The Xavier Zubiri Review, Vol. 15, 2019-2021, pp. 71-105

Artificial Intelligence in Light of Zubiri's Theory of Sentient Intelligence[1]

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

Abstract

Rapid advances in computer technology and what is termed "Artificial Intelligence" in the past 70 years have led to speculation about the ultimate capabilities of electronic devices, including speculation about whether they will make humans obsolete at some future time. Zubiri's distinction between sensible intelligence and sentient intelligence can be applied to understanding of the limitations of AI. Machines can only operate on the sensible intelligence paradigm, which entails limits. Sentient intelligence allows humans to carry out functions that sensible intelligence-based devices can never do. Sensible intelligence-based devices, including AI, will therefore be restricted to amplifying human (i.e. sentient intelligence) capabilities, but never replacing them. Historically, extravagant claims for AI have not been borne out, despite many orders of magnitude increase in computation speed and memory capacity. Theoretical arguments, based on Gödel's Incompleteness Theorem, also confirm this.

Resumen

Los avances rápidos en la tecnología informática y lo que se denomina "Inteligencia Artificial" en los últimos 70 años han llevado a la especulación sobre las capacidades finales de los dispositivos electrónicos, incluyendo especulación sobre si harán que los seres humanos queden obsoletos en el futuro. La distinción de Zubiri entre inteligencia sensible y inteligencia sentiente puede aplicarse a la comprensión de las limitaciones de la IA. Las máquinas solo pueden operar con el paradigma de la inteligencia sensible, que conlleva límites. La inteligencia sensible permite a los humanos realizar funciones que los dispositivos basados en inteligencia sensible nunca pueden hacer. Por lo tanto, los dispositivos basados en inteligencia sensible, incluida la IA, se limitarán a amplificar las capacidades humanas (es decir, inteligencia sentiente), pero nunca las reemplazarán. Históricamente, las afirmaciones extravagantes de AI no se han confirmado, a pesar de que muchos órdenes de magnitud aumentan la velocidad de cálculo y la capacidad de memoria. Argumentos teóricos, basados en el teorema de incompletud de Gödel, también confirman esto.

I. Introduction

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 unbridled speculation about the future of humanity in our technological age. "Artificial Intelligence" is 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 formally reserved to humans alone, thus duplicating human intelligence, and will have much greater capabilities in the future. These areas include:

- Algebraic and symbol manipulation
- Robots and robotic systems
- Game playing
- Theorem proving

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. No less than Alan Turing (1912-1954) has informed us 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 should have to expect the machines to take control.[2]

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 still 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.[3]

In language reminiscent of Turing 70 years earlier, 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.[4]

Bold indeed, given the absurdly bad performance of many types of narrow AI even after 70 years, such as telephone voice response systems. Mechanization of human actions has been going on for millennia. Slaves are basically just automatons, forced to do certain types of repetitive work. Human actions involve a repetitive part, usually the result of training or experience, and creative part, which involves new types/ways of interacting with reality. Bearing this in mind, what, then, is the basis for AI, and how can it be analyzed in Zubiri's noology?

II. Paradigm of AI is Sensible Intelligence

Zubiri divides intelligence into two types: *sentient intelligence*, characteristic of humans, and *sensible intelligence*, of animals.[5] In this paper I wish to argue that machines based on AI, such as robots, operate on the paradigm of sensible intelligence; they do not sense reality. Insofar as they have "intelligence", they utilize sensors that receive some type of input (stimulation) from their environment. The inputs from all of their sensors (if more than one) are processed according to some algorithm, and then additional algorithms instruct the machine to take some action. At no point are they "conscious", understand reality, or even know about the "external world". I claim that this implies fundamental limits for these devices. Indeed, the limitations of AI and AI-based devices mirror the failings of epistemological theories based on sensible intelligence, such as Locke's theory of ideas, Hume's theory of the mind, and Kant's theory of synthesis.

Zubiri has explained the difference between sensible and sentient intelligence, with respect to knowing reality:

> Classical philosophy…believed that there are two acts: the act of sensing gives "to" the intelligence what it is going to work on, i.e., to know intellectively. But this is not the case. The difference between "to" and "in" is essential. That difference expresses the difference between the two concepts of the intelligence. To say that the senses give "to" the intelligence what it is going to work on is to suppose that the intelligence has as its primary and suitable object that which the senses present "to" it. If this were true, the intelligence would be what I call a "sensible intelligence". A sensible intelligence is an intelligence "of" the sensible. On the other hand, to say that the senses sense what is sensed "in" the intelligence does not mean that the primary and suitable object of intellective knowing is the sensible, but rather something more than that, viz. that the very mode of intellective knowing is to sense reality.[6]

Hence, it is something of a misnomer to speak of "intelligence" in connection with AI. What electronic devices perform is not real intelligence because it does not go beyond content, and even content only in the sense of taking inputs or "stimulations" and combining them according to pre-programmed algorithms. As Zubiri has noted:

> Still less can one speak—as is commonly done today—of artificial intelligence. In both cases what is carried out, whether by the animal or some electronic apparatus, is not intelligence because what they operate on and are concerned with is just the *content* of an impression, but not its *formality of reality*. What these animals or machines have are impressions of content, but without the formality of reality. It is for this reason that they do not have intelligence.[7]

Hence, there is no question that Zubiri rejected the idea that machine-based capabilities can ever duplicate human intelligence.

But this brings up the following questions:

- What is the major distinction between the way that sensible intelligence and sentient intelligence interact with the world?
- What empirical/observable implications of this dichotomy should we be able to observe?
- What does the history of computers and AI tell us?
- How does Zubiri's distinction apply to the traditional mind/body problem?
- What theoretical reasons support the idea of AI limitations based on the sensible intelligence/sentient intelligence distinction?

III. Major Distinction Between How the Two Types of Intelligence Interact with the World

One major distinction between automation devices and systems, including AI, and sentient intelligence, is that *contact with reality enables a creative ability to deal with the unknown.* Humans are very good at confronting problems and situations that they have never before encountered. Automated systems can be programmed to react in specified ways to a variety of conditions, but not to deal with reality in a way that requires creative thought. As we shall see in connection with neural networks, even the most sophisticated AI technology falls short in this respect.

The technologies involved with AI and other efforts have turned out to be ways to *enhance* human capabilities, which still require integration by a sentient intelligence to confront reality. Anything that can be reduced to a set of rules is subject to automation, whether requiring digital computers or not. Computers have made the process of creating equipment and processes utilizing rules and rule-based processes much easier and faster, but have not fundamentally changed the goals or means. This applies equally to autonomous systems, such as robots. They can be programmed with pattern recognition capabilities and feedback control, allowing for a degree of goal-seeking—obviously guided missiles have had this ability for decades—but however they are programmed, they have limits. Humans, who perceive reality, can look for vulnerabilities. This will be discussed further below, in connection with autonomous vehicles.

For the sensible intelligence paradigm, on which machines operate,

> Formalization is, as we have seen, independence, autonomization. And that which is apprehended in a merely stimulative manner is independent of the animal but only as a sign. This independence and, therefore, formalization, is merely stimulative. The distinct sensed qualities as mere stimuli are distinct response signs. Every sign is a "sign-of". The "of" is a response, and this "of" itself pertains formally to the manner of being situated and sensed signitively. Thus heat is a thermic response sign, light a luminous response sign, etc. [8]

That is, sensible intelligence operates on the basis of signs, utilizing a type of stimulation paradigm. This exactly describes AI. In light of Zubiri's philosophy, it is reasonable to infer that sentient intelligence, because of its direct contact with reality, is able to do what cannot be programmed, cannot be reduced to rules, and requires creative interaction with reality, such as love, friendship, formulation of scientific hypotheses, creation of art, and other tasks that do not fit the stimulus/response feedback paradigm.

Intellective knowing, through sentient intelligence, involves the ability to think, which is a search for reality in a way that is not operational but truth-seeking:

> In thinking there is always a moment of reality and therefore a formal moment of intellective knowing…Above all, thinking is an intellective knowing which is open through the real itself, i.e., it is the search for something beyond what I already intellectively know. Thinking is always *thinking beyond*. If this were not so, there would be neither the possibility nor the necessity of thinking. But it is necessary to stress that this beyond is a beyond in relation to the very character of reality. We are not dealing only with the search for other things—that animals do as well—but with searching for real things. What the animal does not do is to investigate, so to speak, the reality of the real. But *we investigate not just to find real things, but also to find in these same real things, already known intellectively before thinking, what they are in reality.*[9] [italics added]

This is always the goal of human knowing, 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. This is well-known even to those unfamiliar with Zubiri:

> 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.[10]

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.

IV. Examples of Unique Capabilities of Sentient Intelligence

Is there any evidence for claims about machines and sentient intelligence? Let us review some examples of the capabilities of sentient intelligence not shared by sensible intelligence.

A. Personal causality.

Between persons (and only between them) there is a strict causality, which in turn implies moral obligation and moral responsibility. Thus when we say, "John murdered Bill," or "John robbed the bank," we are making statements that go beyond a simple report of observations—a report that would exhaust the sensible intelligence paradigm. In the first case, we are saying that John knew what would happen when he pulled the trigger; he knew that he would produce a certain reality—namely the death of Bill—and that he willed this to happen. This, clearly, indicates a perception of reality as given in sentient intelligence:

> …it is essential that we introduce a type of what we might call 'personal causality'. The classical idea of causality (the four causes) is essentially molded upon natural things; it is a natural causality. But nature is just one mode of reality; there are also personal realities. And a metaphysical conceptualization of personal causality is necessary. The causality between persons *qua* persons cannot be fitted into the four classical causes. Nonetheless, it is strict causality. As I see it, causality is the functionality of the real *qua* real. And personal functionality is not the same as "meaning". Persons find themselves functionally linked as personal realities…[11]

This type of causality is not just a simple application of classical notions of causality to persons, though it is built on that idea. In its most general form, it goes beyond that type of causality, and is irreducible to the causality of classical metaphysics with respect to rational explanation of the world, and still less reducible to the concept of a scientific law because it operates at a more direct level, that of primordial apprehension. This is what Zubiri refers to as *personal causality*: "And however repugnant it may be to natural science, there is…a causality between persons which is not given in the realm of nature."[12] That is, personal causality is not reducible to anything accessible to sensible intelligence or AI.

