Security Framework of Artificial Intelligence System

Rong Yu*
Xi'an University of Finance and Economics, Professor, No.360 Changning street, Xi'an, Shaanxi, China

*Corresponding author email:yurong21@126.com

Abstract. Artificial intelligence technology not only brings people great satisfaction with the progress of science and technology, but also may bring some social security threats. Therefore, in the face of these social security threats, it is urgent to establish a security framework of artificial intelligence to improve the security of artificial intelligence system. This paper, based on the analysis of social security threats caused by artificial intelligence technology, referring to the traditional security model, proposes an artificial intelligence security framework with security boundary as the core.

1. Introduction
With the rapid development of artificial intelligence which has embedded intelligent algorithms or programs and has human like intelligence, and even self-consciousness, its applications and products have been integrated into all aspects of human life. While making people enjoy the great satisfaction of scientific and technological progress, it is also constantly breaking through the limits of human intelligence, blurring the boundaries between human and machine, and challenging human's subjectivity in society and then it may bring a series of social security threats to human society. Therefore, facing these social security threats, it is urgent to establish an artificial intelligence security framework to improve the security of artificial intelligence system.

2. Social Security Threats Brought by the Development of Artificial Intelligence
Compared with the technological breakthrough of any era, artificial intelligence, whether it is power, steam engine, computer or Internet, seems to attract more attention because artificial goods or products based on artificial intelligence technology are no longer lifeless machines. They not only have human appearance, emotion and emotion, but also may have human character or human attributes. Moreover, they seem to be far more intelligent than human beings. Then in the future development of artificial intelligence, what kind of security threat will it bring to human society?

2.1. Personal Privacy
As the artificial intelligence system based on deep learning and big data needs a lot of data to train the learning algorithm, it will inevitably lead to the problem of privacy leakage. Although there is a consensus on the protection of personal privacy in modern society, artificial intelligence based on Internet, big data, Internet of Things and Cloud Computing may bring unprecedented threat to the privacy of individuals because in the intelligent society, all personal information, including gender, age, height, weight, health status, study and work experience, home address, contact information, and even Internet browsing records, chat content, travel records, shopping records, medical records, bank accounts, etc., may be collected and stored in detail as big data. By mining and analyzing a large amount of personal data that seems to be fragmented, data collectors can generate or restore a person's...
life portrait, and individuals unconsciously lose control of their privacy. Once the personal sensitive information mastered by artificial intelligence is stolen, leaked, traded, or even illegally used for commercial purposes, then, for individuals, the consequences of the violation of personal privacy will be unimaginable. Therefore, it is urgent to solve the contradiction between the protection of personal privacy data and the development of artificial intelligence.

2.2. Technical Security and Responsibility
On the one hand, it is necessary for people from all walks of life to gather intelligence based on cloud or Internet artificial intelligence research and development platform because the complexity of most artificial intelligence technologies based on deep learning is far beyond the research and development capability of a single scientific research or business entity. But the loopholes of Internet and artificial intelligence technology itself may cause huge security risks. On the other hand, even the developers can not control the decision-making or execution process of artificial intelligence completely and accurately because the external perceptor and internal algorithmic program that the artificial intelligence system relies on are completely different from the human perception and internal logical thinking mode. If the loopholes of artificial intelligence technology itself or the negative consequences that its developers are difficult to predict happen, who will bear the corresponding responsibility? In other words, the security problems and the responsibility problems caused by artificial intelligence technology need to be solved.

2.3. Ethical Algorithm Embedding
Although the algorithm of artificial intelligence seems to be a set of program code and it seems to be irrelevant to ethical value, if the designer, owner or user of the artificial intelligence system has subjective bias, the algorithm results of the artificial intelligence that always depends on the given data and algorithm model may have bias, or even "Algorithm discrimination" due to the different positions or needs of the designers, owners and users of artificial intelligence system and it will put forward different value concept requirements for artificial intelligence system. And after deep learning, such prejudice and discrimination may be further strengthened in the algorithm. At present, the embedding problems of this ethical algorithm exist in some fields, such as crime risk assessment, credit risk assessment, prediction analysis, etc.

2.4. Unemployment
As an extension of human intelligence, artificial intelligence is penetrating into every industry and every kind of work for it can greatly promote the development of productive forces, which will change the mode of human labor production and survival profoundly, comprehensively and thoroughly. It may be used in labor-intensive, simple and repetitive, highly process oriented and other types of work because compared with human labor, artificial intelligence has many advantages: it can not only reduce labor costs but also be tireless; it not only has strong self-learning, optimization and upgrading ability but also is durable; it can not only complete high-precision tasks in specific scenarios but also has high accuracy. This may cause some employees to face the threat of unemployment, such as assembly line workers, car drivers, couriers, waiters and so on. That is to say, the development goal of substitute, assistant, surpass human of artificial intelligence may bring about the change of labor structure or the replacement of labor force. Then, once the pure technology development and labor force training are out of touch, it will lead to serious social problems.

