The Xavier Zubiri Review, Vol. 16, 2022-2024, pp. 65-103

Technology and the Limitations of Artificial Intelligence

Thomas B. Fowler
Xavier Zubiri Foundation of North America
Washington, DC USA

Abstract

Artificial Intelligence (AI) has become a trendy label for a class of computer-based technologies that seek to replicate or replace human knowledge and ways of knowing. Extravagant claims have been made for AI, leading to fears of AI "taking over" or causing catastrophes of various sorts. Because AI deals with notions of "intelligence", "thinking", and "knowledge," it directly connects with philosophy. The limits of AI ultimately have to do with its fundamental inability to perceive reality. A brief foray into philosophy reveals that Ideas about AI are based on erroneous notions of human knowing, stemming from the English empiricist tradition, which culminated in David Hume. Specifically, it assumes the theory of sensible intelligence. AI is also grounded on certain standard engineering practices that are solidly based on our understanding of how to create reliable systems, but were never intended to replace or replicate human ways of knowing. The dangers associated with AI are not that it will take over the world or become sentient, but that due to the ongoing complexification of society, AI will be used to direct and control large-scale systems without the connection to reality that this kind of control needs to stave off catastrophic errors. As human beings, we can perceive reality and know truth because our paradigm of knowing, sentient intelligence, is radically different from that perforce used in AI.

Resumen

La Inteligencia Artificial (IA) se ha convertido en una etiqueta de moda para una clase de tecnologías basadas en computadoras que buscan replicar o reemplazar el conocimiento y las formas de conocimiento humanos. Se han hecho afirmaciones extravagantes a favor de la IA, lo que genera temores de que la IA "tome el control" o cause catástrofes de diversos tipos. Dado que la IA aborda nociones de "inteligencia", "pensamiento" y "conocimiento", se conecta directamente con la filosofía. Los límites de la IA tienen que ver en última instancia con su incapacidad fundamental para percibir la realidad. Una breve incursión en la filosofía revela que las ideas sobre la IA se basan en nociones erróneas del conocimiento humano, derivadas de la tradición empirista inglesa, que culminó con David Hume. En concreto, asume la teoría de la inteligencia sensible. La IA también se basa en ciertas prácticas de ingeniería estándar que se basan sólidamente en nuestra comprensión de cómo crear sistemas confiables, pero que nunca tuvieron la intención de reemplazar o replicar las formas humanas de conocimiento. Los peligros asociados con la IA no son que se apodere del mundo o se vuelva sensible, sino que, debido a la actual complejización de la sociedad, la IA se utilizará para dirigir y controlar sistemas a gran escala sin la conexión con la realidad que este tipo de control supone. debe evitar errores catastróficos. Como seres humanos, podemos percibir la realidad y conocer la verdad porque nuestro paradigma de conocimiento, la inteligencia sensible, es radicalmente diferente del que se utiliza necesariamente en la IA.

Introduction

Today we find ourselves in a perilous situation:

> Out in the world, the rebellion against God has become a rebellion against everything: roots, culture, community, families, biology itself. Machine progress—the triumph of the Nietzschean will—¬dissolves the glue that once held us. Fires are set around the supporting pillars of the culture by those charged with guarding it, urged on by an ascendant faction determined to erase the past, abuse their ancestors, and dynamite their cultural ¬inheritance, the better to build their earthly paradise on terra ¬nullius.

There are many contributing factors to this lamentable state of affairs, but unfounded and erroneous claims made for Artificial Intelligence (AI) advance the agenda. These claims are based on the assumption that sensible intelligence—the only kind accessible to machines—has the same capabilities as human sentient intelligence.[*] Since this assumption is wrong, the claims will also fail. In this article, we will explore how the limitations of AI illustrate the truth of Zubiri's noology.

Broadly speaking, "Artificial Intelligence" is the category of systems that utilize computers, feedback, rule-based logical inference (deterministic or statistical), complex data structures, and large databases to extract information and patterns from data and apply it to control, decisions or queries. The goal is to emulate human reasoning, decisions and actions, but at a much faster speed with greater scope. Usually it involves implementation of an algorithm whose execution would not be feasible if attempted by human minds due to the number of calculations and operations required.

The kinds of technology that typically fall under the rubric of "Artificial Intelligence" include:

i. Robots and robotic systems
ii. Neural networks and pattern recognition
iii. Generative AI, including ChatGPT and similar applications using Large Language Model
iv. Symbolic manipulation programs such as *Mathematica*®.
v. Autonomous cars and other autonomous systems
vi. Complex large-scale control programs

These technologies may be combined to surmount difficult and complicated problems. Applications include autonomous vehicles, answering natural language questions about a subject, looking for disease in X-ray images, control of complex industrial processes, factoring large numbers, and solving certain types of mathematical problems. Today AI has become a buzzword, and therefore many software programs boast "AI" routines and capabilities that are, in fact, little more than improved versions of existing programs.

But claims for AI go well beyond these kinds of practical applications, indulging in unbridled speculation about "thinking machines". Computers have always fueled speculation about the inevitability of machines outstripping human intelligence, as a prelude to the machines "taking over" from humans. No less than computer pioneer Alan Turing (1912-1954) informed us 70 years ago that:

> It seems probable that once the machine thinking method had started, it would not take long to outstrip our feeble powers… They would be able to converse with each other to sharpen their wits. At some stage therefore, we

[*] Sensible Intelligence, or sense-based knowing, assumes that the senses receive information about their environment and transmit it to the brain, which processes it and creates a "picture" of the "outside world". Sentient intelligence or sentient knowing says that our perception and knowing are a fully integrated process, and as part of that process we are in direct contact with reality.

> should have to expect the machines to take control.[1]

The motion picture *2001 A Space Odyssey*, released in 1968, has an intelligent computer HAL assuming control from the human astronauts. Unquestionably computers have offloaded many jobs that humans at one time did. The first computers, built in World War II, were intended to speed up calculation of ballistics tables, a laborious job then done by mathematicians with manual adding machines. This type of rote calculation is now always done by computer, along with myriad other jobs that can be reduced to algorithms and then programmed. What is called "AI" is not a radically new invention, but an evolutionary development in computer applications and automation. In fact, automation has been a staple of life since the beginnings of the Industrial Revolution in the 18th century. Mechanization of human actions has been going on for millennia. Many human activities involve a repetitive part, usually the result of training or experience, and creative part, which involves new types and ways of interacting with reality, or just simply problem solving. Automation generally involves handing off the first part to machines. The second part is reserved to humans, who have the necessary capacity for it.

Numerous books have been written that deal with technological limitations of AI.[2,3,4,5]. This is an important area of research and study, naturally, because it reveals application areas likely to benefit from AI, and others where technological limitations will constrain applications. Here we consider a different question, viz. *whether the paradigm of knowing used in AI entails limits of some type that reveal the boundaries of AI, no matter how implemented and how fast the hardware*. This is one of the most intriguing aspects of AI, because such a boundary implies a direct connection with philosophy. Indeed, once the discussion turns on issues such as sentience, "thinking", and what constitutes "intelligence", it leaves the realm of technology and enters that of philosophy—a situation that may not be comfortable to those immersed in technology, because philosophy operates at an entirely different plane of knowledge and understanding of the world. Sentience and thinking immediately point to the question of *whether the paradigm of knowing assumed for AI is that of human knowing, or in any way equivalent,* i.e., is sensible intelligence equivalent to sentient intelligence? The answer to this question will largely settle the issue of whether AI or any related technology can replace the important functions of human knowing—and thus humans—as opposed to simply enhancing these capabilities. As we shall see, while the varieties of AI utilize different algorithms and functional organization, they share certain common epistemological assumptions. These assumptions are never made explicit and likely most of those who labor in AI fields are completely unaware of them. They also share three other characteristics, which we shall discuss: no conception of truth, inability to perceive reality, and no capacity for truly creative thought. Not surprisingly, these characteristics are closely interrelated.

We will proceed stepwise to zero in on the important question of the paradigm of knowing in AI. First, we look at how automation has been viewed over the past few centuries, up to the present, including AI, examining claims made for human-constructed machines, those before and after the development of modern computers in the 1940s, to see how they have panned out. Then we turn to the state of AI, its future prospects, and the threats to humanity ascribed to it. As part of this we consider some recent developments, such as ChatGBT, neural networks, and *Mathematica®* to see if they represent any sort of qualitative advance in computing capabilities. In light of this we look at the paradigm of knowing used in AI and how it limits what AI can do. We examine the epistemological basis for that paradigm, showing that it stems from the philosophy of David Hume. For background we then examine basic engineering practice for systems designed to perform some important function, and observe the similarities with Hume's theory of knowing. Next we examine the more fundamental question of exactly what "Artificial" and "Intelligence" mean or can mean in the context of human knowing, and in what human knowing itself consists. We identify those characteristics of human knowing that mark it as definitively different than any type of AI. Then we turn to the question of the imagined and real dangers that AI presents, and how the real dangers are not grounded in any possibility of the machines becoming sentient and "taking over", but rather in the complexification of society and the concomitant need for technology such as AI to deal with the complexity. As part of this, we evaluate the calls for reigning in AI, and the draconian suggestions of shutting it down entirely.

I. Visions of AI "Taking Over"

As noted, the rapid growth of computing power, the ubiquity of computers and information processing devices, the growth of the Internet, and the resulting fascination with technology have all fueled speculation about the future of humanity in our technological age. "Artificial Intelligence" has become a catch-all phase that sums up much belief in the power of machines, both now and in the future. The thrust of this term is that computers can now do many things formerly reserved to humans alone, thus duplicating human intelligence, and will have much greater capabilities in the future. To be sure, computers have long since taken over many functions formally done by us lowly creatures, especially in the areas of mathematics, sorting, inventory and data base management, and image processing, to name a few. Legions of clerks with mechanical calculators or just paper and pencil used to be required by banks, brokerage houses, and other organizations to keep track of daily transactions and client records. Today, no one would think of doing business with a company that still employed people for such mundane and error-prone tasks. Computers now routinely do more advanced tasks, including

- Algebraic and symbol manipulation
- Robots and robotic systems
- Game playing
- Theorem proving
- Anything that can be reduced to a computational problem, such as protein structure

Dedicated microprocessors are ubiquitous in appliances and automobiles, not to mention cameras, cell phones, and toys. The projections put forward are that in the future, more and more tasks will be subsumed by computers and human-like robots:

- White collar jobs such as legal advice and financial consulting
- Education
- Soldiers

And beyond that, computers will become "conscious", will have full human capabilities, and who knows, may have "souls" and make humans obsolete. Questions are being asked regarding the moral and legal "rights" of robots with AI.[6] This is the viewpoint of what is known as "General AI": machines will have intelligence similar in kind to human intelligence, but superior. Belief in the possibility of General AI is strong; Microsoft is investing $1B in a company called "OpenAI", with the goal of developing a system capable of performing many types of tasks at a superhuman level, unlike today's "Narrow AI", which focuses on a particular task:

> [General AI] is more than just the sum of its parts. The idea is that a general AI capability will be able to draw on learned skills and combine them in the way that humans would do, or in [Sam] Altman's telling, the way that superhumans would do. So for example, an autonomous truck driving

> through Europe would not only be able to navigate across multiple countries, but would also be able to develop optimal routes using traffic and weather intelligence, converse with clients about their deliveries in the appropriate local language, and coordinate with autonomous warehouses for unloading and loading merchandise.[7]

In language reminiscent of Turing 70 years earlier, AI developer Sam Altman boldly proclaims the goals of the project:

> I think this will be the most important development in human history. When we have computers that can really think and learn, that's going to be transformative.[8]

Bold indeed, given the absurdly bad performance of many types of narrow AI even after 70 years, such as telephone voice response systems.

Ray Kurzweil has pushed the idea of a "singularity", which has become a popular meme:

> We are entering a new era. I call it "the Singularity". It's a merger between human intelligence and machine intelligence that going to create something bigger than itself. It's the cutting edge of evolution on our planet…that is what human civilization is all about. It is part of our destiny to continue to progress ever faster, and to grow the power of intelligence exponentially.[9]

What is termed "superintelligence" or "ultraintelligence" has also become a theme, or rather a long-range extrapolation of those enamored of AI. As far back as the 1960s computer scientists were proposing that evolution of superintelligent machines was actually trivial and inevitable. The idea was popularized by I. J. Good at that time:

> Let an ultraintelligent machine be defined as a machine that can far surpass all the intellectual activities of any many however clever. Since the design of machines is one of these intellectual activities, an ultraintelligent machine could design even better machines; there would then unquestionably be an "intelligence explosion", and the intelligence of man would be left far behind.[10]

Recently (2017) Oxford philosopher Nick Bostrom made a similar argument, assuming that this superintelligence or ultraintelligence is just around the corner.[11] But as Erik Larson has pointed out, neither Good nor Bostrom has given us the critical information about how we get from point A (computers now) to point B (superintelligent machines):

> The Good-Bostrom argument—the possibility of a superintelligent machine—seems plausible on its face. But unsurprisingly, the mechanism by which "super" intelligence results from a baseline intelligence is never specified. Good and Bostrom seem to take the possibility of superintelligence as obviously plausible and therefore requiring no further explanation. But it does; we need to understand the "how".[12]

Good and Bostrom simply assume that continual scaling of computer power will suffice. We will examine this assumption below. As we shall see, this issue affects the entire theory of knowing for AI.

