

Artificial Intelligence: Risk Assessment and Considerations for the Future

Balu N. Ilag
AI and UC & Collaboration Expert
Silicon Valley, CA

Yogesh Athave
Sr. Software Engineer
Silicon Valley, CA

ABSTRACT

The rapid growth of artificial intelligence is fast impacting other sectors like governance, security, economics and health. Artificial intelligence and associated machines have the potential to change the world as people know it through improvements the human condition and better decision making. With the growth of artificial intelligence, however, there is an increasing need to assess risk factors and develop mitigating plans. New technologies replacing human intervention pose a threat to existing weaknesses in areas social structures, economic systems and general human defenses. This paper looks at the transformative nature of artificial intelligence and investigates the need for a more proactive risk assessment framework in the development of technologies in this area.

General Terms

Artificial Intelligence, Future consideration, Risk Assessment, Methodology, Sampling, analysis.

Keywords

AI, Artificial Intelligence

1. INTRODUCTION

Artificial intelligence as a discipline in the academic world has existed for quite a while and has been evolving. Early scientists in the field theorized that computers and machines could be programmed to undertake more humanlike and intelligent activities if they could solve fundamental mathematical problems ("Artificial Intelligence: Machines, man and intelligence," 2018). Initial developments in the area included machine learning and heuristic search with broader targets like robotics, natural language processing and computer vision (Bartoš, 2015). A pure scientific interest in understanding intelligence and creating it within computers and other machines has significantly motivated research and development in the field.

Over time, the growing sophistication of computer software and hardware and the use of sophisticated artificial intelligence calculations being brought into a range of practical utility by computers, the field has gained prominence with a few commercial appliances (Harris, 2010). Computer vision, for instance, has dramatically advanced medical imaging with automated diagnostics (Bartoš, 2015). Domestic applications, robotics with industrial capabilities as well as predictive business analytics are some of the other examples of advancements in artificial intelligence. New applications in the field are now focusing on cutting costs, increasing efficiency, analyzing goals, gathering of data and classification (Huang, Nijholt, Pantic, & Pentland, 2007). This research focuses on the usefulness of artificial intelligence and analyses the risk factors around its continued deployment in cognitive processes.

2. LITERATURE REVIEW

The goal of keeping the impact of artificial intelligence positive to society is rapidly driving research in technical areas such as control, security, validity and verification (Johnsen, 2017). Studies in these areas are critical given that, in the age of hacking and cyber insecurity, gadgets with artificial intelligence are fast learning how to control things like power grids, trading systems, pacemakers, aeroplanes, and cars (Waser, 2010). There is also a need for preventing a jeopardous arms race in autonomous weapons ("Artificial Intelligence: Machines, man and intelligence," 2018). The pressing question that requires answers by various scientists pursuing advancements in this area, however, is what happens in the age that artificial intelligence becomes better than a human being at cognitive tasks (Bartoš, 2015). The creation of systems that operate on artificial intelligence is itself a cognitive task (Harris, 2010). There is the adverse potential of such systems undergoing self-improvement leading to an explosion of intelligence that will outpace human intelligence (Kurki & Pietrzykowski, 2017).

Technological revolutions, however, might be instrumental in the eradication of poverty, disease and war (Kurki & Pietrzykowski, 2017). Some scientists have postulated that the invention of artificially intelligent systems might be the most significant human achievement in history (Huang, Nijholt, Pantic, & Pentland, 2007). Many proponents of artificial intelligence claim that artificial intelligence systems are guaranteed to be beneficial to the human form (Johnsen, 2017). Other experts have also observed that there is need to tame the growth of artificial intelligence since there is a risk of superintelligence that will replace or colonize human intellect (Kurki & Pietrzykowski, 2017). The possibility of artificially intelligent systems that can cause harm to human civilization either intentionally or unintentionally is alive (Waser, 2010). There is a growing need for research in the area of artificial intelligence to maximize potential benefits and avoid or mitigate against potentially adverse consequences to human life (Huang, Nijholt, Pantic, & Pentland, 2007).

