Assessing China's Policy thinking on ai development

Nathália Viviani Bittencourt[I]

Karla Godoy da Costa Lima[II]

Abstract: The rapid technological development of China, especially when it concerns to its strategic interests in artificial intelligence, has often been assessed as a disruptive threat to the global order and to the balance of international power. However, recent documents released in China seem to present a view that deviates from the simple winner-takes-all framework. Indeed, they illustrate an innovative approach to ethics and governance in artificial intelligence, which demonstrates China's intention to build a legal basis for dialogue with the international community. Thus, this article aims to answer the following question: What is the Chinese policy view on the development of artificial intelligence technologies? This paper proposes to use both qualitative and quantitative methods to answer the research question. First of all, we are going to carry out a brief document analysis of the Chinese documents that dealt with national plans to booster artificial intelligence industry from 2015 until 2017. Second, to provide a systematic analysis of Chinese thinking on the subject, we will do a content analysis of the documents dealing with issues of ethics, security, and governance in the use of artificial intelligence produced in 2018 and recently in 2019. Contrary to the zero-sum game that results from a skeptical look at China's technological ambitions, these documents show that there may in fact be many areas of mutual interest the international community should discuss and build a legal framework regarding safe and ethical AI. Many issues discussed in the paper, such as algorithmic model defects and training data bias, are the same issues raised in the democratic discourse around safe and ethical AI. Advancing global norms and standards in these areas may, therefore, stand to be a win-win relation.

Keywords: Artificial Intelligence (AI), China's Strategy, AI's governance, Ethics.

Artigo recebido em 11/11/2019 e aceito em 27/11/2019.

1 Introduction

Since the 1950s, with scientist Alan Turing's provocations of machine thinking as well as the Darthmouth technology conferences, the term artificial intelligence (AI) has begun to take shape. In fact, the central idea is that it is a decision-making system that responds, acts and enhances the chances of solving a problem.

Today, the evolution of software engineering techniques considers AI as an umbrella term that also encompasses machine learning and deep learning, whose skills, especially in relation to the latter, bring to light ethical discussions about its increasing

autonomy and the difficulty of scientists to develop AI models and decision chain explainable.

Given their ability to facilitate individuals' choices and enable accurate answers to any type of problem, AI agents have come to be widely used by businesses, governments and individuals as a mechanism for finding rapid responses from systems immersed in digital data. However, some impacts from its high capillarity are already being observed, such as structural unemployment, increased state surveillance, loss of privacy and cybersecurity.

Artificial intelligence has become a strategic technology for many states. Since 2016, national AI strategy papers policies have begun to emerge. From this perspective, in 2018, Prakash (2018) wrote the book "Go.AI (Geopolitics of Artificial Intelligence)", in which the author states that AI production race reformulated politics in many ways, especially with regard to geopolitics and changes in the balance of power.

From this perspective, China was one of the first countries to release policies for AI development in a number of strategic sectors. Aiming to make the nation prosperous and with competitive human capital in this area, the country has devoted itself heavily to the massive investment project in innovation by training experts and subsidizing companies to create an environment of unified national growth.

Moreover, it is noted that the Middle Empire's ambition is not only to adopt cutting-edge technologies, but also to set international technology standards and to become a global reference in AI performance. Thus, through a strategic plan to expand its use for service optimization, China began to spread it to almost all social sectors, including health, state surveillance, education and the military.

Due to this massification of AI development, Chinese researchers, as well as the government itself, have begun to raise concerns about the impact this technology reverberates on social and international relations. Thus, a number of challenges began to be assessed, such as the need to develop robust models that translate AI principles into practice. In this sense, many Chinese documents address the relevance of building steps to data chain verification from the early stage of the AI creation process to evaluate whether the project follows principles of non-discrimination, accountability, transparency, explainability and legality.

Given this scenario, the present article intends to make a document and content analysis of some important documents that relate the Chinese AI strategy and its developments in the ethical and security field. The paper seeks to answer the following question: What is the Chinese policy view on the development of artificial intelligence technologies?

In addition, the paper will briefly access how this technological race became a key element of a possible new Cold War between Sino-American relations. Beyond the description of the documents, it is essential to evaluate China's ambition through the thinking of scholars and policymakers, as well as the view of some International Relations schools about this new world phenomenon. In order to demystify the zero-sum game through the technological race, we propose that these two great powers engage with the international community to ensure that global standards of security and ethical development in AI are built.

This article is divided into four parts. The first is devoted to theoretical analysis, whose function is to identify the factors that delimit the new technological race along the lines of Revolution 4.0 and how it affects the international balance of power and China-US relations. The second, in turn, deals with our methodological approach and the stages of access to the texts and the search for the central ideas of each one of them, while the third delves into the documents, specially into the ones that regards Chinese approach to AI security, governance and ethics. The fourth, finally, illustrates the results and conclusions that this research made it possible to find.

2 Theory

According to professor Xuetong (2014), the top strategic interest of a rising power is to establish a new world order. It is also argued that Chinese diplomacy has shifted from keeping a low profile (KLP) to striving for achievement (SFA) since the arrival of President Xi Jinping. In this sense, it is interesting to note that this kind of shift aims at increasing China's strategic credibility by providing security protection and enhancing mutual development.

In this context, it is essential to highlight what Chinese scholars actually think about the country's aiming at leading role in the international arena and move away from

the commonplace of western-theories oriented. Thus, Zhang (2012), for instance, coined the term moral realism (or Confucian Realism), which combines political determinism of Chinese traditional philosophy with modern realist theory of international relations. It should be noticed that moral realism categorizes international leadership under humane authority, which is observed under the principles of fairness, justice, and civility, in accordance with Xuetong (2016).

On the other hand, offensive realist, such as Mearsheimer (2001), argue that all great powers have to employ offensive strategy to maintain their dominant positions in the anarchical international system. In contrast, moral realism skeptically observes this approach and believes that there are other strategies available to improve a nation's power (Xuetong, 2016). Given this, it is interesting to look at different perspectives on how the field of realism theory, for example, is evaluated differently from the traditional by Chinese researchers. In this sense, the Chinese technological rise must be broadly interpreted, not only by the western view.

Besides, part of the driving force that guides China in its quest for a prosperous nation and as an international reference comes from the trauma of the humiliating century. In a recent article, WU (2019) demonstrates how science has always permeated Chinese culture as an essential sector for saving China to build a modern society. From this perspective, this fact becomes important to understand how China demonstrates its interests through the foreign policy and how researches assess the country's approach in international affairs.

Thereby, the rise of China as a new hegemon is one of a major International Relations puzzles for the 21st century, and the technology race is just one of the intricacies of this scenario. Minghao (2019) promotes a Chinese perspective on US–China strategic competition by pointing out that the economic and technologic realm have become more salient and could potentially exert fundamental impact to the global order. It concludes that b*oth sides should deal effectively with the transition to a relationship wherein there is a balance regarding the competition in order to avoid the so-called Thucydides trap.*

In addition, as technology has become vital not only to China's quest for strength and wealth but also to long-term competition between the US, Some researchers claim that the present resembles in many respects the Cold War scenario. In this sense, Minghao (2019) asserts that this possible contemporary Cold War has similarities of ideological

divergence and competitiveness, but the new guise stands out with the interdependent economic structure and the impact of social media and other technologies.

Concerning the strategic advancement in AI for the Sino-American competition, Wang and Chen (2019) argue that AI development has been widely recognized as a manifestation of a country's power. However, China's rapid advance in this area still has a considerable gap with the United States, especially in producing talent reserves, innovation systems, and computer algorithms. The authors understand that AI race has become a new dimension of Sino-U.S. competition because major breakthroughs in AI technology and application can change the balance of military and economic power between both countries, as well as being part of the trade war through protective measures.

In the light of all the above, Ying (2019) investigates the influence of AI for International Relations. The paper believes that the security and governance challenges brought by AI should be faced together by a multi-stakeholder approach. Countries should look at the problem from the perspective of building a community of human destiny, and discuss the international norms of artificial intelligence from the perspective of common security. In her words,

> If we look at the world as a zero-sum game and a pursuit of absolute security, there is no doubt that AI will, similar to the way atomic bombs and satellites were in the 1940s and 1950s, become the new focus of competition among big countries and a driving force of two or multiple parallel global orders. However, if we adopt the perspective of a Community of Shared Future for Mankind and view the problem through the concept of common security, it is not difficult to realize that the security and governance challenges brought about by AI technology are problems that all people face together. In this way, it should not be difficult for us to jointly explore the norms acceptable to all stakeholders in the spirit of equal consultation. If so, will AI become the "Mars invasion" challenge that unites China, the United States, Russia and the rest of the world?[III]

3 Methodology

The purpose of this section is to describe the main methodological procedures used in conducting the research in such a way that it increases transparency and ensure the replicability of results (King, 1995; Janz, 2015). Intending to answer the question

proposed for this research (what is the Chinese policy view on the development of artificial intelligence technologies?) we use both qualitative and quantitative methods.

First of all, we are going to carry a brief document analysis of the following Chinese documents that were the first to deal with the development of artificial intelligence technologies until 2017:

Figure 1: AI Policy Progress in China (2015-2017)

Source: own elaboration, 2019

Document analysis allows adding the dimension of time to the understanding of social and political events. It favors the observation of maturation or evolution of individuals, groups, concepts, knowledge, behaviors, mentalities, practices, among others. (Cellard, 2008). That is our objective by reviewing those three documents.

Second, in order to provide a more in-depth analysis of the Chinese thinking on AI security, ethics and governance, we carry out a content analysis of the documents produced in 2018 and 2019. For Olabuenaga and Ispiúza (1989) content analysis is a technique that allows the researcher to interpret a vast class of documents, whose purpose is to acquire knowledge about aspects and phenomena of social life. Weber (1990, p. 19), states that Content analysis involves the development of a series of procedures for making inferences from texts. In this article, we perform the content analysis - qualitative and quantitative - using *NVivo software, version number 1.*

The content analysis is performed in two documents in order to understand which are the areas of main concern of AI and its proposals for future international legal standards since 2018, as following:

Figure 2: AI Policy Progress in China (2018-2019)

Source: own elaboration, 2019

Use used the translated version provided by the projet *Digichin*[IV] to access The Governance Principles for Responsible AI (2019). On the other hand, there was no translated version for the full The AI Security White Paper (2018) document, so we had to do the translation ourselves. For reasons of transparency and accountability, the translated document will follow as an appendix. To translate the document we used Google Translate and prior Chinese knowledge of the authors.