Anyone who has experienced deep friendship, or seen how the unselfish actions of a good person can radically transform others, will immediately grasp the concept. This is causality in the sense of production of reality—the key component of "classical" causality. Real changes are produced in other people, whose lives are often radically altered by their experience of contact with the good person whose life, works, and example inspired them in ways that no rational argument could do. This cannot be reduced to stimulus-response, or any type of feedback-mediated goal-directed behavior—the realm of AI and sensible intelligence, and indeed, makes no sense in that context.

B. Science

Next let us consider some examples of hypothesis and theory formulation in science, to illustrate the link between doing science and reality. Obviously, science operates at the level of reason; but it still is rooted in logos, which in turn is rooted in primordial apprehension of reality. So science, like all rational knowledge, has the goal of reaching reality. Recall Zubiri's words, reason is "measuring intellection of the real in depth."[13] Zubiri also noted,

> The moment of affection and the moment of otherness in an impression cannot be split apart (as we have already seen). Being impressions of ours does not mean being unreal, but rather being a reality which is impressively present. The determination of what these qualities are in the world beyond what is formally sensed is precisely the task of science.[14]

Thus science as an intelligible enterprise does require contact with reality. And this contact with reality is the same as that of the artist. In the words of CERN physicist Savas Dimopoulos:

> The thing that differentiates scientists is purely an artistic ability to discern what is a good idea, what is a beautiful idea, what is worth spending time on, and most importantly, what is a problem that is sufficiently interesting, yet sufficiently difficult, that it hasn't yet been solved, but the time for solving it has come now.[15]

That is, science requires a creative vision of reality, allowing the scientist to zero in on problems and manage the solution space in a way that makes no sense under any type of AI or sensible intelligence paradigm.

1. Special Relativity: Einstein and Electrodynamics

At the turn of the 20th century, Maxwell's equations for electrodynamics were well known. But there was a problem: the principle of relativity recognized at that time said that the laws of physics applied equally in any reference frame of uniform motion, and that Galilean transformations from one to another gave the correct answers. But it did not seem to apply to electrodynamics. Consider a train with observers, and stationary observers on the ground. A charged body on the train would generate a magnetic field as seen by the stationary observers because it appears to them to be moving, but would not generate a magnetic field for the observers on the train because as they see it the charge is stationary. Now consider the case where a wire loop is on the train, and a magnet is on the ground. (See Figure 1).

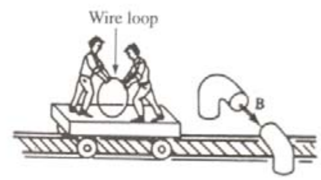


Figure 1. Moving wire loop and stationary magnet[16]

As the loop on the train passes through the magnet's field, observers on the ground see an electromotive force (EMF) generated,

$$\mathscr{E} = -\frac{d\Phi}{dt}$$

due to the magnetic field from the magnet acting on the magnetic field due to the moving charges in the loop as it goes by. But someone on the train, also applying the laws of electrodynamics, sees no magnetic force generated on the charges, but as the magnet goes by, the magnetic field seen by the loop changes, and by Faraday's law, an electric field is induced, so the electric force would produce an EMF in the loop

$$\mathscr{E} = -\frac{d\Phi}{dt}$$

the same as before, even though the physical interpretation of events, as seen by the train observers, is wrong. Einstein was bothered by this coincidence. The *formulae*, seemingly by coincidence, give the right answer, but the *reality* appears to be different.

> To Einstein's predecessors, the equality of the two EMFs was just a lucky accident; they had no doubt that one observer was right and the other was wrong. They thought of electric and magnetic fields as strains in an invisible jellylike medium called "ether", which permeated all of space. The speed of the charge was to be measured *with respect to the ether*—only then would the laws of electrodynamics be valid. The train observer is wrong, because that frame is *moving* relative to the ether.[17]

Einstein realized that it was absurd to take one frame of reference as absolute; all should be on a equal footing for electrodynamics as well as mechanics. This absence of any preferred frame was the outcome of the Michaelson-Morley experiment, performed in 1887. *The transformation from one frame to another had to correspond to an underlying reality that allowed for both interpretations (the train observers and the ground observers) to be equally valid.* Einstein realized that the only way this can work is if the underlying reality is that the speed of light is the same for all observers—which became the Special Principle of Relativity—and that the Lorentz transformation, not the Galilean transformation, is correct. This immersion in reality, combined with a creative imagination, allowed Einstein to solve a difficult problem and postulate not just a new theory, but one that revealed more of reality to us. This is what no amount of pattern recognition or pure symbol manipulation can do.

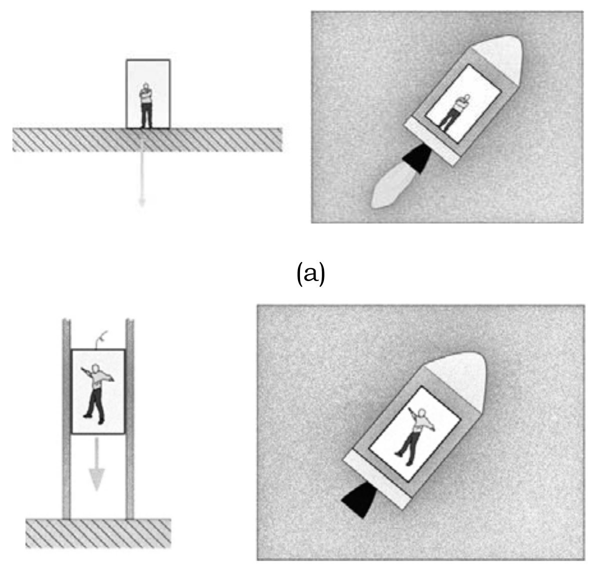
2. General Relativity: The Principle of Equivalence

We return to Einstein for another example. It had been known for more than two centuries that inertial mass (that which appears in Newton's second law, $F = ma$) is the same as gravitational mass (that which appears in Newton's law of gravitation):

$$F = \frac{GmM}{r^2}$$

Einstein knew that this could not be a mere coincidence, but had to reflect some underlying reality, because it is totally different from all other forces in nature, which depend on the composition of the objects involved. To determine that reality, he resorted to one of his famous thought experiments, this time, an elevator under various scenarios (see Figure 2).

If the elevator rests on the earth, an observer inside would feel gravity pulling him and any objects inside down toward the floor.



(a)

(b)

Figure 2. Einstein's "thought experiment" for gravity[18]

If the elevator is out in space away from any gravity, but is undergoing uniform acceleration, the observer will feel exactly the same type of force (Figure 2a). If the elevator is falling down a mineshaft, the observer feels no gravity (i.e., is weightless). If the elevator is in space but moving with uniform velocity, the observer again feels no gravity (Figure 2b). From this kind of thought experiment, Einstein realized some key things about physical reality, the most important of which is that gravity "is an aspect of the geometry of spacetime, not of the composition of the object moving through spacetime", and also that "the laws of physics in a uniformly accelerating frame should be the same as in a frame in a gravitational field".[19] This latter is the strong equivalence principle, and together with former is a statement about reality that was revolutionary, because it dispensed with the notion of force in this context, as Zubiri himself has noted.[20] Again we see that Einstein succeeded in his work because he was always in contact with reality and was seeking to understand reality at a very fundamental level. This, obviously, goes far beyond any conceivable machine capabilities.

C. Art, Literature, and Music

"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. Zubiri 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.[21]

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, 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 of the two main characters. 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 metanovel. 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 a those of Richard Wagner (1815-1882).

D. Mathematics

Zubiri discussed mathematics in some detail, especially in connection with Godel's Theorem, treated in Section VI. He emphasized that mathematics deals with reality, though obviously not the reality of chairs and tables. Mathematical objects are real by postulation; mathematics:

> …is a system of necessary truths about an object which, in its way, has reality before the intelligence. What the postulates postulate is not "truth" but "reality"; what is postulated is the reality of that about which one postulates.[22]

For Zubiri, of course, reality is not a zone of things, but a formality, the "in its own right" or *de suyo*. But mathematical objects, though "in their own right", are "irreal", as are characters from literature. Mathematics is a construction according to concepts:[23]

> The affirmations of mathematics and fictional literature thus refer to something un-real which is realized (made real) by constructive postulation, whether in the form of construction according to concepts (mathematics) or construction according to percepts and fictions (fictional literature). The intelligence is thus not limited to apprehending what "is already" in it, but also realizes (makes real) its concepts, its fictions, and its percepts in it, or rather, before it. What is intellectively known "is" not then before the intelligence but is something "realized" by the intelligence before itself.[24]

As an example for the present purposes, we may observe that the mathematician deals with reality in a creative way. Formulation of mathematical theorems is a very creative activity in which the mathematician acts and speaks in terms of the reality of that about which he is formulating his theorems. "How can one deny the essential function that creative fantasy has in mathematics!"[25] Mathematics is essentially creation, but not arbitrary creation. Zubiri notes that "mathematical construction is a free construction of mine in the realm of reality,"[26] though subject, of course, to certain constraints. The mathematician then explores the reality thus created, though Zubiri speaks of *reality in construction* rather than *construction of reality*:

> …the irreal inexorably has its "own" properties about which it is possible to debate. As I see it, this can only happen because the "created" is always and only the character of a content of physical reality itself. Physical reality actualized in a free system of ideas and previous affirmations can and does have more properties than those determined by the logical content of said ideas and said affirmations. And this is inexorable. Creation, then, radically and primarily concerns reason itself as intellection of the ground of something in depth.[27]

As an example, consider first a simple existence theorem, the well-known Intermediate Value Theorem:

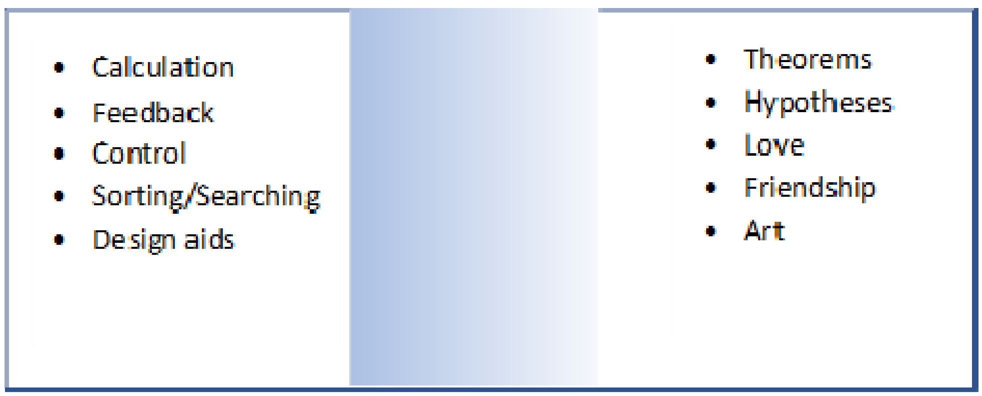
> Let $f(x)$ be a continuous function defined on the closed interval $[a,b]$. For any L, $f(a) \le L \le f(b)$ there exists a number c, $a \le c \le b$ such that $f(c) = L$.