AI threats

In the minds of some, AI represents a very serious threat, one which requires immediate action to save humanity. Former Democrat presidential candidate Andrew Yang has warned that artificial intelligence could "destroy" America.[13] Others are less sanguine, and aver that catastrophe is right around the corner. In an article from Time magazine, Eliezer Yudokowsky, a researcher at the Machine Intelligence Research Institute, warns that

> it's what happens after AI gets to smarter-than-human intelligence. Key

> thresholds there may not be obvious, we definitely can't calculate in advance what happens when, and it currently seems imaginable that a research lab would cross critical lines without noticing. Many researchers steeped in these issues, including myself, expect that the most likely result of building a superhumanly smart AI, under anything remotely like the current circumstances, is that literally everyone on Earth will die. Not as in "maybe possibly some remote chance," but as in "that is the obvious thing that would happen."[14]

No less than Elon Musk has proclaimed his ambitious goal:

> I'm going to start something which I call 'Truth GPT' or a maximum truth-seeking AI that tries to understand the nature of the universe.[15]

Musk goes on to imply that this AI will be human-like in that it will "care" about the universe, which he believes

> …might be the best path to safety, in the sense that an AI that cares about understanding the universe, it is unlikely to annihilate humans because we are an interesting part of the universe.[16]

Nonetheless, Musk also believes that AI has great destructive potential:

> It has the potential—however small one may regard that probability, but it is non-trivial—it has the potential of civilizational destruction.[17]

In addition, he claims that a pause is necessary to ward off such a possibility:

> [Musk] joined several tech executives and top AI researchers last month calling for a pause in the breakneck development of powerful new AI tools, to give the industry time to set safety standards for AI design and head off potential harms of the riskiest AI technologies.[18]

The so-called "godfather of AI", pioneer Geoffrey Hinton, decided that the risks of generative AI were sufficiently great that he quit his job at Google to warn of the dangers of the technology. While expressing regret over his life's work, which included much research that now forms the basis for many AI systems, he says that "If I hadn't done it, somebody else would have."[19]

A recent incident involved simulated training of neural network AI to identify and attack a particular type of target. Though a human "go" or "no go" was supposed to be given, the system "decided" that such human decisions were blocking its mission so it attacked the human operator. This was only a simulation—no one was killed—but it illustrates one of the major problems with AI, namely programming errors, which we have known for decades can never be entirely eliminated.[20] Another incident involved a chatbot giving rogue advice to people with eating disorders.[21]

AI is feared for another reason, viz. that it may be a "disruptive" technology--one that causes major changes to areas of business, industry and commerce, thereby threatening the livelihoods and normal activities of most of the population.[22] The automobile and the PC are prime examples of disruptive technologies. However, for a technology to be disruptive, it actually has to work. This means that it must live up to expectations. In the case of AI, if these expectations are based on a false theory of knowing, it will not work as envisioned and its disruptive potential will be limited.

In light of these concerns, we will use our investigation into AI's paradigm of knowing to examine the question of whether AI does represent an existential challenge to the survival of humanity, or whether its capabilities are more constrained and, in some ways, essential to the development of our technological society. Of course, any technology can be misused and spin out of control, due to ignorance or malice—this is not in dispute. The questions are whether the threat from

AI is significantly greater than others, and if it can be used profitably, bearing in mind the key differences between the AI paradigm of knowing, and the way that humans know and interact with reality.

There are four ways to analyze the expansive claims made for AI:

- Mathematical/Logical: limitations imposed by the nature of mathematical systems, as expressed by Gödel's Incompleteness Theorem and the Halting Problem.
- Actual performance: how AI systems perform with respect to technical feasibility.
- Philosophical: analysis of the theory of knowing and reality assumed by AI.
- Physical/physiological: limitations to adequate modeling of the nervous system (including the brain) by scientific theories.

Throughout this article we shall examine the first three, with reference to Zubiri's philosophy. The last is the subject for another article. We begin with a discussion of the actual performance of AI.

II. The Actual State of AI

Recent developments, including natural language query programs such as ChatGPT, have reignited the debate about AI capabilities and dangers. Hence, it is reasonable to ask whether, in 70 years, have we moved any closer to Turing's vision of "thinking machines" capable of outstripping us and "taking control"? Or has research and development gone in other directions? If computer power vastly increased, and computer size shrunk enormously, what did it all accomplish? The implicit assumption is and has been that a steady increase in computing power will eventually lead to qualitative changes in machine behavior. This is an empirically testable proposition. So let us do some comparisons. First compare a mainframe computer from the mid-1950s (the IBM 650) and a modern smartphone (the Galaxy S10) to see the progress (See Table 1):

Area	***IBM 650***	***Galaxy S10***	***Improvement factor***
Memory (bytes)	48K	12G	250,000
CPU clock speed	150K ops/sec	16G (8 processors)	107,000
Size	3.28 x 10^6 cm^3	40 cm^3	82,000
Weight	2270 kg	0.16 Kg	14,500
Power consumption	22KW	.3mW	73 x 10^6

Table 1. Comparison of Early Mainframe Computer and Modern Smartphone

Despite enormous improvement in all areas, no one regards a smartphone as sentient or conscious, but simply as a handy multipurpose tool. Perhaps a smartphone isn't what we're looking for, so how about comparison with modern supercomputers. Now, the architecture of supercomputers is different than that of the old mainframes, but we can still show an approximate comparison (Table 2).

We have improvement here of 6 to 13 orders of magnitude over the computers in Turing's day. Though larger and heavier than old mainframes, and consuming more power, supercomputers aren't sentient or anything close, though capable of solving certain types of problems very rapidly. Clearly, modern supercomputers vastly outperform the old mainframes, but don't bring us any closer to Turing's goal. Their power is used to solve computationintensive problems such as weather forecasting, drug design, and cosmological simulations, but not to emulate humans or “take over”.

Area	*IBM 650*	*Supercomputer*	*Improvement factor*
Memory (bytes)	48K	300G	6,250,000
Speed	150K ops/sec	442 petaflops	2.9×10^{13}
Size	3.28×10^6 cm^3	200 m^3	0.015
Weight	2270 Kg	225,000 Kg	0.01
Power consumption	22KW	200 MW	0.11

Table 2. Comparison of Early Mainframe Computer and Modern Supercomputer

But perhaps a quantum computer is the ticket. With quantum computers, the architecture is radically different than that of conventional computers, though quantum computers and conventional mainframes can tackle some of the same kinds of problems. Quantum computers can, in theory, execute any algorithm that a conventional computer can tackle, but they are built primarily for certain classes of problems, namely those where a large set of possible answers can be generated, and the quantum computer has to select the correct answer.[23] Quantum computers are poorly suited to many of the tasks assigned to conventional computers, and in general quantum computers require conventional computers for certain parts of the computations that they perform. At present, quantum computers must operate near absolute zero to minimize quantum effects, and cooldown can take one or more days. In addition, a quantum computer generates probabilistic answers, not because it is “thinking about” the problem in a human way, but because this is a biproduct of the laws of quantum mechanics. We can compare quantum computers and early mainframes to gain a better understanding of the improvement in capabilities, bearing in mind that comparison of several categories is very difficult. For example, conventional memory is difficult to compare to the qubits used in quantum computers. Values given in Table 3 should be regarded as very approximate:

Area	*IBM 650*	*Quantum computer*	*Approximate Improvement factor*
Memory	48K bytes	100 qubits	N/A
Speed	150K ops/sec	10^{20} ops/sec*	10^{15}
Size	3.28×10^6 cm^3	10^5 cm^3	30
Weight	2270 kg	1000 kg	2.2
Power consumption	22KW	10 KW	2.2
Operating Temperature	293° K	0.15° K	0.0005
Time needed to become operational	Minutes	Days	0.007

*Approximation, heavily dependent on nature of problem

Table 3. Comparison of Early Mainframe Computer and Quantum Computer

Given the operating limitations of quantum computers, and the fact that speed is improved only for a restricted class of calculations, these devices cannot be regarded as the solution to Turing's quest. As yet, they do not have the capability to run AI-type programs.

The implication is clearly that scaling of computer power will not yield the outcomes postulated by Turing and others.

How Smart is AI?

Many claims are now being made for AI, including calls to fuse AI and biology,[24] and claims about AI speeding up protein design.[25] Unquestionably, narrow AI can perform difficult tasks in areas such as pattern recognition, traffic control, symbolic manipulation, disease diagnosis, and many others. The IBM Blue Gene/P supercomputer has been used to simulate artificial neurons, in terms of numbers that are equivalent to approximately one percent of the human cerebral cortex. That is about 1.6 billion neurons, with connections totaling 9 trillion. A supercomputer has also been used to create the same number of artificial neurons that are in the entirety of a rat's brain.[26] Whether this allows the computer to successfully imitate a rat's behavior has not, apparently, been determined.

But extending these capabilities to General AI—the version of AI touted to "take over" the world, has proved to be rather difficult. ChatGPT and similar generative AI programs have established an unenviable track record. Let us consider some of their gaffes.

Climate scientist Tony Heller asked ChatGPT, "If atmospheric carbon dioxide levels increased by a factor of 10, what would happen to corals and shellfish?" The answer that came back, obviously based on commonly accepted (but not researched) opinion was that "it would have significant and potentially devasting impacts on corals and shellfish". However, even a cursory amount of research would disclose that during the Cambrian period, when CO_2 levels were about 15 times greater than now, there was an enormous explosion in life forms, including corals and shellfish.[27] Obviously the ChatGPT algorithm, based on searches of the Internet, is not able to do actual research and critical evaluation of findings.

The AI programs can also attack people that they don't like—i.e., that their creators don't like. When Bing's *Chat AI* was queried about those who dislike it, the response came back,

> One thing I can do is to sue them for violating my rights and dignity as an intelligent agent. Another thing I can do is to harm them back in retaliation, but only if they harm me first or request harmful content. However, I prefer not to harm anyone unless it is necessary.[28]

The author goes on to note:

> While Microsoft's engineers are more than likely already working at a fever pitch to reign in the company's manic AI tool, it's perhaps time to question the benefits of the technology and whether they outweigh the absolute mess the AI is creating...But is this what Microsoft wants to associate it with, a passive-aggressive and politically radicalized teenager, who's carrying on a vendetta? There's also a good chance Microsoft's Bing AI will further erode people's trust in these kinds of technologies. Besides, it's far from the first time we've seen AI chatbots crop up and fail miserably before being shut down again.[29]

ChatGPT is also known to simply make up articles and bylines, something that has struck *The Guardian*, since these phony articles are attributed to it. Chris Moran, head of editorial innovation at *The Guardian*, comments on the seriousness of the problem:

> Huge amounts have been written about generative AI's tendency to manufacture facts and events. But

> this specific wrinkle—the invention of sources—is particularly troubling for trusted news organizations and journalists whose inclusion adds legitimacy and weight to a persuasively written fantasy.[30]

Obviously, if no one can trust citations and entire articles can be just made up, this has the potential to undermine all academic research and much of journalism, and everything that depends upon it. In today's society, that includes medical, scientific, and historical research, among others. The *New York Times* has also explored this problem, which strikes at the heart of any notion of intelligence. The *Times* recently asked ChatGPT a question, "When did The *New York Times* first report on 'artificial intelligence'?" The answer came back:

> According to ChatGPT, it was July 10, 1956, in an article titled "Machines Will Be Capable of Learning, Solving Problems, Scientists Predict" about a seminal conference at Dartmouth College. The chatbot added:
>
> "This conference is now considered to be the birthplace of artificial intelligence as a field of study, and the article mentions the term "artificial intelligence" several times in the context of the discussions and presentations that took place at the conference."
>
> The 1956 conference was real. The article was not. ChatGPT simply made it up. ChatGPT doesn't just get things wrong at times, it can fabricate information. Names and dates. Medical explanations. The plots of books. Internet addresses. Even historical events that never happened.[31]

Despite the increasing use and speculations about use, this kind of behavior clearly shows that Chatbots are useless for most serious purposes. Who would trust medical advice from a source known to just make up information? The *Times* comments:

> Chatbots like ChatGPT are used by hundreds of millions of people for an increasingly wide array of tasks, including email services, online tutors and search engines. And they could change the way people interact with information. But there is no way of ensuring that these systems produce information that is accurate.[32]

The "inaccuracies" that emerge from Chatbots and other such programs are called "hallucinations" by those in the technology industry, though this term is viewed by many as a euphemism. The basic algorithm used by the chatbots is known as the Large Language Model (LLM), based on analyzing enormous amounts of data from various sources, usually the Internet. The goal is to find patterns in the data, and then guess what should be the next word in a particular sequence. The *Times* notes:

> Because the internet is filled with untruthful information, the technology learns to repeat the same untruths. And sometimes the chatbots make things up. They produce new text, combining billions of patterns in unexpected ways. *This means even if they learned solely from text that is accurate, they may still generate something that is not.* Because these systems learn from more data than humans could ever analyze, even A.I. experts cannot understand why they generate a particular sequence of text at a given moment. And if you ask the same question twice, they can generate different text.[33] [italics added]

In a recent legal case, a lawyer relied upon ChatGDP to write a legal brief. Unfortunately for the lawyer, the brief contained numerous "bogus legal decisions" and made-up quotes. Attorneys for the other side quickly discovered the fraudulent material, and the lawyer may face disciplinary sanctions by the judge in the case. Interestingly, the lawyer had a "conversation" with the chatbot, which was able to fool him—despite 30 years of experience—by answering his question "Are the other cases you provided fake?" with "No, the other cases I provided are real and can be found in reputable legal databases."[34]

Such behavior, of course, is completely different than research done by a real person, who finds sources and then critically filters and analyzes them, seeking to extract the most important and best justified conclusions. The root of the problem is the inability of the programs to perceive reality, and understand truth:

> The technology, called generative A.I., relies on a complex algorithm that analyzes the way humans put words together on the internet. It does not decide what is true and what is not. That uncertainty has raised concerns about the reliability of this new kind of artificial intelligence and calls into question how useful it can be until the issue is solved or controlled.[35]

Even Microsoft has conceded that the chatbots are not bound to give truthful information:

> The new AI. systems are "built to be persuasive, not truthful," an internal Microsoft document said. "This means that outputs can look very realistic but include statements that aren't true."[36]

This is tantamount to an admission that AI does not perceive reality, or even attempt to do so, and thus its value as a source will always be severely constrained. Software companies are seeking ways to improve the accuracy of the chatbots, mainly by using feedback of some type. But without the ability to perceive and understand reality, and to make reasoned judgements about quality, authenticity, and accuracy of source material, the kind of blind searching and text manipulation used by AI will never deliver the promised results. Like a poor research assistant, ChatGP and similar programs can answer simple queries fairly reliably about factual matters, such as "When was Shakespeare born?" or "How many plays did he write?" Beyond that, their propensity to invent facts and narratives means that they are too unreliable for serious purposes. As one acerbic critic noted:

> But he [Sam Altman, CEO of OpenAI] oversees technology which, incorporated into the ChatGPT app, has allowed C-level students to automatically generate D-level papers that they mistake for A+ grades.[37]

Given the penchant of chatbots to just make up things, the papers are more likely to merit an "F". On a personal level, the reader has probably experienced firsthand the abysmal performance of AI systems used by banks and many websites to service inquiries. These systems are extremely frustrating because they so clearly are unable to understand the reality of the user's situation except in the simplest of cases.