We chose to perform a deeper analysis in those documents, as it was produced by bodies associated with the Chinese government and they are the most current document produced related to AI security, ethics and governance. In short, this is our research design:

Table 1 – Research Design

Document	Technique	Procedure	Source

Made in China 2025, 2015	Document Analysis	Identification of individuals, groups, concepts, knowledge, behaviors, mentalities, practices, among others.	China's State Council
New Generation Artificial Intelligence Development Plan, 2017.	Document Analysis	Identification of individuals, groups, concepts, knowledge, behaviors, mentalities, practices, among others.	China's State Council
Higher Education IA Innovation Action Plan, 2017	Document Analysis	Identification of individuals, groups, concepts, knowledge, behaviors, mentalities, practices, among others.	Chinese Ministry of Education.
AI Security Paper	Content Analysis	Frequency distribution, words cloud, cluster analysis, and graphics.	China Academy of Information and Communication Technologies.
Governance Principles for Responsible AI.	Content Analysis	Frequency distribution, words cloud, cluster analysis, and graphics.	Chinese Ministry of Science and Technology

Source: own elaboration

As we have already established our research design, we had to set coding groups for the proposed content analysis to start exploring the documents. Table 2 summarizes the coding groups to explore these two documents.

The categories used throughout the analysis were created based on Cowls and Floridi (2018), a well-known paper related to AI principles, and in the AI Security White Paper (CHINA, 2018). Cowls and Floridi (2018) divide standards on AI into five

categories, four of them being inspired by bioethics and the fifth inspired by a principle peculiar to the functioning of AI, which are: **(1) beneficence**: promoting well- being, preserving dignity, and sustaining the planet; **(2) non-maleficence**: privacy, security and capability caution; **(3) autonomy**: the power to decide**. (4) justice:** promoting prosperity and preserving solidarity; and, **(5) explicability**: enabling the other principles through intelligibility and accountability.

Likewise, the AI Security White Paper (2018) gave us some definitions for important issues and standards to the development of artificial intelligence in China. After extensive reading and based in the literature showed, we used the definitions that would help deep understanding the documents.

Category	**Description**
Autonomy	The use of AI technologies to improve decision making (Cowls; Floridi, 2018).
Cybersecurity Applications	"The application of artificial intelligence in the security field is the focus of current domestic and foreign enterprise technology and application innovation." (CHINA, 2018: p.18).
Data Management	"Data management applications refer to the use of AI technologies to achieve data protection objectives such as hierarchical classification, leak prevention, and leak traceability." (CHINA, 2018: p. 8).
Information Censorship	"Information censorship applications refer to the use of AI technology to assist humans in undertaking rapid review of various forms of expression and a large volume of harmful network content." (CHINA, 2018: p. 8).
Network Protection	Network protection applications refer to the research and development of technologies and products to use AI algorithms

	for intrusion detection, malware detection, security situational awareness, and threat early warning, etc." (CHINA, 2018: p. 8).
IA as a Governance Tool	The use AI as an international tool for dialogue and cooperation; with full respect for each country's principles and practices for AI governance.
Financial Risk Control	Efficiency and accuracy of financial risk control work using Artificial intelligence technology.
Intelligent Security	"Intelligent security based on artificial intelligence relies on the learning of massive video data, which can complete the inference and prediction of behavior patterns." (CHINA, 2018: p. 20).
Public Opinion Monitoring	"Public opinion monitoring applications refer to the use of AI technology to strengthen national online public opinion monitoring capabilities, improve social governance capabilities, and ensure national security." (CHINA,2018: p. 9)
Basic Information	Basic information present in the documents (COWLS; FLORIDI,2018).
Actors	To whom belongs the competence to implement the activity.
Identification	Name of the document.
Objective	Objectives of the document.
Year	Year of the document.
Beneficence	The use of AI to promote the well-being and sustainable development of AI technologies (COWLS; FLORIDI,2018).

Concept	The basic concept and development process of artificial intelligence.
Development	The development of artificial intelligence.
Law, Regulation and Policy	Regulations and policies, establish and strengthen corresponding safety management laws and regulations and management policies for key application domains of AI in prominent security risks.
Preserving dignity	The use of AI to improve the adaptability of disadvantaged groups, and strive to erase the digital divide.
Promoting Well-being	AI as a tool for enhancing the common well-being of humanity.
Security Assessment	Mechanisms to clarify the responsibilities of developers, users, beneficiaries and to ensure the human right to know and give notice of possible risks and impacts of AI use.
Standards and specifications	Standards and specifications for AI security requirements and security assessments and evaluations in international and domestic level.
Sustaining the Planet	AI promoting green development and meeting the requirements of environmental friendliness and resource conservation. AI promoting controllable ecology and guaranteeing the secure and controllable development of AI.
Talent Corps	Efforts to develop a workforce specialized in Artificial Intelligence.
Technological Methods	Technological support capabilities for security management, emergency response and governance systems.

Justice	Fairness in the acquisition and utilization of information data.
Preserving solidarity	Eliminate bias and discrimination in the process of data acquisition, algorithm design, technology development, product R&D, and application.
National Security	National security refer to the risks to national military security and political system security brought about by risks and hidden dangers from the application of AI in military operations, public opinion, and other fields (CHINA,2018: p. 8).
International Practices	Practices and problems related to countries other than China about their national military security and political system security.
Military Security	Artificial intelligence being used to build new military strike force.
Political System Security	The use of artificial intelligence in political processes.
Security Risks of AI	"AI cyberspace security risks include: cybersecurity risks, data security risks, algorithmic security risks, and information security risks" (CHINA,2018: p. 7).
Algorithmic Security	"Algorithmic security risks correspond to algorithm design and decision-related security issues in the technical layer, as well as security risks such as black-box algorithms and algorithmic model defects." (CHINA,2018: p. 8).
Decision-making security	Preventing AI to induce unfair decision-making results.

Design Security	Artificial intelligence security architecture covering three dimensions of security risk, security application and security management.
Cybersecurity	"Cybersecurity risks involve vulnerabilities in network infrastructure and learning frameworks, backdoor security issues, and systemic cybersecurity risks caused by malicious applications of AI technologies." (CHINA,2018: p. 7).
Learning Framework Security	AI security assessment and management capabilities.
Network Facility Security	Efficiency in network facilities.
Data Security	"Data security risks include training data bias in AI systems, unauthorized tampering, and security risks such as the disclosure of private data caused by AI." (CHINA,2018: p. 8).
Privacy Security	Protection of personal privacy and the individual's right to know and right to choose information content shared.
Training Data Security	Personal data collection, storage, processing, use, and other aspects, concerning AI development.
Information Security	Information security risks include AI technology applied to information dissemination and information content security issues for smart products and applications.
Information Content Security	Security risks brought by malicious applications to cyberspace and data content spread through AI technology.
Information Dissemination Security	The use of AI to accelerate the spread of bad information.

Societal Security	Using AI to safeguard societal security and respecting human rights.
Bodily Security	Personal safety endangered by security risks of artificial intelligence.
Employment Security	Artificial intelligence reducing or even eliminating existing jobs and leading to structural unemployment.
Ethical Security	AI should conform to human values, ethics, and morality.

4 Assessing Chinese documents on ai

First, it is important to note that the list of documents to be analyzed is not exhaustive. All of them have been selected due to their political relevance, social impact and importance for this work, and most of them are already available in English. The only one whose full translation was required was AI Security White Paper, which is available at the appendix of this paper.

4.1 Brief Document Analysis

4.1.1 Made in China 2025 [*中国制造 2025*]: broadly speaking, the first grand strategy to boost Chinese Industry

Made in China 2015 is a state-driven industrial plan to promote a change of value in some key areas, such as robotics, medicine, agriculture, energy, automotive and aerospace industry and information technology. The strategic objective of the document is to promote a national plan to transform the idea of a low quality Chinese product industry into an innovative and high market value model.

In order to become a major manufacturing power in straight competition with the United States, China creates this master plan to reposition the country as an industrial superpower of the future. In this sense, this state-directed plan to support and subsidies

domestic companies intend to enhance the strategic areas to make the country a reference in innovation and technological-driven products.

4.1.2 New Generation Artificial Intelligence Development Plan *[新一代人工智能发展规划]*

This is the first document that outlines a major strategy specifically focused on AI development that sets goals until the year 2030. This policy shows China's seek to build a domestic AI industry worth nearly 10 trillion RMB in the next few years and to become the leading AI power by 2030. It specifies that as AI has become a central field to international competition, China builds a plan to stimulate the development of new industries and to enhance national security through AI, so as to give the country an advantage in this new competition.

It is interesting to highlight that the document also attributes competitive relevance to the advancement of research seeking standardization in the AI development chain, especially in shaping security, ethics and governance policies. In this sense, China is concerned with occupying this space of global leadership as a strategic and political international advantage.

4.1.3 Action Plan for Artificial Intelligence Innovation in Colleges and Universities [高等学校人工智能创新行动计划]

This document released by the Ministry of Education highlights the importance of building national talent corps in AI as a multi-level structure to optimize research innovation systems in fundamental theorems, such as city governance, legal and medical services. Overall, the plan sets three goals for the next 12 years. It looks for providing infrastructure of colleges and universities to adapt to the next generation of AI development, developing high level of theoretical research in AI and making Chinese colleges and universities the world's leading AI innovation centers.

4.2 Content Analysis

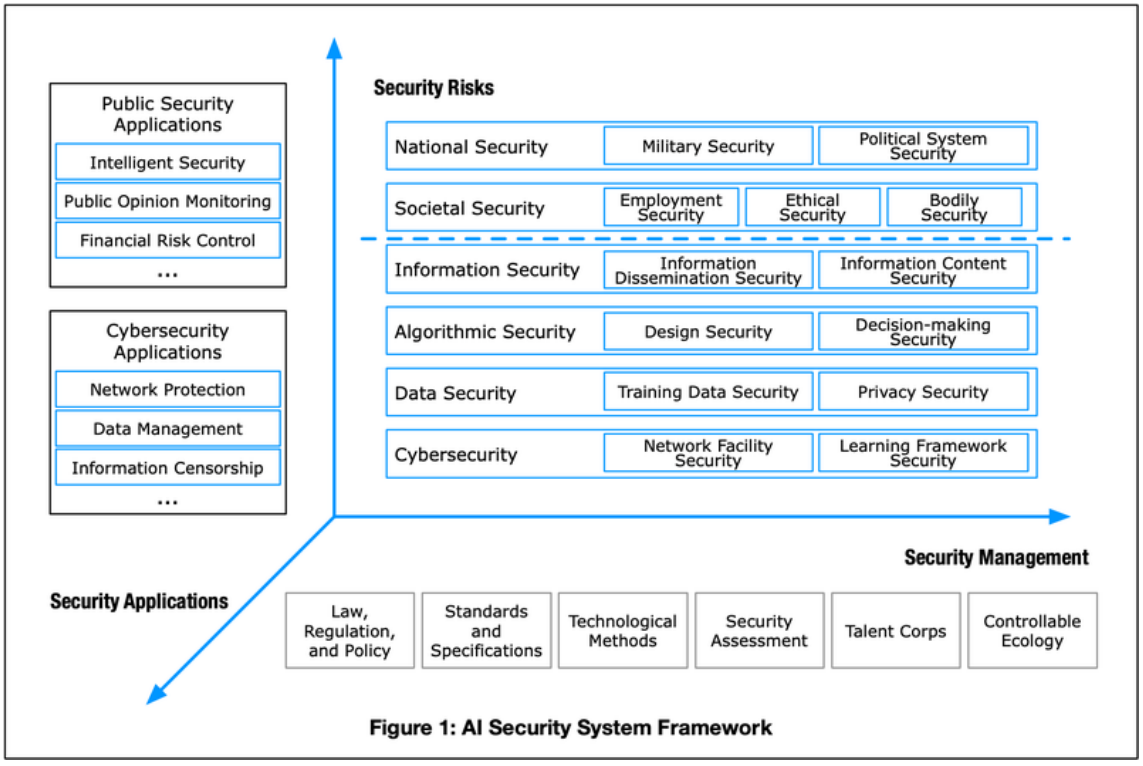
With the help of the categories illustrated in the methodology, it was possible to quantify the two documents below that relate Chinese approach regarding AI ethics, security and governance.

