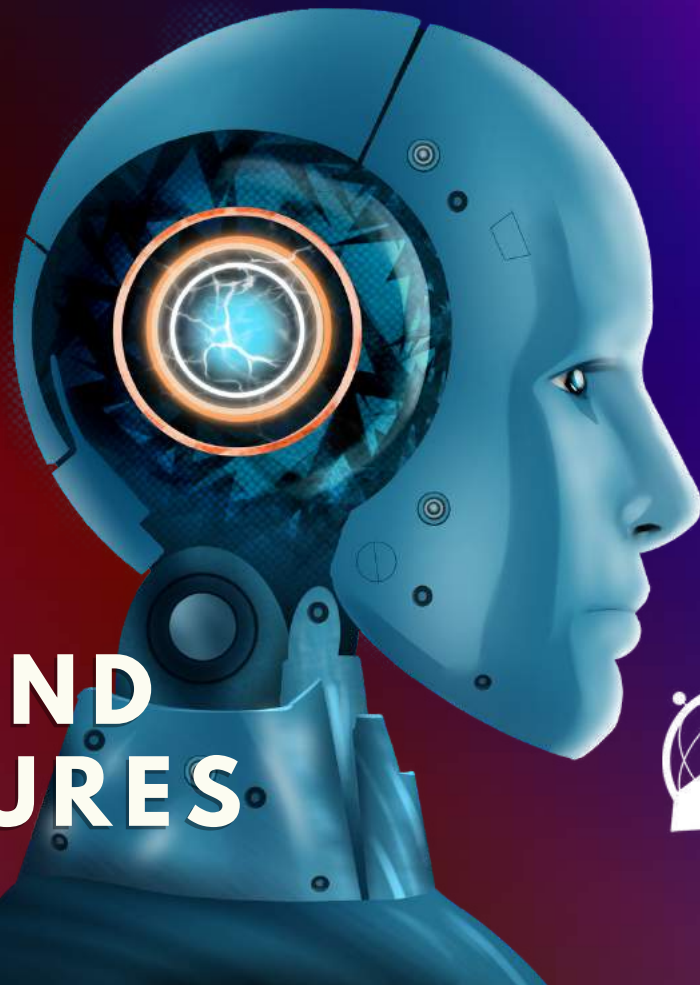DECEMBER 2023

# COMPASS

## THE GLOBAL VOICE OF PROFESSIONAL FUTURISTS

# AI AND FUTURES

ASSOCIATION OF
PROFESSIONAL
FUTURISTS

# QUESTIONING OUR ASSUMPTIONS ABOUT THE FUTURE OF AI

## By Richard Yonck

Today, we work with many different forms of artificial intelligence, a number of which are relatively new. Some are almost mundane in what they do, while others feel like they are performing magic. Relative newcomers such as generative AI, including LLMs, achieve incredible things, while having a range of serious shortcomings. A number of these deficiencies will be overcome in time, while others will probably plague us for decades.

This is the nature of building tools. None of them are great for everything. Part of the trick is understanding what each tool is very good at and if it can be used reliably. Some of the work that must be done then is to better manage the human side of the equation, including our own expectations.

Human beings are creatures of story, of myth, of narrative, and of metaphor. We've used these constructs with tremendous efficiency through the ages: to transfer ideas across vast distances and time, to compress thoughts and insights into manageable nuggets, and to create heuristics that allow us to abstract layer upon layer of accumulated knowledge.

Whether we're talking about the foot of a table, the arm of the law, or the head of a corporation, we routinely use conceptual shortcuts to enable our world of discourse. We do this externally, as we generate and build society and we do it internally, shaping our sense of who we are.

Yet despite their power, these shortcuts remain mere placeholders of what they actually represent. This also goes for the model of our own minds as we apply it to what we imagine the future of artificial intelligence could become. But confusing the model, the metaphor, or the map for the reality itself routinely leads to problems and in few places is this so prevalent as when we talk about AI.

This is becoming increasingly important. We've reached a stage in the development of artificial intelligence where we are seeing a growing number of claims regarding its capabilities that far exceed the reality, even among some of the AI scientists and researchers themselves. While there are many things about AI that should alarm us, machine uprisings and superintelligence will likely remain distant

concerns for some time to come. As I've explored elsewhere, I'm very willing to discuss AI as a new form of intelligence, however, we confuse many of its features with human intelligence and consciousness to our disadvantage and perhaps even at our peril.

## ORIGINS

The term artificial intelligence is the predominant vernacular for a range of computer-enabled technologies intended to approximate human cognition. This term is largely used because of a choice made many decades ago.