Here we see that the existence of a number is required by the theorem, i.e., the theorem says that there *is* a number meeting the conditions—the number is *real*.

As another example, recall 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 a sensible intelligence paradigm. It certainly is not the correlate of any "sensible impression" (in Hume's parlance). Though symbol manipulation programs such as *Mathematica* can operate with infinity, they have only the capability to follow rules for manipulations involving it. As Zubiri has noted,

> 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.[28]

The boundary between AI/Sensible Intelligence functions and SI-unique functions is shown in Figure 3.



Rule based: can be done by humans but faster/better by machines | Boundary hazy | Requires Sentient Intelligence Cannot be reduced to rules

Figure 3. Sensible Intelligence and Sentient Intelligence Capabilities

Sentient intelligence, indeed, is able to perceive more about things than any type of AI or stimulus/response system because it senses more. For a sentient intelligence,

> 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.[29]

It is in this "more" that its capabilities beyond sensible intelligence come into play. Sensible intelligence can only ape what sentient intelligence does. Sensible intelligence can only react 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. Sentient intelligence allows a type of creative vision of reality that is behind theories and literature. This suggests the breakdown shown in Figure 4.

Merely emulating what humans do for some specialized task isn't same as actually carrying out these actions with an understanding and perception of reality.

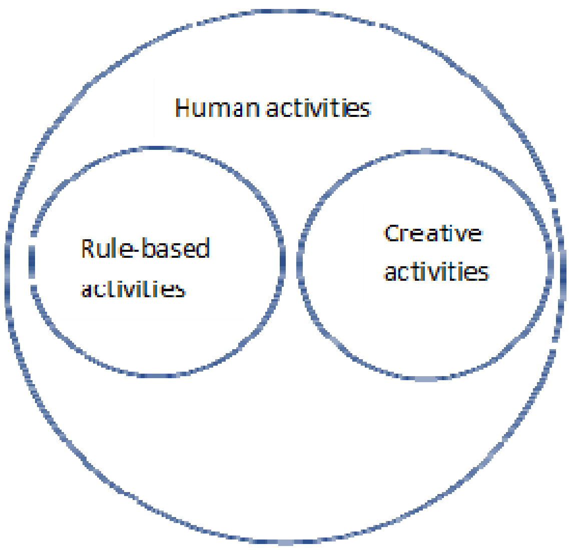


Figure 4. Breakdown of human activities

V. Sentient Intelligence and Sensible Intelligence: Is there an empirically observable or theoretically verifiable difference?

But the foregoing is not enough. It is necessary to go one step further and show that there are limits to what the sensible intelligence paradigm can do. If AI is really based on the sensible intelligence paradigm, a paradigm which is incorrect or at least inadequate, what are the observable consequences? If so, what form would they take? That is, if one attempts to do something by sensible intelligence that requires sentient intelligence, what would happen? Empirically what we observe is that the sensible intelligence paradigm, though initially very successful insofar as machines can be designed to carry out complex tasks that require sensing, data processing, and action, in ways formally reserved to humans, eventually reaches a point where further replacement of humans becomes increasingly difficult. Historically, the goal of AI was changed to that of aiding humans by expanding their capabilities. This, of course, has always been the goal of instruments and mechanical devices from simple machines up to automobiles. The sensible intelligence paradigm, as implemented in machines, works only insofar as machines can be programmed to imitate what sentient intelligence does. Let us examine next the historical evidence regarding machine capability.

A. History of Speculation About Machine-based Capabilities and Artificial Intelligence

Extravagent ideas about machine capabilities are far from new, and indeed speculation about machine intelligence, the brain as a physical device (whether mechanical, pneumatic, electrical, or electronic), and human-like properties of devices goes back at least to Lucretius (99-55 BC). This speculation is about what is now termed "General AI".[30] Blaise Pascal's (1623-1662) design for a mechanical calculating machine, sound but extremely crude by today's standards, triggered an outpouring of speculation. Thomas Hobbes (1588-1679) "described the brass and iron pieces of the machine as being invested with the functions of brain and instructed to perform some of the most difficult operations of mind."[31] Later machines, such Leibniz' (1646-1716) improved model (1694), together with further de development of Newtonian physics, convinced many of the inevitability of the mechanical human, among them Julian Offray De la Mettrie (1709-1751), in his famous book, *L'homme machine* (1748). As Jaki notes,

> De la Mettrie would probably have gladly recognized in the present-day advocates of thinking machines the

> future Prometheus, of whom he wrote that such an extraordinary figure might one day cope with the problem of constructing the mechanism of a talking man. In De la Mettrie's eyes, the problem simply consisted of pushing far enough the complexity and organization of the machine.[32]

De la Mettrie claimed, "thought is so little incompatible with organized matter that it seems to be one of its properties on a par with electricity, the faculty of motion, impenetrability, extension, etc."[33] Substituting conclusion for proof, of course, never works.

In the 1920s, Vannevar Bush (1890-1974) rediscoved the principle of Lord Kelvin's (1824-1907) differential analyzer, an early analog computer. He transformed it into a more modern type of computing device, crude by modern standards. But as usual, any excuse to make extravagant claims. The machine (mostly mechanical) was "a man-made brain which transcends human reasoning", the birth of an "electric thinker".[34] Furthermore, it was "an electrical machine which thinks for itself."[35]

The development of feedback-controlled devices, though based on relatively simple differential equations, also led to speculation about human-like behavior of machines. With the development of electronic digital computers in the 1940s, speculation received another boost, though ironically not from the developers of these new machines.

The Mark I computer, and the ENIAC, both developed in the 1940s to solve a particular type of dynamics problem using numerical methods, triggered an outpouring of extravagant claims: "Robot Einstein", "100 ton brain", "Educated machinery", "It thinks with electrons", and "Electrical Mathematicians".[36] Throughout the 1950s and 1960s, people routinely spoke of "electronic brains". No one uses terms like this anymore, despite the fact that the smartphones they carry have far more memory and computing power than room-sized mainframe computers of those days. As Jaki has noted,

> It is a rather sad aspect of scientific history that the voice of [popular] interpreters often prevails in creating the "momentary scientific consensus" as this is perceived by the general public. Time and again the public is induced to accept as "facts" propositions that are the conjectures of some writers dedicated to the propagation of certain "scientific beliefs." Their tactic betrays itself not only by its silence about weighty opinions to contrary. It also gives itself away by its inability to grasp the bearing of the fundamental propositions of a particular phase of physical science. Such was the case in Pascal's time when Hobbes argued the existence of thinking machines.[37]

Much of the argument associating computers with brains revolves around a corollary of sensible intelligence, namely the principle of isomorphism, which claims that the mind is some type of mirror image of sense perceptions, or their aggregate:

> Isomorphism, a keystone in Hume's thought, is still a cardinal tenet with positivists, empiricists, and sensationalists. It states a one-to-one correspondence between sense experiences and conceptual content, and in its most consistent formulation is equivalent to the proposition that, even without a functioning intellect, sense experiences would eventually order themselves into an intelligible pattern. In this view, the intellect or mind is at best a very useful but not an indispensable factor in the process of transforming sense perceptions into intelligible concepts.[38]

Not surprisingly, this theory has not fared well.

In the area of psychology during the first half of the 20^{th} century, the rage was physicalist theories such as Behaviorism, for which the idea of computers as brains fit well. The theory, according to founder J. B. Watson (1878-1958), viewed man as "an assembled organic machine ready to run", analogous to a car.[39] Psychology was envisioned as a rigorous branch of science, modelled on physics:

> …psychology was to become the study of humans, the future behavior of which, like the motion of billiard balls, could be predicted with absolute certainty. Watson cast psychologists in the role of engineers and physicists with the task of gaining full control over the subject of their investigations.[40]

Alan Turing (1912-1954) devised his famous test, namely that a computer would be equivalent to a human if, by means of interrogation, a human could not tell if he was speaking with another human or a machine:

> "Can machines think?"... The new form of the problem can be described in terms of a game which we call the 'imitation game." It is played with three people, a man (A), a woman (B), and an interrogator (C) who may be of either sex. The interrogator stays in a room apart front the other two. The object of the game for the interrogator is to determine which of the other two is the man and which is the woman. He knows them by labels X and Y, and at the end of the game he says either "X is A and Y is B" or "X is B and Y is A." The interrogator is allowed to put questions to A and B... We now ask the question, "What will happen when a machine takes the part of A in this game?" Will the interrogator decide wrongly as often when the game is played like this as he does when the game is played between a man and a woman? These questions replace our original, "Can machines think?"[41]

In retrospect, Turing's test is much too simplistic. Already by the early 1980s a trivial program, ELIZA, which ran on IBM PCs, could interact with people to do simple-minded psychoanalysis. The program, written almost as a joke, was taken seriously by many people.

Turing, writing 30 years earlier, had no doubts about the progress that would be made in the area of computers and AI:

> …at the end of the [20th] century the use of words and general educated opinion will have altered so much that one will be able to speak of machines thinking without expecting to be contradicted.[42]

In fact a much more difficult type of test is needed. Steve Wozniak, of Apple, proposed the "Coffee Test".