4.2.1 AI Security Paper

This document brings the Chinese view on policy thinking on AI's challenges from cybersecurity to social stability**.** Through a global analysis of the current state of AI secure development, the paper provides different approaches to the social impact of AI and how China should make their access safer and more ethical for the planet.

From this perspective, the paper presents a division of Chinese concerns into three aspects, illustrated by the image below, namely: security risks, security applications and security management

Figure 3 – White Paper mainly concerns

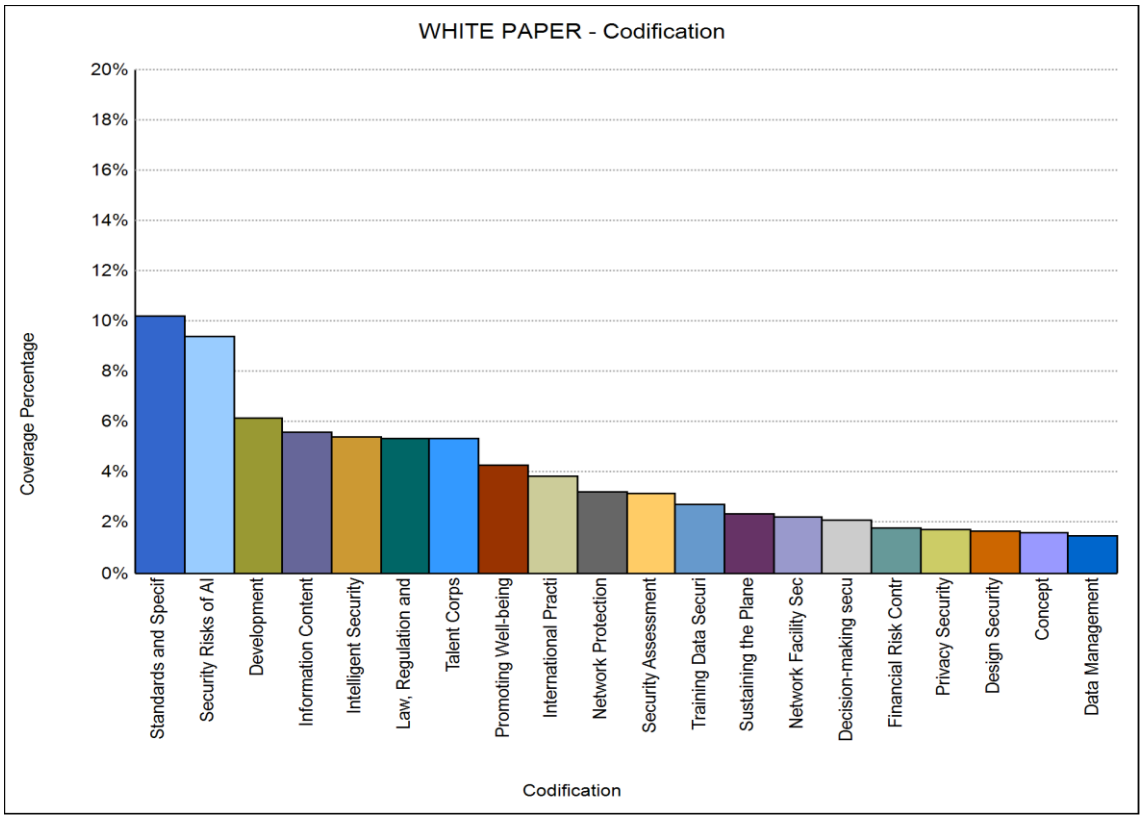


Source: CAICT - translated by DigiChina[1]

[1] Available at: <https://www.newamerica.org/cybersecurity-initiative/digichina/blog/translation-key-chinese-think-tanks-ai-security-white-paper-excerpts/>. Accessed: 9 nov 2019.

In accordance with the figure 3, security risks pervade the military field, political system, employment. In addition, there is Chinese concern about problems that developers of this technology may face, such as the robustness of the model, the need for effective data processing, network security, among others, which should be taken into consideration by all branches exposed in the security management field. In the area of security applications, China raises the use of AI to protect its interests in network protection, safe and unbiased data management, public monitoring and financial risk control. Thus, we can see the wide application of this technology to national security.
Besides, through the categories created, it was possible to make the graph below to illustrate in a quantified way the use of some terms that we've considered important.

Graphic 1 – word emphasis



Source: own elaboration

In this sense, it demonstrates the relevance that the paper is concerned with the issue of national and international regulation for the use of AI, as well as setting standards and specifications to AI development. In addition, according to the widely known principles

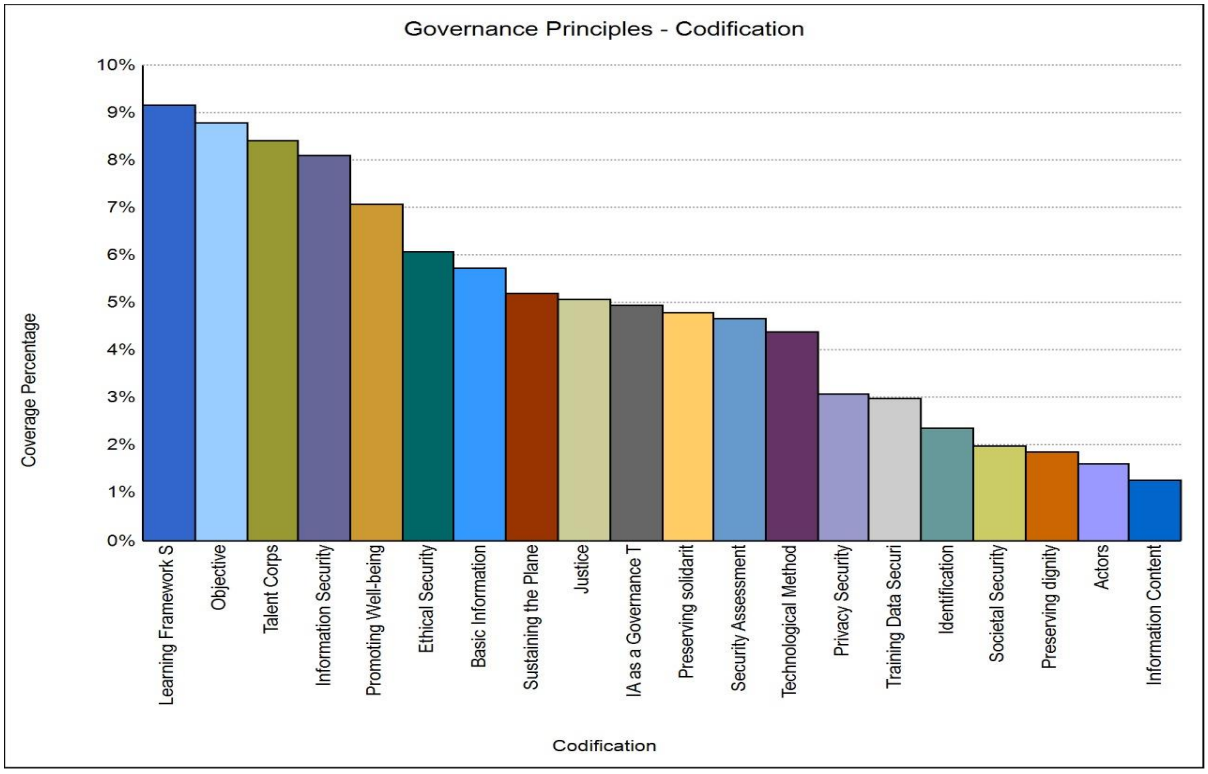
of the AI research community prided by *Cowls and Floridi (2018),* the Chinese also seek to emphasize well-being in AI development.

4.2.2 Governance Principles for Responsible AI

In June of the current year (2019), a Chinese Experts Committee linked to the Ministry of Science and Technology was set up to organize a document on the national principles of AI governance. The core of this document is to show Chinese policies aimed at building a society that develops AI guided by 8 principles: **(1) harmony** (promotion of human-machine harmony and serve human civilization), **(2) fairness** (justice, eliminating prejudice and discrimination in data acquisition, algorithm design), **(3) inclusive sharing** (environmental friendliness and resource conservation), ***(4)*** **privacy** (fully protect the individual's right to know and choose), **(5) safe and controllable use** (regarding robustness and anti-interference of AI models), **(6) sharing responsibility** (accountability), **(7) open collaboration** (multi-stakeholder interaction) and **(8) agile governance that respects the law of AI development** (improve governance systems).

Through these fundamentals, there is a strong relationship with the principles fostered in Cowls and Floridi (2018), which shows a consonance with the major concerns of the international community on the subject of AI governance. In the graphic below, we *illustrate in a quantified way the use of some terms that we've considered important.*

Graphic 2 - *word emphasis*



Source: own ellaboration

According to the chart, the promotion of well-being is one of the central themes explored, as well as the concern to reinforce investment in talent corps and the development of a safety oriented ethics.

5 Results

Through the document analysis, we found that Chinese strategy in Artificial Intelligence development is to ensure they reap all the benefits of this technology and are recognized as an advanced nation in this field. Throughout the content analysis, we could find the Chinese perspectives on AI's ethics, security and governance. We also note a few impressions regarding the security white paper for further analysis.

Indeed, concerning e White Paper, the Chinese Government places a special emphasis on the Security Risks of AI code from the comparison with how other countries have managed AI security risks. From this perspective, this part of the document emphasizes the importance of harnessing the good practices of neighboring countries and how to make the best of it for the national interests. Regarding the issue of monitoring

the public opinion, the document does not provide solutions, but only take as an example how the US uses national public opinion control to prevent crimes. Lastly, the paper also focuses heavily on the dissemination of practices that promote well-being, law and regulation, and especially to set standards and specifications nationally and internationally.

Both documents assessed through the content analysis are concerned with promoting welfare, security assessment, standards and specifications, sustainable planet, talent body and technological methods. This is the first time China's released a document that address the notion of artificial intelligence as a double-edged sword, as its development encompasses factors that enhance human capabilities while bringing new challenges to some areas such as cybersecurity, the future of work and public safety.

Below are the two word clouds from each document evaluated by content analysis to show the main words referred to give some ideas about Chinese concerns regarding AI security, ethics and governance. To illustrate, the figure 3 shows, for instance, a special relevance to security, data application and information standards, while the figure 4 gives more importance to the governance development and the promotion of human principles in AI.