3. METHODOLOGY

3.1 Data Collection and Sampling

Quantitative research was done at a Silicon Valley, California of a prominent technology professional and computer science scholar to determine the potential of artificial intelligence and analyze the perception of risk around new technologies in the area. A sample size of 10 people was determined using (Krejcie & Morgan, 1970) model of defining a sample size. The respondents were mainly technology professional and scholars of computer science and related fields. A convenience sampling technique was used to obtain data from respondents in a cross-sectional research strategy with questionnaire forms. Respondents were required to fill in the questionnaire forms independently and objectively. The

respondents were also asked to complete the forms within ten minutes and submit them with limited biodata to avoid any privacy concerns by some of them.

3.2 Questionnaire

The research used six questions to measure both the potential of artificial intelligence making the world a better place and the adversities that come with some of the technologies. The questions asked inquired of the respondents' perceptions of the impact of artificial intelligence so far, the perceived gains in the continued development in technologies in this area, and the expected future outcomes whether positive or negative. The questions were close-ended clear and sequential to allow for clarity and directness in the collection of responses. Respondents were requested to select their answers on a scale of 0 to 2, 0 being uncertainty, 1 being agreement, and 2 being disagreement.

Table 1: The questions used for the survey are as detailed below;

Construct	Indicators
Artificial Intelligence	Q 1 – Do you think that artificial intelligence has positively impacted the world in its few years of existence?
	Q 2 – Do you think that artificial intelligence has the potential to positively impact the world in future?
	Q 3 – Do you think that current and developing research in the area of artificial intelligence is geared towards positive technologies?
	Q 4 – Do you think that artificial intelligence has the potential to positively complement human cognition?
	Q 5 – Do you think that artificial intelligence has the potential of adversely replacing human intellect?
	Q 6 – Do you think that emerging technologies in artificial intelligence have the potential to undermine human intellect and cause destruction to human civilization through superintelligence?

3.3 Results and Analysis

There was a 100% response rate with 10 out of 10 questionnaires being submitted at the end of the exercise. All questionnaires were duly completed in a legible font. The information collected was analyzed through a detailed review of all the responses.

Table 2: Characteristics and frequency of the Respondents

Characteristic of Respondents	Frequency of characteristic	Expressed as a percentage
Technology Professional Participants	6	60%
Computer Science Scholar Participants	4	40%
Male Participants	5	50%

Female Participants	5	50%
Below 30 Years of Age	4	40%
Above 30 Years of Age	6	60%
Professionals or aspiring professionals in Computer Science and Related Fields	10	100%
Have interacted with Artificial Intelligence Before	9	90%
Have not interacted with Artificial Intelligence Before	1	10%

There was a 6:4 ratio between professional and scholar participants in the survey. There was also gender balance as there were five male participants and five female participants in the research. 40% of the participants were younger than 30 years while 60% of the respondents were older than 30 years. With all of the participants being professionals or aspiring professionals in the field of computer science, 90% indicated that they had personally interacted with artificial intelligence while 10% stated that they had not.

Table 3: Analysis of the results

Construct	Indicators	0	1	2
Artificial Intelligence	Q 1 – Do you think that artificial intelligence has positively impacted the world in its few years of existence?	10 %	70 %	20 %
	Q 2 – Do you think that artificial intelligence has the potential to positively impact the world in future?	20 %	60 %	20 %
	Q 3 – Do you think that current and developing research in the area of artificial intelligence is geared towards positive technologies?	10 %	60 %	30 %
	Q 4 – Do you think that artificial intelligence has the potential to positively complement human cognition?	20 %	70 %	10 %
	Q 5 – Do you think that artificial intelligence has the potential of adversely replacing human intellect?	40 %	50 %	10 %
	Q 6 – Do you think that emerging technologies in artificial intelligence have the potential to undermine human intellect and cause destruction to human civilization through superintelligence?	40 %	50 %	10 %

A majority of the respondents thought that artificial intelligence had positively impacted society in its few years of existence at 70%. 20% of the respondents disagreed with that position while 10% were not sure of the field's contribution. On Q 2, 60% of the respondents thought that artificial intelligence had the potential to impact the world in future positively, 20% disagreed while 20% were uncertain. 60% of the respondents felt that current and developing research in the area was headed in the right direction, 30% did not believe that research in the field was geared towards positive technologies while 10% were not sure. A good portion of respondents thought that artificial intelligence had the potential to complement human cognition positively, at 70%. 20% of the sample size were uncertain of artificial intelligence's potential to complete or complement human action that requires reason and thought processes while 10% strongly disagreed. 50% of the respondents thought that artificial intelligence had the potential of adversely replacing human intellect, undermining human intelligence and causing destruction to human civilization through superintelligence. 40% were uncertain of the potential role of artificial intelligence around these factors while 10% disagreed.