> To pass the coffee test, a robot would have to enter a home it has never seen before, make its way to the kitchen, and then successfully prepare a cup of coffee. For a person, this task would be trivial. For a modern robot it is beyond impossible. The robot would need a generalized sense of what kitchens look like; it would need to navigate potential steps and stairs; it would need a conceptual sense of what "coffee" is; it would need to be prepared to find anything from a Keurig machine to a standard coffee maker to a water pan or a French press; it would need facility with drawers, buttons, knobs, and shelves in any combination; and it would then need an elaborate series of improvised movements to make the actual coffee.[43]

This task is far beyond the capabilities of modern AI.

But, as often is the case where ideology is involved, facts are irrelevant:

> …it is some strange to find today some scientists busy constructing mechanical models, disguised in the language of electronics, of various mental processes, and of consciousness itself. There is something anachronistic in the efforts spent on finding mechanical equivalents of thought processes…It seems, however, that

> scientific thought has never been able to forego the illusory comfort derived from some *Deus ex machina*, the comfort of some cure-all concept that might provide an explanation even for the impossible….Today, it is the feedback that stands in some scientific circles as the symbol of universal intelligibility and explanation.[44]

Thus we learn that "human action is feedback action,"[45] and as a consequence, "man is just a naturally given robot",[46] and moreover there is nothing "superphysical" or beyond physics about consciousness,[47] presumably just a matter of scale. This from a 1963 book by James Culbertson, *The Mind of Robots: Sense Data, Memory Images, and Behavior in Conscious Automata*, whose title tells it all. Given the crude state of computers in those days, the rather long-range extrapolation implied is quite obvious. In 1964, Fred Hoyle (1915-2001) proclaimed that the difference between computers and humans is "one of degree and not of kind".[48]

Often enough metaphor was substituted for solid reasoning about computers, their operation, and capabilities. Even Turing fell into this trap. Commenting on Ada Lovelace's (1815-1852) objection to computers as able only to do what we tell them to do, he argues:

> One could say that a man can "inject" an idea into the machine, and that it will respond to a certain extent and then drop into quiescence, like a piano string struck by a hammer. Another simile would be an atomic pile of less than critical size: an injected idea is to correspond to a neutron entering the pile from without. Each such neutron will cause a certain disturbance which eventually dies away. If, however, the size of the pile is sufficiently increased, the disturbance caused by such an incoming neutron will very likely go on and on increasing until the whole pile is destroyed. Is there a corresponding phenomenon for minds, and is there one for machines? There does seem to be one for the human mind. The majority of them seem to be "sub critical," i.e. to correspond in this analogy to piles of subcritical size. An idea presented to such a mind will on average give rise to less than one idea in reply. A smallish proportion are supercritical. An idea presented to such a mind may give rise to a whole "theory" consisting of secondary, tertiary and more remote ideas. Animals' minds seem to be very definitely sub-critical. Adhering to this analogy we ask, "Can a machine be made to be super-critical?"[49]

That is, will scale result, ultimately, in qualitative change? To answer this, we must consider not metaphors but the growth of computing power and what changes it has in fact wrought.

B. Growth of Computing Power and Modern AI Systems

In 60 years, have we move any closer to the goal of "thinking machines"? Or has research and development gone in other directions? Does anyone aver that their smartphone or laptop can "think" or "have a soul", or any of the other notions common from 50-60 years ago? So what happened? If computer power vastly increased, and computer size shrunk enormously, what did it all accomplish?

A comparison of a mainframe computer from the mid-1950s (the IBM 650) and a modern smartphone (the Galaxy S10) shows the enormous progress made (See Table 1).

By any reasonable measure, these values represent an enormous leap in computing power and capability—and it compares a large mainframe computer to an individual carry-in-the-pocket device, not even to a modern supercomputer. In fact the smartphone can do amazing things, but all of them are designed to enhance the experiences of its human owner, not replace him. No one claims to have a "thinking machine" in his pocket.

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 Smartphone and Early Computer

Modern supercomputers are even more powerful, but are used for highly computation-intensive tasks such as

- Climate modeling
- Cosmological simulations
- Spaceflight calculations
- Molecular modeling

Computation power for supercomputers is usually measured in floating point operations per second, or "flops". Currently supercomputers can perform up to Petaflops, or 10^{15} flops, though not sequentially (these speeds require thousands of slower processors operating in parallel.) Compared to the IBM 650, which could do about 30 flops, this is a speed improvement of a factor of about 6 x 10^9 per processor, assuming 60,000 processors in the supercomputer. No one is claiming anything other than megaflops of computing power for these machines. There is no evidence that scale will lead to quantitative changes, even with (so far) scale changes of 10 orders of magnitude. If this enormous scale change has not only failed to create anything like what was envisioned 60 years ago, but has led to abandonment of the goals by most and a redirection of effort in other directions, there is no reason to suppose that there will be significant changes in the future.

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":

> 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.[50]

So, even with all the advances in computation ability, robots (and AI) are still little advanced from 60 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).[51]

In other words, the dreams of the early days of computers have not materialized, scaling has not brought qualitative changes, and AI has gone on to solve important problems, but not become conscious or even capable of simple human tasks.

This is not surprising: Zubiri noted long ago that humans are a different *kind* of reality.[52] Indeed, the situation with AI and the claims made for it are reminiscent of those made by would-be inventors of perpetual motion machines: success is always one tweak or one more gear away, but never seems to come.

C. Major Directions of Information Technology Research

In light of the foregoing, and the fact that makers of computing equipment want to find profitable uses for their equipment, rather than engage in unprofitable speculation about its long-term capabilities, Information Technology research has moved in two general directions.

1. Most of it has been devoted to applications development, including operating systems. The purpose of these applications is to make human tasks easier and more efficient. This falls under the rubric of "Narrow AI".[53] For example, modern word processors do not write sentences or essays, but they assist writers by formatting their work, correcting spelling and grammar mistakes, providing online dictionaries and thesaurus, and enabling copy and cut-and-paste operations, just to name a few. Similarly, spreadsheets facilitate operations with numbers, but do not tell users what number to use, or what calculations they should do. There are thousands of specialized applications for computers and smartphones, designed to automate a task, do complex searches, get information otherwise unavailable, enable faster communication, control equipment, or perform difficult calculations. *Mathematica* can solve extremely complicated problems at very high speed, but it uses rules developed by human mathematicians; it does not do theorem proving, which except in trivial cases requires a creative imagination. The recent proof of Fermat's Last Theorem by English mathematician Andrew Wiles in 1995 used computers, but as a way to do exhaustive search following instructions of the mathematicians. Brute force enumeration of all possible cases has always been a way to prove some types of theorems, but is not applicable for most since the number of possible cases can be infinite. Essentially, all of these applications utilize human-developed rules to process data and present the results to humans for action. Unquestionably, these programs and applications can do many tasks much faster and more accurately than humans, but no one believes that such capabilities make them human. These information technology devices and systems may displace human workers—that is a separate social problem—but they do not perform any tasks that, in principle, a human could not do. Many applications of AI also fall into this category.

2. The second area of computer development has been that of autonomous devices of various types, generally grouped under the name of "robots", though sometimes called "robotic assistants". The purpose of these devices is to perform complicated tasks in a manner similar to humans, only "better" in some way (or cheaper), and without human intervention. There is very little talk today about these machines "thinking" or assuming other human attributes.

D: Advanced AI Capabilities: Neural Networks, Artificial Intelligence, and Cognitive Computing

We must briefly review 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 sensory 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, which allows predictions, akin to regression analysis. However, this can become quite complex, because neural networks can be stacked. Neural networks are comprised of layers of nodes, which emulate neurons in the brain. A typical node looks like Figure 5:

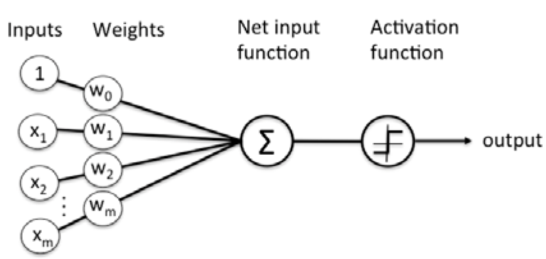


Figure 5. Typical node in a neural network.[55]

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. In practice, multiple layers are used, and each subsequent layer trains on a distinct set of features, using output from the previous layer, as shown in Figure 6.

With each subsequent layer, more complex features can be discerned. For example, see Figure 7. Such "Deep-learning networks" can extract features without human intervention or data labeling, unlike more conventional machine-learning algorithms, though they have significant limitations.

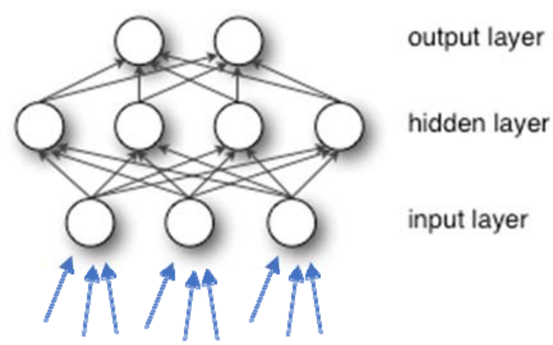


Figure 6. Neural network layers[56]

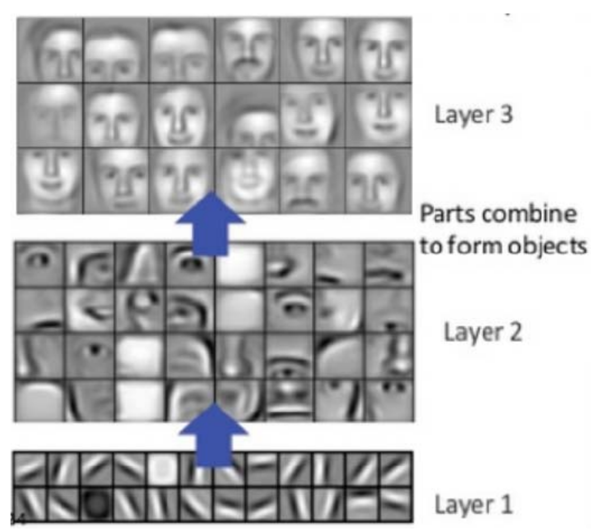


Figure 7. Recognition of higher-level features by neural network layers

At bottom, neural networks are not fundamentally different than other types of programmed machines:

> 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.[57]

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. Though in this respect, they are, again, basically just an ultra-high speed abacus. This is because they work on the basis of vectors in an abstract space:

> In deep learning, everything is a vector, i.e. everything is a *point* in a *geometric space*. Model inputs (it could be text, images, etc) and targets are first "vectorized", i.e. turned into some initial input vector space and target vector space. Each layer in a deep learning model operates one simple geometric transformation on the data that goes through it. Together, the chain of layers of the model forms one very complex geometric transformation, broken down into a series of simple ones. This complex transformation attempts to maps the input space to the target space, one point at a time. This transformation is parametrized by the weights of the layers, which are iteratively updated based on how well the model is currently performing. A key characteristic of this geometric transformation is that it must be differentiable, which is required in order for us to be able to learn its parameters via gradient descent. Intuitively, this means that the geometric morphing from inputs to outputs must be smooth and continuous—a significant constraint.[58]

Differentiability is a significant constraint because reality can often serve up rapid, discontinuous changes, for example, on a battlefield or in an accident situation. Humans can handle these situations, often admirably. The deep neural networks cannot. (Interestingly, mathematicians have developed mathematics to handle functions that are discontinuous everywhere, something that, obviously, deep learning robot mathematicians could not conceive).