Figure 4 – Word Cloud – AI Security White Paper

Figure 5 – Word Cloud – Governance Principles

Source: own elaboration

6 Conclusions

The documents showed above bring an innovative approach to IA that deserve attention to the international community. Indeed, they may be able to demonstrate the intentions of Chinese researchers and other entities to engage in global dialogues on AI ethics and governance.

Contrary to the zero-sum game that results from a skeptical look at China's technological ambitions, these documents show that there may in fact be many areas of mutual interest the international community should discuss and build a legal framework regarding safe and ethical AI. Many issues discussed in the paper, such as algorithmic model defects and training data bias are the same issues raised in the democratic discourse around safe and ethical AI. Advancing global norms and standards in these areas may, therefore, stand to be a win-win relation.

APPENDIX[V]

ARTIFICIAL INTELLIGENCE SECURITY WHITE PAPER
CHINA INSTITUTE OF INFORMATION AND COMMUNICATION SECURITY
SEPTEMBER 2018

ASSESSING CHINA'S POLICY THINKING ON AI DEVELOPMENT

NATHÁLIA VIVIANI BITTENCOURT E KARLA GODOY DA COSTA LIMA

PREFACE

As a strategic technology that leads the future, artificial intelligence has increasingly become an important engine driving the economic and social fields from digitalization and networking to intelligent acceleration. In recent years, the explosive growth of data volume, the remarkable improvement of computing power, and the breakthrough application of deep learning algorithms have greatly promoted the development of artificial intelligence. New technologies, such as autonomous driving, intelligent service robots, intelligent security, and smart investment have emerged in an endless stream, profoundly changing human production and life, and have a wide and far-reaching impact on the development of human civilization and social progress.

However, the advancement of technology is often a "double-edged sword". As a general purpose technology, artificial intelligence provides new means and new ways to ensure national cyberspace security and enhance human economic and social risk prevention and control capabilities. At the same time, in the process of technology transformation and application scenario, artificial intelligence brings about problems such as impacting network security, social employment, legal ethics, etc. due to technical uncertainty and wide application, and the national political economy and social security belt. There are many risks and challenges. The major countries in the world regard artificial intelligence security as an important part of the research and industrial application of artificial intelligence technology, vigorously strengthen the forward-looking research and active prevention of security risks, actively promote the application of artificial intelligence in the security field, and strive to create a new round of artificial intelligence. In the wave of development, we took the lead and won the initiative.

Based on the connotation of artificial intelligence security, this white paper puts forward the artificial intelligence security architecture for the first time. On the basis of systematically combing the artificial intelligence security risk and security application, this paper further summarizes the current status of artificial intelligence security management at home and abroad. Artificial intelligence security risk response and future development recommendations.

ASSESSING CHINA'S POLICY THINKING ON AI DEVELOPMENT

NATHÁLIA VIVIANI BITTENCOURT E KARLA GODOY DA COSTA LIMA

【Table of Contents】

(5) Promote industry collaboration and promote technology security applications.
(6) Increase personnel training and improve the employment skills of personnel.
(7) Strengthen international exchanges to address common security risks.
(8) Increase social propaganda and scientifically handle security issues

I. Artificial Intelligence Security Connotation and Architecture

(1) The basic concept and development process of artificial intelligence

1. Basic concept of artificial intelligence

Alan Turing, the father of computers, put forward "machine intelligence" and the famous "Turing test" in the 1950's "Computers and Intelligence": If more than 30% of the testers can't determine whether the testee is a human or Machine, then this machine passed the test and is considered to have human intelligence. In 1956, at the Dartmouth meeting in the United States, scientist McCarthy first proposed "artificial intelligence". Artificial intelligence is to make the behavior of the machine look more like the intelligent behavior exhibited by human beings. When the concept of artificial intelligence[VI] is proposed, the scientists mainly determine the intelligent discriminant criteria and research objectives, but do not answer the specific connotation of intelligence.

Afterwards, famous scholars including Winston, Nelson[VII] and Zhongyixin[VIII] of China all put their own opinions on the content of artificial intelligence, reflecting the basic ideas and basic contents of artificial intelligence: how to apply computer simulation of human intelligence behavior in basic theory, methods and techniques. However, due to the continuous evolution of the concept of artificial intelligence, there is no uniform definition.

In combination with industry experts, the project team believes that artificial intelligence is the use of artificial manufacturing to implement intelligent systems on intelligent machines or machines to simulate and extend human intelligence. Theories, methods, and techniques for perceiving the environment, acquiring knowledge, and using knowledge to get the best results.

2. The development of artificial intelligence

The development of artificial intelligence has experienced many downturns, and this round of development has shown an acceleration. Artificial intelligence has

been in existence for more than 60 years since its birth in 1956. In its development process, it has formed a number of schools, such as symbolism, connectivityism, and behaviorism, and has achieved some milestone research results. However, due to the limitations of scientific cognition and information processing at various stages, the development of human intelligence has experienced many rounds of ebb and flow, and has repeatedly fallen into a trough.

Since the beginning of the new century, with the development of cloud computing and big data technology, it has provided super power and massive amount for artificial intelligence data.

In addition, with the introduction of the 2006 deep learning model, the artificial intelligence core algorithm has made major breakthroughs and continuous optimization. At the same time, the development of the mobile Internet and the Internet of Things has provided a rich application scenario for artificial intelligence technology. The combination of algorithms, data and application scenarios has stimulated a new wave of artificial intelligence development, and artificial intelligence technology and industrial development have accelerated.

At present, artificial intelligence is still in the stage of weak artificial intelligence, mainly for specific areas of special intelligence. From the perspective of overall development, artificial intelligence can be divided into three stages: weak artificial intelligence, strong artificial intelligence and super artificial intelligence.

Weak artificial intelligence is good at simulating and extending human intelligence in specific areas, limited rules; strong artificial intelligence is capable of thinking, planning, problem solving, abstract thinking, understanding complex concepts, learning fast and learning from experience, and other human-level intelligent work; super artificial intelligence is a machine that greatly surpasses human intelligence in all fields. Although artificial intelligence has experienced many rounds of development, it is still in the stage of weak artificial intelligence, but it is only specialized intelligence to deal with specific domain problems. There is no consensus in the industry on whether or not it can achieve strong artificial intelligence.

(2) Artificial intelligence security content

Because artificial intelligence can simulate human intelligence and achieve replacement of the human brain, in every wave of artificial intelligence development,

especially when technology is emerging, people pay great attention to the safety and ethical impact of human intelligence. From 1942, Asimov proposed "Three Laws of Robots" to Hawking in 2017, and Musk participated in the "Asiloma Artificial Intelligence 23 Principles". How to promote artificial intelligence to be safer and more ethical has always been a long-term thinking and constant human deepening proposition.

At present, with the rapid development of artificial intelligence technology and industrial explosion, artificial intelligence security has received more and more attention. On the one hand, the immature nature of artificial intelligence technology at this stage leads to security risks, including technical incompatibility, data strong dependence and other technical limitations. And human-induced malicious applications may bring security risks to cyberspace and national society; on the other hand, artificial intelligence technology can be applied to cybersecurity and public security, perception, prediction, early warning information infrastructure and social economic operation, active decision-making response, improve network protection capabilities and social governance capabilities.

Based on the above analysis, the project team believes that the artificial intelligence security content includes: First, reduce the immatureness of artificial intelligence and the security risks brought by malicious applications to cyberspace and national society; second, promote the depth of artificial intelligence in the field of network security and public security. The third is to build an artificial intelligence security management system to ensure the safe and steady development of artificial intelligence.

(3) Artificial intelligence security architecture

Based on the understanding of the security content of artificial intelligence, the project team proposed an artificial intelligence security architecture covering three dimensions of security risk, security application and security management.

Among them, security risk is the negative impact of artificial intelligence technology and industry on cyberspace security and national social security. Security application is to explore the specific application direction of artificial intelligence technology in the field of network information security and social public security. Effectively control the artificial intelligence security risk and actively promote the application of artificial intelligence technology in the security field, and build an artificial intelligence security management system.

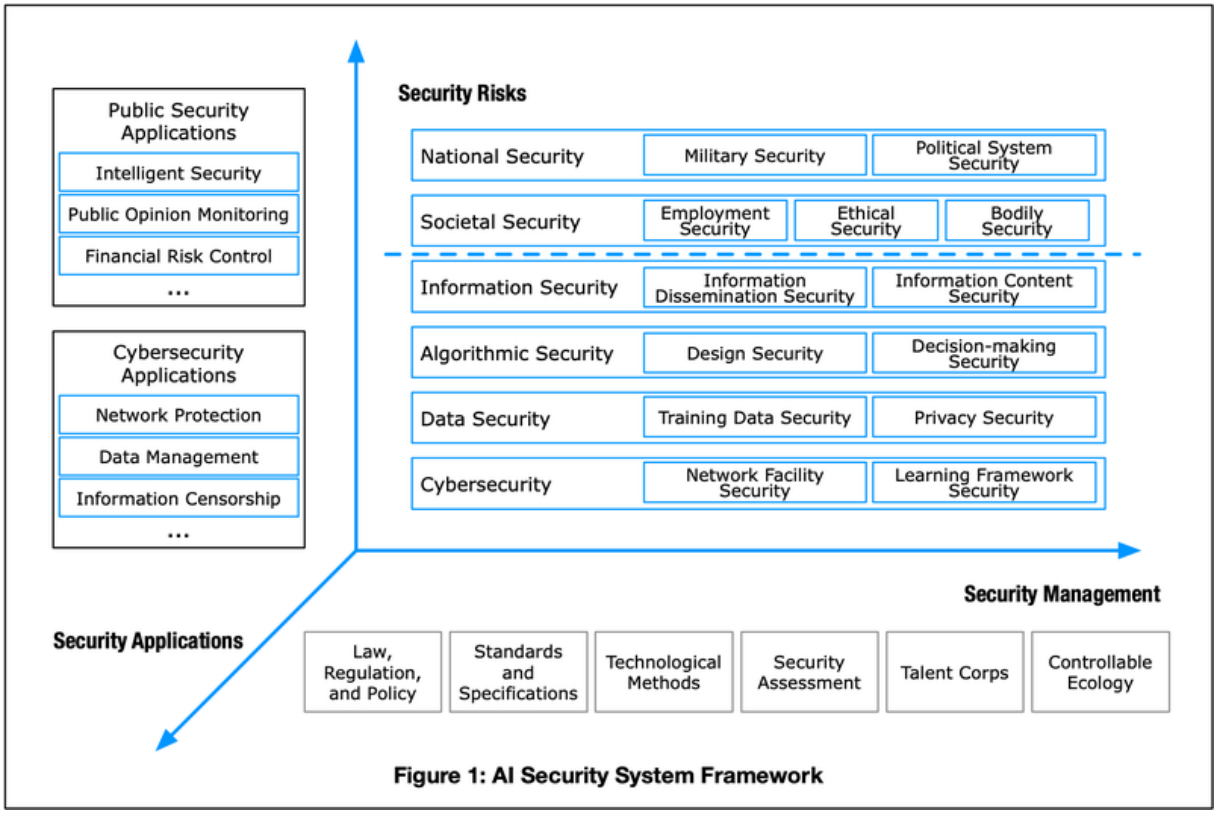


Figure 1: AI Security System Framework

1. Security Risks of AI

As a strategic and transformative information technology, AI has introduced new uncertainties into cyberspace security. AI cyberspace security risks include: cybersecurity risks, data security risks, algorithmic security risks, and information security risks.

