Vidyankur: Journal of Philosophical and Theological Studies XXI/2 July 2019 | ISSN P-2320-9429 | **17-27**
https://www.vidyankur.in | DOI: 10.5281/zenodo.4128032
Stable URL: http://doi.org/10.5281/zenodo.4128032

Human Existence with Artificial Intelligence: Friendly or Frightening?

Rajesh Vijayakumar
Prabhodhana, Pallottine Institute of Theology and Religion, Mysuru, Karnataka

*Abstract***:** Technological advancements have transformed the human way of living, with its pros and cons. It has reached a point, where Man has started to become a mere spectator. It has replaced man from many of his traditional areas of expertise and has paved the way for many new innovations. Today a life devoid of technology cannot be comprehended nor be lived. Thus the technological evolution poses a fundamental question to us 'Is our existence with artificial intelligence is going to be friendly or frightening?' This essay tries to address this question by analyzing the possibilities, effects and challenges of Artificial Intelligence and carving out an optimal solution to it.

Cite APA Style: Vijayakumar, Rajesh. (2019) Human Existence with Artificial Intelligence: Friendly or Frightening? Vidyankur: Journal of Philosophical and Theological Studies July-Dec 2019 XXI/232 www.doi.org/10.5281/zenodo.4128032 17-27.

Keywords: Artificial Intelligence (AI), Friendly AI, Anthropomorphic Bias, Cyber Crime, Surveillance, Value-Loading, Goal alignment.

"The rise of powerful AI will be either the best or the worst thing ever to happen to humanity. We do not yet know which." -Stephen Hawking

Introduction

This article analyzes the impact of Artificial Intelligence on human society and to derive the modus operandi for the human society for a better future, that makes the optimal use of this

Evolution is the invisible truth of our visible universe. What we see today is the product of 13.8 billion years of evolution. From giant galaxies to tiny life forms, all have undergone it. Some could not keep pace with it, while others adapted to its ever-increasing demands and evolved. Humans too are the products of this indispensable and imperative process. While evolution occurred in other species was mostly one-dimensional (biological), humans exhibited a multi-dimensional evolution. Over the last 66 million years, humans have witnessed biological, social, spiritual, moral, economic, legal, political, and technological evolution and they still keep evolving. From the dawn of the 18th century, technical developments have spearheaded this process. The world today has already witnessed the phenomenal shift from Steam Engines to Automated factories, from the ENIAC Computers to Big Data and from Neural Net devices to Self-driving Cars. Technically we have witnessed two main revolutions: the Industrial Revolution

and Digital Revolution. While the former revolutionized our work life and amplified our outputs, the latter has revolutionized our standard of living and communication systems. Today we are witnessing yet another revolution, AI Revolution, which is practically revolutionizing all our mental tasks and slowly replacing ourselves. Future may reveal whether it is a boon or a curse.

AI: The New Sensation

Beyond the realm of computers, the concept of AI is familiar to us as a fictitious entity with apocalyptic implications. The Hollywood blockbusters like Terminator, The Matrix, Avengers, etc., have showcased this picture of the AI system. But AI is no longer a fictitious entity of the future. It is already part of our world and making rapid strides in our everyday life. From the Amazon recommendations to Spam filters of our email accounts, from virtual assistants like Siri, Cortana, and Alexa to airline autopilot and self-driving cars are all powered by AI system. Space exploration, Manufacturing, Transportation, Energy, Healthcare, and Communication are a few other fields that have highly benefitted from AI. Thus, the way we live today is very much defined by AI systems.

AI: What Is It?

In common parlance, AI is the field in the computer world, which is devoted to developing systems that can learn to make decisions under specific circumstances, based on the available data. These systems will be able to perform tasks that would require human intelligence, like visual recognition, language translation, decision making, etc. While its humble beginning can be traced back to early 1950s, the emergence of 'Big Data' and advanced technologies of robotics and sensing have fuelled an explosion of interest in AI and its applications (Mathew and Neupane, 2018). Basically, AI systems can be classified as 'Weak' and 'Strong'. Weak AI systems are the

machines that are designed and developed to respond to specific tasks or situations. They cannot think for themselves. Strong AI systems are those machines that are able to think and act just like humans. These are able to learn from their experiences. To date, there are no real-life Strong AI systems. The best representation would be the Hollywood portrayal of robots.

AI's Impact on Human Society

Most of us have experienced media-friendly AI applications like speech recognition, web recommendations etc., but AI has other widespread applications too. In the Healthcare domain, it fills the gap in human expertise to improve productivity and enhances disease surveillance. Automated diagnostic systems can give higher and accurate results at a faster rate so that the healthcare providers can work more efficiently. AI systems are capable to track and provide early warnings of possible epidemic outbreaks. In the administration domain, it improves the understanding and implementation of e-government applications and services. This drastically optimizes the government's service delivery and maximizes the services cost-effectively provided to the citizens. AI applications, like Artificial Intelligence for Disaster Response (AIDR), are efficient to plan and mitigate natural disasters. These applications process the overwhelming amount of information from different sources and sift out the necessary for rescue operations. In the Agricultural domain, AI applications help with disease identification in corps, Water management and drought monitoring by image analysis. Another field of application is Education. AI systems, like Intelligent Tutoring Systems (ITS), can take up the roles of tutors, teachers and administrators and thereby improving the teaching and learning process. It

makes education student-centred and personalized. This list can go on into other fields like Economy, Defense, Market, Surveillance, etc (Mathew and Neupane, 2018).