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. But is it? 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 creation 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. Much more is involved than simple application of rules. 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 the limitations of this technology:

> 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.[59]

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.[60]

This quotation references the image in Figure 8.

American Technology guru George Gilder quotes Silicon Valley technology pioneer Keith Diefendorff on the capabilities of AI. Diefendorff has an array of patents in microprocessor architecture, reduced instruction set computing (RISC), optical interface technology, and other areas. He led the team that created the PowerPC micro-family for IBM and Motorola and later Apple. He also served for nearly a decade as editor in chief of the *Microprocessor Report*. He knows the researchers working in the AI area. According to Diefendorff:

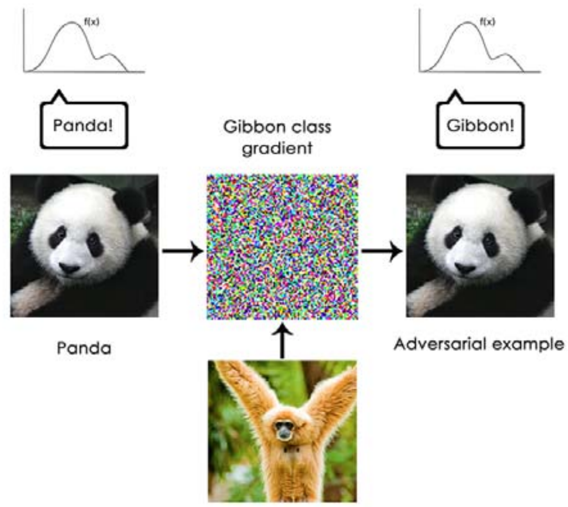


Figure 8. Example of how a deep learning neural network can be easily fooled in ways that would not fool a human observer[61]

> [the AI researchers] are getting nowhere with general purpose tasks. AI is proving good for specific niches… Games, in fact, are what they do best.[62]

Gilder observers that though niches are important, e.g., recognizing faces, interpreting speech, 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 is their misunderstanding of what computers do. Computers shuffle symbols. As philosopher Charles Peirce observed more than a century ago *the links between compu-*

> *tational 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.[63] (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.[64]

But in the real world, of course, the map is never the territory. 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.[65]

Obviously, human recognition of things, as realities, based on primordial apprehension, does not have this problem.

As an example, the author shows an example in which a deep neural network has given a hilarious caption for a photo (See Figure 9).

Cognitive computing is a subset of AI research that seeks to simulate or imitate human thought and reasoning processes utilizing a computerized model. The idea is to combine data mining, pattern recognition, and natural language processing to allow humans to interact with computers by normal speech, and then have the computers use their enormous processing power to answer questions. IBM's Watson is an example of a cognitive computing system. Such a system could be deployed

> …in a healthcare setting to help collate the span of knowledge around a condition, including patient history, journal articles, best practices, diagnostic tools, etc., analyze that vast quantity of information, and provide a recommendation. The doctor is then able to look at evidence-based treatment options based on a large number of factors including the individual patient's presentation and history, to hopefully make better treatment decisions.[66]

Figure 9. Example of misinterpretation of photograph by deep learning network

The goal of such a system is not to be a substitute doctor or replace the doctor, but expand the doctor's knowledge and range of treatment options by processing far larger amounts of data than any human could do in the time allotted. As the discussion of neural networks above indicated, however, a human doctor would always have to review any diagnosis or suggestion from the cognitive computing system for "sanity", since such systems can give absurd answers.

E. Autonomous Cars

The best-known example at present of a system designed to replace humans and duplicate their ability to interact with the outside world, and probably the most advanced, is autonomous or "self-driving" robotic cars, sometimes termed "robotic cars". Considerable effort is underway around the world at the present time to make such vehicles a reality. All the major automobile manufacturers, Google, and many others have research programs. Such vehicles, were they to become a reality, would suggest that AI and machines could in fact duplicate human intelligence in a very general way, even though driving as a human task does not involve creative thinking in the same way as painting or music composition. It does, however, involve constant interaction with reality and with continually changing, sometimes unfamiliar circumstances and actions. To understand the issues and difficulties of the autonomous car program, it is necessary to review the 5 levels of automation defined by the automotive engineers and the USDOT (Table 2):

Level	Capabilities	Status
0: Zero Automation	Driving as usual. A human driver is required to operate the vehicle safely at all times	Now
1: Driver Assisted, Function Specific	Intelligent features add a layer of safety and comfort. Human driver required for all critical functions. Car can alert driver to conditions, environment and obstructions. Can also offer assisted performance and driving capabilities	Now
2: Partial Automation	Key automated capabilities become standard but driver still in control. At least 2 simultaneous tasks are managed by the vehicle in specific scenarios	Now
3: Conditional Automation/Limited Self-Driving	Car becomes a co-pilot. Vehicle manages most safety-critical driving functions in known, mapped environmental conditions. Human driver present and expected to manage vehicle operation.	Now
4: High Automation (Highest NHTSA level)	Vehicle capable of performing all safety-critical driving functions while monitoring environment and conditions in defined use cases. Driver expected to take control if needed	?
5: Full Autonomous (SAE only)	Vehicle is completely driverless. Full-time automated driving in all conditions.	? Maybe never

Table 2. Levels of Automation in Automobiles[67]

At present most cars have level 1 capabilities, some have level 2. There are level 3 cars, and some claim level 4 in testing. But the key is the phrase "known environmental conditions" and "defined use cases". That is, the autonomy does not extend to unknown or un-programmed conditions or cases. That is, the car cannot deal with reality in the broadest sense, only with what it has been programmed to do. Some have claimed that level 5 and possibly level 4 will require mapping of roads to a level of some 10 or so centimeters. Others have stated that—comically—a human in a central location will monitor vehicle cameras, and be able to take control of the car in certain situations. Programming cars to deal with mechanical failures—something humans routinely do—is not even on the above list. A tire blowout at 60 m/hr (100 km/hr) can cause highly erratic movements that very enormously from case to case. A human driver can usually bring the vehicle under control; but given the way deep learning neural networks operate, this type of malfunction could cause instability. As noted above, one of the main problems with these networks is that no one knows exactly how they are programmed, so it is impossible to know how they will act under circumstances that deviate from their training.

As an example, a security robot went crazy and threw itself into a pool of water in a Maryland shopping center (Figure 10). This is comical but would not be funny if it happened to an autonomous vehicle at 60 m/hr (100 km/hr) on a crowded highway.

Figure 10. Security robot after committing suicide in Maryland shopping center

Autonomous cars attempt to reduce driving to set of rules, which can be implemented in such a way that the car can react faster than humans, and thus reduce or avoid accidents. The neural networks are in place to do pattern recognition of objects in the path of the car, and to assist with recognizing road marks and signs. But they cannot deal with new situations for which no rule has been devised. Reality is broader than what rules, however written, can cover. Even those with no knowledge of Zubiri have conceded this point:

> Every driver makes hundreds of daily driving decisions that, strictly speaking, break driving laws (for example, crossing the yellow line to pull around a double-parked vehicle). It all works out fine because of something called "human judgment." But what company is going to program its driverless cars to break the law? And what regulators will approve that product, knowing that it has been programmed to break the law?[68]

Early efforts were able to reach level 1 fairly quickly, and then on to level 2. At that point, it became more difficult (and more expensive) to get to level 3. Level 4 has proven to be quite difficult, and there is some doubt about level 5. The date for it has been repeatedly pushed back, now reckoned to be decades in the future. A quick summary of some major problems is useful:

> [Autonomous cars] don't track the center line of the street on ill-maintained roads. They don't operate in snow and other bad weather because they can't "see" in those conditions…One self-driving car was spotted going the wrong way down a one-way street. The software apparently didn't reflect that the street was one-way. The cars are easy to confuse because they rely on the same mediocre image recognition systems that mislabel pictures of black people as gorillas. Most autonomous vehicles use…neural networks, which can be confused simply by putting a sticker or graffiti on a stop sign. GPS hacking is a very real danger for autonomous vehicles as well. Pocket-sized GPS jammers are illegal, but they are easy to order online for about $50…Self-driving cars navigate by GPS; what happens when a self-driving school bus speeding down the highway loses its navigation system at 75 mph [110 km/h] because of a jammer in the next lane?[69]

These problems are easily handled, or would never arise, with a human driver. 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 language reminiscent of Zubiri:

> 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. [70]

The thrust of this article is that the sensors and systems designed for fully autonomous vehicles are being used on human-driven cars to make them safer and more efficient, and ultimately better than autonomous cars. The implication is that level 5 will never be attained; this is the level that would indicate that machines had equaled humans in a task that directly involved perception of reality in a fairly general way (though driving still involves only a limited set of rules).

Others have made the same point about sentience:

> The core problem is sentience. Because there [is] no way to program theory of mind, the 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.[71]

To get around the formidable sentience problem, 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.[72] [Italics added]

Video games, however, are at best a pale reflection of reality, and quite unable to subsume all aspects of it.

