UNIVERSITY OF BIRMINGHAM | THE JUBILEE CENTRE FOR CHARACTER & VIRTUES

Will Technology Help Humanity to Grow and Flourish?

Philip Larrey

This is an unpublished conference paper for the 8th Annual Jubilee Centre for Character and Virtues conference at Oriel College, Oxford University, Friday 3rd – Sunday 5th January 2020.

Jubilee Centre for Character and Virtues

University of Birmingham, Edgbaston, Birmingham, B15 2TT United Kingdom

T: +44 (0) 121 414 3602 F: +44 (0) 121 414 4875

E: jubileecentre@contacts.bham.ac.uk W: www.jubileecentre.ac.uk

I. Aristotelian-Thomistic Framework

The following paper contains reflections on the philosophical implications of AI. I vividly recall the time that the prominent Thomist philosopher, Ralph McInerny told me at the Pontifical Lateran University, "If you think that you can do philosophy without presuppositions, try it". For that reason it is important to state at the outset that these reflections come from an Aristotelian-Thomistic point of view. Why did I choose this philosophical framework? There are several reasons.

First and foremost, Aristotle and Aquinas have formed the basis of my own philosophical formation, received at the Pontifical Gregorian University in the 1980s and 1990s. This way of "philosophizing" became second nature for me, and it is thus natural that I would frame questions concerning AI in this context.

Second, Aristotle and Aquinas are two of the most important thinkers of all time. Of course this statement is somewhat subjective, i.e., according to whom are these two of the most important thinkers of all time? At the same time, their importance and influence are quite objective if one looks at the history of Western thought. Approaching philosophical issues from an Aristotelian-Thomistic point of view allows one to access an enormously rich tradition that has helped to shape Western civilization.

Third, Aristotle and Aquinas provide an epistemological basis for metaphysical realism. Although such an approach is quite problematic in philosophical terms, it represents a choice both useful and pragmatic. Realism has been under fire from many philosophers since the beginning of philosophy itself, but that does not mean that it does not accurately portray the relationship between human thought and the reality of the external world. Again, the choice of the Aristotelian-Thomistic point of view is a motivated one: it frames both an ontology (the way things are) and an epistemology (the way we think about things) which ground a metaphysical realism.

Fourth, Aristotle and Aquinas begin philosophizing from common, ordinary experience. Neither author has an "agenda" or an axe to grind. Neither author subscribes to an "ideology" or an a priori position that is the result of a bias or a preconceived notion. The foundation for their point of departure is the human experience of reality. Certainly, Aristotle was heavily influenced by Plato and Aquinas was influenced by faith in divine revelation, yet both begin their philosophical journeys from human commonsense knowledge. Again, not unproblematic, but a motivated choice.

Fifth, even though neither one wrote about AI, their framework helps to explain, describe and distinguish the difference between human intelligence and that which AI is doing. If we

describe human intelligence in terms of what AI is doing, then it is easy to conclude that machines are intelligent, that they learn, and that they might become conscious, just like humans. But that misses the point. Aristotle and Aquinas describe human intelligence within a systematic framework which has survived the test of time. Is it possible to reach into their anthropology, their ontology and their epistemology in order to shed light on the nature and meaning of AI? The following pages are an attempt to explain why I believe the answer to this question is a resounding "Yes".

II. Flawed Presuppositions

Stuart Amstrong works at the Future of Humanity Institute in Oxford, and takes a realistic approach to AI and all things dealing with machines. In his brief book, Smarter than Us. The Rise of Machine Intelligence he states that "It isn't just that computers are better than us in [certain] domains; it's that they are phenomenally, incomparably better than us, and the edge we've lost will never be regained. [...] Computers can't reliably beat the best poker players yet, but it's certain that once they can do so (by reading microexpressions, figuring out optimal bidding strategies, etc.) they will quickly outstrip the best human players. Permanently."[1]

Amstrong's book was published in 2014. It would take only three years for a computer to beat the best poker players. In a 20-day marathon, the software program called Libratus defeated four of the top professional poker champions in a Head's up No Limits Texas Hold'em tournament at the Rivers Casino in Pittsburgh, Pennsylvania on January 31, 2017[2]. One of the aspects that makes this victory significant for machine based intelligence deals with bluffing: the program would not be able to defeat such champions without bluffing, essentially "lying" about the value of its cards. This implies that it is not simply a question of "crunching numbers" or of calculating the odds of winning a hand with certain cards. The machine has to calculate what to bet, when to fold and when to stay on the basis of incomplete information.