3. Construction of the Security Framework of Artificial Intelligence System
In view of the social security threat caused by artificial intelligence, it is urgent to establish a security framework of artificial intelligence system to ensure the security, reliability and controllability of artificial intelligence. The establishment of this security framework of artificial intelligence should ensure that artificial intelligence will not cause possible damage to human life, property and environment, and will not be attacked by system damage, information theft and misleading decision-making. At the end of the 20th century, American Internet security systems company proposed security models such as P2DR (Policy Protection Detection Response), P2DR2
(P2DR+Restore). This security model is suitable for artificial intelligence system. However, artificial intelligence system also has its own characteristics, such as whether the traditional security detection is certain security, whether the security response means can ensure timely closed-loop, and whether the security strategy is concerned about certain accuracy. Considering the above characteristics, this paper holds that the concept of security boundary should be added to the artificial intelligence security framework based on P2DR model (Figure 1). This kind of security boundary means that the security rules of artificial intelligence system should be defined in a standardized paradigm, that is, to formulate the active space that artificial intelligence system can allow or to define the width of the security boundary of artificial intelligence system.

Figure 1. Security framework of artificial intelligence.

3.1. Establishment of the Security Embeddedness Mechanism of Artificial Intelligence
As a tool and means developed by human beings and aimed at promoting human development, artificial intelligence may develop independent consciousness and will, surpass human ability and become a brand-new individual out of human control. Therefore, at the beginning of the development of artificial intelligence, it must be established the security embeddedness mechanism. That is, on the basis of safety regulation on the developers, owners and users of artificial intelligence, the relevant stakeholders can correctly grasp the development direction of artificial intelligence technology, enhance their sense of social responsibility, abide by the corresponding basic safety standards and fulfill the corresponding safety obligations. Simultaneously, on the premise of safety risk assessment for the expected research and development objectives, behavioral characteristics and application scenarios of artificial intelligence system, the laws, ethics and other norms and values of human society are digitized and coded in the way of safety judgment. Then, the algorithm language is embedded with moral logic, cognitive logic and behavioral logic, and the intelligent system with comprehensibility of operation logic, predictable behavior consequence and traceability of operation process and discontinuation of unexpected error is developed. Through the top-down or bottom-up form of artificial intelligence system implantation, and after these norms and values are implanted, the norms and values that have been embedded in the artificial intelligence system are professionally evaluated according to the social security standards that are consistent with the norms and values of human society.

3.2. Formulation of the Safety Design of Artificial Intelligence
As the development and application of artificial intelligence may substantially change people's living practice and specific interpersonal relationships, it is necessary to improve the system design of "people-centered". This kind of system design should first protect the individual's right to know about their own data collection, that is, for the research, development and application of data that may involve people's physical and mental integrity, personality and dignity, as well as people's legitimate rights and interests, only with the understanding and consent of the parties can they collect, store or use the data. And in case of unexpected consequences endangering the life, physical and mental integrity or other legitimate rights and interests of the parties in the process of collection, storage or
use, the design should be re-authorized. Secondly, in line with the principle of equality for all, on the basis of efforts to eliminate the digital divide, the gap between the rich and the poor, economic inequality and social polarization between the rich and the poor, this design should put an end to the harm that "capital logic" or "technical logic" may cause to people, and through the establishment of a sound education and training, social welfare and security system, carry out the protection and assistance of "digital poor areas" or "digital poor".

3.3. Improvement of the Safety Rules for the Decision-making of Artificial Intelligence Algorithm
In order to prevent the known or potential risks brought by artificial intelligence, it is particularly important to improve the relevant safety rules to clarify the nature and ownership of its responsibility. On the basis of establishing a strict safety review system of artificial intelligence at the national level, these safety rules should not only be able to evaluate the security of artificial intelligence rationally but also effectively regulate the design, research, development and application of artificial intelligence; they can not only determine the rights, responsibilities and obligations of different participants in the process of research, development, application and management of artificial intelligence, and predict or prevent its adverse consequences, but also seriously hold the responsible person accountable once the fault event occurs.

4. Conclusion
In the face of a series of social security threats that artificial intelligence technology may bring, based on the traditional security model, this paper proposes an artificial intelligence security framework with security boundary as the core to improve the security of artificial intelligence system.

References
[1] Chen Yufei, Shen Chao, Wang Qian, et al. Security and privacy risk of artificial intelligence system. Journal of Computer Research and Development, 2019, 56 (10)
[2] Murray G. Stoic Ethics for Artificial Agents. Canadian Artificial Intelligence, Edmonton, Alberta, (2017)
[3] Daniel Thompso. Program Good Ethics into Artificial Intelligence. Nature, October, p. 291. (2016)
[4] House of Commons. Robotics and Artificial Intelligence. London: Parliament, pp.12-16. (2016)
[5] Orseau,L. and M.S. Armstrong. Safely Interruptible Agents. Association for Uncertainty in Artificial Intelligence, 2016
[6] Bostromn. Superintelligence: Paths, Dangers, Strategies. Oxford: Oxford University Press, (2014)