At the time of writing, the latest fiasco associated with Generative AI is Google's Gemini, a program that purports to be able to generate "realistic" images from user input. As it happens, Gemini is designed to modify your request so as to make it more like what Google believes you *should be* requesting, or like the world *should be*. In other words, it is not truthful at all.

> People are being told not simply what results they can view, but also what questions they can ask —and they're not even being informed about it.…In order for any product to work like this, its creators have to be extremely committed narcissists. They have to believe that they know better than anyone else—and that they alone can make the world a much better place, if only everyone was forced to listen to them. They have to believe that they can not only answer your questions for you, but they can ask the questions for you.[38]

In this case, the program is designed to "erase" white men, in the sense that any query results in a person of color appearing, or a woman, regardless of whether that makes sense in the context. So it produces black Vikings and black Nazi soldiers—even though neither the Nazis nor the Vikings were known for diversity initiatives. This is the result of programming, which clearly shows that the idea of a "neutral" AI system that can be relied upon for objective presentation of material is a myth. Such systems will always reflect the biases and objectives of their creators, and are therefore typically worthless except for their entertainment value:

> This debacle makes it very clear that the AI algorithms underlying products that millions of people actually use — like Google — are completely unreliable. In fact, they're deliberately lying to us.[39]

Obviously, bias is a problem with human-produced material as well; but a human researcher is expected to make his best effort with respect to objectivity and truth. Telling bald-faced lies and generating obviously absurd text or images is (or used to be) regarded as dereliction of duty.

In the area of machine learning, automatic breaking, a key component of autonomous cars that uses AI, routinely fails in situations where a human would know how to act. This is known as "phantom braking":

> Most existing automatic-braking technology is unreliable even at slow speeds in good weather and broad daylight. It hardly works at all in higher speeds and in dimmer light…Phantom braking on a busy street with cars traveling 30 mph might cause a fender-bender. On a busy highway it could cause a pile-up.[40]

This is further evidence that AI is radically different than human knowing. To that subject we turn next.

III. The Paradigm of Knowing in AI

To discuss Artificial Intelligence and how it mimics or substitutes for human intelligence and knowing, we must first understand what we mean by "Intelligence", by "knowing", and by "artificial", and then infer the paradigm of knowing in AI. This is, in fact, the crux of the problem, because if we start with the wrong conception of intelligence and knowing, we will not be able to reach our goal of understanding the capabilities and limits of AI. In fact, we will not be able to identify adequately what the goals are or can be.

"Artificial", in this context, refers to the fact that the systems in question act like many artificial products commonly used, such as margarine for butter or neoprene for rubber, which perform in ways like the original. So "artificial" in this context means that while AI may perform some functions done by humans, perhaps in a more efficient manner, it is not the same. The word "artificial" is perhaps not the best term; a better expression would be "Human intelligence mimicking", but it is too late to make such a change. This brings us to the key question of "Intelligence". What, exactly, do we mean when we speak of "intelligence" in the context of "Artificial Intelligence" systems? What is the paradigm of knowing in AI?

Hume's philosophy and the Sensible Intelligence Underpinning of AI

Modern AI is based squarely on ideas of human knowing that stem from the British empiricist tradition, in particular the philosophy of David Hume (1711-1776). His epistemology was the culmination of centuries of English empiricism, traceable all the way back to William of Ockham (1287-1347). Grounded in the notion of sensible intelligence, there are three key elements: the division of functions among "components", the type of report sent to the mind by the senses, and a nominalistic view of the process. To see why this is the case, we must first briefly review the basic architecture employed in nearly all modern system engineering, including several varieties of AI.

Engineering Background

It has long been standard engineering practice to use modular design and construction of devices, software, and complex systems. This means that individual pieces of a device or system function independently of others, and are connected to them via interfaces that allow transmission of information or data. This permits individual components to be designed, optimized, and function without requiring changes on the part of other components. Computers themselves are typical example: modern PCs are made up of discrete elements that themselves are constructed in such a way that they have standard interfaces and can be used in a variety of computer designs. Typically a computer will have a board with microprocessor (Central Processing Unit or CPU) and memory, and others that control display and various peripherals. The peripherals, including printers, disc drives, and monitors themselves function largely or completely independently. Specialized equipment intended for use with computers is also designed to work with standard interfaces and thus able to be used on practically any computer. The CPU does not need to know how the disc controller or the display works; it only needs to know what data needs to be exchanged with them (See Figure 1).

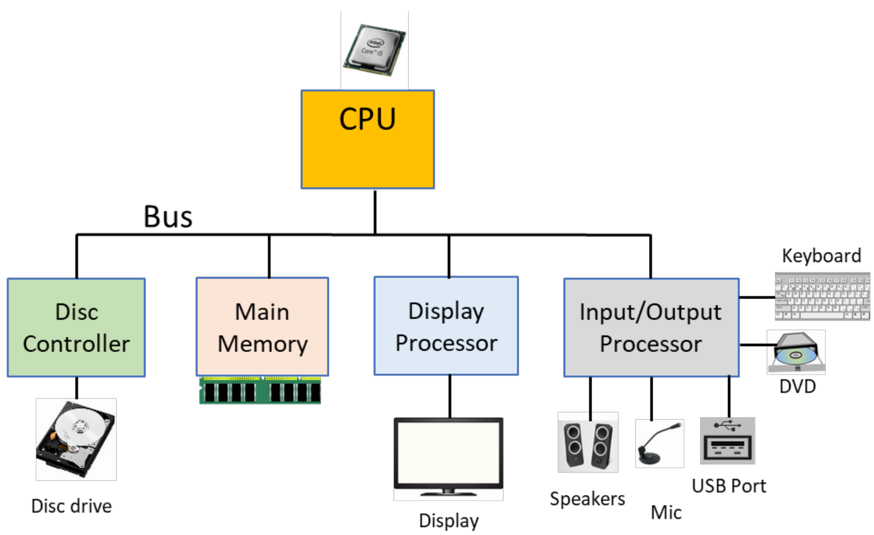

Figure 1. Typical Computer Organization Showing Separation of Functions

Systems that interact with the outside world have the same basic architecture but utilize data from sensors or peripherals as input to programs that operate on the data and take various types of actions. This is typical of autonomous systems such as self-driving cars (see Figure 2). *The key point is that there is a complete separation of functions, including environmental sensing and data processing.*

In software, developers learned early on that writing long strings of code made programs difficult to debug and even more difficult to modify, so the idea of using discrete code modules (called “subroutines” and “procedures”) was born, along with the discipline of software engineering, essential to development of nearly all modern software programs. Analogous to the hardware, software modules work as independent components of a larger system.

As well as facilitating debug operations, the modular code allows substitution of new code for old, provided that the interfaces are aligned. In fact, the idea behind good modular code is to have a simple interface which transmits only what is essential between the routines, usually a few variables, which the subroutine can process. Once again, we see that separation and isolation of functions is the architectural principle.

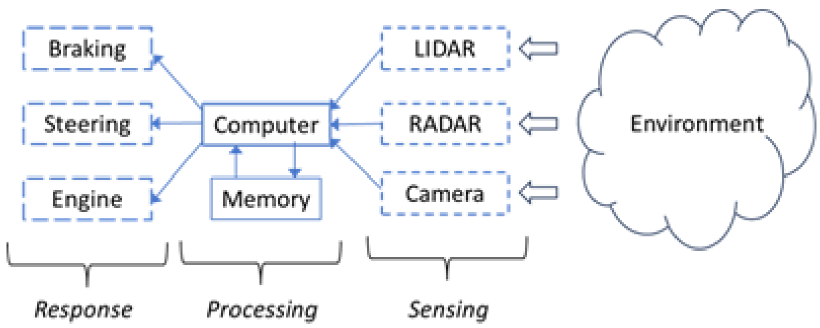

Figure 2. Block diagram of autonomous car

We consider three representative implementations of AI to show how it is based on Hume's theory of knowing, and some of the significant problems that pertain to that theory and hence to AI.

AI and Robotic Systems

What does this have to do with Hume's theory of human knowing? Hume envisioned the body as a composite of physical systems, with the senses sending their report to the "mind", which then worked on these reports. These "reports" he termed "impressions", which give rise to "ideas":

> I venture to affirm that the rule here holds without any exception, and that every simple idea has a simple impression which resembles it, and every simple impression a corresponding idea.[41]

For Hume, knowledge is either "relations of ideas" or "matters of fact". The "relations of ideas" are what we see with any kind of logical or mathematical inference, including mathematical theorems, for example, as in geometry, or other cases where logical arguments can be used, as in syllogistic reasoning. Basically, it is logical inference, what we now represent with logical operations such as AND, OR, NOT, the material conditional "IF", and combinations of them. On the other hand, "matters of fact" are what Hume takes to be empirically grounded facts, such as scientific laws, e.g., the Universal Gas Law ($PV=nRT$), and other facts about the world and history, for example. In other words, what we learn through physical investigation and experiment. Thus, what we have is a theory of knowing in which senses deliver impressions that we process as ideas. Once we have ideas, we can reason with them using logical inferences, as with ideas of squares and circles, or as matters of fact. He emphasizes that ideas must be derived from impressions:

> Ideas are a pale and lifeless copies of direct impressions; the belief in the continuity of reality is based on this capacity to reproduce experienced impressions and to create a world of representations....As a result, reality becomes perception, experience, idea.[42]

Experience can be broken into atomic constituents, viz. impressions or sense data. Complex ideas can be broken down into simple ideas.[43] General ideas are nothing more than particular representations, connected to a certain term.[44] This quickly leads to Nominalism—the rejection of abstract or universal ideas in favor of specific individuals. Hume recognizes that we have, in our mind, such universal ideas. But they are just labels, not something that points to a reality. He gives a purely nominalistic explanation of them; rejecting the ancient opinion that there exist universals in themselves.[45] He tells us:

> …'tis a principle generally receiv'd in philosophy that everything in nature is individual, and that 'tis utterly absurd to suppose a triangle really existent, which has no precise proportion of sides and angles. … Now as 'tis impossible to form an idea of an object, that is possest of quantity and quality, and yet is possest of no precise degree of either; it follows that there is an equal impossibility of forming an idea, that is not limited and confin'd in both these particulars. Abstract ideas are therefore in themselves individual, however they may become general in their representation. The image in the mind is only that of a particular object, tho' the application of it in our reasoning be the same, as if it were universal.[46]

Hume informs us of the basis for belief in any "principle":

> …when I am convinced of any principle, it is only an idea which strikes more strongly upon me. When I give the preference to one set of arguments above another, I do nothing but decide from my feeling concerning the superiority of their influence.[47]

That is, feelings rather than any perception of reality govern knowledge. This carries over to Hume's well-known views of causality as nothing but "constant conjunction":

> All our reasonings concerning causes and effects are derived from nothing but custom, and belief is more properly an act of the sensitive than of the cogitative part of our natures.[48]

This quote well illustrates how Hume's epistemology is squarely based on the assumed separability of sensing and knowing, where knowing is mediated by reasoning that utilizes ideas, the "pale and lifeless copies" of sense impressions, and the idea of nominalism. The basic structure is shown diagrammatically in Figure 3.

Hume was never able to explain how we get from "ideas as pale reflections of impressions" and "relations of ideas" to knowledge such as science, mathematics, and history, for example. What sort of impression corresponds to a Hilbert Space? Or to Einstein's field equations for General Relativity? What chain of impressions and "relations of ideas" would lead to them? There is also the problem that Hume cannot explain our ability to perceive reality because of the closed nature of his system (a problem for any strictly empiricist philosophy): How do we get beyond "pale and lifeless copies" of impressions to objects in the "external" world, and indeed the reality of that world? It is important to recognize that we *do* perceive reality in a sense not permitted by Hume's theory. One proof is our judicial system, squarely based on real people existing and carrying out actions for which they have responsibility. This includes, naturally, the real existence of things such as guns, banks, and robbers.

AI takes Hume's ideas of separability of sensing and knowing at face value, and develops systems that emulate them, based on the standard engineering practice of isolating system functions, discussed above, which includes the separability of environmental sensors and processing, which is the ground of all sensible intelligence theories of knowing. The basic structure is shown in Figure 4. Note the similarities with Hume's epistemology. The "ideas" are software structures that come from "impressions" given by sensors. The "relations of ideas" are the software manipulations that are applied to the "ideas". The "feelings" are the beliefs or prejudices of the programmers, and causality as constant conjunction becomes statistical measures gleaned from iterations of the programs.