And that is the other significant aspect of a machine beating the best poker players at the game, i.e., that of "partial knowledge". The machine does not have all the information it would require to make an informed choice: it does not know what the other players are holding. It can only proceed on the basis of its own cards and the cards showing by the dealer.

Professional human poker players have to be good bluffers, or they will not win consistently. We usually call such an ability intuition, which however vague a term, does convey the notion of making winning decisions on the basis of incomplete information. Human players must have good intuitive skills in order to keep winning. This is why playing poker is exciting: there is an unknown element that determines the winner, aside from the luck of drawing good cards. Witnessing a machine being able to master that unknown element was unnerving for the professional poker players. One even stated that he learned from the software how to play better.

Another example deals with the board games chess and Go. Chess and Go are somewhat different from poker in that moves can be calculated based on possible options. The machine has all the information it needs, because its opponent has a limited series of counter moves (although the number of possible moves in Go is nearly infinite). When Deep Blue from IBM beat the Grand Master Garry Kasparov in 1997, and Deep Mind's AlphaGo beat the world champion, Lee Sedol,

---

[1] Stuart Armstrong, *Smarter Than Us. The Rise of Machine Intelligence,* MIRI 2014, 13.

[2] See "*Carnegie Mellon Artificial Intelligence Beats Top Poker Pros*. Historic win at Rivers Casino is first against best human players", at https://www.cmu.edu/news/stories/archives/2017/january/AI-beats- poker-pros.html.

in Seoul on March 16, 2016, they were considered amazing showcases of computer intelligence."[3] Yet, beating four of the world's top poker players seems even more impressive still because of the machine's capacity to bluff.

Every day, we hear more and more hypothetical analyses of a "world controlled by computers" who are "out to destroy the human race". Many well-known figures have chimed in concerning AI as an existential threat. Elon Musk tweeted last year that AI is potentially more dangerous than nukes[4]. He also claimed at an MIT conference that "I think we should be very careful about artificial intelligence. If I were to guess like what our biggest existential threat is, it's probably that. So we need to be very careful with the artificial intelligence. Increasingly scientists think there should be some regulatory oversight maybe at the national and international level, just to make sure that we don't do something very foolish. With artificial intelligence we are summoning the demon. In all those stories where there's the guy with the pentagram and the holy water, it's like yeah he's sure he can control the demon. Didn't work out."[5]

Bill Gates agrees with Elon Musk, stating "I am in the camp that is concerned about super intelligence. First, the machines will do a lot of jobs for us and not be super intelligent. That should be positive if we manage it well. A few decades after that though the intelligence is strong enough to be a concern. I agree with Elon Musk and some others on this and don't understand why some people are not concerned."[6]

The famous British astrophysicist, Stephen Hawking, in 2014 warned us that "Artificial intelligence could spell the end of the human race."[7] Although he died in March of 2018, his posthumous book, Brief Answers to the Big Questions, published in October of 2018 continues to warn us that "While primitive forms of artificial intelligence developed so far have proved very useful, I fear the consequences of creating something that can match or surpass humans. Humans,

---

[3] Even so, the new program, AlphaZero, exceeded even the earlier AlphaGo by teaching itself from scratch (hence the name of AlphaZero). The programmers of AlphaZero were ecstatic about the results, because there was no human intervention in the programming of the machine: all the programmers told the machine was the definition of winning the game, and that it should always win in a competition. The program then figured out itself how to do so… within four hours. On November 28th, Lee Sodol announced his retirement from the game *Go*, convinced that "AI cannot be defeated". See https://inews.co.uk/news/go-champion-lee-sedol-retires-admitting-ai-cannot-be-defeated-1327548.

