



THE AMERICAN UNIVERSITY IN CAIRO
الجامعة الأمريكية بالقاهرة

Ahmed Zewail Prize Essay February 2018

The Biggest Event in Human History

By: Lobna Dawoud

لبني داود

The Biggest Event in Human History: The Renaissance of Artificial Intelligence

The introduction and prominence of artificial intelligence (AI) today sparks the beginning of a new renaissance in human history. AI is a field of science that aims to simulate the human mind and human intelligence in a machine, with the purpose of transforming our lives and substantially elevating our existence. After years of research and trials in this domain, scientists have succeeded in creating the first humanoid robot both able to introduce itself in front of an audience and gain human citizenship from the Kingdom of Saudi Arabia, Sophia. Her creation is the largest breakthrough invention in the history of the debated field of AI and a manifestation of its latest advancements. The event sparked human spectators to wonder whether their minds are nothing more than mere substance that can simply be emulated by a network of wires or rather a distinct non-physical being that is the essence of who they are. Their wonder begs the question: is AI the phenomenon that could end the human race or is it the biggest, most significant event in human history? In light of the important advancements in the field of AI, contemporary scientists are more inclined to celebrate the advances that they tend to overlook the philosophical ramifications involved as they are continuously degrading the mighty human mind to an architecture of wires and relentlessly feeding the seeds of an existential threat.

Going down the history lane, our ancestors faced the daily challenges of finding food and shelter then they marvelously evolved that they travelled through space, landed the moon and marked it with their footprints. Man has evolved from a caveman to a doctor who can treat disease, a scientist who can create algorithms and robots, a trader who can profit on financial markets and the list goes on. All these professions have one mutual element; they all rely significantly on AI, but what exactly is AI? Coined by the father of

AI, John McCarthy, AI is the science of designing and engineering computer-controlled machines and software with human-like intelligence to enable them to execute all cognitive tasks that the human mind can perform such as learning, reasoning, decision-making and problem-solving (Hutter). Till now, only weak AI has been developed where machines are human-coded and programmed to perform specific tasks. However, scientists are aiming to develop “artificial general intelligence” (AGI) which is a long-term far-fetched goal of developing machines that would outperform humans at every cognitive task in all domains present on planet Earth (Tegmark). Most scientists believe that the convergence of AGI with almost every industry will eventually cure all disease, eradicate war, end poverty and eliminate aging and even death; in a nutshell will radically transform our lives.

The accelerated progress in AI and the flourishing of several industries that has come along with it cannot go unnoticed. We have witnessed a boom in the medical industry as AI has become able to perform complex surgeries and diagnose cancer (Brynjolfsson and McAfee) and another boom in the field of astronomy as AI allowed space travel to become a reality. Furthermore, AI contributed in making military operations safer by using robots to reach dangerous and inaccessible zones (Hutter). On a commercial and economic front, dozens of companies use AI to prevent money laundering and make financial and economic decisions when trading on Wall Street (Brynjolfsson and McAfee). Moreover, we witnessed machines beat the world’s best human players in games like poker and chess which left us in awe. We have already lived the day when AI advances to the extent that cars will be self-driven, and drones and robots with artificially intelligent image, voice and speech systems will replace and enhance the job performance of security guards,

housekeepers, chefs and others. Advancements never thought possible about a decade ago are currently an on-going episode in the series of AI.

Despite all the advances, contemporary scientists have been under heavy scrutiny by many philosophers for their conviction that the human mind is a complex tangled web of brain cells, neurons and dendrites which can be replicated by wires and silicon forming “artificial neural networks” (ANNs) (Frings, Maschke, and Timmann). Inserting ANNs inside a machine is perceived as the only method to enable machines to have a human-like mind, thus become intelligent like us. For millennia, the theory of “reductionist physicalism” adopted by AI enthusiasts and the idea of attributing intelligence to machines have tantalized many philosophers and raised a host of philosophical arguments. Many have rejected the idea that machines can acquire human-like intelligence by simply emulating the human mind, refusing to believe that the mind is radically nothing but a biological computer. Advocates of René Descartes’s theory of Dualism consider the concept of “reductionist physicalism” that references the reduction of the human mind that thinks, understands, reasons, wills and desires to physical substance - the brain- as utterly absurd. The fact that this theory equilibrates mental states and phenomena to the chaotic movement of atoms of the physical brain (Frings, Maschke, and Timmann) diminishes the efficacy of the brilliant human mind to nothing but a mere working machine. Descartes reached the conclusion of “I think, therefore I am”, arguing that thinking is the essence of who we are. He believed that “I” which represents the mind, is the only thing he could not cast any doubt on since his consciousness proved his existence. Meanwhile, his body which can be easily doubted, does not exist. Thus, the mind, as a thinking thing, is considered a separate entity from the physical body which is an extended thing, since the mind can exist

with or without the body. Nevertheless, both mind and body interact together in commingled unity founding our nature (Descartes). Therefore, the pure and conscious thinking thing that is the human mind, based on this argument, cannot be explained by or degraded to physical substance like a network of wires in an AI system.