AI: The Dark Side

But is AI all saintly, or does it have a dark side too? The experience and current research suggest some potential risks. Anthropomorphic Bias and lack of Transparency are the two key areas of concern. Being developed and tested on data generated by humans, AI systems are also susceptible to biases. Such systems can exhibit discrimination systematically in critical areas of human life. Ad targeting is a typical example. Due to the complexity of AI algorithms, it is very difficult to isolate why a particular choice was made. Transparency becomes a deception when the system makes an unfair or unethical choice, leading to various liabilities.

A highly advanced AI system can result in human rights violation. Surveillance is processing personal data for the purpose of care or control, to influence or manage people. The use of technologies is leaving a huge amount of digital data both online and in the physical world. AI can process these endless data to keep track of our interests, locations, whereabouts, etc. This results in privacy erosion, which violates human dignity, personal autonomy, freedom of expression, freedom of choice, and freedom of movement. Our over-dependence on social media for information can be systematically manipulated by AI through automated Chatbots, fake news, and highly targeted misinformation. It can distort and manipulate public opinion on a very large scale and can lead to war, discrimination, hostility, and violence. Thus, AI can be a threat to the right to life and the right to equality. Cybercrime can increase due to AI misuse. It can lead to a proliferation of efficient forms of malware and spyware, through which identity, data and e-fund thefts can happen.

Cybercriminals with AI tools can become a risk to individuals, organizations and nations as such. But the greatest threat associated with AI systems is automation, which leads to Job loss. With the arrival of AI system, automation has become part of all domains. Today, many jobs which were traditionally considered to be impossible without human aid are no more so. For example, Competent Deep Learning algorithms have replaced dermatologist in cancer detection. AI-based automation will result in widespread disruption of labour markets and a major shift in the very nature of work (Mathew and Neupane, 2018). Existential philosopher Dr Viktor Frankl holds the view that work or job is one of the main aspects which gives meaning to human life (Fabry, 1988). Thus unemployment due to automation can result in existential problems (Andersen, 2018).

The Human Response to the AI Revolution

Weighing the pros and cons of AI, the future looks perilous for humans. But can AI-enhanced machines, as depicted in the famous movie 'Terminator', take over the world, wiping away the human race? Is that the next evolution that is in pipeline? There are four kinds of human responses to this dilemma. They are Optimists, Pessimists, Pragmatists and Doubters.

The Optimists foresee the future as a science fiction in which man is able to harness the speed and processing power of the AI systems by being directly connected to them. This will empower humans to avoid diseases, to slow down the ageing or even reverse ageing and finally making them immortal. Machines will take over all the work and humans will be left free to choose activities of their choice and time of work.

In Pessimist's view, the AI revolution will lead to such complex and powerful systems, that they will make humans an endangered species. They argue that as the machines become more powerful and intelligent, people will let them make all the important decision, as they will be flawless, unbiased, and produce better results. This will make us more dependent on them and afraid to make our own choices. Eventually, humans will be degraded to second rate status, making us their pets. This will directly impact our basic character to be free. Even though it may lead to a world that works perfectly, humans will no longer be humans.

The Pragmatists hold the view that the impending threats that AI can bring in the future can be mitigated by effective regulations and thus use AI to augment human skills. This will keep humans always a step ahead of AI or at least not in a disadvantageous position. Thus they emphasize research works on intelligence augmentation so that humans are always in the safe zone and capable of harnessing rich dividends from the AI.

The Doubters are sceptical and believe AI is impossible. So it is never a threat to humanity. For them, human intelligence and expertise cannot be replicated or be defined as informal rules. Even if the computers are provided with sufficiently advanced algorithms, they will be never able to replicate human intelligence, as human creativity is not always strictly rule-based. The human decision-making process often breaks the rules and becomes anti-algorithmic, based on the complexity of the situation. Thus the human creativity will never be duplicated by any AI algorithms. They will be always an insurmountable vacuum between the human mind and AI, like the paintings of masters are far superior to those of the millions of average painters (Makridakis, 2017).