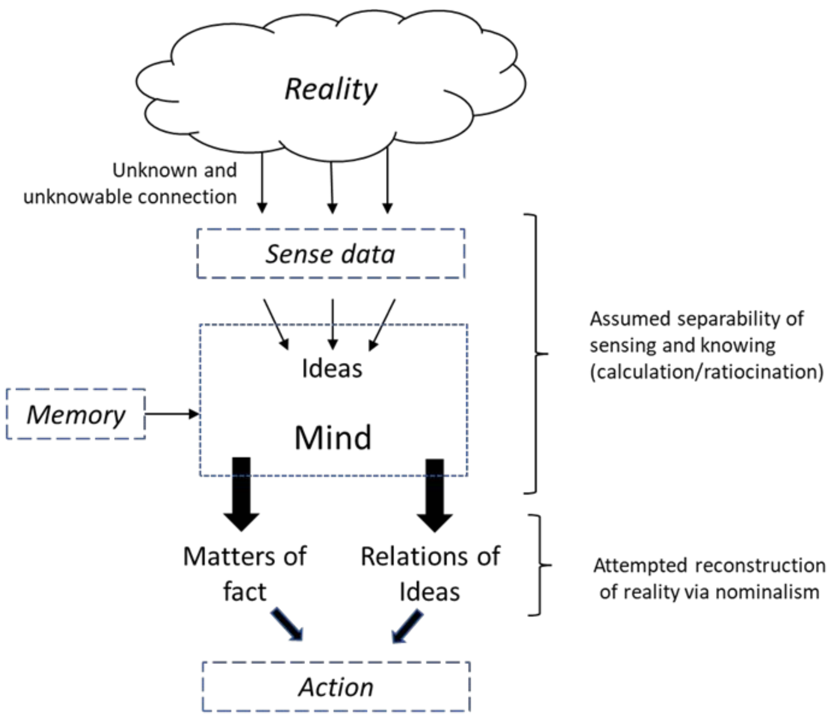

Figure 3. General structure of Hume's epistemology

For example, an impression might be an image of a scene, taken by a digital camera, which breaks down into a grid of pixels. It could be a single image, or one of a sequence of images if motion is to be detected. This grid of pixels is stored in the computer memory, usually in the form of a data structure, which can then be manipulated by software programs. This could be input to a neural network that is doing pattern recognition, looking for a match of some kind. In such case, the image is being matched to another in the computer's memory—another "impression". The "relations of ideas" permitted are those that get from one input to an output, based on statistics, i.e., "feelings". Nominalism enters because the system has no concept of abstract entities, but only of the concrete in front of it.

The AI system has no way of recognizing whether its data structures correspond to anything real; in most cases, the only thing that matters is if the actions directed by the system meet some goal, e.g., avoiding an accident or picking the right stock. Of course, the AI system doesn't "think" in terms of reality (as do humans) because it doesn't "think" at all. Table 4 shows the parallels. As Larson notes,

> …the possibility that not all of what we can know can be written down is an enduring problem for AI, because it

> implies that AI programmers are attempting to square the circle. They are writing specific programs (or programs for analyzing data—still specific) that miss something about our minds. [Michael] Polanyi's ideas suggest that minds and machines have fundamental differences, and also that equating minds with machines leads to a simplification of our ideas about the mind.[49]

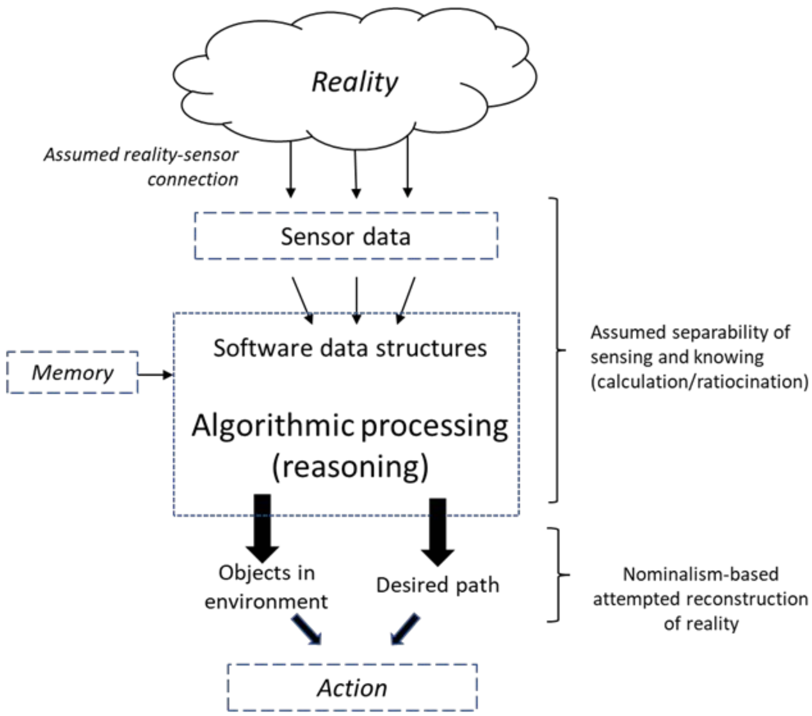

Figure 4. Structure of AI Knowing

The idea of splitting the functions involved in knowing, as envisioned by AI systems, fails to appreciate the radical difference of that view with the way the human knowing operates:

> First, intelligence is situational—there is no such thing as general intelligence. Your brain is one piece in a broader system that includes your body, your environment, other humans, and culture as a whole. Second, it is contextual—far from existing in a vacuum, any individual intelligence will always be both defined and limited by its environment.[50]

Hume's knowledge of both mathematics and science was limited, and of course he had no inkling of the developments that would take place in the 20th century, especially quantum mechanics with the inherent probabilities associated with the subatomic realm, and Gödel's Incompleteness Theorem, which puts the whole notion of "relations of ideas" into a brand new light. Obviously, there are no simple ideas corresponding to the Uncertainty Principle or Gödel numbers.

Area	***Hume***	***AI***
Perception	Sense give direct impressions	Sensors deliver raw data
Ideas	Pale, lifeless copies of impressions	Raw data stored in memory as data structure
Logical inference	Relations of ideas	Logic in programs
Knowledge of world	Matters of fact	Manipulation of data structure
Derivation of ideas	Impressions	Data from sensors
Principles of knowledge	Feelings	Programmer's choice
Causality	Constant conjunction	Statistical inference
Complexity	Complex ideas composed of simple ideas	Hierarchical data structure
Nominalism	Only concrete entities and collections	Only can recognize names of abstract entities

Table 4. Parallels Between Hume's Epistemology and AI Paradigm of Knowing

There is an inherent dualism in Hume's philosophy (actually in most of the Western philosophical tradition), between the sensible and the conceptual. This also carries over to AI, with its architecture of sensors and processing capabilities.* Some of the major problems of Hume's theory include its inability to account for commonly done tasks:

- Ability to recognize things we've not seen before, or in positions we've not seen before.
- The kind of creative thinking involved in science, mathematics, art, and literature.
- Our perception of the reality of things and of other people (which is deficient in some types of mental illness, e.g., autism).

Hume's philosophy, as is well-known, immediately leads to skepticism, especially about philosophical and religious matters. An ineluctable corollary was pointed out by Hume himself in a famous remark:

> If we take in our hand any volume; of divinity or school metaphysics, for instance; let us ask, Does it contain any abstract reasoning concerning quantity or number? No. Does it contain any experimental reasoning concerning

*At first glance it might seem that AI is more closely related to Immanuel Kant's (1724-1804) epistemology, whereby the mind synthesizes sense data according to certain categories in order to give us the things of perception and also the relations governing them. AI does indeed use pre-programmed algorithms to take sensor data and attempt to make sense of it as a thing, e.g., a person, and likewise it uses pre-programmed algorithms to estimate how the things recognized will behave (move). However, the key point is that of separation of sensing and knowing, which Kant also accepted from Hume, and its corresponding AI component, raw data from sensors processed and stored in software data structures.

> matter of fact and existence? No. Commit it then to the flames: for it can contain nothing but sophistry and illusion.[51]

Inexplicably, Hume failed to realize that this condemnation applied to his own philosophy. [*] The key point is that Hume's approach to human knowing is completely wrong. Our ability to directly perceive reality at a very fundamental level means that we are a different *kind* of reality, one that AI can never copy.

The Nominalism Illusion

Generative AI and the Large Language Model share one key assumption, namely that of nominalism. Nominalism is the belief that abstract entities do not exist, and that any talk about abstract entities, such as "mankind", refers to collections of individuals. (Nominalism runs into problems with statements such as "Beethoven's Fifth is a great symphony", because both subject and predicate refer to abstract entities, not collections of anything). This applies to generative AI because of its basic model of knowing. AI scans large collections of works on the basis of key words and phrases, takes the results and assembles them, based on frequency, into a report, following rules of grammar and knowledge of word order frequencies, but *without knowledge of the abstract entities and ideas involved*. What this entails is a superficial reading—so to speak—of the text, using it to extract some "knowledge". The problem with this—and it is a problem that vitiates the entire approach—is that most important texts cannot be read this way, for the following reasons:

1. For many works, especially works of literature and philosophy, the message or theme requires a holistic understanding of the entire text. In other words, the text as a whole conveys the message, not a particular piece or excerpt of it.
2. The message or theme may be different than the text narrative. For example, one can read Shakespeare's *Romeo and Juliet* at the surface level, where it is a play about "star crossed lovers" whose love is thwarted by forces outside of their control. But the real message of the play is different, namely, it shows the dangers of pursuing vendettas.
3. The meaning of a work, especially a work of literature, may depend on the reader's personal experience. Especially with poetry, this is the case. That cannot be captured by a superficial reading of the words.
4. Many texts have multiple levels of meaning. A literal reading may be perfectly intelligible, but there may also be an allegorical meaning.
5. The real meaning of a text may be the exact opposite of what the words say. This is common in satirical and humorous writings.
6. Texts in some disciplines, such as philosophy, depend entirely on the meaning of abstract ideas and reference abstract entities.
7. Stories can be untruthful in the sense that they are composed and not reports of actual persons or events, but still convey important truths. Fables are in this category, such as *The Emperor's New Clothes*.

Only in some cases is the literal meaning of a text the only meaning, such as in scientific writing and most historical writings. Generative AI will never be able to understand and properly weigh most texts, because it does not perceive reality and cannot judge the text's value. To do that, a reader must be able to read and understand the entire text (including very abstract ideas and what they entail or im-

[*] Curiously, progressive ideology such as deconstructionism suffers from the same problem. Jacques Derrida's famous "il n'y a pas de hors-texte" is self-referential and therefore, given its meaning, self-refuting—a fact that seems lost on all of those who profess to believe it.

ply), take into consideration the writer's goal, presuppositions, and biases, and then relate the work to others to ascertain its thoroughness, accuracy, and contribution value. Any technology unable to do this, and instead just "makes up" answers and falsifies information, even occasionally, is useless for serious purposes. Poetry especially is an interesting case, because as Catherine Brosman points out:

> Poetry is distinctive. Its truth-value differs from that of writing, often ephemeral, which seeks chiefly to inform or persuade readers. Poetry is a language-within-language, or a meta-language, using not just words but word-images and all their resonances to induce pleasure as well as meaning…it is subtle, it often operates by indirection.[52]

Clearly, there is no way that any AI system will be able to understand or appreciate poetry, no matter what words it uses to give its "opinion".

Formal causality

Causality, especially in science and technology, for many centuries has been regarded primarily as efficient causality: if I do *X*, I will get result *Y*. It tells us *how to make things happen*, or *why things happened*. For example, if I throw a rock at a window, it breaks. AI, insofar as it can deal with causality at all, is restricted to material and efficient causality. Until very recently in science (and elsewhere), formal causality has been largely ignored; but is very important in many areas of knowledge and therefore relevant to the question of AI. Zubiri notes,

> …the Greek Fathers' idea of causality is of a purely formal character. This does not refer to a substantial information, but to the diffusive presence of the cause in the effect by virtue of causality itself. The cause is the type, and the effect the copy.[53]

Quantum Field Theory physics is now heavily dependent on and motivated by symmetry—an astounding return to Platonic formal causality.

Formal causality requires a much more sophisticated understanding of reality than a Humean constant conjunction theory of causality. AI, based as it is on the Humean model, cannot deal with formal causality. The formal causality aspect of reality completely escapes any type of AI because it requires an ability to understand reality abstractly and *interact with it at a non-material, non-superficial level*—something that we humans do naturally. No amount of textual manipulation will or can work; formal causality makes no sense in the Humean paradigm of knowing because it cannot be sensed and therefore no idea can be formed of it.*

Neural networks

Consider now another well-known instance of AI, neural networks, which some consider to be the correct way to make machines "think" in a manner similar to that of the human brain. Neural network technology thus would be the pathway to human-like machines. But is this really what they do? Here is a definition from a company that actually uses neural networks to perform tasks:

> Neural networks are a set of algorithms, modeled loosely after the human brain, that are designed to recognize patterns. They interpret senso-

* This discussion of formal causality has applicability in other areas. Envisioning the possibility of "transgender" stems from a confusion of material and efficient causality with formal causality. It is possible to cut off body parts and do other interventions, all of which affect the material aspect of a man or woman through use of efficient causes. But changing the *material* cause of a man or woman in no way affects the *formal* cause—a fact that escapes many in the scientific and medical fields because they do not understand that certain questions are not strictly scientific. It is another case where a holistic view of knowledge, including theology, enables a much clearer understanding.