Cybersecurity risks involve vulnerabilities in network infrastructure and learning frameworks, backdoor security issues, and systemic cybersecurity risks caused by malicious applications of AI technologies.

Data security risks include training data bias in AI systems, unauthorized tampering, and security risks such as the disclosure of private data caused by AI.

Algorithmic security risks correspond to algorithm design and decision-related security issues in the technical layer, as well as security risks such as black-box algorithms and algorithmic model defects.

Information security risks mainly include AI technology applied to information dissemination and information content security issues for smart products and applications

Considering the deep integration of AI and the real economy, its full security risks in cyberspace will be more directly transmitted to society, the economy, and national

politics. Therefore, from the overall considerations above, AI security risks also involve societal security risks and national security risks.

Societal security risks refer to the structural unemployment brought about by the application of AI and its industrialization, which will seriously affect ethics and morality and may even cause damage to personal safety.

National security risks refer to the risks to national military security and political system security brought about by risks and hidden dangers from the application of AI in military operations, public opinion, and other fields.

2. Security Applications of AI

Because of its outstanding data analysis, knowledge extraction, autonomous learning, intelligent decision-making, automatic control, and other capabilities, AI can have many innovative applications in network information security and societal public security fields including network protection, data management, information censorship, intelligent security, financial risk control, and public opinion monitoring.

Network protection (网络防护) applications refer to the research and development of technologies and products to use AI algorithms for intrusion detection, malware detection, security situational awareness, and threat early warning, etc.

Data management applications refer to the use of AI technologies to achieve data protection objectives such as hierarchical classification, leak prevention, and leak traceability.

Information censorship applications refer to the use of AI technology to assist humans in undertaking rapid review of various forms of expression and a large volume of harmful network content.

Smart security applications refer to the use of AI technology to upgrade the security field from passive defense toward the intelligent direction, developing of active judgment and timely early warning.

Financial risk control applications refer to the use of AI technology to improve the efficiency and accuracy of credit evaluation, risk control, etc., and assisting government departments in the regulation of financial transactions.

Public opinion monitoring applications refer to the use of AI technology to strengthen national online public opinion monitoring capabilities, improve social governance capabilities, and ensure national security.

3. AI Security Management

Combining AI's security risks and its applications in the field of cyberspace security, the project team proposed AI security management ideas that include six aspects: laws, regulations, and policies; standards and specifications; technological methods; security assessments; talent corps; and controllable ecology. Achieve effective control over AI security risks; actively promote the overall objectives for AI technology in the security domain.

With regard to regulations and policies, establish and strengthen corresponding safety management laws and regulations and management policies for key application domains of AI and prominent security risks.

With regard to standards and specifications, complete the formulation of international, domestic, and industry standards for AI security requirements and security assessments and evaluations.

With regard to technological methods, build technological support capabilities for security management, such as AI security risk monitoring and early warning, situational awareness, and emergency response.

With regard to security assessment, accelerate the research and development of indicators, methods, tools, and platforms for the evaluation of AI security assessments, and build third-party security assessment and evaluation capabilities.

With regard to the talent corps, increase the education and training of AI talent, form a stable talent supply and an sufficient talent pool, and promote the secure and sustainable development of AI.

With regard to controllable ecology, strengthen research and inputs at bottlenecks in the AI industrial ecology, enhance the self-guiding capability of the industrial ecology, and guarantee the secure and controllable development of AI

II. Artificial intelligence security risk analysis

(1) Cybersecurity risks

Artificial intelligence learning frameworks and components present a risk of security vulnerabilities that can cause system security issues. At present, the research and development of domestic artificial intelligence products and applications are mainly based on artificial intelligence learning frameworks and components released by technology giants such as Google, Microsoft, Amazon, Facebook, and Baidu.

However, due to the lack of strict test management and security certification for these open source frameworks and components, there may be security risks such as vulnerabilities and backdoors. Once exploited by an attacker, the integrity and availability of artificial intelligence products and applications may be compromised, and may even result in major property damage and adverse social impact. In recent years, the research team of domestic network security companies have repeatedly discovered security vulnerabilities in software frameworks such as TensorFlow, Caffe and their dependent libraries. These vulnerabilities can be exploited by attackers to tamper with or steal artificial intelligence system data and information, resulting in system decision errors and even collapse.

Artificial intelligence technology can enhance network attack capabilities and pose threats and challenges to existing network security protection systems.

First, artificial intelligence technology can improve the efficiency of network attacks. Artificial intelligence technology can dramatically increase the automation of malware writing and distribution. In the past, the creation of malware was largely done manually by cybercriminals, by manually scripting them to make up computer viruses and Trojans, and using rootkits, password grabbers, and other tools to help distribute and execute.

However, artificial intelligence technology can automate these processes by inserting a portion of the antagonistic samples, bypassing the detection of security products, and even implementing the software and automatically changing the code and signature form in each iteration, based on the detection logic of the security product, and automatically modify the code to evade the detection of anti-virus products while ensuring that their functions are not affected.

In March 2017, the first case of machine-based malware creation appeared in a paper report on "Generating Hostile Malware Samples for GAN-Based Black Box Testing," based on Generating Against Network (GAN) algorithms to generate anti-

malware Software samples that bypass the machine learning-based inspection system. In August 2017, security company EndGame released an artificial intelligence program that modifies malware bypass detection, through which a slightly modified malware sample can bypass the security system's defense detection with a 16% probability.

Second, artificial intelligence technology can increase the damage of cyber attacks. Artificial intelligence technology generates intelligent botnets for scalable attacks. In its 2018 Global Threat Forecast, Fortinet said that artificial intelligence technology will be widely used in Hivenet and Swarmbots in the future, using self-learning capabilities to attack vulnerable systems on an unprecedented scale. Unlike traditional botnets, networks and clusters built using artificial intelligence technology can communicate with each other and act on shared local intelligence. Infected devices will also become more intelligent, without having to wait for the botnet controller to issue commands to autonomously execute commands, while automatically attacking multiple targets, and can greatly hinder the attacked target itself. Solution and response measures are implemented. This essentially means that intelligent IoT devices can be controlled to scale and intelligently attack vulnerable systems.

(2) Data security risks

Reverse attacks can lead to data leaks within the algorithm model. Artificial intelligence algorithms can capture and record the details of training data and runtime acquisition data. The reverse attack is to obtain some preliminary information of the system model by using some application programming interfaces (APIs) provided by the machine learning system, and then carry out reverse analysis of the model through the preliminary information to obtain the training data and the data collected at the runtime. For example, Fredrikson et al. can recover a patient's genetic information through a patient's drug dose in a black box-only access to an artificial intelligence algorithm for personal drug dose prediction.[IX]

Fredrikson et al. further developed a reconstruction of a specific facial image in the training data set by using the gradient descent method for the face recognition system[X]. Artificial intelligence technology can enhance data mining and analysis capabilities and increase the risk of privacy leakage. The artificial intelligence system can obtain more information related to user privacy through deep mining analysis based on its collection of seemingly unrelated data segments, identify individual behavior

characteristics and even personality characteristics, and even artificial intelligence systems can pass data.

The re-learning and re-inference leads to invalid security protection measures such as current data anonymization, and personal privacy becomes more easily exploited and exposed. Cambridge Analytics, the protagonist of the Facebook data breach, obtained a wealth of information about US citizen users through association analysis, including skin color, sexual orientation, intelligence, personality traits, religious beliefs, political opinions, and the use of alcohol, tobacco, and drugs. In this way, various political propaganda and illegal profit-making activities are implemented.

(3) Algorithm security risk

Mistakes in algorithm design or implementation can result in inconsistent expectations with noxious results. The design and implementation of the algorithm may not achieve the designer's pre-set goals, causing the decision to deviate from expectations or even result in harmful results.

For example, in March 2018, Uber self-driving cars were not recognized in time by the machine vision system, causing pedestrians to collide with pedestrians, causing death. Google, Stanford University, Berkeley University and OpenAI research institutions were based on mistakes. The stage of the algorithm model design and implementation of security issues are divided into three categories.

The first category is that the designer defines the wrong objective function for the algorithm. For example, the designer does not fully consider the common-sense constraints of the operating environment when designing the objective function, which causes the algorithm to adversely affect the surrounding environment when performing tasks.

The second category is that the designer defines an objective function with a very high computational cost, so that the algorithm cannot be completely executed according to the objective function during the training and use phase. It can only perform some low-cost alternative target function at runtime, thus failing to meet expectations. The effect or adverse effect on the surrounding environment.

The third category is that the selected algorithm model has limited expression ability and cannot fully express the actual situation. As a result, the algorithm may face erroneous results when faced with a new situation different from the training phase.

The algorithm has hidden bias and discrimination, which may lead to unfair decision-making results. Artificial intelligence algorithms have been applied to personalized recommendations, precision advertising, and credit, insurance, financial and other financial areas and judicial risk assessments for risk assessment, which may result in discrimination and bias. For example, when using the crime risk assessment algorithm COMPAS developed by Northpointe, blacks were incorrectly assessed as having a high risk of crime twice as many as whites[XI].

Algorithmic discrimination is mainly caused by two reasons. **First,** the algorithm is essentially "in the form of mathematical or computer code." The design goal of the algorithm, model selection, data usage, etc. are subjective choices of designers and developers. And developers embed their own biases into the algorithm system. **Second**, the data is a social reality response. The training data itself is discriminatory. The algorithm model trained with such data naturally hides discrimination and prejudice.

The algorithm black box causes the artificial intelligence decision to be unexplainable, causing the supervisory review dilema. When the society is functioning and people's lives are increasingly dominated by intelligent decision-making, it is important to supervise and review the decision-making algorithms. However, the "algorithm black box" or algorithm opacity leads to the supervisory review dilemma. The algorithm black box or algorithm opacity is mainly caused by three reasons:

First, the company or individual with the decision algorithm can claim trade secrets or private property on the decision algorithm, and refuse to publicize it.

Second, even if the source code of the decision algorithm is published, the general public cannot understand the inherent logic of the decision algorithm because of insufficient technical capabilities.

Third, because the decision algorithm itself is highly complex, even the programmer who develops it cannot explain the basis and reason for the decision algorithm to make a decision. Therefore, it is very difficult to effectively supervise and review the decision algorithm.

