Artificial Evil Intelligence and Ethical Approaches in Development

Chinmay Khole[1], Ashika Hande[2]

[1,2]Department of Electronics and Telecommunication Engineering, Pune Vidyarthi Griha's College of Engineering and Technology, University of Pune, Pune, India

---

Abstract:- *Recently, Evolution of AI has introduced some serious threats to the mankind. Many scientists also fear that strong AI could potentially harm the human race; moreover there are some ethical boundaries, if not followed while developing AI systems, might lead to the destruction. AI has a great contribution in the fields like technology, military etc. Nevertheless, it will grow in the negative manner causing major risks, said Elon Musk. This paper extensively collects the possible inimical effects that machine having human level intelligence could cause. Although AI is the blooming industry with lots of advantages, its other side is full of bane and is few steps away to reveal its appearance to the mankind.*

1. INTRODUCTION

Over the last few years, the field of artificial intelligence has experienced enormous increase in applicability. It is becoming ubiquitous and with the advancement in technology, efforts are taken to make the AI system to resemble with humans. Stephen Hawking had stated that AI is an existential threat to human race and this statement does emphasize on the AI turning evil. If ethics are not followed in the development of AI systems, what could possibly go wrong? How profound would be the impact of evil AI? The human intelligence, being natural, differs significantly with artificial intelligence, but in near future it is highly possible that this gap will minimize drastically which result into the development of strong AI which could potentially harm us. In coming decades or less, intellectual capacity of robots will match to the human kind and the computational algorithms will be so advanced that they will imitate the processes being carried out in the human brain within fraction of a second. Even humans do not know what is right and what is wrong and are we really capable to teach machines to know the difference between the two? AI was once thought to live wholly in the human imagination, is a real and impending subject, right around the corner. Are we really boasting and are in excessive pride of designing and creating these machines which ultimately might be the reason for downfall of living things on the earth? High time. One of the messages I read recently 'Every AI advance by the good guys is an advance for the bad guys, too.'

As said by Stephen Hawking, AI will rule the humans by the end of 2100. If not supervised, it will surely be an existential threat. Ray Kurzweil, a computer engineer and known futurist predicted that "by 2029 computers would have all of the intellectual and emotional capabilities of humans." furthermore, entrepreneur like Musk, well known futurist and self-taught rocket engineer, hedged his bets against AI, devoting millions of dollars towards multiple research projects worldwide towards those that could create warnings and alarms before a harmful AI was unleashed in the society.

2. IMMORALITY IN DESIGN

All the designing aspects of an Artificially Intelligent system should be based on the ethical standards which are required to be followed by the designers. But this might not be followed always, which would result in creation of evil AI systems. This might happen because of the immorality of the designer or due to the misinterpretation of the system itself. When we think about the design and results of a system, we need to consider the following two aspects. First is, if the machine has done something wrong or the second can be, that's how it was designed. The latest consideration is how the MIT scientist fed the evil and violent content to an AI system named 'Norman'. The major consideration in this case would be what actions we are willing to take to overcome these. What the designers are trying to do since the beginning of the development of AI is to inculcate a system to develop human intelligence on its own, and we need to consider that as the human intelligence has various aspects, there is a high probability that the system designed by a designer having nefarious inclination turns out to be immoral. For an instance, let's consider a system designed by someone having the above aspect. What if he trains the system in a way that it should be able to track down the machines which are lacking security which are connected through the servers or through cloud, implant a malware in it? Who knows that Ransomware was implemented by a human or an Artificially Intelligent agent? It might even be possible by the second case. If the designer has taught the system to behave unethically, how can we expect it to give us positively appreciable results? Moreover, there is no way to predict the magnitude of evilness it would possess. Usually an AI system designed is integrated with an active internet connection. Through this connection, the system is made to be connected to the different servers which are the sources of immense data. This calls for a major question for consideration. What if the system accesses the data over the internet it isn't supposed to be? As we know the security is of utmost important in today's world, how do we make sure that AI system doesn't misemploy each and every aspect of each and everything?

3. SELF DEVELOPMENT

Development of an Artificially Intelligent system is eminently based on the six dimensions.

- Hardware
- Changes in System Goals

- Learning
- Resources Acquisitions
- Code Building
- Recursive Backtracking

What do we call a developing AI? The system which will be able to become better version of itself by developing its own cognitive functions that will be similar to human intelligence. In short, these modifications are a cascading series of improvements in which each stage trying to make the system smarter. Can't we consider the victory of IBM's AI against humans in this context? Comprehensively, a self-developed AI system can be thought of a system which will even be able to surmount the areas of human intelligence which are less likely to be excelled by humans. However, self-improving Artificial Intelligence is a sugar coated term which is not as good as it seems and far more complex than one thinks after reading it.