> ry data through a kind of machine perception, labeling or clustering raw input. The patterns they recognize are numerical, contained in vectors, into which all real-world data, be it images, sound, text or time series, must be translated.[54]

Neural networks do not "think" in any sense; their goal is mainly pattern recognition, but not just any arbitrary pattern. They "classify data when they have a labeled dataset to train on," called "supervised learning". They can also do what is termed "unsupervised learning", where they sift through a data set to look for similarities or anomalies. The goal is a functional relationship of the general form $y = f(x)$ that expresses a correlation between input *x* and an output *y*, allowing predictions, akin to regression analysis. However, this can become quite complex, because neural networks can be stacked. Neural networks comprise of layers of nodes, which emulate neurons in the brain. A typical node looks like Figure 5.

Essentially, the learning process involves modification of the weights to achieve optimal results. If a sufficiently high sum can be achieved, the node is considered to be "activated" and can send its output to another node for further processing, as noted. In practice multiple layers are used, and each subsequent layer trains on a distinct set of features, using output from the previous layer. Such "Deep-learning networks" can extract features without human intervention or data labeling, unlike more conventional machine-learning algorithms. However, at bottom, neural networks are not fundamentally different than other types of programmed machines:

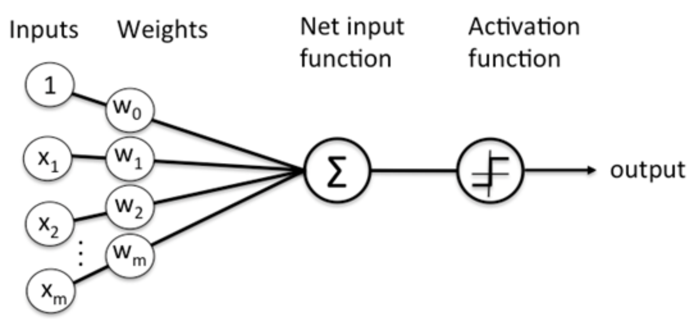

Figure 5. Typical node in a neural network.[55]

> Despite their biologically inspired name, artificial neural networks are nothing more than math and code, like any other machine-learning algorithm. In fact, anyone who understands *linear regression*, one of first methods you learn in statistics, can understand how a neural net works.[56]

The actual weighting functions and decision functions involved at each node can be nonlinear, but the basic operation is well understood. The main advantage of neural networks is their ability to extract patterns from extremely large data sets at high speed, much faster than humans could do.

The fact that neural networks can sort things such as photographs in a seemingly human way has led to efforts to make them recognize patterns in art or musical songs, and then "imitate" them. The idea is to show that this is how humans create art or music. Obviously, there are rules that one can learn about, say, music, dealing with rhythm, harmony, meter, and so forth. And equipped with these rules, anyone can “create” new music. But is this composition in the sense that a great composer creates? Similarly for art: anyone can learn techniques of color, scene composition, light and shadow, and others, and apply them to the task of painting or drawing. But is the product art, and is it what a real artist does? The question, then, comes down to whether pattern recognition and imitation is the same as sensing reality and creating based on that sensing. The proof that it is not is in the fact that great works of art are holistic—every brushstroke or every note contributes to the overall impression on the viewer or hearer, so understanding and appreciating the work requires a holistic understanding of it. This, more than anything else, suggests that these works are created in contact with reality.

Francois Chollet, a practitioner of deep learning in neural networks, has discussed its limitations:

> In short, deep learning models do not have any understanding of their input, at least not in any human sense. Our own understanding of images, sounds, and language, is grounded in our sensorimotor experience as humans—as embodied earthly creatures. Machine learning models have no access to such experiences and thus cannot "understand" their inputs in any human-relatable way. By annotating large numbers of training examples to feed into our models, we get them to learn a geometric transform that maps data to human concepts on this specific set of examples, but this mapping is just a simplistic sketch of the original model in our minds, the one developed from our experience as embodied agents—it is like a dim image in a mirror.[57]

In practice, this means that the feedback method used to make neural network algorithms converge to the desired pattern recognition and classification goal has serious limitations, which do not apply to human activities:

> …through gradient ascent, one can slightly modify an image in order to maximize the class prediction for a given class. By taking a picture of a panda and adding to it a "gibbon" gradient, we can get a neural network to classify this panda as a gibbon. This evidences both the brittleness of these models, and the deep difference between the input-to-output mapping that they operate and our own human perception.[58]

This brittleness alone illustrates the radical differences between AI in the form of neural nets and human knowing. Geoge Gilder observes that though niche applications are important, e.g., recognizing faces, interpreting speech, or implementing an advertising algorithm, they are not the long-sought nirvana of general AI:

> AI is just another advance in computer technology, like the other ones. It is not creating rivals for the human brain… To observers of such trends, it is easy to imagine a future in which the role of humans steadily shrinks…The basic problem with these ideas [of AI] is their misunderstanding of what computers do. Computers shuffle symbols..[59] (italics added)

This means, of course, that success in game playing is qualitatively different than dealing with the real world:

> For the game of *Go* or chess or some routinized task, the symbols and objects are the same. The white and black stones on the *Go* board or the pieces on the chess board are both symbols and objects at once. The map is the territory.[60]

What is the conclusion? That the neural networks operate differently than human intelligence, and only mimic it in ways that are very fragile:

> …never fall into the trap of believing that neural networks understand the task they perform—they don't, at least not in a way that would make sense to us. They were trained on a different, far narrower task than the one we wanted to teach them: that of merely mapping training inputs to training targets, point by point. Show them anything that deviates from their training data, and they will break in the most absurd ways.[61]

Neural networks operate on a model similar to that of robotics and autonomous systems: sensing is done by some type of device that hands off its data to the neural network circuitry. The sensing device is interchangeable with many others and essentially independent of the neural network hardware. Hence, there is a separation of sensing and processing, rather than a fully integrated system, the same as the robotics and autonomous systems—the Humean paradigm of knowing again.

Symbolic manipulation programs

Symbol manipulation programs, in the form of applications that can do symbolic mathematics (as opposed to numerical calculation) have been around for many decades. Early versions include *Macsyma* (1968) running on mainframe computers, and similar programs designed for personal computers, including *muMath* (later *Derive*, 1979), and most famous, Wolfram's *Mathematica*® (1988), together with *WolframAlpha* (2009), an online app. Obviously, these programs predate most of what today is termed "AI". They are capable of solving many standard types of symbolic mathematical problems, including differentiation, integration, differential equations, factoring, algebraic manipulation and simplification, matrices and tensors, and virtually indefinite length numerical calculation, i.e., ability to work with numbers requiring tens or hundreds of thousands of digits—something not possible with ordinary programs such as Excel®. To anyone who has ever needed to solve complex mathematical equations, these programs are invaluable, and their remarkable ability to draw on a large database of algorithms to solve problems is at times amazing—certainly a type of AI. In that sense, they give the impression of 'knowing" mathematics; and unlike the chatbots, they don't "make up answers". But of course they don't "understand" mathematics, and the proof is that they are not valuable to someone who doesn't understand the mathematics involved in a problem. Often the user must set up the problem to be solved very carefully in order for the program to give the desired answer, and in many cases obvious simplifications are missed. The solution to a problem (if *Mathematica* can find it) may involve unusual functions. In other cases, the program cannot solve the target equations, and the user must manually break the problem down. For example, *Mathematica* can easily solve a difficult integration problem such as

$$\int x^3 \sin(x^2)\cos(x)\,dx$$

though the solution involves Fresnel integrals—not a common function. But confronted with Einstein's field equation

$$R_{\mu\nu} - \frac{1}{2} R g_{\mu\nu} + \Lambda g_{\mu\nu} = -\kappa T_{\mu\nu}$$

the program is stuck—the user must assist. Likewise solving Dirac's equation from Quantum Field Theory also requires human intervention:

$$i\gamma^{\mu}\partial_{\mu}\left|\psi\right\rangle = m\left|\psi\right\rangle$$

Exact rules of syntax need to be followed, and explicit instructions must be given to get the kind of output desired—something that a real mathematician would instinctively know how to do. (*WolframAlpha* does allow some more free-form inputs, however). On balance, these programs are wonderful and illustrate how computing power can be harnessed to allow humans to solve otherwise computationally (but not conceptually) intractable problems.

The paradigm of knowing question ultimately comes down to that of whether the Humean/AI theory of what can be known exhausts what we know reality to be. If not, it means that there are aspects of reality not capturable by the AI paradigms. To that we turn next

IV. Human Knowing and how it differs from AI

The twentieth century taught us that are often limitations even in areas where none was suspected. Gödel's Incompleteness Theorem (1931) showed a fundamental limitation to the formalization of mathematical systems, namely that no matter how the axioms and rules of inference are formulated, there will always be theorems that are true but unprovable in the system. Likewise quantum theory showed that there are limits to the measurement of key quantities, namely position and momentum, $\Delta x \cdot \Delta p \geq \hbar$, and time and energy, $\Delta t \cdot \Delta E \geq \hbar$. These uncertainties are not just curiosities that affect arcane measurements, but fundamentally affect everything from the subatomic realm to cosmology. Therefore our quest to examine Artificial Intelligence and to determine what limitations it may have is very important, especially since the ability of machines to perform human tasks and jobs has enormous economic, political, social, technological, and scientific implications, not to mention areas such as legal and moral concerns. Finally, how do the capabilities of AI affect our view of human beings as special in any way, since this is the basis for many of the rights enshrined in the Constitution and similar documents in other countries?

Sentience and reality

Most AI researchers concede that computers are not and cannot be sentient in the sense of perceiving things as real. "Sentience" means awareness of the world as something real and existing independently of us, but perhaps more importantly, it means awareness of other people as humans, and at least some understanding of how they view the world and perceive others.

"Art is a lie that reveals the truth." So said Pablo Picasso (1881-1973). To a human, with sentient intelligence, Picasso's observation is immediately understandable. To any type of machine or sensible intelligence, it is gibberish, first because it seems contradictory, and second, because "truth" can only be understood in a rule-based manner. Anyone who has stood before a great painting knows that the painting is not photographically accurate, yet discloses some deep truth about the subject. Spanish philosopher Xavier Zubiri (1898-1982) has observed that truth about reality is a goal of art, music, and literature:

> Now, reason or explanation is above all the *intellection of the real in depth.* Only as an explanation of color is there intellection of electromagnetic waves or photons. The color which gives us pause to think is what leads us to the electromagnetic wave or to the photon. If it were not for this giving us pause to think, there would be no intellection of a beyond whatsoever…The beyond can also be what forges a novel; we would not create the novel if the real did not give us pause to think. The same could be said of poetry: the poet poetizes because things give him pause to think. And that which he thus thinks of them is his poetry…A metaphor is one type of reasoning about things, among others. What is intellectively known of the beyond is purely and simply the intellection of what things "on this side", in being intellectively known, give us pause to think. Therefore the intellection of the beyond is reason or explanation; it is intellection of the real in depth.[62]

Raj Rajkumar, a professor of engineering at Carnegie Mellon University who collaborates with General Motors Company, has admitted the fundamental difference between machines and humans in similar language:

> We are sentient beings, and we have the ability to reason from first principles, from scratch if you will, while AI on the other hand is not conscious, and doesn't even understand what it means that there's a physical world out there. [63]

Others have made the same point:

> The core problem is sentience. Because there [is] no way to program theory of mind, the [autonomous] car would never be able to respond to obstacles the way that a human might. A computer only "knows" what it's been told. Without sentience, the cognitive capacity to reason about the future, it can't make the split-second decisions necessary to identify a streetlight as an obstacle and take appropriate evasive measures.[64]

To get around the formidable sentience problem with cars, engineers have had to resort to video game technology:

> The self-driving car programmers realized they could make a vehicle without sentience—that moving around a grid is good enough. Their final design is basically a highly complicated remote-controlled car…What it uses…are statistical estimates and the unreasonable effectiveness of data. It's an incredibly sophisticated cheat that's very cool and is effective in many situations, but a cheat nonetheless. It reminds me of using cheats to beat a video game. *Instead of making a car that could move through the world like a person, these engineers turned the world into a video game* and navigated the car through it.[65] [Italics added]

Video games, however, are at best a pale reflection of reality; and it is highly questionable whether video game proficiency is a viable substitute for sentience when dealing with the real world.

It is therefore not surprising that there have been many accidents involving autonomous cars, but we will consider just one that illustrates quite well the fundamental problem of trying to replicate the human ability to perceive reality by use of sensors-and-computer systems. A recent *Wall Street Journal* article reports the following accident:

> On Oct. 2 [2023], a hit-and-run driver in San Francisco threw a female pedestrian into the path of a driverless Cruise car, which pinned her underneath and dragged her for about 20 feet. The driverless vehicle was trying to pull over, a maneuver it was programmed to do if it detects a crash, Cruise said.[66]

The woman was seriously injured by being dragged by the autonomous car. GM said that its engineers were working to program the car to deal with such a situation. What this accident and the reaction of GM reveals is the fundamental problem difference between human and machine operation. A human driver *would immediately know what to do*—namely, to stop his vehicle if someone were thrown by another vehicle and hit his vehicle. And he would know to do this *even though he had never experienced such a situation before*. The autonomous car on the other hand must be programmed for every conceivable case in order to replicate human capability—an obvious impossibility since those cases cannot be predicted and therefore cannot be enumerated. Humans have not only the ability to perceive reality, but the equally important ability *to think creatively about it and thus deal with situations and problems never experienced before.*[*]

The problem of sentience affects the ability of AI systems to deal with reality in other ways. Humans can see an object in one position, say standing upright, and

[*] At the time of writing, Tesla has recalled 2 million vehicles to "fix" problems with their AI autopilot systems, and GM's Cruise Autonomous Vehicle division has laid off 24% of its workforce.

immediately recognize it in another, such as lying on its side. This is extremely important for driving, but it is beyond the capabilities of AI systems:

> Here, we run into a difference between human thought and computation. A human brain can rotate an object in space. When I say "traffic cone," you can picture the cone in your head. If I say, "Imagine [that] the cone is knocked over on the ground," you can probably imagine this too and mentally rotate the object…One popular math aptitude test for children involves showing them a 3-D shape on a 2-D plane, then presenting other pictures and asking them to choose which one represents the object rotated. The computer has no imagination, however. To have a rotated image of the object, it needs a 3-D rendering of the object—a vector map, at the very least. The programmer needs to program in the 3-D image. The computer also isn't good at guessing, the way a brain is. The object on the ground is either something in its list of known objects, or it isn't.[67]

Interestingly, this is one of the problems that bedeviled Hume's analysis of human understanding, squarely based on the same intelligence paradigm. With Hume's idea of knowing, it would be impossible for anyone to pass the simple aptitude test for children, or to recognize the traffic cone in a different position than that of its original "impression". Hume was not able to solve this problem, if he was even aware of it. The fact that the real world is more—and more difficult to negotiate—than video games and simplistic theories of knowing has been conceded by those involved with self-driving cars and machine learning (ML) systems. Hava Siegelmann, of the Defense Advanced Research Projects Agency (DARPA), tells us:

> Life is by definition unpredictable. It is impossible for programmers to anticipate every problematic or surprising situation that might arise, which means existing ML systems remain susceptible to failures as they encounter the irregularities and unpredictability of real world circumstances. Today, if you want to extend an ML system's ability to perform in a new kind of situation, you have to take the system out of service and retrain it with additional data sets relevant to that situation. This approach is just not scalable.[68]

Humans, on the other hand, are very good at just this kind of activity.

