



# Magical thinking and the test of humanity: we have seen the danger of AI and it is us

David Morris<sup>1</sup>

Received: 21 July 2023 / Accepted: 28 August 2023 / Published online: 14 September 2023  
© The Author(s), under exclusive licence to Springer-Verlag London Ltd., part of Springer Nature 2023

The danger of general, generative AI is not the technology itself. The danger is that we, and especially the leaders of the AI industry, have already taken thinking to be the sort of ‘intelligence’ that can be generatively stockpiled by AIs. To paraphrase Voltaire, human thinking created AI and AI, in kind, reinvented human thinking as a stockpile of information to be strip-mined for valuable, entrancing, or deceptive and dangerous, information patterns. These patterns are marketed in products that steal and eviscerate human labour. The patterns are circulated as completed end results abstracted from their origins in human dialogue and community; this contributes to the erosion of two capacities crucial to processes of thinking: attention and questioning.

I use “general, generative AI” (GnAI) to designate publicly available systems, designed to be used by non-experts, to generate textual, graphical, video, and sonic, etc., outputs, in response to an open-ended and indefinite set of prompts (albeit with some guardrail limits set by the providers). I distinguish this from, say, “specialized AI,” designating AI systems trained on specific data sets to assist with specific problems, used by experts who understand the problem domain—even if experts may have no detailed idea how the AI produces its answers (say AlphaFold for protein folding).

Savvy statisticians routinely share stories of statistical software lending itself to misuse, mistake, and misunderstanding by researchers not expert in stats, conducting misleading claims that sometimes get published. GnAI allows removal of: the need to learn software; the expertise of the researcher about the problem domain; the expertise required to understand good or bad reasoning; peer review of results; and it gives only limited attention to good or bad motives of the user. Results are often replicated in viral media engines powered by audiences affectively charged to the point of

combustion by the social pistons of likes. Social media is already driven by amplifying the heuristics and biases that Tversky and Kahneman showed skew human judgement. Should we turbocharge these explosive pumps with AI engines promising expertise without understanding? Do we need to boost our inborn overconfidence bias? Or more straightforwardly, should we connect social media misinformation engines to AIs, allowing users with no regard at all for expertise (or hatred of expertise) to generate persuasive texts, which may full well be what I’d dub ‘fabulloshit’? Bullshit is speech intended to persuade without regard for truth. GnAIs hallucinate and fabulate persuasive patterns without regard for either truth or falsity, since the algorithm as such has no concern for either, just producing patterns that fit. Fabulloshit can look fabulous to us. Off the shelf, GnAIs are a bit like genies—but with cautions about what you wish or ask for dialed down, turned off, or forgotten, in the rush to get them out of the bottle, to capture markets.

However, we must distinguish dangerous technological consequences of GnAI from what enables GnAI systems to have this possibility in the first place. The enabling precondition is our already having equated thinking with information processing (despite even Claude Shannon emphasizing that his theory of information has nothing to say about meaning). This is amplified by plugging our bodies and lives into mobile networks of information flows that monetize and algorithmically mediate our attention. These preconditions, in turn, hinge on at least two interrelated errors: magical thinking about thinking and forgetting to ask what humanity and thinking are when it comes to our fascination with Turing tests and AI intelligence.

Magicians hack our attentional, perceptual, and cognitive tendencies to make us perceive and believe what is not there. Twin A goes into the door on stage left and twin B exits the door on stage right; we perceive teleportation. The magician’s patter further lulls us into the magic. To think AIs are intelligent is to be hacked by your experience of reading and

✉ David Morris  
David.Morris@concordia.ca

<sup>1</sup> Department of Philosophy, Concordia University, 1455 de Maisonneuve Blvd. West, Montreal, QC H3G 1M8, Canada

writing as thinking, and further lulled by marketing and the AI moniker. It is a trick! (Incredibly impressive, mind you.)

Imagine a Large Scent and Taste Model (LaSTM) that has crunched databases of molecular structures of scents and flavours humans have evolved and enculturated to like. LaSTM can produce formulae for novel molecules with scents and tastes we like. Would anybody in their right mind say LaSTM has a sense of taste or smell? No, it has never tasted or smelled anything at all! LLM generators of word patterns do not need to possess the intelligence of what created the training data set to crunch out patterns of what we take to be intelligent.

The view that LLM generators are intelligent is as cogent as the view that an advanced weather prediction/pattern generator system, LWM, is stormy. LWM generates *artificial-weather*. Why don't we similarly think that artificial intelligence precisely generates.... *artificial-intelligence*, intelligence as artifice, craft, or fabulloshit deception? LWM predicts what moisture and pressure patterns a real weather system would generate next. GnAI predicts what word patterns thinking human beings would generate next. If you train GnAI on garbage nonsense, it would generate that just as well. It did not do the work of evolving what works as intelligent, we did that. It did not need to make a living by thinking, we made it. Why would we think that its pattern making is itself intelligent?

My bet is that what inclines us to this confusion is a cultural tendency (by no means universal, globally or historically) to textualism, legalism, and so on: to take words on pages as powers and powers of thinking. (That is also quite a trick: "You can't stay on your land because that is what the written law says!" "Say what?! Marks on paper don't say anything, *you're* the one saying things and forcing me off my land!").