During the early days of computing, John McCarthy, Marvin Minsky, Claude Shannon and Nathaniel Rochester wrote a proposal for the now-famous Dartmouth Summer Research Project on Artificial Intelligence. The two-month workshop, which took place between mid-June and mid-August of 1956, attracted about two dozen scientists and mathematicians, the early pioneers of this then-nascent field. The stated purpose of the gathering was to create "thinking machines" capable of using language, forming abstractions, and solving problems normally considered to belong exclusively in the realm of human reasoning. They hoped to achieve much of this during that summer, a goal that was decidedly optimistic.

The various participants were focused on different approaches that used the new computing technologies that were themselves still very much in their earliest stages. Utilizing logic and applied mathematics these pioneers were developing fields like automata theory,

cybernetics and complex information processing. In an effort to avoid conflict among the different factions in attendance, McCarthy settled on the seemingly neutral name of "artificial intelligence." Unfortunately, this term carried with it all kinds of unhelpful associations, especially for the public and popular press in the years that followed. It became far too easy to say these programs could think or had minds of their own. But the name stuck and continues to skew the conversation to this day.



Image source: Canva

## SOCIAL BEINGS

In the course of human evolution, we learned to recognize patterns in our environment, especially among our fellow human beings. This allowed us to anticipate events and choose our future actions accordingly. In the course of this we also developed a tendency to perceive human-like qualities in other parts of our world. This animistic trend likely had the advantage of helping us make decisions that promoted our survival in a harsh environment. Today, we still do this to varying degrees – anthropomorphizing the weather, our pets, and especially technology. This is even more prevalent with certain kinds of devices, especially those with interfaces that have become increasingly natural, allowing us to interact with them almost as if they were another person.

In one of the early efforts to study human-machine communication, in 1965 Joseph Weizenbaum of MIT developed one of the earliest chatbots, a script-driven program called ELIZA. One of ELIZA's scripts, known as DOCTOR, was written to emulate a Rogerian psychotherapist. To do this, it used pattern recognition and word substitution to create an illusion of intelligence and understanding. Weizenbaum developed ELIZA intending to demonstrate the superficiality of such communications but was alarmed to find users engaging much more fully than he'd expected. This was underscored when one day Weizenbaum entered his office to find his assistant using the program. To his chagrin, the assistant asked if he could come back later since she was in the middle of a private conversation with the therapist.

Half a century later, we've entered the era of transformers, large language models (LLMs) such as OpenAI's GPT-4 and Google's LaMDA, which are the basis of chatbots such as ChatGPT and Bard. These draw from vast collections of text, sometimes referred to as a corpus. While these conversational AI's are far more capable than ELIZA, they have little more awareness than did earlier chatbots. However, LLM-based chatbots apply sophisticated statistical techniques to their underlying corpus in order to converge on an answer. They do this so capably that it often feels like there is an underlying intelligence behind them. In a sense, this is absolutely true, because given the corpus they are working from, that intelligence is our own.

When we look at the output of an LLM, it is the product of more knowledge than any person can truly comprehend. In some respects, it can be said they nearly contain the sum of human knowledge. Because what these programs do can sometimes feel like magic, it's not surprising that some people believe we've finally built the perfect thinking machine. But what we've actually done is create a mirror that reflects our own knowledge back on ourselves.

*...it's not surprising that some people believe we've finally built the perfect thinking machine. But what we've actually done is create a mirror that reflects our own knowledge back on ourselves.*

This is a great step forward, to be sure. But we would be mistaken to say these systems actually think in the sense that people do. LLMs do not have self-awareness or a real contextual understanding about their output. They don't have the ability to make complex value judgements or assess right from wrong beyond what they can process based on prior human output. This is certainly a huge advance from where AI was only recently, but there is still a long way to go. You may have heard of these system beings referred to as "stochastic parrots." It's a term that accurately reflects the statistical nature of what they do.

