

Analysis on Artificial Intelligence Security and Its Countermeasures

Man Qi

Wuhan University of Science and Technology
Wuhan, China

Wenzhang Tang

Wuhan University of Science and Technology
Wuhan, China

Abstract—The rapid development of artificial intelligence technology has aroused widespread concern about its security. The artificial intelligence may surpass human intelligence, and it is very necessary to study the security of artificial intelligence. Viewing from the internal approach to solve the security problem of artificial intelligence, we have ethical design at least, limited scope of application, limited degree of autonomy and intelligence, and so on. From the perspective of external approach, we should emphasize the social responsibility of scientists and international cooperation. It should guide people to accept artificial intelligence, as well as the way of safety evaluation and management of artificial intelligence. Only by taking practical and effective measures to ensure the safety of artificial intelligence can it bring happiness to human beings instead of harm.

Keywords—artificial intelligence; security; solving ideas

I. INTRODUCTION

From March 9 to March 15 in 2016, Google artificial intelligence "Alpha Go" and the Korean professional Lee Se-dol had a "man-machine war". Lee's defeat made people again astonished by the power of artificial intelligence. Therefore, people are worried about the safety of artificial intelligence.

II. POSSIBLE HARM AND CONSEQUENCES OF ARTIFICIAL INTELLIGENCE

Undoubtedly, the level of human intelligence as a whole as well as many sci-fi films and literary works far exceeds that of other biological intelligence. Due to this, human beings become the ruler of the earth. Therefore, it is natural to conclude that if artificial intelligence transcends human intelligence, artificial intelligence may no longer follow the instructions of human beings. And it will fight for the domination with human beings. Many scholars have demonstrated the possibility and feasibility of the development and progress of artificial intelligence from the perspective of philosophy and futurology. From the perspective of the history of science and technology, many of the impossible prophecies of science and technology have become the reality later. For example, some famous scientists and engineers once thought that it was impossible for an airplane to fly into the sky. They believed that it was a fantasy to let a mechanism heavier than the air fly. However, the facts have proved that they are wrong. Therefore, when

scientists make negative predictions about certain science and technology, they should be more cautious. Of course, it is almost impossible to predict the development and application of particular science and technology accurately. However, from the development and progress of related technology and the degree of attention to artificial intelligence in the world, artificial intelligence is likely to develop rapidly in a period of time in the future. The rapid development of computing speed and storage capacity is obvious to all. In recent years, the development of deep learning, cloud computing, super-computing and big data technology will also promote the progress of artificial intelligence. Obviously, the success of Google Alpha Go has led to the widespread belief that artificial intelligence is bound to have rapid development. Reportedly, after the "man-machine war" between Alfa Go and Lee Se-dol, the South Korean government announced a special plan with a total investment of about \$840 million is used to speed up the development of the artificial intelligence industry. Many countries in the world naturally do not want to lag behind. And they all hope to seize the high ground of research and development of artificial intelligence. It increases the research funds and personnel investment. From this point of view, it is almost impossible to stop the development of artificial intelligence.

In view of some twists and turns in the development history of artificial intelligence and the understanding of the current situation, most artificial intelligence experts hold a cautious or even negative attitude about whether artificial intelligence can surpass human intelligence (in a short time). Japanese artificial intelligence expert Matsumoto believes: "Artificial intelligence conquers human beings. Artificial intelligence creates artificial intelligence — this possibility does not exist at this stage. And it is just imagination." In his view, we do not need to worry about artificial intelligence conquering human beings. There are also some scientists concerned about the future of artificial intelligence. For example, at the end of 2014, the BBC reported that famous theoretical physicist Hawking said: "All-round development of artificial intelligence may lead to the extinction of human beings." Hawking worries that artificial intelligence may catch up, or even overpass humans one day. Famous people such as Bill Gates have similar concerns.

However, artificial intelligence as a whole does not exceed human intelligence. The uncontrolled one-sided artificial intelligence may also bring harm to human beings.