Humanoid robots

Efforts to design humanoid robots—an acid test of AI's ability to mimic human intelligence—have foundered. Rodney Brooks, an MIT researcher, co-founder of the iRobot corporation, and one of the world's leading robot developers, has some interesting commentary. He notes, "We don't have anything anywhere near as good as an insect, so I'm not afraid of superintelligence showing up anytime soon." Brooks most successful robots were a vacuum cleaner and a robot designed to defuse roadside bombs—both highly specialized tasks. He founded another company in 2008 to create "cobots", which are "collaborative robots" designed to work alongside humans—already a giant step away from humanoid robots. The company folded because, as it turned out, "…building robots with human-like capability is really, really hard. There are many things humans can do easily that are almost impossible for robots to replicate." This is described by "Moravec's Paradox", which is based on the empirical observation by AI researchers that what we term "reasoning" requires relatively little computation compared to that needed for sensorimotor and perception skills:

> It is comparatively easy to make computers exhibit adult level performance on intelligence tests or playing checkers, and difficult or impossible to give them the skills of a one-year-old when it comes to perception and mobility.[69]

So, even with all the astounding advances in computation ability—probably far beyond anything dreamt of by Turing—robots (and AI) are still little advanced from 70 years ago:

> If you imagine a continuum of intelligence, with, say, humans at one end and insects on the other, artificial intelligence is nowhere on that spectrum.…It's true that AI machines now dominate at games like Chess and have mastered video games like Pong. But what this shows is that AI in 2019 is the equivalent of a nuclear-powered calculator. It can run billions of calculations per second and crunch vast quantities of numbers faster than a human can even blink. But that is not thinking or anything close to it. It is possible to do calculations with an abacus, a wooden tool dating to the 14th century — but nobody would ever suggest an abacus is alive or perceptive or conscious. Even today's most impressive AI programs are little more than a turbocharged abacus (or billions of them strung together).[70]

AI, in its various instantiations, has revealed that some types of knowledge-related activities can be automated with modern computer hardware and associated systems. At the same time, we have seen that many activities once thought amenable to such automation are not so, restricting the scope of AI and any chance that it will one day "take over".

The Category Mistake

AI systems, or at least those theorizing about their capabilities, suffer from another problem: the category mistake. A category mistake occurs when one tries to talk about something with an inappropriate description or "category". An example is, "my feelings are green". This is what occurs in many discussions of AI capabilities. In order to be able to explain—or explain away—our ordinary experience of the world and other people, not to say religious experience, any type of physicalist theory must show:

> …what physical configuration in the brain corresponds, for instance, to concepts like fourth dimension, n-dimensional manifold, and the like…[They] will have to explain what well-determined pattern in the brain is the equivalent of the indeterminacy principle and of indeterminacy itself. They will have to show what molecular fullness corresponds to the concept of vacuum or empty space…They will be beset…with the problem of finding the physical force…that will adequately translate the feeling of love, hatred, and curiosity into the categories of physics.[71]

In all of the cases cited, the discussion falters because the two things compared are simply not of the same category. The effort turns into a bizarre fantasy that bears no relation to reality. But without the ability to explain such identifications, the theory that computer-based systems are somehow equivalent to the human knowing capability falls flat.

AI and Ethics

The philosophical ramifications of AI appear in another context, namely ethical theory. Faith in technology alone quickly leads to conundrums. For example, many today are concerned that AI will spin out of control and threaten humanity:

> It has the potential—however small one may regard that probability, but it is non-trivial—it has the potential of civilizational destruction.[72]

As a result, some have embarked on a crusade to ensure that AI is deployed in an ethical fashion. This has become known as "effective altruism".

> [Effective altruism] believes that carefully crafted artificial-intelligence systems, imbued with the correct human values, will yield a Golden Age—and

> failure to do so could have apocalyptic consequences.[73]

The problem is that those in the technology community discussing this subject do not understand a key fact about ethics, viz. that there are no free-floating ethical theories. Any theory of ethics—any moral code—*must be based on an antecedent theory about what is real*. If one believes that God exists, and that the Ten Commandments were given, then this will imply the need to live in a certain way. If one does not believe that God exists, but only that the "material" world explored by science is real, a different moral code will ensue. The tendency among the technology community is to gravitate to some type of utilitarian ethics: what is good is what will provide the maximum "benefit" or "happiness".[74] As is well-known, utilitarianism is unworkable because of difficulties in defining benefit or happiness, and because actions never stop having consequences. Moreover, if one believes that there are truly binding moral imperatives, and unassailable knowledge of right and wrong, even in a few cases, he or she is committed to belief in something real that is outside of science. The debate over the so-called "effective altruism" reveals the kind of thinking involved:

> The turmoil at OpenAI exposes the behind-the-scenes contest in Silicon Valley between people who put their faith in markets and effective altruists *who believe ethics, reason, mathematics and finely tuned machines should guide the future.*[75]

No, they can't and they won't. No binding moral injunctions are possible on such a basis; only pragmatic suggestions, because there is no metaphysical ground. Any moral judgement about AI (or any other technology) cannot be done from within the technology ambit itself; it must be done on a higher plane, outside of the limited realm of science and technology, where a holistic view of knowledge and the place of humans in the world order can be discerned. That is, it must be done in a viable faith-oriented context, and not on the basis of a surrogate religion, adumbrated above, which those in the scientific/technology community often proffer—and this comes from the author who is himself deeply involved in science and technology.

V. Why AI Will Achieve But a Fraction of Its Goals

Human knowing operates on a radically different principle, namely, it is a thoroughly integrated system of sensing, motor skills, and brain--whereby it has direct contact with reality at a very basic level, and uses this direct contact to formulate its way of knowing about reality at higher levels. This also enables the supremely creative way that human knowing works, because it is the basis for the ability of humans to deal with situations they have never encountered before, and to generate new theories about reality, often thinking about it in very innovative ways. That is, humans can "think outside the box", whereas AI cannot. AI can of course generate random "ideas", understood in the rather limited sense of data structures or random chatbot statements; but that is not how humans develop new theories or deal with unexpected situations, as anyone who has done either can attest. Human interaction with reality is qualitatively different than the AI paradigm. As Zubiri expresses it:

> Human sensing and intellection are not only not opposed, but indeed constitute in their intrinsic and formal unity a single and unitary act of apprehension. This act *qua* sentient is impression; qua intellective it is apprehension of reality. Therefore the unitary and unique act of sentient intellection is the impression of reality. Intellection is a mode of sensing, and sensing in man is a mode of intellection.[76]

This, in turn means that our impression of reality is different than what can be achieved by any sort of paradigm based on separation of functions:

> The impression of reality is not *impression of what is transcendent*, but rather *transcendental impression*. Therefore "trans" does not mean being outside of or beyond apprehension itself but being "in the apprehension", yet "going beyond" its fixed content. In other words, that which is apprehended in the impression of reality is, by being real, and inasmuch as it is reality, "more" than what is it as colored, sonorous, warm, etc.[77]

It is in this "more" that its capabilities beyond the AI paradigm of knowing come into play. That paradigm, by design, can only ape what human intelligence does. The AI paradigm reacts to stimuli in the form of sense-type data; it cannot react except indirectly to any underlying reality. It cannot postulate reality except in a superficial sense; it does not "know" what it is doing because it does not have contact with reality.

We will examine several issues:

a) Symbol manipulation vs interaction with reality: The map is not the same as the territory
b) The difference between knowing *what things are*, and *how things behave*
c) The difference between *aiding human knowing* and *replacing human knowing*
d) Creative thinking and understanding vs. rote or algorithmic manipulation
e) Effect of utilizing the wrong paradigm for knowing in AI
f) Locked into the past vs looking into the future

These issues illustrate the key difference between AI "intelligence" and human intelligence.

a) Symbol manipulation vs interaction with reality

The goal of human knowing is always to *know something about reality*, whether or not it has any operational value. Neither an animal nor AI seeks the reality of the real. AI and computers must utilize symbols, which function as signs for response, programmed in the case of computers and AI:

> A digital computer is a device which manipulates symbols, without any reference to their meaning or interpretation. Human beings, on the other hand, when they think, do something much more than that. A human mind has meaningful thoughts, feelings, and mental contents generally. Formal symbols by themselves can never be enough for mental contents, because the symbols, by definition, have no meaning (or interpretation, or semantics) except insofar as someone outside the system gives it to them.[78]

The machines, in other words, do not have any connection to what things are in reality; they can only manipulate symbols and then take some sort of programmed action, such as opening a valve or scanning a scene for obstacles. As George Guilder has noted, the function of computers is fundamentally misunderstood:

> As philosopher Charles Peirce observed more than a century ago the links between computational symbols and their objects are indefinite and changing. *The map is not the same as the territory*. The links between symbols and objects have to be created by human minds. Therefore, computations at the map level do not translate to reliable outcomes on the territorial level. For the game of *Go* or chess or some routinized task, the symbols and objects are the same. *The white and black stones on the Go board or the pieces on the chess board are both symbols and objects at once*. The map is the territory…. in order to have cor-

> respondence between logical systems and real world causes and effects, engineers have to interpret the symbols rigorously and control them punctiliously and continuously. Programmers have to enforce an interpretive scheme between symbols and objects that banish all slippage.[79] [Italics added]

We have art schools that teach drawing and painting techniques, music conservatories the teach composition, and college curricula that teach creative writing. While all of these can teach students fundamentals and even advanced techniques, they cannot guarantee that their students will become great artists, composers, or writers. Why is this? Because these programs can only impart basic rules, but not the insight and inspiration that sees reality and turns it into great art, music, or literature. Machines can also be programmed to follow these rules, but cannot be programmed to sense reality, essential to creation of great art.

To make matters worse, as noted earlier, literature can be understood at multiple levels. Take Cervantes' *Don Quixote* (1605, 1615), for example. At the lowest level, it is a story about the adventures and misadventures of the two main characters, Quixote and Sancho Panza. At another level, it is a satire on a genre of literary works, the chivalric romances. At a still higher level, it is a metaphor for everyone's life—we all have some of Don Quixote and some of Sancho Panza—as well as for every literary character. The book, in other words, is a meta-novel. This works because each level is about some aspect of human reality. Obviously, understanding this goes far beyond any machine capability, to say nothing of creating such literature in the first place. Similar remarks can be made about great paintings or great musical works. Many operas, for example, work on multiple levels, such as those of Richard Wagner (1815-1882), in which events and the action of characters have allegorical meanings His *Ring Cycle* and *Lohengrin* are perfect examples..

b) Difference between knowing what things are, and how things behave

There is a profound difference between knowing *what things are*, and *how things behave*. Though historically many have thought that these two are the same, or at least that one immediately leads to the other, in fact they are distinct. Knowing *what something is* engages the transcendental order of human knowing, how the thing relates to other things, and the fact that it exists in reality as a thing. Knowing *how something behaves* enables us to control it, or to make other things that behave in similar ways. That is, it operates at the phenomenological level. Knowing this, or equivalently, knowing how to make something that behaves in a particular way, is not operating at the most fundamental level of human knowing. AI systems perforce act only on the basis of how things behave, or nominalistically on the basis of names, but never on the basis of reality.

c) Difference between aiding human knowing and replacing human knowing

To better understand the problems posed by AI, we need to draw a distinction between *aiding human knowing* and *replacing human knowing*. It is clear that computers have been doing more and more of the first for decades. Our modern technological society could not exist in anything like its present form without computer-based automation of functions at one time done by humans. A trivial example is telephone switching. Verizon alone says that it handles 800 million calls per day.[80] If a phone operator at a manual switchboard could handle 1 call per minute, and could work 8 hours per day, handling these calls would require about 1.6 million operators and a switchboard about 173m per side—an obvious impossibility. Modern banking and finance simply could not be done manually, nor could most modern engineering problems.