Furthermore, most philosophers have been skeptical of and were never satisfied with the idea of attributing intelligence to machines. Their argument arises to there not being a surefire method to determine whether these machines have developed human-like intelligence. In 1950, Alan Turing, the renowned English mathematician proposed the famous Turing Test to prove AI machine intelligence. The test entails that if the moderator fails to distinguish between the responses of the human and the AI machine tested simultaneously, the machine can be considered intelligent. Turing believed that intelligence was merely a display of behavior and performance, debating that over word meanings of “intelligence” and “thinking” were “fruitless disputes” initiated by “armchair philosophers” (Vellino). However, Turing’s negligence of the verbal issue led to heavy criticism by many philosophers as it masqueraded the philosophical dilemma that intelligence is intertwined with understanding and consciousness; hence, it cannot be assessed by performance only. Therefore, even if an AI machine passes the Turing Test, the idea that it can be considered intelligent is entirely incoherent as consciousness emanates from the human mind that is lacked by an AI machine.

Since there is no legitimate proof that AI machines which diagnose cancer can understand what cancer is and those that beat the best players in chess can perceive the meaning of the game, it is illogical to call these machines intelligent. Intelligence as Descartes indicated, comes from the mind which is a thinking thing and thinking is the

umbrella term for understanding, perception, reasoning, judgement, will, doubt, sensation and imagination. All these qualities are the definition sapience which makes us the distinct Homo sapiens. Therefore, an AI system that simply manipulates input binomial symbols of 0's and 1's producing a certain output, can neither understand the input it took in nor the output it produced. All it does is operate using hard-wired codes without the will to make decisions or the understanding of the generated output's meaning. As contemporary philosopher John Searle puts it, a AI system can form Chinese sentences like a native Chinese speaker by using rule books without understanding a single Chinese character- let alone what country speaks Chinese. Searle cuts through the noise with this famous and widely respected refutation to AI, known as the Chinese Room Argument (CRA). The bottom line is that it is impossible for an AI system which runs on currents, voltages and ANNs to “produce syntax from semantics; meaning from mere marks [and symbols]” (Cole). Therefore, attributing human-like intelligence and cognition to an AI system would be a huge fallacy regardless of the output it generates or the task it performs.

To further prove an AI system's lack of intelligence, when a programmed robot detects a flying animal through its AI senses, it can automatically label it as a bird. However, if the same robot detects a penguin, will it be able to realize that it is also a bird? The answer is no and that is because a robot can only observe without perceiving, since the capacity of perception relies on the faculty of judgement which originates from the mind (Descartes). Given that an AI robot lacks consciousness and sapience, it will not be able to judge and perceive the penguin as a bird. Even the sensory data that a robot receives through its image, sound, speech and other recognition systems are insufficient as they still lack “Qualia”- the human ability of getting emotional mental states like pain, pleasure, joy,

...etc. (Bostrom). An AI Robot can detect heat, but will neither feel the pain of burning nor perceive that it has been set on fire and will have neither the will nor the judgement to take action and put out the fire. No dualist including Descartes can deny the interaction between the mind and the body holding the brain or the way both are tightly joined, but that does not imply that the outstanding human mind can be dehumanized and reduced to the brain which is further reduced to multiple networks of wires by AI scientists in their quest for AGI.

Not only does AI devalue our mind, it also threatens our existence. Despite how fascinating all these new discoveries led by AI may seem; our fascination can lead us down the wrong path if we mistakenly overlook the potential existential threat of AI and the gun barrel it is pointing at the entire human race. Gurus like Stephan Hawking, Bill Gates and Elon Musk are not losing sleep over AI because of the possibility of machines becoming conscious and more intelligent than us since as aforementioned, AI machines can achieve neither. They are not anxious about Sophia turning into a red-eyed malevolent robot determined to destroy the human world. One crucial potential risk they are wary of is constituted in the creation of devastating AI machines such as autonomous weapons of mass destruction which would result in a “nuclear holocaust” if they fell in the wrong hands (Bostrom). Another risk lies in AI machines developing disastrous methods to achieve their final goals (Bostrom). Nick Bostrom perfectly describes how this could occur with his famous paperclips example where an AI machine is asked to make some paperclips. To achieve its goal, the machine would wreak havoc with the entire planet and exhaust its resources, turning it into a scrapyards of paperclips. Not only this, but also the machine would destroy any human who would try to pull the plug on it as they would be eyed as a

threat to machine's mission (Natalie Wolchover). The threat of goal misalignment in AI is frighteningly inevitable as controlling the machine's decision criteria and aligning its goals with ours is strikingly difficult.