The Pessimist view looks too naive. History has already proved that humans have an incredible survival instinct. We have adapted and evolved. When the industrial revolution resulted in automation, we did not stop working, but rather we created new jobs and reinvented others. Thus, technologies and innovations have always improved us and brought out the better in us. Similarly, AI too will bring out our better version. Let us not forget that the arrival calculators did not result in the extinction of mathematicians, but made them more efficient. They will free us from reiterating many tasks and thus reducing our work hours. AI can work and learn only within constrained parameters. So, they always will lack emotional intelligence and aesthetic sense, thus requiring a human mind guiding it to perform creative tasks, unless Singularity is achieved. Thus, it would be a folly to believe in the machine-Armageddon theory (Franklin, 2018). But rather it will take us beyond what we are today, making us superhuman. But we cannot also take the side of optimists, completely ignoring the potential threats. Proper measure in design and development need to be drawn so as to make AI as friendly as possible. Thus, we need to take the Pragmatists safe ground, to harness the advantages of AI to the maximum and reduce risks to the minimum. This calls for 'Friendly AI'.

Human Future and Friendly AI

According to Fyodor Dostoyevsky, the mystery of human existence lies not in just staying alive, but in finding something to live for (Tegmark, 2017). Dr Viktor Frankl would define man as 'Meaning Seeking Being', which in turn is oriented to a goal or a purpose (Fabry, 1988). Goals are power sources of our life, something that makes humanity move ahead. Being 'Goal-Oriented', all our basic activities and complex works have goals behind

them. The very goal of inventing new technologies is to make our life more comfortable. This is the aspect of man has been passed on to machines too. The only difference is that machines fulfil their goals to utmost perfection until they are properly fueled or they break down, while man's efficiency varies, based on different conditions.

AI-safety pioneer Eliezer Yudkowsky defines Friendly AI as those systems whose goals are aligned to those of humans (Tegmark, 2017). These systems will peacefully co-exist with humans and work side by side. If goals are not aligned, then their ability to accomplish goals can turn into a threat, as they will prioritize their goals over ours. So the real risk with AI system is not the malice it has, but its competence. Thus, goal alignment is the priority in the AI world. It involves making AI learn, adapt and retain human goals. To learn our goals, AI must figure out why we do something and not what we want to do. For this, the machine requires the knowledge of our preferences that go unstated in our requests. The machine needs to observe our behaviours and thus study our preferences. This poses two challenges. Firstly an efficient way to encode and store a system of arbitrary goals and ethical principles and secondly to make the machine capable to figure out which system best matches the observed. Once this is achieved, machines will efficiently understand the what-why-how of something we are in need of.

But this is not enough. The mere knowledge of our goals will not make the machine adopt them. We may have to persuade them to choose our goals. This is termed as Value-Loading, similar to teaching children moral values. The process becomes harder as the systems grow in intelligence and become smarter. So the time window available to load our goals is quite short. Even if humans are able to resolve the problem of goal adopting, there is one more hurdle to cross. It is to make sure that the AI system retains the already loaded

human goals and does not displace them, as it grows more intelligent. Optimist argues that, if we can get our self-improving AI to become friendly with us by the process of goal-learning and goal-adopting, then we can guarantee retention of our goals, as it will try best to remain friendly to us. But there is always a possibility that a self-improving AI system, as it is growing in intelligence will evolve new goals and finding our goals contradictory to them, will replace them.

Thus the human future very much depends on the problems of goal-alignment, which are yet to be resolved. All of these are undergoing active research today and thus requires a significant resource allocation. Humanity as a whole has to collaborate and participate in this process, as our future heavily depends on it.

Conclusion

There is no doubt that the AI in future will transform and revolutionize our social and institutional structures. But at the same time, it can also be a Pandora's Box. Thus the major challenge before us would be to maintain the status quo and move towards Friendly AI. While aiming for more intelligent and smarter AI systems, there should be also proper efforts of research and legislation to ensure the fair and appropriate use of AI systems. We need to develop local and global values that can be incorporated into the AI system, which will prioritize ethics and transparency, through interdisciplinary and international collaborations. Our future is destined to be shaped by AI, and the question is whether we are prepared to make them Friendly AI.

References

Andersen, Lindsey. *Human Rights in the Age of Artificial Intelligence*. New York: AccessNow, 2018.

Fabry Joseph. *Guideposts to Meaning.* Oakland: Harbinger,1988.

Makridakis, Spyro. *The Forthcoming Artificial Intelligence Revolution: Its Impact on Society and Firms*. Neapolis: Hephaestus, 2017.

Smith, L. Mathew, and Sujaya Neupane. *Artificial Intelligence and Human Development*. Canada: IDRC, 2018.

Tegmark, Max. *Life 3.0: Being Human in the Age of Artificial Intelligence*. New York: Borzoi, 2017.

Thomas Franklin, '5 Reasons I won't replace humans. It will make us superhuman', https://hackernoon.com/5-reasons -ai- wont-replace-humans- it-will-make-us-superhuman, July 8 2018, (accessed 16 February 2020).

Rajesh Vijayakumar is a student of Theology (First Year), at Prabhodhana, Pallottine Institute of Theology and Religion, Mysuru, Karnataka. He belongs to the Prabhu Prakash Province of the Society of Catholic Apostolate (SAC). Email: 19raj88@gmail.com

Article Received: May 24, 2019; Accepted: June 3, 2019: Words: 2730

© by the authors. This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license. (http:// creativecommons.org/ licenses/ by /4.0/).