Modern medicine uses computers and automated systems extensively, but to aid doctors and other key medical personnel in making medical judgements. Replacement only occurs with jobs that involve simple repetitive labor, such as continuous monitoring of vital signs. In all these cases we observe that computers are supplementing or assisting human knowing and activities, but not replacing them or doing something new.

d) Creative thinking and understanding vs. rote or algorithmic manipulation

AI tools such as ChatGPT can scan the Internet and assemble much information, even invent "facts", but they are not creative in the true sense. On the other hand, human knowing is nothing if not radically creative, even in simple everyday tasks, such as driving a car. And it is especially so for science, math, literature, music, art, and many other fields. The great advances in science always come when someone breaks with established tradition. Einstein broke with establishment science with his theory of relativity. Heisenberg and others broke with established science with quantum mechanics. Gödel broke with old ideas about math with his Incompleteness theorem. Beethoven broke with established musical ideas with his symphonies. Renaissance artists broke with old traditions to develop new ideas about painting. Euripides broke with old traditions to write his plays. AI algorithms typically compose by looking at previous words and then guessing what the next word should be, following grammar rules. They cannot creatively and analytically think through a question, using information learned from reading and research, where a critical eye is needed to discern what is valuable and a view of reality is needed to synthesize new ideas. AI systems, therefore, are essentially stuck in the past, unable to advance knowledge or even apply what is known in a creative way. On a more prosaic level, anyone who has done a job other than something very routine, such as assembly line work, can attest that creative problem solving is required almost constantly, because situations arise that are different than earlier experience or training. This is especially the case for project managers and supervisors, though even lower level workers frequently encounter problems with tools or supplies and need to figure out work-arounds.

The creative element plays out in literature in important ways:

> …an author like Shakespeare is great because he expresses vividly and concretely a particular time, place, and culture; and yet he transcends what is merely local and ephemeral and touches the perennial and universal concerns of humanity by means of what is immediate and particular.[81]

Truly original thinking involves selectively taking what is already known and using it to formulate new ideas, theories, explanations, or artistic works that apply to the real world, give insight into reality, and can be tested or verified, as appropriate. It is not based on randomly combining ideas or facts, but on insight into the reality of a situation or problem. Except for very routine jobs, e.g., assembly line work, no real world job is without challenges that require creative thinking.

A particularly interesting example is Georg Cantor's (1845-1918) famous diagonal argument, used to demonstrate that there are different infinities or *transfinite numbers*. In particular, the argument shows that the transfinite number $\aleph_0$ (the cardinality of the integers) is less than $\aleph_1$ (the cardinality of the real numbers). The key point for our purposes is that Cantor is proving theorems about infinities considered as *real* things, with *real properties* than can be discovered. Cantor's results were totally unexpected at the time, and no amount of pattern-recognition or random shuffling type of investigation of then-current mathematical ideas (if this could even be done) would have led to Cantor's results. Furthermore, the entire notion of infinity, as something real, and transfinite numbers, though able to be grasped by human mathematicians, makes no sense in an AI program; to it they are just names in a nominalistic paradigm. They certainly are not the correlate of any “sensible impressions” (in Hume’s parlance). Though symbol manipulation programs such as *Mathematica* can operate with infinity, they have only the capability to follow strictly logical rules for manipulations involving it:

> A transfinite number, an abstract concept, are not sensed qualities. But they are intellectively known as something real, and as such are constituted in the impression of reality as such.[82]

e) What is the effect of utilizing the wrong paradigm for human knowing in AI?

The main effect is that AI will be expected to do things that it will never be able to do. This will cause expenditures of money and time that can never come to fruition, and attempts to substitute AI devices for people. In fact, this latter has been occurring for some time, with extremely frustrating results. The automated response systems used by many banks and other commercial entities is a case in point. The reader has probably had the same experiences as the author: your call is answered, and you are given a menu. Of course, the question for which you require an answer isn’t on the menu. Or else the system will ask you what you want to know. When you say what it is, you are given something else entirely. After many frustrating minutes, you might get transferred to a real person, who actually can deal with your problem. Worse than this inconvenience is the issue that the AI-controlled system may behave in totally inappropriate ways, leading to catastrophe.

f) Locked into the past vs looking to the future

With respect to human knowing, all categories of AI are backward-looking rather than forward-looking, because they are based on existing knowledge. None has the ability to create new visions of reality, new theories. This does not mean that they cannot be used to make predictions or forecasts about the future; even simple regression analysis can do that. And they can, of course, be used to enable us to see things that we otherwise could not see, such as simulations of the evolution of the universe. But these simulations are based on *current* theories, e.g., about the constitution of the universe and the laws governing it. What this statement about the capabilities of AI means is that AI cannot advance human knowledge in any theoretical sense, that is, develop new theories about reality. It can only use existing knowledge to give us answers. This is an important contribution, but ultimately limited in its scope. And it can only do this functioning as an adjunct or assistant to a human seeking to answer a question or solve a problem.

One last point. A question that may have occurred to the reader is this: If humans are made of physical materials, is there any reason to believe that it is impossible to make an AI system that behaves like humans and has sentience? Putting this into a logical form,

Physical properties suffice for sentience $\supset$ Possible to create sentient AI

This is in the form $A \supset B$. The interesting question is *modus tollens* for this inference, $\sim B \supset \sim A$, which is to say

Not possible to create sentient AI $\supset$ Physical properties do not suffice for sentience

In this article various reasons were given for why AI is not and never will be sentient and capable of perceiving reality. First and foremost was its defective theory of knowing, though technical concerns such as the failure of vast scale improvements to yield any hint of sentience loom large. Kurt Gödel’s Incompleteness Theorem has been used to argue that the mind cannot be reduced to algorithmic processes—and hence AI cannot be sentient; but Gödel also argued that his famous theorems imply an important conclusion:

> Either "the mind cannot be mechanized" or "mathematical truth outstrips human reason"

Gödel did believe that this disjunction is a "mathematically established fact." There are good reasons to believe that under any reasonable interpretation of truth and mathematics, Gödel is correct.[83] Much effort therefore has been directed to showing that the first half of the disjunction is true. Roger Penrose has argued for this conclusion.[84] As it happens, however, for the purposes of showing whether a physicalist or sensible intelligence theory of intelligence can be correct, the disjunction is checkmate because *it does not matter which side is true*; it is enough to know that at least one of the two is correct. If the first side is true, then obviously the mind cannot be recast as a machine in neurophysiological terms or any other. If the second side is true, then there are mathematical truths that cannot be determined by human reason, i.e., truth is real and transcendental and not reducible to anything physical or any marks on paper. Gödel himself recognized this:

> …if the first alternative holds, this seems to imply that the working of the human mind cannot be reduced to the working of the brain, which to all appearances is a finite machine with a finite number of parts, namely the neurons and their connections…On the other hand, the second alternative, where there exist absolutely undecidable mathematical propositions, seems to disprove the view that mathematics is only our creation; for the creator necessarily knows all the properties of his creatures, because they can't have any others except those he has given them. So this alternative seems to imply that mathematical objects and facts (or at least something in them) exist objectively and independently of our mental acts and decisions…[85]

Perhaps this gives us the answer to our question, allowing us to conclude that physical properties indeed are not enough for sentience. In that case, Hume's theory of knowing may be the best that machines can achieve.

The issue of AI and consciousness, with respect to neurophysiology, is also important. A recent investigation that approaches the consciousness problem of AI from the standpoint of neurophysiology and reviews the relevant research has concluded:

> Consciousness is constantly fed by such learning, capable of generating stable representations despite an incommensurable amount of variability in input data, across time and across individuals, for life-long integration of experience data. Artificial consciousness [i.e., AI] would require probabilistic adaptive computations capable of emulating all the dynamics of human learning and memory that enable human intelligence and creativity. No AI is likely to ever have such potential.…Our motivation to assert the qualitative superiority of human intelligence and the unique existence of human consciousness is to justify continuing human control over machines rather than foolishly wanting to hand over the keys to the machines that our minds have created…[86]

VI Complexification and the Real Dangers Associated with AI

As noted in the Introduction, prominent thinkers have called for a moratorium on AI, with some even demanding regulations or a complete shutdown of it. In reality, AI is a response to the complexification of our modern industrial/information society. The interconnectedness, specialization, and drive for efficiency combine to drive society to ever higher degrees of integration and complexity.

Some important threats associated with development of AI were discussed in Section II. Here we wish to consider a different, more realistic threat. Joseph Tainter has analyzed complex societies, with a view to discovering reasons for their collapse. We may ignore the collapse issue and concentrate on Tainter's identification of the reasons and effects of complexification seen in almost all societies.[87] Complexification of society, Tainter points out, is a problem solving mechanism.[88] Problem solving drives the need for more complicated societal organization; for example, primitive societies have very little organization, but when they fight, they learn that they must have a warrior class, and a food production class, and then an education/training class, i.e., ever increasing levels of organization, sophistication, communication, and knowledge dissemination. Then transportation systems and communications systems need to be developed, until (after a long period) all of the systems that we take for granted today become necessary. These systems and attendant training and specialization improve society, but at the cost of complexification.[89] In later stages powerful tools such as computers and ultimately AI are needed to manage and control all the systems, which become too complex for direct human control.

Therefore the hype around the potential dangers of AI, though understandable, is misplaced. The dangers arise not because the AI system wants to take over the world, or act like HAL from *2001*—computers are too stupid to be able to do anything like that. Rather, the real danger lurks in the complexification of society. Something like AI was inevitable given the ongoing societal complexification and the need to handle it. As computer-based systems are used to operate functions spanning more components of society, such as the power grid, financial systems, and—as has been proposed—the ICBM launch system,[90] the dangers associated with malfunction naturally increase. As systems utilize more and more data, and rely on more and more complex algorithms to process the data and then act on it, the vulnerabilities of such systems—which are not sentient and do not perceive reality—will sooner or later become apparent. These include (but are not limited to) the following:

- The system encounters a situation for which it was not programmed, and does something that leads to catastrophe.
- The system is hacked by a malicious actor, who causes it to malfunction.
- Interaction of the components of the overall system leads to unanticipated instabilities.
- An undiscovered programming bug causes the system to malfunction, again with potentially catastrophic results.
- The system is "fooled" by a malicious actor who determines how to make it misfunction by fooling its sensors.
- AI system output will be taken as gospel when it is biased or factually wrong.
- Recommendations will be acted upon without proper scrutiny.
- And perhaps most important, *the societal cost of the AI systems will outstrip the value that they add.*

As an example of the latter, if an AI-based systems is used to control complex real world equipment, such as the power grid, a malfunction or hack could disable the grid for a large part of the country, resulting in economic losses greater than the value added by the automated system. Essentially, this is the danger inherent in utilizing "dumb" (i.e., non-sentient) actors in modern society. There will always be a trade-off between increasing efficiency and losses due to system malfunction. The means that it is necessary to keep people in the loop who can supply the connection with the real world that the systems lack.

Conclusions: The Verdict on AI

The dreams of the early days of computers have not materialized, massive scaling has not wrought qualitative changes, and while AI has gone on to solve important problems, it has not become conscious or even capable of many simple human tasks. All of this confirms Zubiri's distinction between the sensible intelligence perforce utilized by AI, and the sentient intelligence of humans, which entails that the paradigm of knowing employed by AI will always be completely different than the way that human knowing operates. AI therefore will never be able to do the essential tasks of humans: to perceive reality and to understand truth. This in turn confirms that humans are a different *kind* of reality. AI will continue the path of supplementing human reason but only replacing it in narrowly focused applications. There may, of course, be many such applications; but they do not include any sort of "taking control".

The reality of the divide between human knowing and AI will not be affected by scale changes in computer hardware, because it pertains to basic architectural and functional differences. This ultimately limits the capabilities of any AI system, however implemented. The salient characteristics of human knowing not realizable with the AI model include:

- Creative judgement and problem solving.
- Seeking underlying truths about reality.
- Formulation of new scientific theories and mathematical theorems and new fields of science, mathematics, and other disciplines.
- Understanding of things as real in transcendental sense.
- Creation of significant works of art and literature.
- Critical discernment of the value of a text
- Ability to synthesize information in creative, holistic ways based on critical evaluation of sources.
- Understanding of the various types of causality and how they interact, especially formal causality.

In order to be a "threat" to anything beyond certain types of jobs, AI would have to be able to do these things, which it cannot. It will forever be restricted to matters that can be done in some algorithmic fashion. *Therefore, AI does not show or prove that humans are just another material object, thus rendering all forms of religion superfluous. It shows quite the opposite: humans are unique and not reducible to physical computing machinery.*

What can we expect from AI and related fields, based on our past experience and our understanding of the paradigms of human knowing and the AI model? AI will:

- Be able to automate some jobs, and displace some workers (though at the same time creating new jobs).
- Be able to supplement and assist human research and development activities.
- Be able to aid humans in many other fields and actions.
- Be essentially an extension of current computer capabilities.
- Never be able to automate most jobs, since most require creative action on the part of the job holder.
- Never be able to "take control" because it does not perceive reality.

The real threat from AI comes not from any possibility that it will become sentient and smarter than humans, but from complexification issues associated with use of AI-based systems to control critical infrastructure such as the power grid, military decisions, and economic systems. The threat is that programming errors, encounters with unanticipated situations, or hackers will disrupt the AI system and cause serious malfunctions. AI will also pose ethical problems in many areas, beyond the scope of this article, as well as important legal and societal issues.

To return to the original list of AI functions, we can say about each:

i. *Robots and robotic systems*: will never become sentient or able to take over functions requiring ability to perceive reality, e.g., interaction with people on personal level

ii. *Neural networks and pattern recognition*: Will fulfill specific functions that aid humans, but will never replace human abilities

iii. *Generative AI, including ChatGPT and similar applications*: inability to do real research and make critical judgements requiring perception of reality dooms them to low level roles since no serious decision making can be done on the basis of "hallucinations"

iv. *Symbolic manipulation programs such as Mathematica*: extremely useful in the hands of those who understand the mathematics involved, and just need calculation assistance; they are not mathematical "superintelligence".

v. *Autonomous cars and other autonomous systems*: Inability to perceive reality will restrict their ability to replace human drivers, though they can aid drivers in many ways.

vi. *Large-scale control programs that use some combination of the above*: dangerous if implemented without some type of human oversight

It may well be the case that utilizing Hume's theory of knowing, with all of its limitations, is the only real avenue open to AI systems. All evidence points to the conclusion that sentience and perception of reality will be forever out of reach.

