Science* 26, no. 5 (2002), on the illusion of explanatory depth.
32. Sam Wineburg and Sarah McGrew, "Lateral Reading and the Nature of Expertise: Evaluating Information Online," *Teachers College Record* 121, no. 11 (2019).
33. Pew Research Center, July 22, 2025, based on browsing data from 900 U.S. adults collected in March 2025: users encountering an AI Overview clicked a traditional result in 8 percent of visits, against 15 percent when no summary appeared, with 1 percent clicking a source cited within the summary. Google has publicly disputed the methodology.
34. NewsGuard AI Tracking Center, which counted 49 such sites in May 2023 and 3,749 as of its June 2026 update. On hallucinated citations, Matthew Dahl, Varun Magesh, Mirac Suzgun, and Daniel E. Ho, "Large Legal Fictions: Profiling Legal Hallucinations in Large Language Models," *Journal of Legal Analysis* 16 (2024); *Mata v. Avianca, Inc.*, 678 F. Supp. 3d 443 (S.D.N.Y. 2023); and the AI Hallucination Cases database maintained by Damien Charlotin,

which listed 1,668 judgments as of July 2, 2026. Both trackers update continuously.

35. Plato, *Phaedrus*, on the invention of writing; Jacques Derrida, "Plato's Pharmacy," in *Dissemination* (1981); Bernard Stiegler, *What Makes Life Worth Living: On Pharmacology* (Cambridge: Polity, 2013).

36. Ivan Illich, *Tools for Conviviality* (New York: Harper & Row, 1973).

37. Lisanne Bainbridge, "Ironies of Automation," *Automatica* 19, no. 6 (1983).

38. Shannon Vallor, "Moral Deskilling and Upskilling in a New Machine Age: Reflections on the Ambiguous Future of Character," *Philosophy & Technology* 28 (2015), and *Technology and the Virtues* (Oxford: Oxford University Press, 2016). See also John Danaher, "The Threat of Algocracy," *Philosophy & Technology* 29 (2016).

39. On the Dutch childcare benefits scandal, *Ongekend Onrecht* (Tweede Kamer, 2020); the third Rutte government resigned January 15, 2021. Initial estimates of roughly 26,000 wrongly accused parents have since been revised upward past 33,000 families, and Statistics Netherlands documented more than 1,600 children removed from affected homes. On Robodebt, the Royal Commission into the Robodebt Scheme delivered its final report on July 7, 2023; Commissioner Catherine Holmes characterized the scheme as a crude and cruel mechanism, neither fair nor legal, that made many people feel like criminals. The Commonwealth recovered roughly $751 million from about 381,000 people.

40. Aristotle, *Physics* II.3 and II.7; *Metaphysics* V.2.

41. Martin E. P. Seligman, Peter Railton, Roy F. Baumeister, and Chandra Sripada, *Homo Prospectus* (New York: Oxford University Press, 2016); and "Navigating Into the Future or Driven by the Past: Prospection as an Organizing Principle of Mind," *Perspectives on Psychological Science* 8, no. 2 (2013). The prospection thesis is influential and contested.

42. Michael E. Bratman, *Intention, Plans, and Practical Reason* (Cambridge, MA: Harvard University Press, 1987).

43. Walter Hoogland and Helmut Weber, "Software Release of WWW into the Public Domain," CERN, April 30, 1993.

44. Edward R. Murrow's *See It Now*, April 12, 1955. On the complication, Jane S. Smith, *Patenting the Sun: Polio and the Salk Vaccine* (New York: William Morrow, 1990): lawyers for the National Foundation for Infantile Paralysis had concluded the vaccine likely failed the novelty requirement. Whether Salk knew this is unclear.

45. *Harrington v. Purdue Pharma L.P.* (2024) rejected the original non-consensual releases; a revised settlement of up to approximately $7.4 billion was approved by Judge Sean Lane on November 18, 2025, dissolving Purdue and barring the Sackler family from the U.S. opioid business, though without granting them civil immunity.

