Safe AI Systems

Ananta Bhatt
Fanshawe College
13 Hyatt Avenue, Ontario, Canada

ABSTRACT
This Report highlights the development in AI both the positive and negative aspects introduced with this development. Ultimately pointing out the importance of development of safe AI systems.

As you know, it is inevitable to cease the advancement in AI because of the quest for developing systems that learns from experience and comply with the human capabilities. The leading ground-breaking impacts offers both positive and negative aspirations and are addressed as part of this research. To meet with the evil effects of the AI systems, is now the need of the hour to direct the focus on creating safe AI systems. The safe system must satisfy with 2 majors- applying the security measures and analysing the system. Moreover, system must fulfil the risk and safety values- intelligence, goals and safety in order to minimise the destruction that may occur due AI technology. However, further research is needed on the actual implementation of the safe AI systems in order to proceed with this research.

General Terms
Research Article on Artificial Intelligence and importance of safe system thereby defining safety and risk values.

Keywords
Safe system, risk and safety values.

1. INTRODUCTION
Artificial Intelligence (AI) field is growing at an unpredictable rate. Even with optimistic approach towards this development, the adverse approach still follows and cannot be overlooked. Through this technology, many industries can benefit ranging from the medical field to eliminating wars. AI plays a key role in defining systems that benefit in achieving modular activities.

This Report outlines the development pros and cons made so far with this technology and illustrates about the past, present and the future growth in Artificial Intelligence. With the early growth that initiated in this field involved neural networking concept around 1950s, further stirred excitement for machine learning directives increasing the possibility of creating 'human-like-thinking' feature in machines. Moreover, demonstrating the present approach followed in AI comprises of deep learning that lead to boom in this field. With such never ending progress through AI, significantly dominant approach can be to consider developing safe AI systems that will help minimise the malicious use of AI technology. The importance of creating safe AI system is the major focus of this research.

The findings of this research are as follows:

- AI helps achieve greater end of accuracy- Systems such as Google Photos, Alexa assistant provide accurate data as per required by the user. These systems are based on the deep learning that is a current followed approach.
- AI technology can be put to use to overcome the limitations of the human. For future scope AI technology can be applied to explore the deep oceans. This technology can also be used to find the fuels.
- If the control of AI goes in wrong malicious hands, this may lead to may disastrous events. It can be programmed to create dangerous and devastating effects.
- Apart from all the positive aspirations, there is still a fear that AI technology will help robots takeover the human world. Though humans are the master of the machines, it doesn't take time to turn the tables with the self-learning ability that the robots possess.

Research revolves around the following recommendations:

- The Risk that this technology possesses, enable us to focus our attention on bringing safe AI systems. The standards created for marking the danger in machines must be enlarged to increase the scope of verifiability and operational lifetime.
- Defining the risk and safety values clearly in order to distinguish and identify between the mishap activities. Following a triangle approach can help minimise the disastrous events that can happen- Intelligence, Safety, Goals.

2. RELATED WORK
AI involves collaborative actions and enhancements leading to the goals as requested by the user. Machines increasing the capabilities to match the natural human intelligence, creates new factor of good and evil concepts for the proceedings. With the initial development in AI progressed its through the 1950s and now holds a significant recognition. This Report helps understand the development achieved with AI from the past till the predicted trench in the AI field. Moreover, helps distinguish the factors that increases its scope and usability. However, AI development also cringes towards damaging malicious factors like robots taking over humans. But its not possible to neglect the positive advancement through AI, which leads to creating safe and secure AI systems.

3. METHODOLOGY
All the Information collected for this report are obtained primarily from 4 sources-, 1. Article by Thomas G. Dietterich, 2. Article by Theresa Cramer 3. Article by PC Quest and 4. Article by Ross Bentley. All the sources are listed in the below reference list. Moreover, there are many reference websites which will be used for this report.

4. SCOPE
The ultimate reason for development of AI technology is helping to establish goals that are difficult to achieve by humans such as full time labour work. However, the underlying factor is the impulsive direction that can lead by this technology. The positives and negatives addressed in this report will help provide a clear understanding of its

objectives. Analysing these factors will helps us be prepare for the wrong outcomes through this creation.

This Report is a structured document directed to help minimise the futuristic catastrophic measures. Through the development of a safe bounded AI system, it will help minimise the evil implications that may occur. The importance of such safe AI system will help appreciate this technology and led to much wider extend of development. Greater aspects of applications can be covered with the safe AI systems such as fraud detection machines.

5. FINDINGS

5.1 Positive Aspects

5.1.1 HIGHER END OF ACCURACY

AI systems are trained in order to recognize data and predict the data. This strategy is used in place for processes such as voice recognition, image recognition. From the first program created under AI, the checkers game version that self- learned to play. The predictability of the system to give the desired output helped increase the efficiency and scope of the system.