It can be said that such amalgamating of prior knowledge is not so different from how many people produce their own work and in a sense that's true. Human knowledge is built upon a vast foundation of ideas that precede us, which is exactly why society is possible. But we sell ourselves very short if we say this small subset of cognitive emulation is equivalent or even superior to our own intellectual processes. Within specific limits, these systems can be said to better human abilities, but this is exactly why we build most technologies – to be the tools we need for our next chapter of progress.
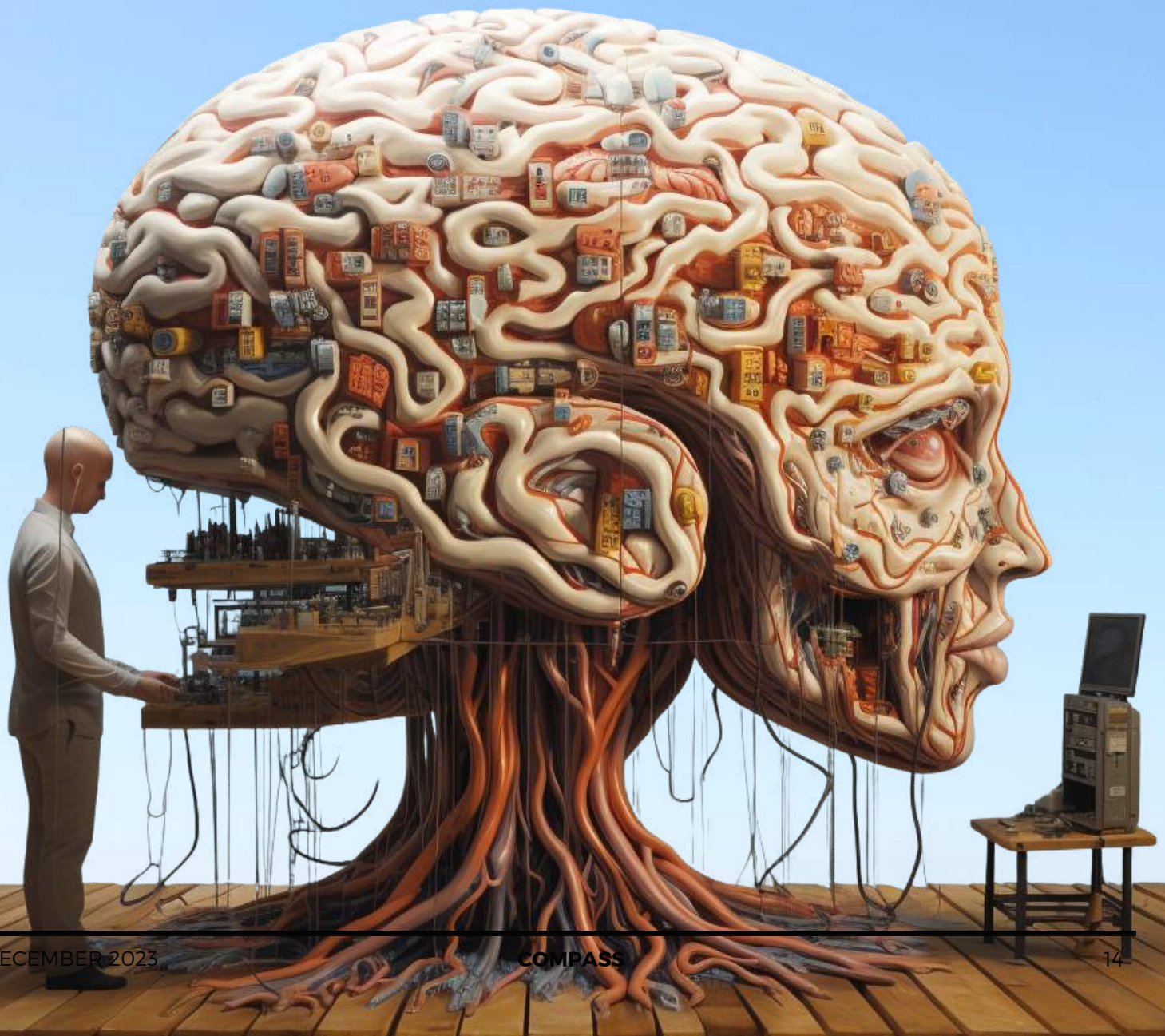


Image source: Mid journey created by Richard Yonck

## THE MYTH OF GENERAL INTELLIGENCE

Artificial general intelligence, or AGI, has been a goal and a fear, a dream and a nightmare, for decades. A product of science fiction, AGI is sometimes deemed the holy grail of what machine intelligence is moving toward. Except for one thing – general intelligence doesn't exist.

Usually when we talk about general intelligence, we're using it as shorthand for human intelligence. Which is just our anthropocentric way of saying ours is the best kind of intelligence. However, all forms of intelligence found in nature are basically just the optimum qualities and skills appropriate for the particular ecological niche an organism occupies. For instance, our lack of natural flying skills and echolocation would make human beings ill-suited to exploit the ecological niche occupied by bats.

The "cognitive niche" we humans occupy is one we manage to take tremendous advantage of, but it's safe to say there are many areas in which we fall short. Furthermore, even within our own unique range of abilities, there are cognitive biases that prevent us from using our minds optimally. This leads to mistakes and errors that are themselves part of the feedback loops we have to navigate as we go through life.

One of the general trends in computing is that the more specialized and optimized a system is for a particular task, the faster and better its performance. Conversely, as a system is made more capable of dealing with a broader range of tasks or capable of learning and adding to an already optimized skill set, it tends to slow down and become less efficient. Since we build

technology to be our tool, it serves us to have it perform tasks far better than we can, rather than hobbling it so it can perfectly replicate our own skills. (If this was even possible, which is debatable given the very different origins of biological and machine intelligence.)