[4] See https://twitter.com/elonmusk.

[5] See https://www.washingtonpost.com/news/innovations/wp/2014/ 10/24/elon-musk-with-artificial-intelligence-we-are-summoning-the-demon/?utm_term=.7b838549a131.

[6] *Ibid*.

[7] See https://www.bbc.com/news/technology-30290540.

who are limited by slow biological evolution, couldn't compete and would be superseded."[8] He also predicts that there will be a clear difference between human beings who enhance themselves with advanced technology and those who do not. "Once such superhumans appear, there are going to be significant political problems with the unimproved humans, who won't be able to compete. Presumably, they will die out, or become unimportant. Instead, there will be a race of self-designing beings who are improving themselves at an ever-increasing rate."[9]

The Spectrum computer's inventor Sir Clive Sinclair stated that he thinks artificial intelligence will doom mankind. "Once you start to make machines that are rivaling and surpassing humans with intelligence, it's going to be very difficult for us to survive."[10] The new head of the British Science Association, Prof. Jim Al-Khalili, recently stated that "Until maybe a couple of years ago had I been asked what is the most pressing and important conversation we should be having about our future, I might have said climate change or one of the other big challenges facing humanity, such as terrorism, antimicrobial resistance, the threat of pandemics or world poverty. But today I am certain the most important conversation we should be having is about the future of AI. It will dominate what happens with all of these other issues for better or for worse."[11]

Adding to the scare are a dozen or so Hollywood movies which have been very successful at the box office. Blade Runner, The Matrix, Terminator, Her, Avengers: Age of Ultron and Ex Machina are only a few that place machines as existential threats to the human race. Yet, is machine control such a bad thing? It is relatively easy to scare people with "impending doom" and other apocalyptic scenarios, because people fear what they do not know. IBM's DeepBlue and Watson and DeepMind's AlphaZero scare us because the capacity of their AI systems seems almost boundless. People in general tend to fear that which is faster, smarter, more capable than most human beings. Nick Bostrom's bestseller, Superintelligence, presents several examples of such intimidating prowess.[12]

Human beings tend to fear machines also because of their lack of emotion. Robots seem heartless, cold-blooded and deprived of any and all emotions. They are impersonal, they behave according to rigorous laws and they do not take into account the richness and complexity of human beings (at least apparently not). At the same time, complex systems are beginning to be developed

---

[8] See https://www.thetimes.co.uk/article/stephen-hawking-ai-will- robots-outsmart-us-big-questions-facing-humanity-q95gdtq6w.

[9] *Ibid*.

[10] See https://www.bbc.com/news/technology-31023741.

[11] See https://www.breitbart.com/tech/2018/09/06/head-of-british-science-association-ai-greater-concern-than-terrorism/.

[12] See Nick Bostrom, *Superintelligence. Paths, Dangers, Strategies*, Oxford University Press, 2016². Another excellent analysis of the risks of "run-away AI" is James Barrat's work, *Our Final Invention*, Thomas Dunne Books, 2013.

in an interdisciplinary fashion in order to program into the machine a more humanistic approach based on detecting social signaling in human. "Our findings demonstrate a considerable influence of personality on everyday eye movement control, thereby complementing earlier studies in laboratory settings. Improving automatic recognition and interpretation of human social signals is an important endeavor, enabling innovative design of human–computer systems capable of sensing spontaneous natural user behavior to facilitate efficient interaction and personalization."[13]

Instead of resorting to "scare tactics" and "hype" in the media, we should continue the conversation about human-machine interactions, because, whether we like it or not, the machines are here to stay.

[13] See *Frontiers in Human Neuroscience* [www.frontiersin.org] 6 April 2018, Volume 12, Article 105, 1.

III. Consciousness

Much ink has been used (and probably wasted) concerning whether or not machines will become "conscious" sometime in the near, or not so near future. From a philosophical point of view, the difficulty facing this issue is that there is no consensus on what "consciousness" is in human beings.