Current innovative system created to predict the lifeline of a person show great aspect of accuracy. About 69% of the predicted cases showed correct mortality rate, the rate at which the system can predict the lifeline of people. AI system can predict it at the rate of 69% within 5 years of span.

Our communication method has drastically moved by AI and will continue to progress in the near future. Moreover, AI combined with virtual reality (VR) brings new level of efficiency and productivity in our communication methods.

Further, there are many examples of systems that have higher accuracy rate such as Google Maps, Google Photos, Alexa Assistant, systems that detect cancer where the system was provided with a slide of lymph node was tested to detect the cancer. This automated system showed correct data about 92% of time. (Center, healthcare-in-europe, 2016).

Another example is of a computer software that defeated the chess champion Garry Kasparov. The computer software is called Deep Junior created by two Israeli programmers Shay Bushinsky and Amir Ban. The precision of the computer to respond to the chess move shows the accuracy that the system can achieve. (Bentley, R., 2003, March 20).

5.1.2 Overcoming Human Limitations

The main aspect of building an AI system is to overcome and assist humans. Development has been started in this sector to help create a better future.

All the energy and direction has been turned in order to create a plausible future.

AI System can be programmed to perform tasks such as"

- Dig and search for fuel
- Search for mining purpose
- Explore the ocean to make new discoveries
- Labour Work
- Space exploration

These factors are taken into consideration and currently systems are being incorporated to enable AI exploration. Mars 2020 mission will incorporate AI technology to venture the space exploration.

Various measures such as digital payments offer personalised experience to consumers leverage the development allowing consumers to perform various processes. For example, cheque photo to transfer or withdraw money.

Government is also taking various measures and steps to encourage the use of AI technology such as self-driving cars, robot surgery. Moreover, companies like Google, IBM are creating robust systems encouraging our use of AI.

(Dietterich, T. G. 2017, Fall).

5.2 NEGATIVE ASPECTS

5.2.1 AI IN WRONG HANDS

The fate of our race is entirely dependent on scientists adding new features in order to make our life easier. However, every coin has two sides depending on the side you chose will analyse the growth and progress. There are various organisations and leaders that have come up to spread awareness about the evil fundamentals of this approach. The Future of Life, an Institute also warned that the creation of this technology for areas such as autonomous weaponry. This system will not only help in decreasing the potential number of soldiers that died in the battle but also can cause mass destruction.

5.2.2 ROBOTS TAKEOVER

AI is simple ground-breaking tool developed to benefit our race. However, with the self-learning capabilities that the robots/machines possess makes it inevitable to deny that robots will someday takeover humans.

Structurally, the neural networking system is designed as such that it learns from the experience gained and tries to improvise on it. Live example is where the humanlike robot Sophie jokes about taking over the world. (SULLEYMAN, 2017).

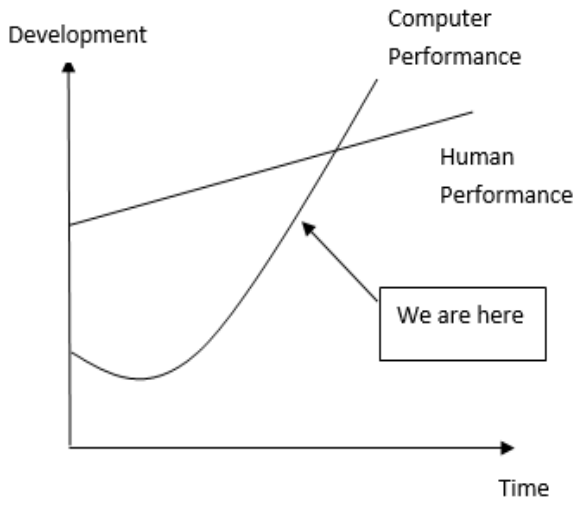

Fig 2: Robot Intelligence taking over human Intelligence

(Urban, 2015)

In the above Fig 2 shows, the computer/ machine intelligence takes over the human intelligence. It is not far when the robots

will actually takeover our world. Artificial Intelligence allows machines to learn and grow in the environment.

6. RECOMMENDATION

6.1 Safe Systems

The core liability lies at creating safe systems. Methods can be designed to continuously monitor the actions of the machines. It will help analyse the cause of the action and help prevent it.

Safer systems will reduce the probability of negative behavior.

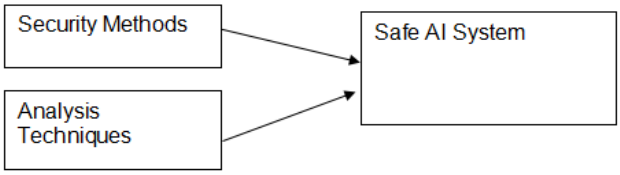

Fig 3: Safe AI System

As shown in the above Figure, Fig 3 two governing factors for creating a safe AI system are:

- Applying the Security Methods
- Analysing the system

The pace at which the development in AI is taking place is undeniable and requires attention to developing strong and safe AI systems without rendering the natural language processing development. (Cramer, T. ,2018, Winter).