For years, AI experts have been surveyed about when we will achieve AGI. The results have tended to converge somewhere between the middle and end of the 21st century. Some outliers have answered this will happen in the next few years, while others say it won't occur for centuries, if ever. Given the degree to which human intelligence and consciousness are themselves understood, I would say that when we achieve AGI will have a great deal to do with how we define it.

## SENTIENCE, SAPIENCE AND AI – OH, MY!

Last year, a Google engineer named Blake Lemoine publicly claimed the LaMDA large language model he was working with had become sentient and that it even had a soul. The popular press jumped on this news and Lemoine was fired not long afterward, a victim of the all-too human willingness to ascribe consciousness to our creations.

To begin, Lemoine and the journalists covering the story are using the term *sentience* incorrectly. From the Latin, "sentir," sentience actually means the ability to feel. Because of this, many animals are said to be sentient. *Sapience* on the other hand is the ability to think and reason as we *Homo sapiens* do. The frequent misuse of the word in science fiction is often considered the reason for the confusion. Based on his statements, Lemoine presumably meant he felt LaMDA had become sapient.

For the reasons previously mentioned, it's extremely unlikely that LaMDA, GPT-4 or any other LLM has or will become sentient or sapient. While there have been all kinds of attempts to explain human thought, self-awareness and consciousness over the millennia, much about these processes remains a mystery. Recent efforts to describe causal mechanisms, including Orchestrated Objective Reduction (Orch OR) and Integrated Information Theory (IIT), have yet to demonstrate anything substantive.

My own belief is that the highly integrated modular nature of our brains leads to continuous crosstalk that gives rise to the emergent metacognitive properties we refer to as consciousness. This follows some of the ideas put forth by Minsky, Gazzaniga, Hawkins and others. These are not an actual explanation of the processes, but merely pointers toward the possible underlying mechanisms. Given the general nature of emergence, it may be that we will never fully unravel the mysteries of how we are able to ponder all of this in the first place.

But looking at the differences between biological and electronic minds, I suspect our electronics are still many orders of complexity away from producing anything like the different forms of meta-awareness that we experience. Until this is overcome, we probably won't be able to replicate such phenomena in our machines. Coming back to the idea of maps and models, we long ago fell into the trap of equating transistors and neurons. Yet even the spiking neural networks and neuromorphic circuits being developed to emulate neurons in recent years are vastly less complex than the biology they seek to reproduce.

Like the hydraulics, pneumatics, and clocks of centuries past, we now look to computers, the internet and quantum theory as metaphors to describe the workings of the physical brain and its emergent properties of mind. LLMs are only the latest in this long parade of explanatory models and like all of these it will almost surely fall short.

## WHERE WE GO FROM HERE

There are so many amazing developments ahead of us when it comes to artificial intelligence. As in the past, these will continue to transform our world both positively and negatively, bringing opportunities and challenges with each stage of advancement.

This is where we need to prioritize our attention. While we should also be thinking about the dangers artificial superintelligence could potentially lead to, we need to prioritize the immediate threats AI presents as a tool and weapon wielded by human intent.

Anticipatory surveillance, algorithmic influence, predatory marketing, computer profiling, AI colonialism and many other anti-societal developments should be the focus for much more of our concern as we develop the AI tools of tomorrow.

**REFERENCES:**

1.Wired: "LaMDA and the Sentient AI Trap" by Khari Johnson, June 14, 2022

2.Gazzaniga, M., The Consciousness Instinct: Unraveling the Mystery of How the Brain Makes the Mind. Farrar, Straus and Giroux, 2018.

3.Guthrie, S. E., Faces in the Clouds: A New Theory of Religion, Oxford University Press, 1995.

4.Hawkins, J., A Thousand Brains: A New Theory of Intelligence. Basic Books, 2021.

5.Minsky, M., The Society of Mind. Simon & Schuster, 1988.

6.Yonck, R., Future Minds: The Rise of Intelligence from the Big Bang to the End of the Universe. Arcade, 2020.

**Richard Yonck**

*Richard Yonck is a Seattle-based futurist, author and keynote speaker, who helps organizations and audiences explore, anticipate and plan for future change. He's the author of two books about the future of artificial intelligence: Heart of the Machine and Future Minds. He's also written for a wide range of publications including Scientific American, Fast Company, Wired, GeekWire, World Future Review, The Futurist, Salon, and many others. He's a member of the Association of Professional Futurists, the World Future Studies Federation, the National Association of Science Writers and a TEDx speaker.*