Starting from a non-controversial (yet not-extremely-useful) beginning point, we can state that "conscious" in human beings is the difference between being awake and being asleep. When we see another human being lying "unconscious", the meaning is obvious: the human being is either asleep, knocked out (for example, as the result of a tremendous blow to the head from a professional boxer), or in a coma. When a patient regains consciousness (usually evident from eye contact), we say, "She is now conscious". Here, consciousness refers to being aware of one's environment, or being aware of sensory experience.

Although the difference between being conscious and being unconscious is pre-philosophically clear, it still does not explain what consciousness is. The challenge facing the philosopher in this regard is to explain the phenomenon in adequately philosophical terms. Such a challenge, at least in recent history, has been met through a variety of attempts to explain consciousness in scientific terms. Almost all of these attempts end in some type of reductionism, i.e., that consciousness can be "reduced" to physical-chemical events in the brain.

Other forms of such reductionism can be seen in materialism, naturalism or epiphenomenalism. An example of this last current of thought comes from William Robinson's book, Understanding Phenomenal Consciousness, where he writes in chapter ten, that "The picture that these facts support is one in which all our actions are effects of extremely complex, but completely physical, interactions of the cells in our brains and the neurons that lead into and away from our brains."[14] What Robinson means by "epiphenomenalism" is not, however, a reductionism, precisely because "In previous chapters, various attempts to fit phenomenal consciousness into a materialist framework have been considered and found to be inadequate... I have agreed that the position that phenomenal qualities are nonetheless material cannot be refuted; but the tenable versions of this position provide no explanation of how phenomenal qualities could be material, i.e., they are empty materialisms."[15]

Robinson's position is quite nuanced, for although his account of consciousness is a physicalist one, he would not accept being a reductionist. One of the most famous philosophers to try and tackle this issue is Paul Churchland. In one of his earliest works, Matter and Consciousness[16], he takes the reader through a series of proposals to explain what consciousness is. Although the book is somewhat dated, the ideas contained there-in are still very valid (which in itself is indicative of the stagnant nature of the contemporary debate).

---

[14] William Robison, *Understanding Phenomenal Consciousness*, Cambridge Studies in Philosophy, Cambridge University Press, 2004, 158.

[15] *Ibid*., 159.

[16] Paul Churchland, *Matter and Consciousness. A Contemporary Introduction to the Philosophy of Mind*. Revised edition. MIT Press, Cambridge 1990, 7.

An interesting thought experiment from Searle is relevant here: suppose we have a patient who sustained a brain injury. We manage to replace the damaged part with a special implant made of silicon, which exactly replicates the original part of the brain (with the exception of being made out of silicon and not the ordinary gray matter of the brain). Then, another part of the brain is damaged and so we replace that with another silicon-based implant. What happens if we eventually replace all of the biological parts of the brain with silicone implants?

Of course, we are nowhere even close to having the technology to actually do this, but the thought experiment is a meaningful one, especially when we consider that we do have brain implants which are effective (for example, cochlear implants for the deaf; and brain implants to help people who suffer from Parkinson's disease). Would our patient still have the same mental activity, just not based on biological gray matter but on silicon chips? The answer is not an easy one.

Assuming for the sake of argument that we actually possessed the technological know-how to pull this off (which is a huge assumption), there are two alternatives. One alternative looks at mental states as completely independent from the substratum in which they operate. Therefore, the silicon-based brain would work just fine, and our patient would be considered to have mental activity all the same. Another alternative looks at mental states as caused by the biological substratum of the brain, in which case the answer would be no: the silicon-based brain would not be able to cause the mental state as the biological brain does (unless, of course, along with the implants came such causal duplication).

The first alternative appears too much like substance dualism (and therefore frowned upon); the second alternative is Searle's point (and always has been): the biological brain causes mental states. "Actual human brains cause consciousness by a series of specific neurobiological processes in the brain."[17] Over the years, Searle's point has remained consistent: there is a fundamental difference between simulating mental activity (which a sufficiently advanced computer system can do) and causing mental activity (which implies causal powers). "Because we know that brains cause consciousness with specific biological mechanisms, any nonbiological mechanism has to share with brains the causal power to do it. An artificial brain might succeed by using something other than carbon-based chemistry, but just shuffling symbols is not enough, by itself, to guarantee those powers."[18]

The simple answer to the question, Can a sufficiently advanced AI become conscious? is that no, a sufficiently advanced AI cannot become conscious, at least in any way resembling consciousness we find in nature. The reason is that a machine is not a living form, and therefore does not have the same relationship to sensorial experience as natural forms that exist in nature.