This tendency bends us to what I call 'magical thinking about thinking' or 'magical literalism': if a word pattern looks like a thought, we take it to literally be a thought, because we equate thinking with records and information, detach records from thinkers, and then run that equation in reverse. As the philosopher Merleau-Ponty would have observed, magical literalism conflates recorded results of an expressive process of thinking, with the thinking process itself. Magical literalism thus prepares us to gasp at general AIs as if new divinities have descended amongst us, versus "stochastic parrots." Magical literalism is conducted by hacking our longstanding attitudes to powers of texts and speech that have led us down rabbit-holes of conspiracy and into the sway of autocrats and other con artists. The autocrat who mesmerizes audiences with empowering incantations is hacking our patterns of desire, fear, FOMO, needs, need to know, and our susceptibility to seek ready answers spelled out in easy formulas, to make us believe that there is power or thought in words themselves. The marketing of

(and media hubbub around) the AI industry and AI outputs plays on similar needs and wants in ways that misdirect us to take words as themselves thoughts and answers, misdirect us from thinking about the questions to which these words would be answers—and above all misdirect us from asking what thinking is and how it matters in the first place.

There is no better emblem of this industry tendency or indicator of the underlying problem than the "Modern Turing Test." Turing's test is now all too easily passed by AIs. The Modern Turing test would require an AI to "successfully act on this instruction: "Go make \$1 million on a retail web platform in a few months with just a \$100,000 investment." Studied in its own intellectual and historical context, what Turing himself was seeking to test is: what would pass for intelligence in a *novel, purely information processing domain*. The test is as much about what we take intelligence to be as what it takes for a machine to produce the informational pattern that passes for this.

What the Modern Turing Test exemplifies is a tendency in the AI industry and related industries to think that human intelligence is distilled in the sort of ability required to make money and run businesses. We end up with an industry in which at least some captains of industry take human intelligence to be the sort of intelligence that decides to press the button on unleashing AI without asking whether it is good. Isn't that the danger of AI in a nutshell, or bombshell? Not what AIs themselves do but what we take to be intelligent, the way we market that, and our readiness to fall under the spell of word patterns as if they provide answers, without asking questions? Not to mention our repeated pattern of marketing information systems without asking important questions first. (Is it really secure to market routers or IoT devices with default passwords published for any hacker to find?).

If AI destroys us (maybe) or makes misinformation and human life worse (check your email, hasn't the spear phishing already improved via AIs?) it is not because of AIs themselves. It is because we are building them to mirror an intelligence that does not bother to take time to ask who we should be. Any decent general AI tasked with the Modern Turing test is going to find the pattern that the best return on money is offered by scams. The author of the test is putting humans in the loop on the AI pulling the triggers, but this I think is missing the danger. If your criterion of intelligence is making killings on the market, you need to acknowledge that your AI is likely to turn into LOGaN-RoY 2.0.

Word pattern generators do not need to understand what the questions are about or whether these are good questions to ask: all they need to do is to produce patterns that a human would take to be an answer. And the end-user can take that pattern as an answer without working on the question, and fire it off into viral replication. We are letting GnAI insert itself into what most of all makes us human, which is

meaningful communication with one another. As Aristotle might have said, AI systems may generate and communicate information, but only human information systems communicate about the good (see his *Politics*).

Instead of working out Turing tests for whether machines can think, we should be working out tests for what thinking is such that it is human. Kant defined enlightenment as “humanity’s emergence from self-incurred immaturity.” Isn’t reliance on autocrats to save us or stochastic parrots to do our thinking the height of self-incurred immaturity? Do we not infantilize ourselves to some degree by allowing ourselves to be trained to incant slogans, or like, retweet, follow, autocomplete, and so on, to let ourselves get sucked into vortices of information that are so easily automated and manipulated? Is self-incurred immaturity not part of the product being sold when the general AI industry offers to write our replies to other human beings for us?

Instead of worries that the AI technology is itself an existential threat, we need to look ourselves in the face and turn our thinking to testing our own humanity. We need a Kantian worry about humans being ready to launch AI (and other) industries upon us without first of all asking basic questions about what I can know, what I ought do, what I should hope for, and what a human being is. The tech industry track record is all too often: claim to know it all; move fast and scale up; hope for millions; and not bother much about the human (or environmental) impact.

The danger of AI is not the technology but a mindset of magical thinking that takes textual residues of thought as another resource to be strip-mined, without sufficient thinking about effects on human life, attention, and thought, combined with a tendency to conflate intelligence with calculation or profit-making craft. What is dangerous about AI is not the machines as such but the way it amplifies and mirrors our killer intelligence, to let us rapidly build an industry that inserts itself into and threatens our ability to clearly communicate and think with one another about what is good for life.

**Curmudgeon Corner** Curmudgeon Corner is a short opinionated column on trends in technology, arts, science and society, commenting on issues of concern to the research community and wider society. Whilst the drive for super-human intelligence promotes potential benefits to wider society, it also raises deep concerns of existential risk, thereby highlighting the need for an ongoing conversation between technology and society. At the core of Curmudgeon concern is the question: What is it to be human in the age of the AI machine? -Editor.

**Data availability** There are no associated data.

## Declarations

**Conflict of interest** The corresponding author states that there is no conflict of interest.

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.