Just as Nick Bostrom did in his paper clip thought experiment: if an artificial intelligence system is set to maximize the yield of paper clip. The artificial intelligence system will have an unsatisfied appetite for material and ability. And it would go on the path of turning the earth first and then most of the observable universe into a paper clip. The technology of internet and internet of things make the security of artificial intelligence more complicated. On the one hand, the network resources make the development and evolution of artificial intelligence and the available resources is boundless. On the other hand, the technology of internet and internet of things makes hackers, viruses and other human factors are a great threat to artificial intelligence products. Even if artificial intelligence is not as intelligent as human intelligence, network technology is likely to turn our dependence on artificial intelligence into a disaster. For example, if hackers control child-care robots, help-the-elderly robots or other intelligent systems in people's homes, the consequences will be unthinkable.

Viewing from the society theory of risk which is very popular in recent decades, it is necessary to study the security problem of artificial intelligence. As one of the representatives of the society theory of risk, Ulrich Beck believes: "The concept of risk shows that people have created a kind of civilization. Their decisions will have unpredictable consequences. And then, it would control the uncontrollable things. Through intentional preventive action and the corresponding institutionalized measures, it would overcome the side effects." And different schools of risk society research have different views on the definition and prevention of risk and other basic issues. However, they believe that the understanding of high-tech will lead to high risks. Therefore, the theoretical results of the study of risk society are of great significance to the study on the security problems of artificial intelligence.

Generally speaking, in view of the possibility that artificial intelligence surpasses human intelligence, and the possibility and the severity of the harm caused by artificial intelligence, and the inherent uncertainty of science and technology, these factors are sufficient to constitute the necessity that we study the safety problem of artificial intelligence. In fact, the study of humanities and social sciences should go beyond the development of natural science. We can't wait to study the ethical and social problems of the clone until the birth of the clone. We must be well prepared before the emergence of security problems of artificial intelligence.

III. COUNTERMEASURES TO THE SECURITY PROBLEM OF ARTIFICIAL INTELLIGENCE

Generally speaking, the negative effects of technology (security problems) are mainly caused by technology or by human factors. And the corresponding solutions can be drawn from these two aspects. Therefore, we can roughly divide the security problem of artificial intelligence into internal and external ways. From an internal perspective, we have at least a few solutions.

A. Ethical Design

Ethical design of artificial intelligence products is one of the basic ways to solve its security problems. In recent ten years, more and more western scholars have paid attention to the ethical problem of robot. The main goal of ethics research of robot is to make robot have certain moral judgment and behavior ability in the process of interaction with human. And the behavior of robot would accord with the moral standard that people preset. Theoretically, the robot that makes moral decision according to the moral principles preset by human beings can become the "moral model". To a certain extent, it can avoid the improper use and malicious use.

B. Limited Scope of Application

At present, the advantage of artificial intelligence mainly lies in clear rules of the field, such as a variety of chess competition. The success of Alpha Go shows that the artificial intelligence has the ability of super-computing and in-depth learning is fully capable of dealing with the great amount of information. In this respect, human intelligence has lagged far behind artificial intelligence. However, although Alpha Go has strong learning ability, it can only be used to learn Go. It is easy for humans to freely convert their learning experiences in different fields. (Lv) The ability of artificial intelligence to learn and transform is considered to be an important defect. It is an important guarantee for the security of artificial intelligence. If humans are unable to ensure good control of artificial intelligence, it is wise to control the function of artificial intelligence in a single case. In other words, the study and application ability of the professional artificial intelligence developed for chess competition, expert system, unmanned driving and so on is limited in the scope of its own field. On the one hand, it can ensure that the artificial intelligence can reach a very high level in their respective scope. On the other hand, it can also avoid that the artificial intelligence is too strong to pose a threat to human beings.

From the perspective of the public, people generally hope to use artificial intelligence as an important tool. Especially, they want to make up for the lack of human intelligence in some aspects. And they do not want to develop artificial intelligence as a whole close to more advanced than human intelligence. In the whole process of research and development of artificial intelligence, this positioning should be clear. And limited the scope of its application may be one of the fundamental ways to carry out this positioning.

C. To Limit the Degree of Autonomy and Intelligence of Artificial Intelligence, and Establish the Safety Standards and Norms of Artificial Intelligence

The root of security problem of artificial intelligence is not whether it can really surpass human, but whether it is a safe and reliable tool, whether human has full control over it. Just like high-speed rail, planes and other vehicles, although they are far faster than humans, people have absolute control. People believe that they are safe. In order to achieve its control of the goal, the degree of autonomy of artificial intelligence should be limited. Although artificial

intelligence develops rapidly, human intelligence also has its own advantages. The current cognitive ability of artificial intelligence is far less than that of human intelligence.