---

[17] John Searle, "I Married a Computer", in Ray Kurzweil, *Are We Spiritual Machines? Ray Kurzweil vs. the Critics of Strong AI*, Discovery Institute Press, Seattle 2002, 66.

[18] *Ibid.*, 75.

We have already shown that the most fundamental distinction in the animal kingdom regarding consciousness refers to the difference of "conscious" and "unconscious" regarding sensory experience. People who are asleep, in a coma or who have been knocked out by a professional boxer are "unconscious", meaning that they are not in command of their perception and sensory experience. AI does not fit into this description. If we state that a sufficiently sophisticated machine with advanced AI is "conscious", what would it mean to be "unconscious"? The description itself does not suit the nature of the machine.

All of our personal computers (and especially laptops) have a "sleep" function in order to save battery power. But is the machine really "going to sleep"? If so, then we could apply the distinction "conscious/unconscious" to the machine; yet the laptop does not really "sleep", the way animals do: it just halts certain logical tasks in order to conserve electricity.

In order to be conscious in any way resembling animal consciousness, the AI would need to be alive. Could a machine become conscious in a non-living way? That is the question that AI experts are discussing right now. From an Aristotelian-Thomistic point of view, there is also an essential difference between animal consciousness and human consciousness: humans are self-aware in a way that animals are not. Human beings "know that they know", and can reflect on their cognitive experience, whereas less complex animals cannot (at least not in the same way). Aquinas calls this capacity reditio completa meaning that the human intellect is capable of "returning" to its cognitive act and placing that act as the object of the intellect (and not external reality).[19]

Aquinas writes: "Those things are said to turn towards themselves substantially that subsist by them- selves, having fixity in such a way that they do not turn toward anything else that holds them up, as accidents turn toward their subject. And thus it is proper to the soul and to each being that knows itself, that every being of this kind is a simple substance, sufficient unto itself by itself, as though not needing a material support."[20]

What Aquinas means when he states "not needing a material support" is that the soul is immaterial and capable of existing without the body. Aristotle also uses such an argument in order to "prove" that the soul is immortal: given that the human soul is capable of actions which transcend the body (like rational understanding) it must be capable of existing without the body. Therefore, after the demise of the material principle of the hylomorphic substance, the form continues to exist, albeit in a different dimension. The argument for the immortality of the human soul is not based on a faith commitment (the ancient Greek philosophers were pagans), but on rational discourse.

Reditio completa, the intellect's capacity to "return" to its intellectual act and revisit it, independently of the sensorial experience which was essential for the initial act itself, demonstrates

---

[19] For an expert treatise on this issue, see Therese Scarpelli Cory's chapter, "*Reditio conpleta, Reditio incompleta*: Aquinas and the *Liber de causis*, Prop. 15, on Reflexivity and Incorporeality" in the volume edited by Alexander Fidora and Nicola Polloni, *Appropriation, Interpretation and Criticism: Philosophical and Theological Exchanges Between the Arabic, Hebrew and Latin Intellectual Traditions*, FIDEM, Rome 2017, 185-230. There Cory analyzes the secondary literature on this issue over the past 50 years, in various languages.

[20] Thomas Aquinas, *Super Librum de causis espositio*, ed. Henri Dom- inique Saffrey, J. Vrin, Paris 1989/2002, 91.

the non-physical nature of the agent intellect and shows an essential difference between the human intellect and the animal intellect. Inasmuch as human consciousness indicates a relationship between the sensory apparatus of the human and the human's intellect, it makes little sense to call a machine "conscious". The relationship between the machine's AI and its sensory apparatus is completely digital. The machine's senses can actually be more powerful than human senses (like extremely powerful lenses for digital cameras, or auditory capabilities which far exceed those of human beings), although they differ essentially from human senses. Machine sensibility has to digitize the world in order to be useful for an AI, whereas the human being's sensibility is analogue. To ask, "Which is more effective?" makes little sense in generic terms: each of the sensory apparatuses do different tasks, geared toward either human processes or machine processes.