The problem of sentience affects the ability of AI systems to deal with reality in other ways as well. Humans can see an object in one position, say standing upright, and 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.[73]

Interestingly, this is one of the problems that bedeviled Hume's analysis of human understanding, also squarely based on the sensible intelligence paradigm.

Locke sought to derive all knowledge from "simple ideas". Hume switches to "impressions", the immediate data of experience.[74] For him, impressions must precede 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.[75]

But in that case 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.[76]

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

Among the current technological problems with autonomous cars are:[77]

- Eye safety because of use of infrared light in LIDAR sensors
- Power consumption of hundreds or thousands of watts for AI equipment
- Potential cross talk from light beams of multiple vehicles

These are engineering problems and may be solved. More serious is the issue of hacking, and the related issue—which directly impinges on the reality aspect of robotic devices—of people attempting to sabotage the vehicles by putting dummies and other obstacles in the path of the vehicle, sending signals to jam or fool the sensors, for example, changing speed limit signs. It would be easy to paint over a 30 m/hr sign to read "80 m/hr". A human driver would immediately realize that this could not be correct, but an autonomous vehicle might not. Even worse would be the case of a hacker or terrorist taking control of one or even a fleet of vehicles. This, essentially, weaponizes a large, 3 tonne vehicle carrying a flammable liquid. It would be foolish to assume that, given the track record of hackers and terrorists, many will not attempt this.

What we have learned, from dealing with systems designed to carry out specific functions, is that nearly all have vulnerabilities because they cannot anticipate all the things that could go wrong, or all the ways in which they might be fooled. Humans, because they sense reality, are usually able to spot and deal with many if not all of these problems.

There is also a whole host of moral questions that emerge and that also indicate the contact with reality aspect of human driving:

> Imagine a person jumps in front of an autonomous car driving at 30 miles per hour in a busy street. In some circumstances, the two options could be to (1) hit the person, or (2) swerve and crash the car causing damage to the driver. Irrelevant as to when the AI component of the car reacts, there will be a braking distance. In some circumstances, there is no alternative solution aside from to injure, poten-

> tially fatally, one of the individuals involved. So which one?....Who is ready to allow the machine to make the decision? Who is ready to hand over control of such a moral conundrum to a computer? Or how do you programme such a decision into an algorithm? What data could the AI access to process such a decision? If there is a fatality, should someone go to prison; the driver who is not responsible for the direction of the vehicle? The person who stepped in front of the car by accident, causing a crash? Or programmer who wrote the algorithm?[78]

In fact, when presented with this type of moral dilemma, people tend to choose the option of fewest casualties, but they then say that they will never buy a self-driving car.[79] Our ordinary contact with reality allows moral questions, such as those dealing with responsibility, to be addressed. The sensible intelligence paradigm, using AI, does not do so.

In the end, this points to the need to use technology responsibly, taking into account its limitations due to sensible intelligence:

> We should really focus on making human-assistance systems instead of on making human replacement systems. The point is not to make a world run by machines; people are the point. We need human-centered design.[80]

The impending failure of autonomous cars to reach Level 5 tells us that fantasies about robots taking over the world are just that—fantasies—and therefore we need to concentrate on use of machines for what they can do to assist humans, bearing in mind that direct contact with reality is and will always be essential, and only humans have it.

F. The Ghost in the Machine

In one of the more curious developments of AI, we now have the situation where AI devices—supposedly autonomous—are backstopped by remote-control human guiders:

> Whether it's terrestrial delivery robots, flying delivery drones, office-patrolling security robots, inventory-checking robots in grocery stores or remotely piloted cars and trucks, the machines that were supposed to revolutionize everything by operating autonomously turn out to require, at the very least, humans minding them from afar…Until the techno-utopian dream of full automation comes into effect—and frankly, there's no guarantee that will ever happen—there will be plenty of jobs for humans, just not ones their parents would recognize. Whether the humans in charge are in the same city or thousands of miles away, the proliferation of not-yet-autonomous technologies is driving a tiny but rapidly growing workforce.[81]

This confirms that actual interaction with reality, using the proxy method of AI, does not work in the way intended. Once again, we see that AI can handle simple problems; but because it does not actually perceive reality, it is limited.

G. Virtual Reality (VR)

Virtual reality is a growing area of computer science in which the ability of computers to draw three-dimensional images is coupled with binocular-type displays to immerse the user in a "virtual" world. Of course, this world is not real, but can seem very real to the observer, who is actually able to interact with it through head and body movements. In virtual reality, we take as real things that are only impressions, and therefore "spectres" in Zubiri's terminology. Things in virtual reality have existence, but not reality in the true sense because they are not "in their own right", *de suyo*. Virtual reality created by computers is different than more traditional examples of spectre-type reality, such as television and motion pictures, because it is entirely computer-generated and it allows the user to interact with it and thereby change the course of the "reality" presented. But can this "virtual reality" ever approach the level of the real world? In other words, can machines fully simulate reality? It would seem that if they could do so, they could do anything that the human mind could do. To this question, Zubiri replies that the answer is no, because reality cannot be captured in any human "formula", and therefore cannot be simulated by machines, however sophisticated. In Zubiri's terminology, reality is constitutively open, an element missing from VR simulations. Spectral phenomena, because they are not real in the sense of "in its own right", cannot create reality as we know it:

> In virtue of this, the formality of reality has, besides its individual moment, a moment of openness toward something beyond the reality considered individually. That is to say, a thing, by being real, exceeds or goes beyond itself in a certain way. This moment of going beyond or excendence is grounded in the openness of the formality of reality. Every thing, by virtue of being real, is what it is; and considered according to its own reality, is in some way being more than itself.[82]

Machine creations—which are really programmed by humans—can only imitate reality, but not recreate it in all its richness.

VI. Zubiri's Distinction and the Mind/Body Problem

What is the impact of Zubiri's distinction between sensible and sentient intelligence on the traditional mind/body problem? This is an issue because "bodies", or at least bodies such as robots made using computers and mechanical equipment, surface the question of whether they have "minds" or, by extension, "souls". That is, can the sensible intelligence paradigm used by AI devices truly mimic the mind, considered as sentient intelligence? And what does that mean for the "soul"? Closely related is the question of what it means for a mind to affect something material.

A. AI, the Mind, and Gödel's Theorem

We proceed by examining the implications of Kurt Gödel's (1906-1978) famous Incompleteness Theorem (1931). Volumes have been written about the implications of Gödel's result for the mind-body problem. This in itself is an index of how critical and relevant the theorem is, in a way that, say, the Periodic Table or the Special Principle of Relativity are not. Roughly the theorem states that for most branches of mathematics, including arithmetic, any attempt to create a deductive system, consisting of axioms and rules of inference, one that can establish only true propositions (theorems) about the subject, will be *incomplete* in the sense that there will be true theorems that cannot be proved in the system. If the system is expanded to allow it to prove all true theorems, then it will also be able to prove *false* theorems, e.g., 1=0, which means that the system has become inconsistent. The immediate implication is that it is impossible to formalize all of mathematics—a notion that had been taken for granted prior to Gödel's result. Another implication is that *truth* in mathematics is not only not synonymous with *provability*, but much broader.

The standard argument against minds as machines, utilizing Gödel's result, is based on the levels of knowledge and reasoning involved:

> What Gödel's proof brings out so forcefully is that rationality, consistency, and anything that forms the bedrock of human reasoning is not merely a set of formal steps but implies the instinctive ability of man to reflect on the correctness of those steps. The fact that the mind cannot derive a formal proof of the consistency of a formal system from the system

> itself is actually the very proof that human reasoning, if it is to exist at all, must resort in the last analysis to informal, self-reflecting, intuitive steps as well. That is precisely what a machine, being necessarily a formal system, cannot do, and this is why Gödel's theorem distinguishes in effect between self-conscious beings and inanimate objects.[83]

Although some have argued that this theorem shows that the mind cannot be "mechanized", i.e., modeled as a strictly mechanical system, Gödel himself did not draw that conclusion. Rather, he argued that his theorems implied a weaker 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.[84] Much effort therefore has been directed to showing that the first half of the disjunction is true. Roger Penrose has argued for this conclusion (see below).

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]

This, incidentally, agrees with Zubiri's interpretation of Gödel's result, to wit, that it is about reality not exclusively about logical operations:

> Mathematical realism is one of the main implications of Gödel's Theorem. In fact…it does not mean the mere limitation of formal systems to express mathematics, but shows the real character of what is constructed according to the axioms and postulates. What is constructed in reality, by virtue of being here-and-now realized, is more than what is postulated when it is made: it has properties in its own right. Its structure is trans-logical or transcendental ("more" than any given content). Zubiri, like Gödel, considers Metaphysics, and not Logic, the foundation of mathematics. It lives, in effect, from the transcendental or metaphysical.[86]

Penrose has used a similar argument to demonstrate that belief is non-computational. He argues further that understanding—or at least mathematical understanding—cannot be algorithmic:

> If it were algorithmic it would have to provide a formal mathematical system

> that was specific to doing sophisticated mathematics. Such things are very far from ordinary human experience…I claim that the Gödel argument demonstrates that whatever understanding is, it is indeed not a computational thing.[87]

Penrose does believe that our brains are rooted in physical activity, but that physical activity cannot be simulated computationally. Unquestionably the physical brain is responsible in some way for our mental processes; this is not in dispute. What is in dispute is whether those physical processes are capable of explaining all of the reality associated with the mental life. The Gödel disjunction demonstrates that it is not.

Analogous to the wave/particle duality, the truth is more likely that even the material has aspects not capturable in the functional relations used by science. As in the case of the wave/particle duality, this is a limitation of the human mind, which cannot perceive reality as a whole, in all of its dimensions. For this reason, it has great difficulty understanding the real, metaphysical causality involved even in purely material interactions. The net result is that attempts to reduce human thought to physico-chemical changes are doomed to failure, and a more holistic approach is required, one which looks at the reality of human experience as a whole instead of concentrating on particular functional relationships. Even this will not suffice to capture all of reality in some type of vision.