Training data with noise or deviation can affect the accuracy of the algorithm model. At present, artificial intelligence is still in the stage of relying on massive data to drive knowledge learning. The quantity and quality of training data is one of the key factors determining the performance of artificial intelligence algorithm model. Artificial

intelligence algorithms trained on more noisy data and small sample data sets have weaker generalization ability. When faced with new scenes different from training data sets, the accuracy and robustness of the algorithm will be greatly reduced. For example, most mainstream face recognition systems use Caucasian and yellow facial images as training data. The accuracy rate will be greatly reduced when identifying black people. MIT researchers and Microsoft scientists tested the face recognition systems of Microsoft, IBM and Geshi Technology, and found that the error rate for white males was less than 1%, while the error rate for black women was as high as 21%-35%[XII].

Countering the sample attack can induce the algorithm to recognize the misjudgment and the wrong result. Previously, the artificial intelligence algorithm learned only the statistical characteristics of the data or the relationship between the data, but did not really obtain the characteristics or the causal relationship between the data reflecting the nature of the data.

The confrontation attack is the above-mentioned defect of the attacker using the artificial intelligence algorithm model. In the prediction/inference stage, an attack method is prepared for the purpose of running the input data to achieve the purpose of escaping detection and obtaining illegal access rights.

Common anti-sample attacks include two types of escaping attacks and imitative attacks. Escape hacking attacks can generate malicious attacks on the system by generating confrontation samples that can successfully evade detection by security systems, for example, Biggio. The research team used the gradient method to generate optimized escape evasion samples, successfully implementing attacks on spam detection systems and malicious program detection systems in PDF files.[XIII]

Imitating an attack by generating a specific confrontation sample, causing machine learning to erroneously classify a sample that humans seem to have a large gap into a sample that the attacker wants to imitate, thereby achieving the purpose of obtaining the rights of the imitator, and currently mainly appears in the machine-based In the image recognition system and speech recognition system of learning, for example, Nguyen et al. use an improved genetic algorithm to generate an optimal confrontation sample after evolution of multiple categories of pictures, and imitate attacks on Google's AlexNet and the Caffe-based LeNet5 network. Deceive DNN to achieve misclassification.[XIV]

(4) Information security risks

Intelligent recommendation algorithms accelerate the spread of bad information. Personalized intelligent recommendation combines artificial intelligence related algorithms, relies on user browsing records, transaction information and other data to analyze and predict user hobbies and behavior habits, and recommend information content according to user preferences.

At present, personalized intelligent recommendation has become a necessary means to solve the overload of Internet information content. Once intelligent recommendation is used by criminals, it will make the dissemination of false information, such as false information, pornographic and illegal speeches more targeted, reducing the possibility of being reported while expanding negative effects.

Once the smart recommendation is used by criminals, it will make the dissemination of false information, such as false information, pornography, and non-compliance, more targeted and concealed, and reduce the negative impact while reducing the possibility of being reported. McAfee said criminals will increasingly use machine learning to analyze a large number of private records to identify potential vulnerable targets, and to deliver customized phishing emails through intelligent recommendation algorithms to improve the accuracy of social engineering attacks.

Artificial intelligence technology can make false information content for fraudulent activities such as fraud. In the case of sufficient training data, artificial intelligence technology can make artificial recordings that are comparable to the original sound. It can also synthesize images based on text descriptions, or synthesize 3D models based on 2D images, and even modify the expressions of people in the video according to the sound clips. Mouth movement, generating audio and video synthesis content with consistent mouth shape. At present, images, audio and video synthesized by artificial intelligence technology have reached a level of realism, which can be used by criminals to implement fraudulent activities. In 2017, many criminals in Zhejiang, Hubei and other places used voice synthesis technology to pretend to be victims. The case of defrauding by relatives has caused adverse social impact.

In February 2018, the University of Cambridge, UK, "The Malicious Use of Artificial Intelligence: Predictive, Preventive, and Mitigation" research reports that it is possible to synthesize voice and video and multiple rounds of fraud technology in the future, based on artificial intelligence-based precision fraud. It will make people hard to

defend. On May 8, 2018, Google's chat bots at the I/O Developers Conference, when talking to people on the phone, are naturally smooth and organized, and have completely deceived humans.

If the relevant post personnel cannot learn through the new skills and realize the post conversion, they will If the relevant post personnel can't learn through the new skills and realize the job change, it will cause a lot of unemployment, which will form a serious social problem.

(5) Social security risks

The industrialization of artificial intelligence will reduce or even eliminate some existing jobs, leading to structural unemployment. Artificial intelligence is recognized as the core driving force of the fourth industrial revolution[XV]. In its integration with traditional industries, it is no longer limited to replacing human hand and foot and physical strength. It can replace the human brain, from making simple repeated manual activities to even knowledge-based industries such as consulting and analysis may face the threat of layoffs.

According to Forrester Research's forecast, artificial intelligence technology will replace 7% of US jobs by 2025, and 16% of American workers will be replaced by artificial intelligence systems. "A brief history of the future" by Yuval Herali predicts that more than 50% of work in 30 years will be replaced by artificial intelligence. If the relevant post personnel can't learn through the new skills and realize the job change, it will cause a lot of unemployment, which will form a serious social problem.

The security risks of artificial intelligence, especially highly autonomous systems, can endanger personal safety. Traditional information systems are mainly used for personal daily life and office assistance. However, artificial intelligence products and systems such as drones, auto-driving cars, and medical robots can replace human decision-making and behavioral control in their personal lives.

Therefore, artificial intelligence security risks not only cause data leakage caused by traditional information systems, but also affect network connectivity and business continuity, and directly threaten personal safety. The abnormal operation of systems such as autopilot and drones may directly endanger human health and life safety. For example, in May 2016, the Tesla car with automatic driving function could not recognize the white truck under the blue sky background, and the driver died in a car accident in the United

States. At the beginning of 2017, there were many UAV disturbances in China which caused the flight to be forced to land.

Artificial intelligence products and applications will have an impact on the existing social ethics system. First, the decision-making algorithms of intelligent systems will affect social fairness and justice. Intelligent systems are biased or discriminated due to training data or decision-making algorithms, and their decision-making results are bound to affect the fairness and justice of human society. For example, Kronos' artificial intelligence employment assistance system makes ethnic minorities, women or people with a history of mental illness more hard to find a job.

Second, the artificial intelligence application lacks ethical constraints, and the capital-based nature of the capital will lead to the infringement of public rights. The enterprise has a natural capital profitability. When using user data to maximize its own interests, it often ignores moral concepts and damages the user community rights. For example: Ctrip, Didi, etc. based on user behavior data analysis to achieve price discrimination against customers; Facebook uses artificial intelligence to specifically target users to games, addictions and even fake dating sites to get huge profits.

The third is that artificial intelligence will cause humans to rely heavily on it, impacting existing interpersonal concepts. For example, intelligent companion robots rely on personal data analysis to better understand individual psychology, close to the needs of users, and extremely considerate and respectful to human beings. This will allow humans to abandon normal heterosexual interactions and seriously impact traditional family concepts.

Fourth, property damage caused by artificial intelligence products and system security incidents, personal injury, etc. are facing the dilemma of irresponsibility. Artificial intelligence systems may produce unpredictable results in human-machine coordination, resulting in property damage or personal injury. Products and applications themselves do not have the ability to assume responsibility and legal subject qualifications, and there are unexplained links in the traceback of problems, which brings serious challenges to the existing legal system and ethical order.

(6) National security risks

Artificial intelligence can be used to influence public political ideology and indirectly threaten national security. Cambridge Analytical Inc., which was deeply

involved in the Facebook data breach scandal this year, was deeply involved in the 2016 US elections by several media reports. The company mainly uses artificial intelligence technology to support the advertising orientation algorithm, behavior analysis algorithm and data mining analysis technology to support the psychological analysis. The predictive model assists in the "election strategy" to help politicians determine the different types of voters' positions on specific issues and guide their language and intonation in campaign advertisements. Albright, a data scientist at Elon University in the United States, pointed out that the use of behavior tracking and identification technology to collect massive data, identify potential voters, and conduct peer-to-peer push of false news can effectively influence the results of the US election.

Artificial intelligence can be used to build new military strike forces that directly threaten national security. The application of weapons can remotely control future warfare, combat precision, miniaturization of the war domain, and process intelligence. At present, major countries have made artificial intelligence an important military change that affects the future world pattern, from strategy to organizational structure. Applying equal angles to increase the investment of artificial intelligence in the military field, or lead to a new round of arms race. For example, the US Department of Defense explicitly uses artificial intelligence as an important technical pillar for the third "offset strategy." The Russian military began to install a large number of robots in 2017. It is planned that by 2025, the proportion of unmanned systems in the Russian military equipment structure will be up to 30%[XVI]. In addition, with the rapid development of artificial intelligence, the price of smart products will fall, and access will be easier. Terrorists will increasingly use artificial intelligence weapons. For example, on August 4, 2018, the Venezuelan president was exposed to na attempt of assassination with a drone bomb attack.

In view of the development status of artificial intelligence, the artificial intelligence security risk is combed and analyzed. Overall, from the perspective of risk, the security risk brought by artificial intelligence is caused by its own **technology immatureness** and **technical malicious application.** Although artificial intelligence security risks exist in many fields of cyberspace and national society, some security issues are still in the forward-looking and initiating stage, and have not really penetrated into the industrial ecology.

At present, the development of artificial intelligence technology is accelerating due to its own interdisciplinary and vertical applicability. It will surely integrate deeply with the traditional industry. With the innovation breakthrough of artificial intelligence technology and the increasing application scenarios, its security risks will also evolve dynamically, and it will become more and more ubiquitous, scene-oriented, and integrated. Production and life, the country's political economy and other aspects have a profound security shadow

III. Artificial Intelligence Security Application

At present, artificial intelligence technology can grasp the critical situation of key information infrastructure and economic and social security operation because of its ability to perceive, predict, and timely grasp group cognition and psychological changes. Active decision-making has an irreplaceable role in safeguarding cyberspace security and effectively maintaining social stability.

Therefore, the application of artificial intelligence in the security field is the focus of current domestic and foreign enterprise technology and application innovation. Combining the practice of artificial intelligence security application, the artificial intelligence-based network information security application innovation is active. At the same time, the integration of artificial intelligence and traditional social public security also promotes the development of security monitoring, financial risk control, and public opinion monitoring.

(1) Network information security application

1. Network security application

The application of network security protection based on artificial intelligence has become the focus of the development of network security industry at home and abroad. As network security evolves toward dynamic defense and active defense, artificial intelligence is an important engine for promoting network security technology innovation with its rapid recognition, response and self-learning potential for network security threats. To a certain extent, the application of artificial intelligence technology has improved the automation and intelligence level of network protection, reduced the workload of network intelligence analysts, and made up for the shortage of network security talents.