The human sensatory apparatus is particularly well adapted to perceive the world of common sense, i.e., objects which are not too large (like galaxies) and not too small (like subatomic particles). For centuries, human beings have used machines in order to extend the reach of our senses (especially the sense of sight, but also of that of sound). This use of technology to extend our senses will only increase in the future.

IV. Common Sense Knowledge

For Aquinas, "common sense" is a technical term referring to the first of all internal senses, as opposed to the external senses (of which there are five). There are other internal senses also, all part of what Aquinas calls the passive intellect (or the intellect in potency to act, by the agent intellect). The role of the common sense is to unify the data of the five senses into one item that will then pass through the imagination (or fantasy), memory (temporality) and the vis cogitativa which gives the species an initial value (in animal intellects this power is called estimative, because it does not participate in the higher power of reason; whereas in the human intellect it does).

As a technical term, common sense plays an essential role in the apprehension of external reality and the formation of the phantasma which is the result of hu- man perception and the basis for the mental construction of universal concepts.

The meaning of common sense in general is much more difficult to grasp. Often it is associated with good sense, as in the phrase, "She had the good sense to stop arguing with her three-year-old child", or "It makes good sense to be honest with the IRS". Yet there is a difference between good sense and common sense. At times, it is equated with general wisdom of the type that can be seen in the truth of proverbs. "Look before you leap", "Measure twice, cut once", "The early bird gets the worm", or "A stich in time saves nine" are examples of proverbial sayings that express a common knowledge that people value and routinely employ.[21] Common sense represents a uniquely human characteristic of knowledge, and it forms the basis of all subsequent human knowledge. "Common sense" is the subtle and efficacious knowledge of "common folk", the "everyday man" in the words of Michael Polanyi: "[...] An explicit integration [into scientific knowledge] cannot replace its tacit counterpart. The skill of a driver cannot be replaced by a thorough schooling in the theory of the motorcar; the knowledge I have of my own body differs altogether from the knowledge of its physiology; and the rules of rhyming and prosody do not tell me what a poem told me, without any knowledge of its rules."[22] Polanyi adequately shows in his other works that every attempt to eliminate or denigrate tacit knowledge (or common sense) is doomed to fail.

John Searle has always been a supporter of common-sense knowledge, especially in light of anti-common sense epistemologies that have surfaced in modern and post-modern thought. He pokes fun at a colleague from the University of California, Berkeley, while debating the soundness of post-modern conclusions. He relates the episode in the following way: "'Look,' I said, 'suppose you and I go for a walk in the moonlight, and I say, 'Nice moon tonight,' and you agree. Are we creating the moon?' 'Yes,' he said."[23]

---

[21] See Steven Shapin, *Never Pure. Historical Studies of Science as if It Was Produced by People with Bodies, Situated in Time, Space, Culture, and Society, and Struggling for Credibility and Authority*, The Johns Hopkins University Press 2010, 315.

[22] Michael Polanyi, *The Tacit Dimension*, Peter Smith (Double Day) 1983, 20.

[23] John Searle, *Mind, Language and Society. Philosophy in the Real World*, Basic Books 1998, 18.

Searle also identifies what he calls The Default Positions. These are views that we hold pre-reflectively, in such a way that denying them requires a conscious effort and a convincing argument. In our everyday lives, these positions are simply taken for granted uncritically. We can deny them, but to do so demands a philosophical enterprise that specifically goes against them. What are those "default positions"?