As Zubiri has noted, knowledge as a human enterprise is both dynamic and limited. It is limited because the canon of reality, like reality itself, can never be completely fathomed. It is limited because as human beings we are limited and must constantly search for knowledge. The ancient goal of knowledge about nature through causes understood in a metaphysical sense was that of an exhaustive knowledge. But that we cannot achieve:

> The limitation of knowledge is certainly real, but this limitation is something derived from the intrinsic and formal nature of rational intellection, from knowing as such, since it is inquiring intellection. Only because rational intellection is formally inquiring, only because of this must one always seek more and, finding what was sought, have it become the principle of the next search. Knowledge is limited by being knowledge. *An exhaustive knowledge of the real* [through causes understood metaphysically] *would not be knowledge; it would be intellection of the real without necessity of knowledge.* Knowledge is only intellection in search. Not having recognized the intrinsic and formal character of rational intellection as inquiry is what led to…subsuming all truth under the truth of affirmation.[88] [italics added]

As in all areas of science, it is important to have realistic expectations, and to recognize that limitations are not something negative, but a signpost of some larger aspect of reality.

At some level, of course, the whole discussion and physicalist position—that the mind is a physical machine—is absurd, because electrons, fields, and networks are radically different from ideas, concepts, and love. In fact, any attempt to assimilate them is an excellent example of what philosophers term a "category mistake". Any physicalist theory based on sensible intelligence (which they all are),

> …cannot ignore the fact that ultimately it will be its task to show what physical configuration in the brain corresponds, for instance, to concepts like fourth dimension, *n*-dimensional manifolds, and the like. Physicalists will have to identify in the brain a non-evanescent molecular group as the vehicle of that "evanescent quality", the instantaneous rate of change of functions…[they] will have to explain what well-determined pattern in the brain is the equivalent of the con-

> cept of the indeterminancy principle and of indeterminancy itself. They will have to show what molecular fullness corresponds to the concept of vacuum or empty space…what relative to the absolute, what concrete to the abstract, and what sort of thing to the nothing. They will be beset, whether they like it or not, with the problem of finding the physical force…that will adequately translate the feeling of love, hatred, and curiosity into the categories of physics.[89]

In a sense, Behaviorists were the only ones who really understood the problem and recognized the fundamental disconnect between mental notions and physical configurations of molecules, electrons, and fields. By rejecting the reality of those mental notions, and ascribing reality only to verbal or other reports of them, they could sidestep the fundamental issues, though at the price of discounting experiences that everyone has.

B. AI, the soul, and computers

Are there implications for the notion of "soul"? Some AI proponents have claimed that machines will reach a level of consciousness and presumably then have "souls", or act like they do. This would be another consequence of the sensible intelligence paradigm. If the human brain is just another physico-chemical entity, fully analyzable with the methods of science, then the soul must be lurking there, or it doesn't exist:

> It has become the custom to discredit the idea of a soul by trotting out poor old dualist Descartes. The question Descartes notoriously failed to answer was how an immaterial soul could affect a material body. His admittedly unhelpful suggestion that it happened somehow or other in the pineal gland is regarded as the *reductio ad absurdum* of the whole idea. But few who pose this question have stopped to ask themselves how it is that a material body can be affected by anything whatsoever, even by another material body.[90]

Essentially, this is asking about the metaphysical basis of causality. But if reality has both spiritual (or mental) aspects, no answer to this question can be given without a comprehensive understand of reality in its totality, which we do not have and are not likely to ever have.

An example of this mystery, even at the level of "matter", is not far to seek. In physics, charged particles interact with each other by means of electromagnetic fields. These charged particles modify the fields as they move, and in turn the fields act on the particles:

> By what "means" or "mechanism" this happens, physics does not say. It simply says that when electromagnetic fields are present, the charges are, in fact, affected as described by a certain equation; and when charges are present, the fields are, in fact, affected as described by another equation. In other words, physics posits two types of entities and mathematically describes [them], but does not otherwise explain their influence on each other.[91]

Even invoking the notion of virtual particles that mediate forces does not solve the problem; aside from the fact that the reality of these particles is in itself very mysterious, how they interact with other particles is never made clear; all we can do is calculate probabilities.

This recognition of one of the limits of scientific explanations goes back to Newton, who was himself mystified by the idea of action at a distance, and thus why his own theory of gravitation worked:

> I have not as yet been able to discover the reason for these properties of gravity from phenomena, and I do not feign hypotheses.[92]

What this means is that science can give us the functional relationships between and among phenomena, but does not attempt to explain causality in the metaphysical sense of one reality acting upon another.

Returning to the question of neuroscience, we observe that it has not solved the problem of how "matter" as epiphenomena can be influenced by "matter" as body. As Barr has astutely noted, neuroscience has effectively punted, because it cannot

> …explain the connection between motions of material particles and mental experiences any better than Descartes was able to do. For neuroscience, in effect, *the entire brain is just Descartes' pineal gland writ large.*[93] [italics added]

Worse, it makes no effort to go the other way, i.e., explain how mental phenomena can affect the "material" body. Since this cannot be done with the canon of reality acceptable to modern neuroscience, the easiest solution is just to say that it doesn't. But, then, there is the minor problem that we all know that it does!

What this shows, once again, is that an insoluble problem arises when we attempt to force reality into one of our limited modes of understanding. And this neuroscience position does not even begin to broach the question of the ontological status of mathematical objects. As Barr explains:

> …a purely materialistic conception of man cannot account for the human power of reason itself. If we are just "a pack of neurons," in the words of Sir Francis Crick, if our mental life is nothing but electrical impulses in our nervous system, then one cannot explain the realm of abstract concepts, including those of theoretical science. Nor can one explain the human mind's openness to truth, which is the foundation of all science…Scientific materialism exalts human reason, but cannot account for human reason.[94]

The net effect of this and of the sensible intelligence paradigm is to make man not "a little lower than the angels", but a lot lower, down to the level of other animals, denying his status as rational being since he reduces all mental capabilities and operations to programming and instinct. By supreme irony, he reverts to a pre-scientific, pagan worldview:

> Thus, like the pagans of old, the materialist ends up subjecting man to the subhuman. The pagan supernaturalist did so by raising the merely material to the level of spirit or the divine. The materialist does so by lowering what is truly spiritual or in the divine image to the level of matter. The results are much the same. The pagan said that his actions were controlled by the orbits of planets and stars, the materialist says they are controlled by the orbits of electrons in his brain. The pagan bowed down to animals or the likenesses of animals in worship, the materialist avers that he himself is no more than an animal. The pagan spoke of fate, the materialist speaks of physical determinism.[95]

And it is all so unnecessary if one is willing to be humble enough to take a holistic view of knowledge and accept that science does not tell us everything, or even everything important, about reality.

VII. Science, the Brain, and Reality

The brain, whether considered as an organ of formalization,[96] or just a fancy computer, is made up of elements from the Periodic Table, in accordance with the laws of physics and chemistry. So, why can't we make an "electronic brain", or at least, an artificial brain, with the same capabilities as the human brain? That is, why ultimately does the sensible intelligence paradigm fail? This question assumes (1) that the brain is a "thing", and (2) that it is possible to fully understand reality through some type of univocal theory—the famous "theory of everything". Zubiri's noology tells us that this is impossible. Reality is not a closed system, but fundamentally open:

> ...reality as reality is constitutively open, is transcendentally open. By virtue of this openness, reality is a formality in accordance with which nothing is real except as open to other realities and even to the reality of itself. That is, every reality is constitutively respective *qua* reality.[97]

This openness has very profound consequences: neither the brain nor any other object of study can be considered a "thing" in the sense of complete and isolated, and therefore necessarily replicable and fully understandable:

> The world is open not only because we do not know what things there are or can be in it; it is open above all because no thing, however precise and detailed its constitution, is reality itself as such.[98]

For Zubiri, the fundamental or constitutive openness of reality means that the search for it is a never-ending quest. Therefore no formula or theory, scientific or otherwise, can capture all of reality, or even all aspects of some part of it.

As Barr indicated in the last section, we have formulae that describe phenomena, but they do not tell us why or how the phenomena happen. That is, they do not tell us everything about the phenomena, on top of which there is an element of probability even in the descriptions that we do have.

Indeed, Zubiri believes that the development of quantum mechanics in the twentieth century has been an example of how our concept of reality has broadened. In particular, it has been broadened to include the concept of person as a fundamentally different kind of reality:

> That was the measure of reality: progress beyond the field was brought about by thinking that reality as measuring is a "thing". An intellection much more difficult than that of quantum physics was needed in order to understand that the real can be real and still not be a thing. Such, for example, is the case of person. Then not only was the field of real things broadened, but that which we might term 'the modes of reality' were also broadened. Being a thing is only one of those modes; being a person is another.[99]

Since being a person means being someone with sentient intelligence, and since this is different than being a "thing", this implies that the brain—or rather the human being as a comprehensive system—is more than the physics and chemistry involved.

Knowledge as a human enterprise is both dynamic and limited. It is limited because the canon of reality, like reality itself, can never be completely fathomed. It is limited because as human beings we are limited and must constantly search for knowledge. The phrase "exhaustive knowledge" is an oxymoron:

> Knowledge is limited by being knowledge. *An exhaustive knowledge of the real would not be knowledge; it would be intellection of the real without necessity of knowledge*. Knowledge is only intellection in search. Not having recognized the intrinsic and formal character of rational intellection as inquiry is what led to...subsuming all truth under the truth of affirmation.[100]

The net result of this understanding of human knowledge is that we cannot assume that we will ever know enough about reality to "replicate" the brain, which in any case would make no sense since the brain is part of a much more complex system required for sentient intelligence. Science itself is fundamentally limited and will never be able to capture reality in a "theory of everything". How the human body system works to allow perception of reality through sentient intelligence will most likely be forever beyond the capabilities of science because even the terms cannot be given meaning in science. Both of the assumptions in the initial question, therefore, are wrong. Though sentient intelligence is required to create and understand science, scientific theories and related engineering efforts are restricted to sensible intelligence, with all of its limits.