We can give full play to the advantages of artificial intelligence in information storage, processing, and so on. We could make it be the advanced human brain in some important events. However, the human could control the final decision. For example, when we apply artificial intelligence to the military field, we can use artificial intelligence to assess the degree of danger. And we can take measures. When it comes to launch a war and make important decisions, the human need to master the control. As Jeff Hawkins said, "We should be careful about smart machines. And we can't rely on them too much."

Viewing from the source, the security problem of artificial intelligence is caused by artificial intelligence technology. It can be seen that scientific and technological research has no forbidden zone. The maturity of technology is the key factor to solve the security problem. There is uncertainty in any technology. And the problems produced by technology cannot be satisfactorily solved by technology. Therefore, it is necessary to give full play to the important role of external strategy. And then, it could solve the security problem of artificial intelligence.

D. The Social Responsibility of Scientists and International Cooperation

Like other high-tech technologies, artificial intelligence is specialized scientific knowledge. Artificial intelligence scientists and engineers are the researchers of artificial intelligence technology. They are the main body of solving the security problem. They should strengthen the professional responsibility of artificial intelligence experts from two aspects: "negative responsibility" and "positive responsibility". Active responsibility emphasizes what artificial intelligence experts should do and how to ensure the safety of artificial intelligence through technical means. Negative responsibilities focus on the people who should bear the responsibility when artificial intelligence has negative impact or serious consequences. From the perspective of positive responsibility, experts cannot pursue economic benefits, or blindly cater to customer needs in the process of research and development.

From the perspective of negative responsibility, when the artificial intelligence system is wrong, the experts should assume the corresponding responsibility rather than blame the responsibility on the uncertainty and complexity of the artificial intelligence. In a sense, the primary factor to solve the security problem of artificial intelligence doesn't lie in artificial intelligence technology. And it lies in the responsibility of artificial intelligence experts.

At the same time, security problem of artificial intelligence is not only a regional or organizational problem. The governments and international organizations should be the organizations to coordinate the security problem of artificial intelligence. At present, countries all over the world are competing to increase the investment on artificial intelligence. It should allocate special funds for the study of

the security of artificial intelligence. The government funds and human resources investment is the key to solve the problem. In addition, international cooperation will play an important role in solving the security problems of artificial intelligence. As mentioned above, the problems such as the development and application limits of artificial intelligence, safety standards and norms can only be meaningful in specific systems. The international organizations can achieve such a goal. From the perspective of ethical responsibility, the responsibilities of the community of scientists, the governments and international organizations should be clearly defined to avoid the phenomenon of so-called "organized irresponsibility". In recent years, the popular research and practice exploration of "global governance" theory provides a good platform and foundation for international cooperation on AI (artificial intelligence) security issues in the world.

E. Public Acceptance and Adjustment of Ideas

In terms of objectivity, artificial intelligence security mainly refers to the security of artificial intelligence technology. And the subjective aspect comes from people's feelings (especially the public) on the intuition, subjective feeling or experience of artificial intelligence security. With the rapid development of information technology, people's awareness of the risks of science and technology and dissemination of speed has been greatly improved. Contrary to the generally optimistic attitude of artificial intelligence experts to artificial intelligence technology, most of the public and humanities scholars have some doubts about artificial intelligence.

We should consider the public's concerns about artificial intelligence and the fear of accepting artificial intelligence as a security issue, and resolve or mitigate it as much as possible through dialogues and discussions. If artificial intelligence products want to be commercialized, they need to be understood and accepted by the public. Scientists have a responsibility and an obligation to explain to the public. Artificial intelligence experts generally believe that the public do not need to worry about the safety of artificial intelligence. And they generally disclose the technical details of artificial intelligence to the public. Also, they would explain the details to the public. Of course, it is not easy for artificial intelligence experts to explain the technical details clearly to the public. The general public has some difficulty in understanding the technical language. In order to make the public really understand the relevant issues of artificial intelligence in the process of technical experts' dialogue with the public, it must establish a relationship of mutual trust, which requires technical experts to pay a lot of effort.