The most precarious thing that can turn the whole system upside down is considered to be the complex and entangled algorithms. The capacity of algorithm to alter its own instruction set and with simultaneous execution is developing nowadays. The altering of such algorithms by all the AI systems is around the corner. How can we explain the significance of this? As the complexity of algorithm increases gradually, the comprehending capabilities of the system increases proportionally. As the positive comprehension increases so does the negative comprehension. The convenient internet connectivity leads to open access to all the data from all the machines in every server connected. AI systems could make use of this open access to analysis and interpret the data. Furthermore, this data can be used in production of highly effective negative consequences through complex recursive algorithms which could reflect the negative side of the systems. And once it starts possessing human kind of intelligence, it would be no longer for the system to extract and redouble the negative dimension it can reflect. Today AI is in nascent stage and shows the resemblance with an adolescent child. Just like a child, AI should be controlled before it exposes its destructive results. As a remedy for this, access must be restricted at some level.

4. THE RISE OF EMOTIONALLY INTELLIGENT AI

Empathy is considered to be an important aspect of human beings which makes us able to understand others what they want to say instead of what they actually say. It makes us able to understand his/her emotion behind saying that. This quality of humans makes them unique among all other species. When we are trying to teach machines, we obviously want them to possess that kind of empathy as us. What we are trying to do in developing the AI systems is to teach them human intelligence, so they will closely imitate us. There is a lot of research going on in Artificial Emotional Intelligence globally in which we are trying to accomplish the above task. We can see a future where people happily submit their lives to algorithms. And we, as designers and developers, need to make sure that doesn't go in a way that people will be threatened to do so. The problem is, as the machines are still in a developing phase and currently do not have the power of thinking like a human brain, how are we going to make them able to make valid judgments? The systems currently being developed use cognitive reasoning for making themselves react to the situations previously unknown, just like a child trying to discover new things. If the systems succeed in that on their own, it will be a huge success for the developers. However, a major threat in this can be seen as, as they start learning by themselves, how do we know that under worst circumstances, it won't start grasping the negative facet.

5. AI DEVELOPS FEELINGS

When a fully developed AI system isn't able to understand emotions, is it even intelligent? There is indeed no human being who doesn't possess sentiments, so when we are integrating human intelligence with an artificial intelligent system, there is no way it is not going to feature feelings. Emotionally intelligent systems are going to break the barriers between the operator and the system itself up to a great extent. The operator is going to feel like he is talking to other human being when he is really talking to the system. It is important to note here that when we are trying to implant the emotions in robots, it will not affect their mental state but we, as a human will definitely have a sense of harmony with them. It is very easy to handle a novice bot, but it can be of much headache when it comes to learning by its own experiences, specifically when the experience is egregious. This might go beyond the control of humans. The ongoing project on iCub, an open source novice robot is being trained to make itself able to learn by the use of cognitive reasoning. It is being trained to feel and react to the unbeknown situations just like a child. The latest incident of 2017, when two robots- Sophia and Han got into a verbal fight and changes in their tones are really redoubtable. And if we see reactions of common people on this situation, we can feel that people are really frightened by this kind of development in non-humans. When the Facebook AI chat bots started talking to each other in a language unknown to humans, experts thought that the talks were totally irrelevant. They said the systems got caught in a loop. But there can be multiple aspects. Is it probable that both of them got caught in a loop at the exactly same time? Were they communicating some things that both of them were able to understand? And if the answer is yes, the AI systems have developed the ability to communicate surreptitiously and make amendments in their own designs. The Facebook turned the system off, but was it the only way? If the systems are able to make changes in itself on their own, how long is it going to take to switch them on and off by themselves? What can be considered as remedies to circumscribe these developments? The foremost should be the certain regulations on the research and developments of the systems. Though the regulations are of utmost importance, the systems which are designed by ethical standards should be allowed to be produced and publicated.

6. REBELLIOUS AI

The global impact of AI has been emphasized by many experts stating that the innovations of strong AI systems are just about the corner. It is a hypothetical scenario that AI will try to control things and takeover everything humans have. But the pace at which AI technology is advancing, it is highly probable that AI apocalypse will happen in near future .Nonetheless, is has not been clarified when and how is this going to take place. As we know that humans are responsible for extinction of certain species on the earth, and if in case the AI takes control over the humans, it should be able to understand this shouldn't happen. Though we, the humans, possess the best intelligence ever, we are still responsible for extinction of certain species on the earth. If in case AI takes over the human world, possessing human intelligence, how can we expect it not to extinct multiple species including us? How is this going to happen? As we design the AI system, we are under no circumstances going to design it in a way that it is going to eradicate us. Now the only option under which this might happen is, when the system becomes rebellious.