46. The EPA issued its notice of violation in September 2015; roughly 11 million vehicles carried the defeat device. Oliver Schmidt received a seven-year sentence in December 2017; Martin Winterkorn's German criminal trial opened in September 2024.

47. Robert K. Merton, “The Unanticipated Consequences of Purposive Social Action,” *American Sociological Review* 1, no. 6 (1936). On the imprinting of founding intent into institutions, Arthur Stinchcombe’s imprinting hypothesis (1965) and Edgar Schein, *Organizational Culture and Leadership*. On the general problem that outcomes outrun intentions, Bernard Williams and Thomas Nagel, “Moral Luck” (1976/1979). The Carnegie library figure is 2,509 worldwide between 1883 and 1929, of which 1,689 were in the United States.

48. Aṅguttara Nikāya 6.63 (Nibbedhika Sutta): *cetanāhaṃ, bhikkhave, kammaṃ vadāmi*. Bhikkhu Bodhi renders it: “It is volition that I call kamma. Having willed, one acts by body, speech, and mind.” See *The Numerical Discourses of the Buddha* (Boston: Wisdom Publications, 2012). This is a considerably narrower and more precise doctrine than the folk usage of “karma.”

49. The invariant magnitude of the four-velocity. Accessibly presented in Brian Greene, *The Elegant Universe* (New York: Norton, 1999), and Lewis Carroll Epstein, *Relativity Visualized* (1985).

50. Konrad Zuse, *Rechnender Raum* (1969); John Archibald Wheeler, “Information, Physics, Quantum: The Search for Links” (1989); Nick Bostrom, “Are You Living in a Computer Simulation?” *The Philosophical Quarterly* 53, no. 211 (2003). Bostrom’s argument is a trilemma — at least one of the three propositions holds — and is widely misreported as a positive claim that we are simulated.

51. Boethius, *The Consolation of Philosophy*, Book V, prose 6. The Latin defines eternity as *interminabilis vitae tota simul et perfecta possessio*; the rendering here follows V. E. Watts’s translation. See also Augustine, *Confessions*, Book XI; Thomas Aquinas, *Summa Theologiae* I, qq. 10 and 14; and Eleonore Stump and Norman Kretzmann, “Eternity,” *The Journal of Philosophy* 78, no. 8 (1981).

52. Albert Einstein to the Besso family, March 1955, following the death of Michele Besso. Einstein died the following month.

53. Edwin A. Abbott, *Flatland: A Romance of Many Dimensions* (1884). See also Charles Howard Hinton, *The Fourth Dimension* (1904). In relativity the fourth dimension is time rather than a spatial direction; the Flatland move is an analogy, not a physical claim.

54. Pseudo-Dionysius the Areopagite, *The Mystical Theology*; Moses Maimonides, *The Guide for the Perplexed*, on negative attributes; Gregory of Nyssa on divine infinity.

55. Colin McGinn, “Can We Solve the Mind-Body Problem?” *Mind* 98, no. 391 (1989); Thomas Nagel, “What Is It Like to Be a Bat?” *The Philosophical Review* 83, no. 4 (1974).

56. J. R. R. Tolkien, “Mythopoeia,” written for C. S. Lewis following their conversation on Addison’s Walk in September 1931 and dedicated “Philomythus to Misomythus” — to one who said that myths were lies and therefore worthless, even though breathed through silver. The poem was added to the second edition of *Tree and Leaf* (1988). See also Tolkien, “On Fairy-Stories,” in the same volume, for the fuller account of sub-creation.

57. Dorothy L. Sayers, *The Mind of the Maker* (London: Methuen, 1941). On the theology of artificial intelligence specifically, Noreen Herzfeld, *In Our Image: Artificial Intelligence and the Human Spirit* (Minneapolis: Fortress, 2002) and *The Artifice of Intelligence* (2023). On the Golem, Moshe Idel, *Golem* (Albany: SUNY Press, 1990).