6.2 Defining Risk And Safety Values

As easy as it sounds creating safety values are measurements developed to ensure that the system is safe to use. Redefining and creating an appropriate set of values will help establish a better safe system.

It's not just about the technicality of the system but keeping in mind that the system exhibits the flexibility to maintain its boundaries. It can be developing the system with the safety measures at hand rather than developing the system and then setting the safety standards.

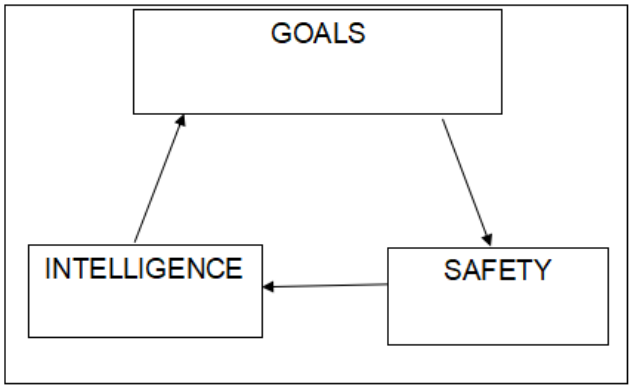

Fig 4: Safety Values

Maintaining the stability between these 3 factors mentioned above in the figure, Fig 4 will help minimise the risk factor associated with the AI technology. Safety governs with checking the goals and matching with the intelligence of the system.

Safety measures that is defined in the above 6.1 section and the subdividing algorithmic knowledge of the machines will be the establishment of the risk values.

7. CONCLUSION AND RECOMMENDATION

Taking all the concerning factors into account, the positives and the negatives the AI technology is growing at a very fast pace. Even with concerning negative effects that may occur, it is inevitable to slow down its progress. With the development of AI that initiated from 1950s has expanded its reach to all the sectors and is still widening for sectors such as medical, space exploration, mining, searching for fuel.

There are subsiding means through which the disastrous effects can be avoided. Safe systems can be created in order to avoid destructive measures. Safe systems however require two major factors to be considered- creating the security measures and analysing the methods. This is will ensure and track any malicious activities that the system can take place.

Second means is to redefine and create the risk and safety values. This is bounded by 3 major values- Intelligence, goals and the safety. Keeping these values in mind will help create a safe to use AI system.

However, further research is required on the actual implementation of the safe AI systems in order to proceed with this research.

8. ACKNOWLEDGMENTS

Special thanks to Mrs. Krista Carson, my lecturer for believing in me and guiding me throughout the journey of this report.

9. REFERENCES

[1] Urban, T. (2015, January 27). WaitButBuy. Retrieved from Ai Revolution

[2] Center, B. I. (2016, July 11). healthcare-in-europe. Retrieved from Breast Cancer

[3] Daws, Ryan (2018, June 18). india-report-ai-uk-japan-germany from artificialintelligence-news

[4] SULLEYMAN, A. (2017, July 13). Independent. Retrieved from Indy/Tech.

[5] Adams, R. L.,(2018, June 10). 10 Powerful Examples Of Artificial Intelligence In Use Today from Forbes.

[6] McClelland, Calum (2017, December 4). The Difference Between Artificial Intelligence, Machine Learning, and Deep Learning from medium corporation

[7] Robitzski, D. (2017, June 2). Inverse. Retrieved from Health.

[8] Artificial Intelligence Transforming the Digital Payments Landscape. (2018, February 21). PC Quest.

[9] Darlington, Keith (2018, August 13). AI Systems Dealing with Human Emotions from openmind.

[10] Cramer, T. (2018, Winter). The State of Artificial Intelligence. EContent, 41(1), 26+.

[11] Wilde, Tom (2018, October 26). Enterprise AI and the Paradox of Accuracy from datanami

[12] Bentley, R. (2003, March 20). Man versus machine. Computer Weekly, 24.

[13] Forbes Technology Council (2018, April 20). Seven Artificial Intelligence Advances Expected This Year from the Forbes

[14] O'Malley, James (2018, January, 10). The 10 most important breakthroughs in Artificial Intelligence from techradar.

[15] Roffel, S., Evans, Ian (2018,July 9). The greatest advances in AI: the experts' view from Elsevier.

[16] Bowman, M., Debray, S. K., and Peterson, L. L. 1993. Reasoning about naming systems. Dietterich, T. G. (2017, Fall). AAAI Presidential Address: Steps Toward Robust Artificial Intelligence. AI Magazine.

[17] Nicholson, Tom (2018, August 20). Google's New AI System Knows When You're Going To Die from Esquire