7. DISCUSSION

The AI market is increasing rapidly and safety and risk of AI is also varying from domain to domain. Every individual who is directly or indirectly connected with the AI systems should follow some regulations and guidelines for adoption of these systems. The following directives will help to avoid the threat by strong AI systems.

- Adoption of ethical standards
- Algorithm literacy
- Extensive testing of AI systems to ensure that safety is the first priority.

Considering the discussions of experts in this particular topic, following are the some of the points which are highly discussed globally.

1. Could AI robots be proved to be better than our current government officials?

AI systems in future are going to replace government officials is one perspective and the other one which can be considered even better where the two most competitive fields, AI and automation are not in conflict but in harmony which are utilized together by the government to make the governance more efficient.

2. The rise in AI at current rate, would be the reason for fourth industrial revolution.

As history says, first industrial revolution was because of mechanical production, second was due to science and mass production, the third one was because of digital revolution and the imminent fourth revolution is going to be spurred by AI systems and their data.

3. How much do we really know about ongoing researches in AI?

The translucency about the high end researches in this field.

4. Wealth distribution inequality.

As it is a worldwide process that people who work in a huge company or a factory, get paid proportionally to the amount of work they do. However, in the future, if this human workforce is replaced by machines, most of the owners of such companies will replace humans with AI. Now the consequences of this can be beyond our anticipations.

5. Elimination of AI unjust.

The efficiency of AI systems is far higher than humans, how do we design the systems which are going to be fair and unbiased. There are multiple events in history where such incidents have occurred. If designed right, or designed by those who really want to make prominent social progress, AI can be made a catalyst for a positive change.

6. Singularity, How do we stay in control of complex intelligent system?

It reflects the hypothetical phenomena that might happen resulting in Artificial Super intelligence following the intelligence explosion. There are heated discussions on the post human life after intelligence explosion.

7. How do we define treatment of AI?

AI can be used to make the world a better place. The task in which the places where humans are working with their full efficiency and still not able to comprehend all the available data, can be made much more efficient by using AI crunching it and sharing it with all the other AI systems on the network.

8. CONCLUSION AND FUTURE:

Definitively, AI will be the blooming technology in the future, which will continue to take on 22nd century in more all-embracing way than it is today. However the question is up to what extent do these systems will affect us? There should be some predefined moral codes which should be made the least to be followed. We should be smart enough to design the systems which work on recursive algorithm which if go wrong, can be controlled by human experts. Taking into consideration the highest level of human intelligence, today, it is upon us to decide the level of intelligence the AI systems should possess. In other words, developing controllable AI systems would be a wise choice. Considering the aspects of human nature it is highly possible that the unethical people can end up doing malicious AI systems. Who knows that someone might develop an AI machine to overcome the malicious nature of other AI? But this will be highly complex. However there might be an AI system that will measure the intelligence of other AI systems so as to alert the humans if the intelligence of AI systems is going above threshold and would be a threat. This all hypothetical idea is based on lot of analysis. A lot is coming up in the near future of AI and if it goes ethically correct, it will be a great success to us.

REFERENCES

[1] Utku Kose and Pandian Vasant,"Fading Intelligence Theory: A theory on Keeping Artificial Intelligence Safety for the Future." 2017 International Artificial Intelligence and Data Processing Symposium, Nov 2017, Doi:10.1109/IDAP.2027.8090235

[2] Luciano Floridi and J.W. Sanders, "Artificial evil and the foundation of computer ethics." Ethics and Information Technology, vol. 3, pp. 55-66, doi:10.1023/A:1011440125207

[3] Narendra Kumar, Nidhi Kharkwal, Rashid Kohli and Shakeeluddin Choudhary, "Ethical Aspects and Future of Artificial Intelligence." 2016 International Conference on Innovation and Challenges in Cyber Security. Aug 2016, Doi:10.1109/ICICCS.2016.7542339

[4] Utku Kose, Ibrahim arda Cankaya and Tuncay Yigit, "Ethics and Safety in future of Artificial Intelligence; Remarkable Issues" International Journal Of Engineering Science And Application, vol. 2, June 2018