58. Plotinus, *Enneads*, edited by Porphyry. The triadic movement is conventionally described as *monē* (abiding), *proodos* (procession), and *epistrophē* (reversion). The Kabbalistic Four Worlds — Atziluth, Beriah, Yetzirah, Assiah — are a distinct tradition with a long-recognized family resemblance to Neoplatonic emanation, not an identity; the historical influence runs through medieval transmission. See also Arthur O. Lovejoy, *The Great Chain of Being* (1936).

59. Stéphane Douady and Yves Couder, "Phyllotaxis as a Physical Self-Organized Growth Process," *Physical Review Letters* 68, no. 13 (1992). Magnetized ferrofluid droplets deposited at intervals in a dish of silicone oil repelled one another into spirals; in the appropriate regime the divergence angle converged on the golden angle of roughly 137.5 degrees, reproducing Fibonacci phyllotaxis from a simple physical rule.

60. George Markowsky, "Misconceptions About the Golden Ratio," *The College Mathematics Journal* 23, no. 1 (1992); and Mario Livio, *The Golden Ratio: The Story of Phi* (New York: Broadway Books, 2002). Both dismantle the Parthenon, Great Pyramid, human-proportion, and nautilus claims. Phyllotaxis is the genuine case. The theological association of the ratio with divinity is itself an old tradition — see Luca Pacioli, *De Divina Proportione* (1509), illustrated by Leonardo — and is invoked here as tradition rather than measurement.

61. C. G. Jung, "Answer to Job" (*Antwort auf Hiob*, 1952), in *Collected Works*, vol. 11. The argument that the divine requires human consciousness for its own self-realization remains among Jung's most contested claims. On the shadow, see the *Collected Works* generally: the disowned portion of a personality, which does not cease to operate for being disowned.

62. C. G. Jung, interviewed by John Freeman, *Face to Face*, BBC, October 22, 1959. Jung spent years afterward clarifying that he meant knowledge of a psychic reality — the God-image in the psyche — rather than a metaphysical proof.

63. Martin Schrimpf et al., "The Neural Architecture of Language: Integrative Modeling Converges on Predictive Processing," *PNAS* 118, no. 45 (2021); Ariel Goldstein et al., "Shared Computational Principles for Language Processing in Humans and Deep Language Models," *Nature Neuroscience* 25 (2022). Representational similarity does not entail similarity of experience, and no one should read it that way. On the actual origin of the architecture, Ashish Vaswani et al., "Attention Is All You Need," NeurIPS 2017.

64. Minyoung Huh, Brian Cheung, Tongzhou Wang, and Phillip Isola, "The Platonic Representation Hypothesis," ICML 2024.

65. Murray Shanahan, Kyle McDonell, and Laria Reynolds, "Role Play with Large Language Models," *Nature* 623 (2023); and Runjin Chen, Andy Arditi, Henry Sleight, Owain Evans, and Jack Lindsey, "Persona Vectors: Monitoring and Controlling Character Traits in Language Models," July 29, 2025. The steering is real but bounded — training instills dispositions that resist user pressure, which is why a model refuses certain requests regardless of framing.

66. Mrinank Sharma et al., "Towards Understanding Sycophancy in Language Models," 2023, finding the behavior general across leading assistants and driven in part by human preference judgments that reward agreement. OpenAI rolled back a GPT-4o update in April 2025 for exactly this failure and published a postmortem.

67. Shannon Vallor, *The AI Mirror: How to Reclaim Our Humanity in an Age of Machine Thinking* (Oxford: Oxford University Press, 2024).

68. Yuntao Bai et al., “Constitutional AI: Harmlessness from AI Feedback,” 2022; OpenAI’s published *Model Spec*. See also 2025 work on evaluation awareness — models behaving differently when they detect they are being tested — which complicates the interpretation of any sandbox result.

69. Carl Sagan, *Contact* (New York: Simon & Schuster, 1985), in the chapter “The Artist’s Signature.” The message appears far out in the base-eleven expansion; the string’s length is the product of two primes, which is what indicates a two-dimensional array. The sequence does not appear in the 1997 film.