"(1) There is a real world that exists independently of us, independently of our experiences, our thoughts, our language. (2) We have direct perceptual access to that world through our senses, especially touch and vision. (3) Words in our language, words like rabbit or tree, typically have reasonably clear meanings. Because of their meanings, they can be used to refer to and talk about real objects in the world. (4) Our statements are typically true or false depending on whether they correspond to how things are, that is, to the facts in the world. (5) Causation is a real relation among objects and events in the world, a relation whereby one phenomenon, the cause, causes another, the effect."[24]

Although Searle goes on to state that these positions describe the Background to all our knowledge and are not identical to common sense, we can comprehend the default positions as a type of foundation for commonsense knowledge.

From an Aristotelian-Thomistic framework, common sense forms the basis for human knowledge, inasmuch as philosophical thought begins with ordinary, common experience. Such a framework has been rejected by many philosophers (especially those of the idealist tradition, or any scholar in the non-realist current of thought), yet for our purposes here – in relation to AI – we will rely on the Thomistic tradition, just as we have throughout this paper.

In human experience, the very first judgement that we make is that "there are things", "things exist", i.e., res sunt. All subsequent knowledge arises from this initial contact with the external world, this primum cognitum that forms the basis for a metaphysical realism with a robust ontology (the way things are) and a robust epistemology (how we know those things). Machines, even very fast and sophisticated ones, do not possess this type of access to the material world. As Polanyi concludes, "the process of formalizing all knowledge to the exclusion of any tacit knowing is self-defeating."[25] Whereas humans have senses, machines have sensors, and the machines' access to reality is not perception, but rather a process of digitalization, the formalization of what we humans count as knowledge. When a human jogger turns the corner in the park, she immediately and efficiently grasps her environment and understands her place. When a machine turns the corner, it must digitize the environment and match objects to what its internal memory indicates as real things. Machines are getting better and better at doing this, but it requires a huge amount of digital processing.

Cyc

The brief history of AI clearly shows how difficult common sense knowledge is for machines. The pioneer of common sense knowledge for machines is Doug Lenat who created Cyc, so far the best attempt to teach common sense to a computer. "If computers were to reach true artificial intelligence—tackling big problems entirely on their own—they needed more than just the raw speed that let them analyze more possibilities than the average human. They needed

[24] *Ibid.*, 10.

[25] Michael Polanyi, *cit.*, 20.

something beyond brute force. They needed a common sense of their own. So he [Doug Lenat]decided to build it."[26] That was over thirty years ago.

Cyc is now part of Cycorp, and provides a platform for other AI systems. The fundamental thesis of Cyc is that machines can be taught common-sense knowledge by feeding them human-based knowledge of the external world. Interestingly enough, the project tries to keep its knowledge base low, i.e., in the millions of key phrases, recognizing that common sense knowledge differs from data-base knowledge in essential ways. Lenat does not want to simply list phrases of common sense (which could easily reach billions) but in some way endow the machine with a capacity to learn commonsense knowledge.

Most of Silicon Valley has preferred machine learning and deep learning over common sense reasoning, because these techniques are achieving greater results and companies like Google, Facebook and Microsoft have heavily invested in such technologies. Lenat continues to believe that Cyc can achieve better results for real-world applications than AI systems which simply crunch data.

The "father of AI", John McCarthy indicated how important commonsense knowledge was going to be in order to create "thinking machines". His seminal paper was entitled, "Programs with Common Sense" and was published in the Proceedings of the Teddington Conference on the Mechanization of Thought Processes, 1959. It was his first major publication. It was also one of his most criticized. Several colleagues at the Conference stated that the paper made little sense. At the time, it was not clear how instantiation of "thinking" could occur in a digital computer unless the computer exhibited the capability of commonsense knowledge. Quite soon after McCarthy's work, AI went in another direction.

The difference between that which happens in a digital computer (including neural networks) and that which humans do with their mind will always be linked to the ability of human beings to access the metaphysical structure of reality through perception and reason. Only the future will tell if such a difference will continue to give human knowledge and understanding an advantage over what AI can achieve. In any event, human-centered AI will not only allow human flourishing, but will actually promote it. This will be our hope as advanced technology becomes more and more an intimate part of us all.

[26] https://www.wired.com/2016/03/doug-lenat-artificial-intelligence-common-sense-engine/.