Real jeopardy lies in AI machines gaining uncontrollable competence and autonomy not malevolence and consciousness. Uncontrollable competence can occur as machines gain the ability to recursively and aggressively self-improve and self-modify their neural architecture at an exponential rate that no human could ever match. Through technology, AI machines can expand their capabilities, download new skills, acquire more data and revitalize their codes; eventually developing "super intelligence" (Natalie Wolchover). Consequently, machines would become superiorly competent and autonomous that they would outperform Homo sapiens to the extent where we would lose control of the world; becoming enslaved and superseded. It is thus not surprising that Elon Musk is spending heavily on AI safety research to mitigate its existential risks. Simply put by the brilliant astrophysicist, Stephen Hawking, "AI could be the biggest event in the history of our civilization. But it could also be the last, unless we learn how to avoid the risks" (Rory Cellan-Jones). Therefore, unless these threats are alleviated, the ripple effect of the AI phenomenon will be like a rising tide that lifts and destroys all boats.

Technology has always been and will remain a double-edged sword. On the one hand, the amount of positive change in human welfare that has been witnessed due to scientific and technological innovation is remarkable. The idea of artificial intelligence was long thought of as a Greek myth or science-fiction, yet through technological advancements, it became a vivid reality that has the potential to help the human civilization further flourish. On the other hand, AI raises philosophical implications as it possesses the

potential to devalue and demonize the human mind and falsely attribute our sacred virtue of intelligence to machines. Humans have evolved to control the planet not because they are the fastest, biggest or strongest but because they are the most intelligent. Thus, if we lose that virtue due to our technological creations, we may one day find that we have developed our own successors in the supremacy of the Earth. We might even end up replaced through natural selection by the fittest robotic race of Sophia. All in all, if we do not take the necessary precautions to mitigate the threats of AI, the day the human civilization comes to an end by the AI phenomenon created by our own hands will arrive a lot sooner than presumed.

References

- Bostrom, Nick. "Ethical Issues in Advanced Artificial Intelligence." *Cognitive, Emotive and Ethical Aspects of Decision Making in Humans and in Artificial Intelligence 2* (2003): 12–17. Web. 6 Dec. 2017.
- Bostrom, Nick. "The Superintelligent Will: Motivation and Instrumental Rationality in Advanced Artificial Agents." *Minds and Machines 22.2* (2012): 71–85. Web. 2 Dec. 2017.
- Bostrom, Nick. "Superintelligence: Paths, Dangers, Strategies." Oxford University Press, 2016.
- Brynjolfsson, Erik, and Andrew McAfee. "The Business of Artificial Intelligence." *Harvard Business Review* 2017: 1–8. Web. 20 Dec. 2017.
- Cole, David. "Alan Turing & The Chinese Room Argument." *The Critique*, 31 Dec. 2014, www.thecritique.com/articles/alan-turing-the-chinese-room-argument/.
- Descartes, René, and Donald A. Cress. "Discourse on Method; and: Meditations on First Philosophy." Hackett Pub, Indianapolis, 1998;1999.
- Frings, M, M Maschke, and D Timmann. "Cerebellum and Cognition - Viewed from Philosophy of Mind." *Cerebellum* 6.4 (2007): 328–334. Web. 14 Dec. 2017.
- Hutter, Marcus. "Artificial Intelligence: Overview." (*Canberra ACT 0200 Australia*) 1–15. Web. 30 Nov. 2017.
- Natalie Wolchover. "Artificial Intelligence Aligned With Human Values | Q&A With Stuart Russell | Quanta Magazine." N.p., 2015. Web. 6 Dec. 2017.

Rory Cellan-Jones. "Stephen Hawking Warns Artificial Intelligence Could End Mankind." *BBC News* 2014: 1–5. Web. 4 Dec. 2017.

Tegmark, Max. "Benefits & Risks of Artificial Intelligence - Future of Life Institute." *Future of Life Institute* 2017: 1. Web. 5 Dec. 2017.

Vellino, André. "Artificial Intelligence: The Very Idea. J. Haugeland, (MIT Press, Cambridge, MA, 1985); 287 Pp." *Artificial Intelligence* 29.3 (1986): 349–353. Web.