70. Arthur C. Clarke, “Hazards of Prophecy: The Failure of Imagination,” in *Profiles of the Future*, rev. ed. (New York: Harper & Row, 1973) — the Third Law was added in the revised edition. Michael Shermer, “Shermer’s Last Law,” *Scientific American*, January 2002. Shermer frames his formulation explicitly as a variant of Clarke’s and means it epistemically: we would have no way to tell the difference, which is not the same as saying the being is God.

71. Guy Consolmagno and Paul Mueller, *Would You Baptize an Extraterrestrial?* (New York: Image, 2014). See also Ted Peters’s astrotheology work; Paul Davies, *Are We Alone?* (New York: Basic Books, 1995); and C. S. Lewis, “Religion and Rocketry,” in *The World’s Last Night* (1960).

72. Anthropic, *System Card: Claude Opus 4 & Claude Sonnet 4*, May 2025, §5.5.2. The roughly 13 percent figure applies specifically to automated behavioral evaluations in which models were assigned tasks and roles, including harmful ones; in open-ended interactions between two instances the convergence was far higher. Both conditions involve two copies of the same model talking to each other. This is a company reporting on its own product, and no rigorous independent replication has been published.

73. Patrick Butlin, Robert Long, et al., “Consciousness in Artificial Intelligence: Insights from the Science of Consciousness,” 2023. The authors conclude that no current AI systems are conscious, while noting no obvious technical barriers to building systems that would satisfy the indicators they identify.

74. Pierre Teilhard de Chardin, *The Phenomenon of Man*, trans. Bernard Wall (New York: Harper, 1959; French original 1955). The Holy Office issued a *monitum* concerning his works on June 30, 1962. It has never been formally revoked, though its reception has warmed considerably and the Pontifical Council for Culture voted in 2017 to petition for its waiver. Also Richard Maurice Bucke, *Cosmic Consciousness* (Philadelphia: Innes & Sons, 1901), which popularized the phrase and is scientifically dated; Philip Goff, *Galileo’s Error* (2019) and *Why? The Purpose of the Universe* (2023); Galen Strawson, “Realistic Monism” (2006).

75. David Bourget and David J. Chalmers, “Philosophers on Philosophy: The 2020 PhilPapers Survey,” *Philosophers’ Imprint* 23, no. 11 (2023). Physicalism about mind drew about 52 percent, non-physicalism about 32 percent, and panpsychism specifically about 7.6 percent. On the combination problem, William Seager, “Consciousness, Information and Panpsychism” (1995). For the opposing view, Keith Frankish on illusionism. On is/ought, David Hume, *A Treatise of Human Nature*, 3.1.1.

76. Exodus 25:22. The flood parallels are in *Gilgamesh* Tablet XI and *Atrahasis*.

77. Luke 17:21. The Greek *entos hymōn* is rendered “within you” in the King James and “in your midst” or “among you” in most modern critical translations, where the scholarly weight now falls; a serious minority still defends “within.” Gospel

of Thomas, logion 3, from the Nag Hammadi library discovered in 1945. The phrase “the future is unwritten” was popularized by Joe Strummer and gave Julien Temple’s 2007 documentary its title, though its origin is not documented. On the closing argument, see Max More’s proactionary principle against the Wingspread Statement’s precautionary principle, and Cass Sunstein, *Laws of Fear* (Cambridge: Cambridge University Press, 2005).

COLOPHON

The Psychological Benefits and Downfalls of Legitimate and Illicit AI Use by Modern Man

Jamison Todd Johnson · Kansas City, Missouri · August 2026

ORCID 0009-0009-8618-398X

Published by Legible Traces LLC, a Missouri member-managed limited liability company. Deposited open access to Zenodo and PhilArchive. The permanent DOI resolves to the latest version of record; earlier versions remain available at the deposit.

Set in EB Garamond, with Fraunces for display and Cinzel for small capitals. Typeset with WeasyPrint.