The changes of modern technology and its applications on human society are often unpredictable. We can only refer to similar technology. Also, we should make full use of our imagination to see the clues. Scholars can explain the characteristics and life style of intelligent society to the public through a variety of ways. And then, they could guide the public to adjust their ideas. The 21st century is an era of intelligence. The interaction between human and smart products will be normalized. In the future, human's reliance

on artificial intelligence is likely to be the same as our dependence on mobile phones and computers. We have to adapt to it.

IV. CONCLUSION

While artificial intelligence brings more happiness to human beings, it may also produce some security risks. In the coming "intelligent society", the security problem of artificial intelligence is no longer a futuristic problem. It should be a very important issue of philosophy and science and technology. The discussion on the security of artificial intelligence in this paper is rather superficial. However, it does not hinder the importance of the issues discussed. Only by paying attention to and solving the security problem of artificial intelligence can artificial intelligence bring bright future to mankind.

REFERENCES

- [1] Ma Junhui. Artificial intelligence is good or bad [J]. Net Friend World, 2011 (11). 马军辉. 人工智能是好是坏[J]. 网友世界, 2011(11)
- [2] Chen Lipeng. Ethical issues of science and technology triggered by artificial intelligence [J]. Literature Education (Vol. 2), 2012(8). 陈立鹏. 人工智能引发的科学技术伦理问题[J]. 文学教育(下), 2012 (8)
- [3] Kang Lanbo. On the philosophical implication of artificial intelligence [J]. Journal of Chongqing University (Social Sciences Edition), 2002(8). 康兰波. 论人工智能的哲学意蕴[J]. 重庆大学学报(社会科学版), 2002 (8)
- [4] Zhang Nan. On the responsibility ethics in the development of contemporary technology [D]. Dalian University of Technology, 2006.. 张楠. 当代技术发展中的责任伦理研究[D]. 大连理工大学, 2006.
- [5] Wang Yan. The ethical fulcrum of harmony between human and nature in environmental ethics [D]. Jilin University, 2008. 王妍. 环境伦理人与自然关系和谐的伦理支点[D]. 吉林大学, 2008.
- [6] Gong Yuan. Philosophical thinking on artificial intelligence [D]. Wuhan University of Science and Technology, 2010. 龚园. 关于人工智能的哲学思考[D]. 武汉科技大学, 2010.
- [7] Jiang Shen. To rule the robots [N]. Wenhui, 2011-6-7. 姜滢. 给机器人做规矩, 要赶紧了[N]. 文汇报, 2011-6-7.
- [8] Hu Yao. Robots must have "virtue" [N]. Economic Information Daily, 2008-12-24. 胡瑶. 机器人也要有“德性” [N]. 经济参考报, 2008-12-24.
- [9] Li Xin. Robots have moral standards [N]. Modern Express, 2012-6-5. 李欣. 机器人也需要道德标准[N]. 现代快报, 2012-6-5.
- [10] Shao Ling. Who needs rules, machines or people [N]. Wenhui, 2011-8-14. 邵岭. 谁更需要规矩, 机器还是人[N]. 文汇报, 2011-8-14.
- [11] Wang Yijun. Artificial intelligence and human life —Academician Zheng Nanning talks about the research status of artificial intelligence [N]. Shanghai Science and Technology News. 2001. 王毅俊. 人工智能与人类生活—郑南宁院士谈人工智能的研究现状[N]. 上海科技报. 2001
- [12] Li Junping. The ethical problems and countermeasures of artificial intelligence technology [D]. Wuhan University of Technology, 2013. 李俊平. 人工智能技术的伦理问题及其对策研究[D]. 武汉理工大学, 2013.
- [13] Chen Lipeng. Ethical issues of science and technology triggered by artificial intelligence [J]. Literature Education (Vol. 2), 2012 (08). 陈立鹏. 人工智能引发的科学技术伦理问题[J]. 文学教育(下), 2012 (08).
- [14] Li Meiqi. Ethical issues of science and technology triggered by artificial intelligence [J]. Shanxi Youth, 2016 (08). 李美琪. 人工智能引发的科学技术伦理问题[J]. 山西青年, 2016 (08).
- [15] Zhai Zhenming, Peng Xiaoyun. "Strong artificial intelligence" change the world — Technological leap of artificial intelligence and prospect of application ethics. People's Tribune, Academic Frontier, 2016, <07>. 翟振明, 彭晓芸. “强人工智能” 将如何改变世界—人工智能的技术飞跃与应用伦理前瞻团. 人民论坛·学术前沿, 2016,<07>.