This can be clarified with reference to Zubiri's discussion of the canon of reality. He notes that science depends on a canon for its theories. We can only accept those things in the canon as viable components of scientific theories. This does not mean that other things, excluded from this canon, are not real, only that they are not allowed as components of scientific theories and explanations. Nor does this mean that the canon cannot change, as indeed it has done historically on many occasions. And it does not mean that things in the canon cannot be subjects of knowledge in other fields as well, including literature, theology, and philosophy. For example, we can discuss the universe in all three of these contexts, as well as in the context of science. Reality is not synonymous with the canon of scientific reality; it is much broader (see Figure 11). This implies that science will never be able to tell us everything about reality—and by extension, that AI will not be able to duplicate all that sentient intelligence can do.

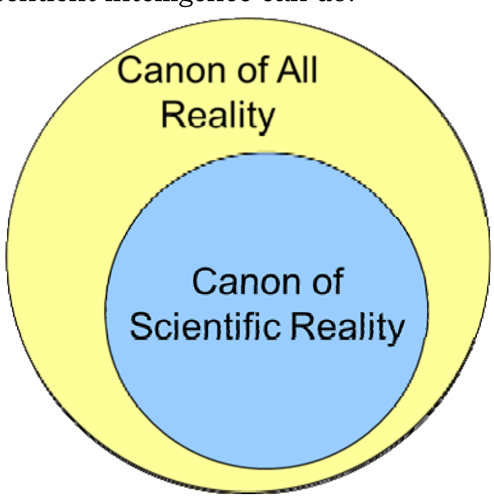


Figure 11. Canon of all reality and canon of scientific reality

VIII. Conclusion

AI is based on the paradigm of sensible intelligence, which limits it capabilities to rule-based behavior that, at best, can imitate some aspects of human behavior. Many other types of human behavior, such as friendship, love, science, and art, have little to do with rule-based behavior and are strictly within the realm of sentient intelligence. Sensible intelligence cannot perceive reality or accomplish any type of creative interaction with reality that requires perception of and direct contact with it. When attempts are made to duplicate behavior that requires contact with reality, at first progress can be made, but then it becomes exponentially more difficult, as in the case of autonomous cars. Empirical evidence shows that though AI can perform many tasks at much greater speed than humans, its primary applications have been in the deployment of functions to support human activity rather than supplant it. In this respect, AI is no different than many technologies developed in the past to assist humans with labor and with perception of the world, such as microscopes and telescopes. Additional evidence is provided by consideration of the mind/brain problem, and by investigation of the consequences of Gödel's famous Incompleteness Theorem. It makes no sense to try to imitate sentient intelligence by imitating brain functions, because sentient intelligence is the result of the entire physiology of the human. The following predictions can be made on the basis of the difference between sensible and sentient intellection:

- Autonomous cars, if marketed at all, will have only a very restricted usage; most of the technology under development for them will serve to make human-driven cars safer and better.
- AI-powered devices will never "take over the world", or even replace humans except in specialized areas.

- Robots will continue to be developed to carry out specific types of jobs, but will always be vulnerable because they do not sense reality.
- Areas requiring creative interaction with reality, such as art, literature, and science, are forever beyond the capabilities of AI, except in a fairly trivial imitative sense.

Notes

[1] Paper presented at the V Congreso Internacional Zubiri, Bari, Italy, September, 2019.

[2] Alan Turing, quoted from https://www.goodreads.com/author/quotes/87041.Alan_Turing.

[33] 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.

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

[5] Xavier Zubiri, *Sentient Intelligence*, tr. Thomas B. Fowler, Washington, DC: Xavier Zubiri Foundation of North America, 1999, p. 33ff (Herafter, SI). Spanish edition, *Inteligencia Sentiente*, Madrid: Alianza Editorial/Fundacion Xavier Zubiri, 1980, p. 82ff. (Hereafter, IS)

[6] SI, p. 34, IS, p. 83.

[7] SI, p.. 34, IS,. p. 85

[8] SI, p. 22, IS, p. 52.

[9] SI, p. 252, Spanish *Inteligencia y razon*, Madrid: Alianza Editorial/Fundacion Xavier Zubiri, 1983, p. 30-31 (Herafter, IRA).

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

[11] SI, p. 363, IRA, p. 339.

[12] [need ref. Not in M&G]

[13] SI, p. 257, IRA, p. 45.

[14] SI, p. 57, IS, p. 154.

[15] Savas Dimopoulos, from the movie *Particle Fever* (2016).

[16] David J. Griffiths, *Introduction to Electrodynamics*, 4th edition, Cambridge: Cambridge University Press, 2017, p. 504.

[17] *Ibid.*

[18] Illustrations from http://www.einstein-online.info/.

[19] Peter Collier, *A Most Incomprehensible Thing*, 2nd edition, Incomprehensible Books, 2014, p. 206-207.

[20] Xavier Zubiri, *Espacio, Tiempo, Materia*, 2nd edition, Alianza Editorial/Fundacion Xavier Zubiri, Madrid, 2008, p. 90-92.

[21] SI, p. 256-257, IRA, p.43-44.

[22] SI, p. 151, Spanish, *Inteligencia y logos*, Madrid : Alianza Editorial/Fundacion Xavier Zubiri, 1982, p. 129 (Hereafter, IL).

[23] Diego Gracia, "Teoría del espectro", *The Xavier Zubiri Review*, vol. 14, p. 99.

[24] SI, p. 152, IL, p. 131.

[25] *Espacio, Tiempo, Materia*, p. 75.

[26] *Ibid.*, p. 74.

[27] SI p. 279, IRA p. 111.

[28] SI p. 271, IRA p. 87.

[29] SI, p. 44, IS, p. 115-116.

[30] Meredith Broussard, *Artificial Unintelligence*, Cambridge, MA: MIT Press, 2018, p. 32-33.

[31] Stanley Jaki, *Brain, Mind, and Computers*, Regnery, 1989, p. 24.

[32] *Ibid*, p. 30.

[33] De la Mettrie, *Man a Machine*, La Salle, Illinois: Open Court, 1961, p. 141.

[34] *Literary Digest*, December 17, 1927, p. 25.

[35] *Ibid.*

[36] See Jaki, *op. cit.*, for citations.

[37] Jaki, *op. cit.* p. 59-60.

[38] Jaki, *op. cit.*, p. 60-61.

[39] Watson, *Behaviorism*, The People's Institute Publishing Company, 1924, p. 216, reprinted by Forgotten Books, 2012.

[40] Jaki, *op. cit.*, p. 151.

[41] Alan Turing, "Computing Machinery and Intelligence", Mind, Vol. 59 (1950), p. 433-60.

[42] *Ibid.*

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

[44] Jaki, *op. cit.* p. 69.

[45] Norbert Weiner, "Some Moral and Technical Consequences of Automation", *Science*, vol. 131 (1960), p. 1357.

[46] J. T. Culbertson, "Some Uneconomical Robots", in C. E. Shannon and J. McCarthy (eds.), *Automata Studies*, Princeton: Princeton University Press, 1956, p. 110.

[47] J. T. Culbertson, *The Mind of Robots: Sense Data, Memory Images, and Behavior in Conscious Automata*, Urbana: University of Illinois Press, 1963, p. 78.

[48] Fred Hoyle, *Of Men and Galaxies*, Seattle: Heinemann, 1965, p. 30-31.

[49] Turing, *op. cit.*

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

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

[52] Need ref on this

[53] Broussard, *op. cit.*, p. 32-33.

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

[55] *Ibid.*

[56] *Ibid.*, with modifications.

[57] *Ibid.*

[58] 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.

[59] *Ibid.*

[60] *Ibid.*

[61] *Ibid.*

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

[63] *Ibid.*

[64] *Ibid.*

[65] Chollet, *op. cit.*

[66] Bernard Marr, "What Everyone Should Know About Cognitive Computing", *Forbes*, https://www.forbes.com/sites/bernardmarr/2016/03/23/what-everyone-should-know-about-cognitive-computing/#1e8a7ae50884, accessed 5 July 2019.

[67] Based on Society of Automotive Engineers (SAE) and National Traffic Safety Administration (NHTSA) classifications.

[68] Steven Hill, "Why Driverless Cars Will Screech to a Halt", *The Observer*, 9 February 2016, https://observer.com/2016/02/why-driverless-cars-will-screech-to-a-halt/, accessed 15 July 2019.

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

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

[71] Broussard, *op. cit.*, p. 129.

[72] Broussard, *op. cit.*, p. 132.

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

[74] Frederick Copleston, *A History of Philosophy*, Vol. 5, New York: Image Books, 1964, p. 68.

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

[76] 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.

[77] Colin McCormick, "Self-driving cars find their way in the world,", *Physics Today*, July, 2019, p. 66-67.

[78] Jamie Davis, "Two reasons why autonomous vehicles won't work for a while", 12 September 2017, *telecoms.com*, http://telecoms.com/484620/weve-got-two-reasons-why-autonomous-vehicles-wont-work-for-a-while/, accessed 12 July 2019.

[79] Broussard, *op. cit.*, p. 147, quoting MIT Media Lab Director Joi Ito.

[80] Broussard, *op. cit.*, p. 147.

[81] Christopher Mims, "The Next Hot Job: Pretending to be a Robot", *Wall Street Journal*, 31 August 2019.

[82] SI, p. 247, IRA p. 17-18

[83] Jaki, op. cit., p. 220.

[84] 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).

[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] Guillerma Díaz-Muñoz, "Esbozo de una filosofía zubiriana de la matemática", *Actas del II Congreso de la Sociedad de Lógica, Metodología y Filosofía de la Ciencia en España*, ed. by Daniel Quesada and Anna Estany (Bellaterra, 6-8 de febrero de 1997), p. 141-145. Author's translation.

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

[88] SI, p. 336, IRA, p. 261-262.

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

[90] Stephen Barr, *The Believing Scientist*, Grand Rapids: Eerdmanns, 2016, p. 113.

[91] *Ibid.*

[92] Isaac Newton, *Philosophiae Naturalis Principia* Mathematica, General Scholium. Third edition, tr. by Bernard Cohen and Anne Whitman's, University of California Press, 1999, p. 943.

[93] Barr, *op. cit.*, p. 114.

[94] Barr, *op. cit.*, p. 10-11.

[95] Barr, *op. cit.,* p. 11.

[96] SI, p. 19, IS, p. 46.

[97] SI, p. 247-248, IRA, p. 19.

[98] SI, p. 248, IRA, p. 19.

[99] SI, p. 261, IRA, p.

[100] SI, p. 261-262, IRA, p.