4. DISCUSSION

From the above findings, the need for regulation and caution in the development of artificial intelligence cannot be overemphasized. Most scholars in the field argue that super intelligent machines may soon become similar to human beings with human-like abilities and emotions like hate and love (Huang, Nijholt, Pantic, & Pentland, 2007). There is also the possibility that human beings will not be able to control how malevolent or benevolent such technologies become ("Artificial Intelligence: Machines, man and intelligence," 2018). Whether future technologies are programmed to carry out devastating or beneficial activities, a risk assessment framework is of utmost importance given that new machines and devices are being produced at a rapid rate.

One of the risk factors in artificial intelligence include autonomous weapons which are programmed to kill ("Artificial Intelligence: Machines, man and intelligence," 2018). If such weapons land in the hands of wrong people or governments, they could be used to occasion devastating mass killings. An artificial intelligence arms race could very possibly lead to an artificial intelligence war that would be difficult to control. The risk of artificial intelligence warfare is narrow with existing technology but is slowly growing with sophistication in the field.

Another risk factor is the development of beneficial artificial intelligence machines which in turn develop destructive means of achieving their goals (Waser, 2010). Such a scenario is possible where human beings fail to align their artificial intelligence goals to those of human civilization. There is a possibility of artificially intelligent machines doing what human beings ask of them instead of what they want to be done. For instance, a human being might ask a self-driving car to take them to the airport, but the vehicle can cause destruction to other vehicles along the way. Another example would be an artificially intelligent machine causing environmental damage in its programmed work of geoenvironmental engineering. All these issues point to an urgent need into the assessment of risk factors in artificial intelligence.

5. CONCLUSION

In conclusion, it is possible to observe from the research findings and the above discussion that the problems in artificial intelligence are competence and risk. Scientists in this area should urgently incorporate risk mitigation plans in their developing work and align artificial intelligence goals to human activity. It is also important not to place human beings at an inferior position to artificial intelligence as this might occasion an end to humanity.

Legal systems and other regulatory frameworks should consider drawing guidelines within which artificial intelligence can operate as new technologies spring up in various corners of the world. The possibility of artificially intelligent machines and devices controlling a world that is currently managed by human beings is real and must be given attention by scientists and other stakeholders alike. The future of artificial intelligence does not, however, lie in the suppression of the field but rather in the promotion of safety measures alongside research

5.1 Considerations for Future Research

Future research in this area should consider tapping into the fears and concerns of people who are not in the computer science as all respondents, in this case, were either professionals or aspiring professionals in the field. Such studies should also employ the use of open-ended questions and other research methods that enhance the qualitative aspects of research.

6. REFERENCES

- [1] Artificial Intelligence: Machines, man and intelligence. (2018). March/April 2018. doi:10.1287/lytx.2018.02.03
- [2] Bartoš, V. (2015). Biological and Artificial Machines. Topics in Intelligent Engineering and Informatics, 201-210. doi:10.1007/978-3-319-09668-1_15
- [3] Harris, M. C. (2010). Artificial Intelligence. Singapore: Marshall Cavendish.
- [4] Huang, T. S., Nijholt, A., Pantic, M., & Pentland, A. (2007). Artificial Intelligence for Human Computing: ICMI 2006 and IJCAI 2007 International Workshops, Banff, Canada, November 3, 2006 Hyderabad, India, January 6, 2007 Revised Selected Papers. Basingstoke, England: Springer.
- [5] Johnsen, M. (2017). The Future of Artificial Intelligence in Digital Marketing: The next big technological break. Maria Johnsen.
- [6] Krejcie, R. V., & Morgan, D. W. (1970). Determining Sample Size for Research Activities. Educational and Psychological Measurement, 30(3), 607-610. doi:10.1177/001316447003000308
- [7] Kurki, V. A., & Pietrzykowski, T. (2017). Legal Personhood: Animals, Artificial Intelligence and the Unborn. Basingstoke, England: Springer.
- [8] Waser, M. R. (2010). Designing a Safe Motivational System for Intelligent Machines. Proceedings of the 3d Conference on Artificial General Intelligence (AGI-10). doi:10.2991/agi.2010.21