From the application scope, the application scenarios of artificial intelligence in network security are increasingly widespread. At present, artificial intelligence has been widely applied from early malware monitoring to intrusion detection, situation analysis, cloud defense, anti-fraud, IoT security, mobile terminal security, security operation and many other fields. For example, in intrusion detection, Israel Hexadite uses artificial intelligence to automatically analyze threats, quickly identify and resolve cyber attacks, and help internal security teams manage and prioritize potential threats. Hillstone Networks develops intelligent firewalls that can be used to help customers discover unknown network threats based on behavioral analysis techniques, providing protection and detection throughout the attack; in terms of terminal security, the United States CrowdStrike's terminal active defense platform based on big data analysis can identify unknown malware of mobile terminals, monitor enterprise data, detect zero-day threats, and then form a set of rapid response measures to increase the risk and cost of hacker attacks. In terms of operation and maintenance, Jask USA uses artificial intelligence algorithms to prioritize and analyze data such as logs and events to help security analysts discover offensive threats in the network and improve the operational efficiency of security operations centers.

From the application depth, the application level of artificial intelligence in network security is still in the early stage of accumulation. In addition to improving the performance of some network security products, the innovation of network security protection system based on artificial intelligence technology is still in the research and practice stage. At present, foreign security companies started earlier. For example, the British DarkTrace company based on Cambridge University's machine learning and artificial intelligence algorithms to bionic the human immune system, is committed to the network to automatically defend against potential threats, and can help enterprises quickly identify and respond to artificially manufactured networks. Attacks can also prevent network attacks based on machine learning. In contrast, the overall solution for network security protection based on artificial intelligence technology in China is still in the research stage, and it is still necessary to explore the innovation and optimization of the overall network security protection system and architecture using artificial intelligence technology.

2. Information content security review application

The application of information security review based on artificial intelligence has entered the initial stage of large-scale application. In recent years, the application of text, image and video recognition based on artificial intelligence technology has become more and more mature, and the global information content security management is increasingly strengthened under the two-wheel drive, and the global information content security management is increasingly strengthening the two-wheel drive to the illegal information. The security review of information content has become the frontier of artificial intelligence in the security field. US Internet giant Facebook not only uses artificial intelligence technology to mark Internet content, but also uses machine learning to develop a tool for real-time monitoring and identification of users' live video content, automatically recording videos in the live broadcast of pornographic, violent or suicide categories.

However, from the perspective of effectiveness, the principle of judging illegal content is still relatively simple, and there are more false positives. For example, the identification of pornographic content is mainly judged by bare skin, which makes some historical and artistic pictures misjudged. Compared with foreign companies, domestic Internet companies started earlier in the automation technology research and industrialization application of information content security review, especially the large Internet companies represented by Ali, Tencent, Baidu and Netease, through the process of security management based on their own business. The accumulated large-scale standard sample database, carrying out modeling training on obscenity pornography, terrorism-related violence and other illegal information identification, has launched an artificial intelligence-based illegal information detection service. According to relevant enterprise research and feedback, the accuracy of using artificial intelligence to identify illegal information of pictures and videos is 99%, and the recognition accuracy of speech and text is as high as 90%.

3. Data security management application

Data security management applications based on artificial intelligence are still in the initial stage of exploration. As data is a key element in the development of the digital economy, the value of data continues to increase, and the importance of data security protection is further highlighted. In addition to external threats such as cyberattacks through artificial intelligence to avoid data breaches, domestic and foreign

companies are actively exploring data security internal threat prevention based on artificial intelligence. At present, the application of artificial intelligence technology in data classification, data leakage prevention and other fields has achieved initial results. Neokami in Germany uses artificial intelligence technology to help customers protect sensitive data on the cloud, local or physical assets. The company's data classification engine is suitable for a variety of business scenarios and has been adopted by many partner vendors. It has creaed value in the top 500 companies. Sky Security, a domestic security company, effectively integrates data within the data through comprehensive use of statistical anomaly analysis, bi-directional cyclic neural networks and other artificial intelligence Technologies, such as security and behavioral identification technologies that enable security alerts and real-time control of people and devices that target high data security risks within the enterprise.

(2) Social public safety application

Intelligent security based on artificial intelligence technology presents a good momentum of rapid global development. The traditional security industry mainly solves the technical problems of optical device resolution video data storage. There are many limitations in the development. For example, traditional security is mostly passive application, used for after-the-fact forensics, and has little effect on pre-emptive action. Video data mining depth is not enough, can not be effectively used, etc. Different from traditional security, intelligent security based on artificial intelligence relies on the learning of massive video data, which can complete the inference and prediction of behavior patterns. It has been developed from passive defense to active judgment and intelligent warning of timely warning. It has been applied to human face. In the system of identification, vehicle identification, etc., the target attribute extraction is performed to achieve intelligent detection, tracking and troubleshooting of the target. In recent years, the intelligent security industry has maintained rapid growth and has become one of the best industries for the application of human intelligence. It is estimated that by 2020, the global industrial scale of intelligent security will reach US$10.6 billion[XVII].

Foreign chip giants seized the opportunity of industry development and stepped up the upstream deployment of the intelligent security industry chain. US chip giant Intel acquired Movidius, a leading-edge computer vision company, in 2016. Later, it launched a number of embedded independent neural computing engines, visual

computing chips supporting edge deep learning inference, and neural computing SDK development kits. Design, providing personalized solutions for major security companies worldwide. In recent years, Intel has further focused on intelligent security as a major business growth point. With the company's advantages in high-end chips, it has actively studied image processing, intelligent analysis, cloud storage and other related technologies, and constantly improved the layout of the upstream links of the intelligent security industry chain.

There is huge room for development of domestic intelligent security industry. With the continuous development of China's safe city, Skynet project, and Xueliang project construction, the security industry has developed rapidly. During the "Thirteenth Five-Year Plan" period, the security industry is gradually transforming into a large-scale, automated, and intelligent transformation. It is expected that by 2020, the total income of security enterprises will reach 800 billion yuan, and the annual growth rate will reach more than 10%.[XVIII] Corresponding domestic intelligent security has entered a period of rapid development since 2016, but it is limited by the high price of intelligent products and the limited application of scenarios. Most security enterprises are still trying to use artificial intelligence. More than 90% of the market share is still occupied by traditional security [XIX], but with the social governance field represented by public security, transportation, and finance, it further drives the rapid application of smart security, and the future market development space is huge.

However, it must be pointed out that although the innovation capability of the domestic intelligent security industry continues to strengthen, **it still needs to make progress towards the upstream of the industrial chain.** At present, the domestic security market competition pattern is dominated by artificial intelligence innovative enterprises and traditional security giants. Among them, artificial intelligence-based startups such as Yun Cong Technology, Shang Tang Technology and Defiance Technology rely on computer vision, data depth analysis and other aspects of technology accumulation, the introduction of smart security products, industrial layout; traditional security giant Hikvision, Dahua shares and other companies in recent years continue to increase investment in research and development, strengthen technological innovation capabilities, and investment acquisitions for start-ups, and gradually improve the intelligence level of security products. The two types of enterprises drive the continuous

and rapid development of the domestic intelligent security market in competition and cooperation.

However, the domestic intelligent security industry is subject to people in the upstream chain of processing chip, sensor equipment and other industrial chains. It is urgent to take advantage of the opportunities of large-scale development of the industry, to use applications as traction, to increase industry integration, to achieve key technological breakthroughs, and to strive to the industry.

2. Financial risk control application

Artificial intelligence technology can be used to improve the efficiency and accuracy of financial risk control work. Traditional financial risk control is often based on a scorecard system, modeling bank borrowing records, etc. The development of finance + Internet has made financial services cover more income groups. The explicit credit data of newly added groups is often lacking. Financial institutions have to use more transaction data, operator data, Internet behavior data and other transaction information for analysis. Although this underlying data change has caused great difficulties for traditional credit scoring, it has provided artificial intelligence technology. For example, in the face of massive heterogeneous data, the deep learning-based feature generation framework has been matured and applied to the wind. In the control scene, deep feature processing extraction is performed on unstructured data such as text, pictures, images, etc., showing an improvement over the imagination of the model. At the same time, human intelligence can greatly improve the decision-making efficiency based on big data analysis, and can remove the subjective view in human judgment, making decision judgment more accurate.

Foreign development is relatively mature and has been applied to financial transaction supervision. US Neurensic companies use artificial intelligence to monitor electronic transactions, identify behaviors that pose a risk to trading companies based on machine learning, and automatically detect high-risk activities from actual regulatory cases. At the end of 2016, Nasdaq and the London Stock Exchange launched artificial intelligence into the market; in the first half of 2017, two exchanges on Wall Street launched an intelligent supervision system. Intelligent supervision can effectively reduce supervision, cost and transaction risk, improve regulatory efficiency, and have broad development space in the future.

The country is in its infancy and still needs long-term market validation. Domestic financial 360, good loan network, creditworthy and other financial companies rely on the influence of the enterprise market, product reputation evaluation, etc. Data collection and effective screening, relying on artificial intelligence technology to achieve quantitative modeling of historical business data and real-time market information, and thus achieve a predictive assessment of various asset risks. However, the market exploration and technological innovation faced by domestic smart investment are still at the stage of exploration. The relevant regulatory and policy environment is still undergoing continuous improvement. The application of artificial wind-based financial risk control is still in its infancy, and its subsequent development is still to be observed.

In addition, artificial intelligence technology has been applied to the field of network public opinion monitoring and analysis at home and abroad. After the 9/11 incident in the United States, the signing of the Patriot Act prompted the legalization of public opinion monitoring in the United States, allowing the government to monitor all communications of potential terrorists, including email, etc[XX]. Artificial intelligence can predict the development direction of network events, strengthen the early warning capability of event evolution, take sensible intervention and guidance beforehand, avoid the occurrence of group public opinion events, and improve social governance capabilities. At present, the main domestic network public opinion monitoring systems try to analyze the original big data on the basis of natural language processing, machine reading comprehension and other related technologies to improve the level of intelligence of the system.

IV. The Status Quo of Artificial Intelligence Security Management

(1) Major national artificial intelligence security concerns

As a strategic technology that leads the future, artificial intelligence has become a new focus of international competition. The major countries in the world regard the development of artificial intelligence as a major strategy to enhance national competitiveness and safeguard national security, and intensify the introduction of planning and policies in an effort to gain ownership in the new round of international science and technology competition. The major countries of the world focus on artificial intelligence security based on their international status and development strategies.

1. United States: Focus on the impact of artificial intelligence technology on national security

With its strengths in talent pool, financial system, IT technology and the Internet, the United States actively seeks global leadership in the field of artificial intelligence, focusing on the impact of artificial intelligence technology on its international leading position. In July 2017, the Kennedy School of Political Science at Harvard University published the "Artificial Intelligence and National Security" report, which summarized the major impacts and response experience of disruptive technologies on national security in the past, and proposed to maintain artificial intelligence technology in the national security field and policy recommendations for effective risk management. On March 20, 2018, the United States National Assembly initiated a proposal to establish the National Artificial Intelligence Security Council and to develop the National Security Council Artificial Intelligence Act of 2018.