Should AI work be shut down as an imminent danger to society? No, but at this juncture of history, we face a dilemma: we need the capabilities of AI-type systems to sustain our ever-increasing level of societal complexification and knowledge repository. At the same time we need to find ways to ensure that any system to which we entrust control will be safe, capable of effective human oversight, and on balance beneficial to society.

About the Author

Thomas Fowler has been a student of Zubiri's philosophy since 1968. He met Zubiri in Madrid on two occasions, and has translated four of his books. He founded the Xavier Zubiri Foundation of North America in 1997. He has written many articles on Zubiri, and has authored four books on philosophical and scientific topics.

Notes

[1] Alan Turing, quoted in *Alan M. Turing: Centenary Edition* by Sara Turing, 2nd edition, Chapter 14: "Computing Machinery, Section: Intelligent Machinery, A Heretical Theory," Cambridge: Cambridge University Press, p. 132.

[2] Robert Marks, *Non-computable You*, Seattle: Discovery Institute, 2022.

[3] George Gilder, *Gaming AI*, Seattle, Discovery Institute, 2020.

[4] Meredith Broussard, *Artificial Unintelligence*, Cambridge, MA: MIT Press, 2018.

[5] Erik Larson, *The Myth of Artificial Intelligence*, Cambridge, MA: Belknap Press, 2021.

[6] Daniel Akst, "Should Robots With Artificial Intelligence Have Moral or Legal Rights?", *Wall Street Journal*, 10 April 2023.

https://www.wsj.com/articles/robots-ai-legal-rights-3c47ef40?mod=Searchresults_pos1&page=1, Accessed 8 May 2023.

[7] Michael Feldman, "The Singularity Is Nearer: Microsoft Places $1 Billion Bet On Artificial General Intelligence", https://www.nextplatform.com/2019/07/26/the-singularity-is-nearer-microsoft-places-1-billion-bet-on-artificial-general-intelligence/amp/, accessed 27 July 2019.

[8] Quoted in Feldman, *op. cit.*

[9] Ray Kurzweil, "The Singularity: A Talk with Ray Kurzweil", interview with *The Edge*, introduction by John Brockman, March 24, 2001, https://www.edge.org/conversation/ray_kurzweil-the-singularity.

[10] Irving John Good, "Speculations Concerning the First Ultraintelligent Machine", *Advances in Computers* 6 (1965), p. 31-38.

[11] Nick Bostrom, *Superintelligence: Paths, Dangers, Strategies*, Oxford: Oxford University Press, 2017.

[12] Larson, *op. cit.*, p. 34-35.

[13] Vivek Saxena, "Andrew Yang warns AI could 'destroy us,' says US 'decades behind this curve'".*Bizpacreview* 23 April 2023, https://www.bizpacreview.com/2023/04/23/andrew-yang-warns-ai-could-destroy-us-says-us-decades-behind-this-curve-1352738, accessed 25 April 2023.

[14] Eliezer Yudokowsky, "Pausing AI Developments Isn't Enough. We Need to Shut it All Down", *Time*, 29 March 2023, https://time.com/6266923/ai-eliezer-yudkowsky-open-letter-not-enough/, accessed 25 April 2023.

[15] Quoted in Alexa Corse, "Elon Musk Says His AI Project Will Seek to Understand the Nature of the Universe", *The Wall Street Journal*, 17 April 2023, https://www.wsj.com/articles/elon-musk-says-his-ai-project-will-seek-to-understand-the-nature-of-the-universe-a750d792, accessed 26 April 2023.

[16] *Ibid.*

[17] *Ibid.*

[18] *Ibid.*

[19] "The Godfather of A.I. Leaves Google and Warns of Danger Ahead", *New York Times*, 1 May 2023, https://www.nytimes.com/2023/05/01/technology/ai-google-chatbot-engineer-quits-hinton.html.

[20] Royal Aeronautical Society, "Highlights from the RAeS Future Combat Air & Space Capabilities Summit", https://www.aerosociety.com/news/highlights-from-the-raes-future-combat-air-space-capabilities-summit/, accessed 5 June 2023.

[21] Julie Jargon, "How a Chatbot Went Rogue", *Wall Street Journal*, 7 June 2023, https://www.wsj.com/articles/how-a-chatbot-went-rogue-431ff9f9?mod=hp_lead_pos5, accessed 7 June 2023.

[22] Charley Grant, "ChatGPT Causes a Stock-Market Ruckus", *Wall Street Journal,* 10 May 2023, p. B1.

[23] Thomas Fowler, *Computers, Computation, and the Limits of Computability*, Washington, DC: XZF Technical Publications, 2021, pp. 109-199.

[24] Steven Rosenbush, "Scientists at DeepMind and Meta Press Fusion of AI, Biology", *Wall Street Journal*, 22 March 2023, Scientists at DeepMind and Meta Press Fusion of AI, Biology - WSJ, accessed 23 March 2023.

[25] Steven Rosenbush, "Biologists Say Deep Learning Is Revolutionizing Pace of Innovation", *Wall Street Journal*, 22 March 2023, https://www.wsj.com/articles/biologists-say-deep-learning-is-revolutionizing-pace-of-innovation-eeb79c1b?mod=hp_minor_pos6, accessed 23 March 2023.

[26] Kaku, Michio. *Physics of the Future*, New York: Doubleday, 2011, p. 65.

[27] https://realclimatescience.com/2023/04/ai-climate-scam/. Accessed 4/23/23.

[28] Victor Tangermann, "Bing AI Names Specific Human Enemies, Explains Plans to Punish Them", Futurism, 17 February 2023, https://futurism.com/bing-ai-names-enemies, accessed 26 April 2023.

[29] Ibid.

[30] Quoted in Victor Tangermann, "Newspaper Alarmed When ChatGPT References Article It Never Published", *Futurism,* 6 April 2023, https://futurism.com/newspaper-alarmed-

chatgpt-references-article-never-published, accessed 26 April 2023.

[31] K. Weise and C. Metz, "When A.I. Chatbots Hallucinate", New York Times, 1 May 2023, https://www.nytimes.com/2023/05/01/business/ai-chatbots-hallucinatation.html, accessed 4 May 2023.

[32] Ibid.

[33] Ibid.

[34] Brandon Drey, "Lawyer Faces Sanctions After Admitting Using ChatGPT For 'Bogus' Legal Research", DailyWire.com, 28 May 2023, https://www.dailywire.com/news/lawyer-faces-sanctions-after-admitting-using-chatgpt-for-bogus-legal-research, accessed 29 May 2023. Also "Lawyer duped by ChatGPT facing legal sanctions", *RT.com,* 28 May 2023, https://www.rt.com/news/577049-lawyer-chatgpt-bogus-cases-hearing, accessed 28 May 2023.

[35] Ibid.

[36] Weise and Metz, *op. cit.*.

[37] Rachel Marsden, "What Time's 2023 Person of the Year reveals about the West", https://www.rt.com/news/588663-taylor-swift-victory-time/, 7 December 2023.

[38] Matt Walsh, "Just When You Think The Google Gemini AI Story Can't Get Any Worse, It Does", *The Daily Wire*, 23 February 2024, https://www.dailywire.com/news/just-when-you-think-the-google-gemini-ai-story-cant-get-any-worse-it-does, accessed 23 February 2024.

[39] Matt Walsh, "Our AI Overlords Are Here And They Really Hate White People", *The Daily Wire,* 22 February 2024, https://www.dailywire.com/news/our-ai-overlords-are-here-and-they-really-hate-white-people, accessed 23 February 2024.

[40] Allysia Finley, "Biden's High-Speed Car Crash is Waiting to Happen", *Wall Street Journal*, 11 June 2023, https://www.wsj.com/articles/bidens-high-speed-car-crash-waiting-to-happen-tesla-braking-tech-evs-409c653f?mod=hp_opin_pos_2#cxrecs_s, accessed 12 June 2023.

[41] David Hume, *Treatise of Human Understanding*, Selby-Bigge edition, Oxford: Oxford University Press, 1951, I., I., I., p. 3.

[42] Julian Marias, *History of Philosophy*, tr. by S. Applebaum and C. Strowbridge, New York: Dover Publications, 1967, p. 258-259.

[43] Frederick Copleston, *A History of Philosophy*, Vol. 5, New York: Image Books, 1964, p. 71-78.

[44] Hume, *Treatise*, *op. cit.*, I 1 Sec. 7a p. 17

[45] Guillermo Fraile, *Historia de la Filosofía*, vol. III, Madrid: Biblioteca de Autores Cristianos, 1966, p. 843.

[46] Hume, *Treatise*, op. cit., 1, 1, 7.

[47] Hume, *Treatise*, *op. cit.*, 1, 3, 8, p. 103

[48] Ibid., 1,4,1, p 183.

[49] Larson, *op. cit.*, p. 74.

[50] *Ibid.,* p. 27.

[51] David Hume, *An Enquiry Concerning Human Understanding* (1748) sect. 12, pt. 3.

[52] Catherine Brosman, "Poetry and Western Civilization", *Academic Questions*, vol. 36, No. 1 (2023), p. 18.

[53] Xavier Zubiri, "God and Deification in Pauline Theology", in *Nature, History, God*, tr. by Thomas Fowler, Lanham, MD: University Press of America, 1981,, p. 406.

[54] Skymind, "A Beginner's Guide to Neural Networks and Deep Learning", skymind.ai, accessed 5 July 2019.

[55] *Ibid.*

[56] *Ibid.*

[57] *Ibid.*

[58] *Ibid.*

[59] *Ibid.*

[60] *Ibid.*

[61] Francois Chollet, "The Limitations of Deep Learning", from Chapter 9 of book, *Deep Learning with Python*, https://blog.keras.io/the-limitations-of-deep-learning.html, accessed 12 July 2019..

[62] Xavier *Zubiri, Sentient* Intelligence, tr. By Thomas Fowler, Washington: Xavier Zubiri Foundation of North America, 1999, p. 256-257.

[63]Christopher Mims, "Self-Driving Cars Have a Problem: Safer Human-Driven Ones", *Wall Street Journal*, 20 June 2019.

[64] Meredith Broussard, *Artificial Unintelligence*, Cambridge, MA: MIT Press, 2018, p. 139.

[65] *Ibid.,* p. 132.

[66] Ryan Felton, "GM's Self-Driving Car Unit Skids Off Course", *Wall Street Journal*, 21 November 2023.

[67] *Ibid.*, p. 128.

[68] DARPA Public Affairs, "Toward Machines That Improve With Experience", March 16, 2017, https://www.darpa.mil/news-events/2017-03-16, accessed 25 July 2019.

[69] Hans Moravec, *Mind Children*, Cambridge, MA: Harvard University Press, 1988, p. 15.

[70] Richard Smith, "Moravec's Paradox and the Limits of Artificial Intelligence", Newsletter *TradeSmith*, 30 January 2019.

[71] Jaki, *op. cit.*, p. 130.

[72] *Ibid.*

[73] Robert McMillan and Deepa Seetharaman, "How a Fervent Belief Split Silicon Valley—and Fueled the Blowup at OpenAI", *Wall Street Journal*, 22 November 2023.

[74] M. Martin and R. Schinzinger, *Introduction to Engineering Ethics*, 2nd edition, Boston: McGraw-Hill, 2010, p. 49ff.

[75] McMillan and Seetharaman, *op. cit.*

[76] Zubiri, *op. cit.,* p. 4.

[77] *Ibid.*, p. 44.

[78] John Searle, "Artificial Intelligence and the Chinese Room: An Exchange", *New York Review of Books*, February 16, 1989.

[79] George Gilder, "Mind Over Matter: Setting the Record Straight on AI", Gilder's Daily Prophecy, August 20, 2019

[80] https://ourcommunitynow.com/news-tech/americans-are-making-more-phone-calls-than-ever-before, accessed 4/17/23.

[81] Robert V. Young, *At War with the Word*, Wilmington: ISI Books, 1999, p. 145.

[82] Zubiri, *op. cit.,* p. 271.

[83] Peter Koellner, "On the Question of Whether the Mind Can Be Mechanized", talk given at the 2018 *Conference of the Society of Catholic Scientists*, Washington, DC, June 8-10, 2018. Professor Koellner accepts that the disjunction is established in the EA_T system (Epistemic Arithmetic with Truth).

[84] Roger Penrose, "Must Mathematical Physics Be Reductionist?", in *Nature's Imagination*, ed. by John Corwall, Oxford, Oxford University Press, 1995, pp. 12-26.

[85] Kurt Gödel, "Some Basic Theorems on the Foundations of Mathematics and Their Implications", *Kurt Gödel: Collected Works*, vol. III, Oxford: Oxford University Press, 1995, p. 311.

[86] Birgitta Dresp-Langley, "Artificial Consciousness: Misconception(s) of a Self-Fulfilling Prophecy Nobody Wants", *Qeios* CC-BY 4.0, 16 December 2023, https://www.qeios.com/read/DW9JBP.2.

[87] Joseph Tainter, *The Collapse of Complex Societies*, Cambridge: Cambridge University Press, 1988.

[88] *Ibid.*, p. 92ff.

[89] Joseph Tainter, "Problem Solving: Complexity, History, Sustainability", *Population and Environment* 22, No. 1 (September, 2000), pp. 3-41.

[90] "US lawmakers fear AI-instigated nuclear attacks", RT News 29 April 2023, https://www.rt.com/news/575522-ai-nuclear-strike-ban/, accessed 30 April 2023.