2. EU and UK: Focus on the impact of artificial intelligence on privacy, employment and ethics

At the end of 2016, the UK released "Artificial Intelligence: Opportunities and Impacts of Future Decision Making". The report focuses on the impact of artificial intelligence on personal privacy, employment, and government decision-making, and suggests ways to address the ethical and legal risks of artificial intelligence. On March 27, 2018, the European Political Strategy Center released the "Artificial Intelligence Age: Establishing a People-Oriented European Strategy", which reported that workers who might be encountered in the process of artificial intellectual development were replaced. It assesses the problem of artificial intelligence bias and put forward the corresponding strategy that the EU should adopt.

3. Russia, Israel, India: Focus on artificial intelligence in the field of defense applications and impact on military security

Although Russia's achievements in the field of artificial intelligence are lagging behind other science and technology powers, with the guidance of its defense needs and the full support of domestic industry, it has achieved many results on the military unmanned platform.

For example, Uran-6 demining robots that have long been used in Syria to clear ISIS traps and explosive devices left in Pamir, Uran-9 multi-robot robots for long-range reconnaissance and fire support, and deployed in the Russian Pacific Fleet Platforma-M detects unmanned vehicles, etc.

After the new century, Israel has developed artificial intelligence related research and development in the fields of network security, communications, password warfare, unmanned traffic, military machinery manufacturing, and aerospace. For example, fully autonomously piloted military vehicles used for border patrols as early as 2016; artificial intelligence algorithms that understand and describe video have been used for battlefield and borderline monitoring; soldier smart bracelets allow commanders to accurately understand battlefield situations, And use artificial intelligence technology to analyze battlefield return data to make informed decisions.

The Indian government announced on May 22 this year that it will use artificial intelligence technology to develop weapons, defense and surveillance systems. As early as April this year, Indian Prime Minister Modi publicly stated that artificial intelligence and robots will become the most important determinants of future military power. India will strive to use artificial intelligence technology to enhance its operational capabilities. At present, the Indian military is developing a road map for artificial intelligence. In the next two years, it will study machine learning in the fields of air force, navy, army, network security, nuclear, biological resources, etc., involving autonomous weapons and unmanned surveillance systems.

4. Canada, Japan, Korea, Singapore: Focus on artificial intelligence talent development, technology research and development, and industrial promotion, etc

The Government of Canada launched the Pan-Canada Artificial Intelligence Strategy in March 2017, which will increase the number of excellent artificial intelligence researchers and technology graduates in Canada as one of the strategic goals. At the same time, Canada has a top-level research team with research centers such as the University of Toronto and the University of Montreal. It attracts a large number of artificial intelligence talents. The government has strong financial support for artificial intelligence, the overall quality of education is high, and the industrial system is mature.

The Japanese government promulgated the "Scenario 5 Science and Technology Basic Plan" in January 2016.It is proposed to establish a super-smart society with artificial

intelligence as the core. n March 2017, the Japanese government formulated a roadmap for the development of artificial intelligence, which clarified the process of artificial intelligence development in Japan with 2020 and 2030 as the time limit. The plan is promoted in three phases. The use of human intelligence to significantly improve the efficiency of the manufacturing, logistics, medical and nursing industries.

South Korea also has a solid foundation in artificial intelligence technology. According to data from the Korea Institute of Information and Communication Technology, South Korea ranked third in the world for artificial intelligence-related patents between January 2005 and the third quarter of 2017, Second to the United States and Japan. In August 2016, the Korean government proposed nine national strategic projects led by artificial intelligence as a new engine to explore new economic growth drivers and improve the quality of national life. In addition, the Korean government has also set a roadmap for the development of artificial intelligence stage goal.

In May 2017, Singapore released the "Singapore Artificial Intelligence Strategy", which plans to invest US$150 million over the next five years to enhance the strength of Singapore's artificial intelligence technology. Singapore combines the goal of "smart country" construction and vigorously promotes the industrial application of artificial intelligence technology. At present, Singapore has adopted artificial intelligence technology very widely. At least one-sixth of organizations use artificial intelligence in various fields, including 60% in information technology, 48% in supply chain and logistics, 49% in customer support, and R&D department. 41%[XXI].

(2) Major national artificial intelligence security regulations and policy development

At present, the artificial intelligence safety management systems of major countries in the world are in the early stage of construction, mainly reflected in strategic planning and reports in the form of ideas and suggestions, and relatively few laws, regulations and management policies are implemented. Some countries have piloted the requirements of artificial intelligence application specifications in the pioneering field of artificial intelligence industry promotion, and tried to carry out security management to constrain the security risks that artificial intelligence may bring.

1. **United States: advocates relying on market forces to reduce algorithmic decision-making risks and strengthen supervision in the field of technology application pilots**

First, advocate the role of the market to reduce the risk of algorithmic decision-making. In the development of artificial intelligence technology and industry, the United States encourages enterprises to increase innovation, and advocates relying on market forces to reduce artificial intelligence security risks. The United States' authoritative science and technology innovation think tank data innovation center released in May 2018 "How policy makers promote algorithmic accountability" report, proposed an algorithmic accountability framework, designed to control algorithmic risk without sacrificing the development of digital economy. The report points out that in most cases, market forces have the ability to prevent most of the flawed artificial intelligence algorithms from being generated, and the regulators do not need to intervene in the algorithm; only when the application of the algorithm is potentially harmful and sufficient to use regulatory review, will trigger the algorithm accountability.

The second is to strengthen supervision of pilot areas such as autonomous driving and criminal justice. In September 2017, the US House of Representatives passed the Automated Driving Act, which proposed 12 safety requirements for autonomous vehicles including system safety, network security, interpersonal interaction, and collision avoidance. In December 2017, the City of New York adopted the "Local Law on the Use of Automated Decision Systems by Government Agencies" to regulate the artificial intelligence automated decision-making system used by government agencies such as courts and police.

2. EU and the United Kingdom: Strengthening the construction of ethical principles led by the government and the constraints of laws and regulations

First, strengthening the construction of basic ethical principles of artificial intelligence. The European Commission for Science and New Technology under the European Commission issued a system statement on artificial intelligence, robotics and "autonomous" in March 2018. In the report, a set of provisions based on the EU Treaty and the EU Charter of Fundamental Rights proposed the basic ethical principles of artificial intelligence values. The principle covers "protection of human dignity", "security, reliability", "responsibility", "sustainability" and so on. On April 16, 2018, the British Parliament issued the "British Artificial Intelligence Development Plan, Capabilities and Aspirations", which proposed that "Artificial Intelligence should serve the common interests and well-being of mankind". "Artificial intelligence should follow

the principles of intelligibility and fairness", "Artificial intelligence should not be used to weaken the data rights or privacy of individuals, families and communities", etc.

The second is to try to establish an artificial intelligence automatic decision-making application specification. The EU's "General Data Protection Regulations", which came into effect in May 2018, set extremely stringent conditions for the legal application of artificial intelligence automated decision-making: with the explicit consent of the user, or between the user and the data controller, the execution of the contract is necessary, or explicitly authorized by the EU or Member State Act; artificial intelligence automated decision-making applications that do not meet the above conditions without human intervention and have legal or similar effects on individuals will be banned. At the same time, the EU General Data Protection Regulations explicitly require data controllers to inform the data subject of the following information when collecting data. The existence of artificial intelligence to automate decision making, meaningful information about the internal logic of automated decision making, the significance of the data subject and the consequences of the assumptions", and encourage data controllers to explain the specific reasons for an artificial intelligence automated decision to the data subject.

3. China: Adhere to the planning and application of norms, explore the construction of artificial intelligence security management system.

Our government takes strategic planning as a guide to increase policy guidance on artificial intelligence security. In July 2017, the State Council issued the "New Generation Artificial Intelligence Development Plan": the development and application of large artificial intelligence will maximize the potential of artificial intelligence; it must also predict the challenges of artificial intelligence, coordinate industrial policies, innovate policies and social policies, achieve coordination between incentive development and rational regulation, and minimize risks. n December of the same year, the Ministry of Industry and Information Technology issued the "Three-Year Action Plan for Promoting the Development of a New Generation of Artificial Intelligence Industry (2018-2020)": To improve the development environment, improve safety and security capabilities, and achieve healthy and orderly development of the industry; Establish an artificial intelligence network security system. Subsequently, Shanghai, Beijing, Zhejiang, Anhui, Guizhou, Jiangxi and other provinces and cities have combined their own industrial development and comparative advantages, issued relevant

implementation opinions for artificial intelligence, and simultaneously proposed to strengthen the key technologies and network security architecture of artificial intelligence information security, such as scientific research, strengthen data security and privacy protection and other key tasks.

At the same time, **around the leading field of artificial intelligence applications, relevant management departments have stepped up the introduction of normative guidance documents**. For example, in the financial sector, on April 28, 2018, the People's Bank of China, the China Banking Insurance Regulatory Commission, the China Securities Regulatory Commission, and the State Administration of Foreign Exchange jointly issued the "Guiding Opinions on Regulating the Asset Management Business of Financial Institutions". Financial institutions use artificial intelligence technology. The use of robotic investment consultants to carry out asset management business should be approved by the financial supervision and management department, obtain the corresponding investment consultant qualifications, fully disclose information, report the main parameters of the intelligent investment model and the main logic of asset allocation and to actively guard against the application of artificial intelligence to the security risks brought by financial investment.

On the whole, China's government departments are more open to artificial intelligence technology and industrial development policies, to promote incentives, and actively use artificial intelligence to develop advantages, and strive to become a new round, technical and industrial change leader. However, how to effectively control The artificial intelligence security risk is still at the stage of exploration. In the future, it is necessary to further improve the forward-looking research and strategic layout of relevant security management work, and promote the artificial intelligence security management work with pragmatic and prudent attitude.

(3) Development of artificial intelligence safety standard specifications at home and abroad

At present, the existing standards of artificial intelligence in the world are mainly the common standards in the application field. The safety-related standards related to artificial intelligence security, ethics, privacy protection, etc., are still mostly in the research stage. The IEEE is developing an artificial intelligence ethics standard that regulates artificial intelligence security design. The IEEE Standards Association is

working to ensure that designers of artificial intelligence technologies prioritize ethics in their work. In March 2017, IEEE published the IEEE Global Initiative titled "Ethical Design for Advancing Artificial Intelligence and Autonomous Systems" in the IEEE Robotics and Automation magazine. The book "Initiatives" help people avoid the fear of artificial intelligence technology and blind worship through ethical design principles and standards, thus promoting the innovation of artificial intelligence technology. Currently, the IEEE working group is developing an ethical standard for ethics in the IEEE P7000 series, which is responsible for the ethical issues in system design, the transparency of autonomous systems, the ethical problem of system/software collection of personal information, and the negative bias of data, children and students data security, Artificial intelligence agente specification.

ISO/IEC established the Artificial Intelligence Trusted Research Group to conduct research on artificial intelligence security standards. ISO/IEC JTC 1/SC 42 The Artificial Intelligence Subcommittee was established in October 2017, and its scope of work is standardization in the field of artificial intelligence. The second research group of ISO/IEC JTC 1/SC 42 is a credible research group whose research scope includes: Methods for establishing trust in artificial intelligence systems through investigations such as transparency, verifiability, interpretability, and controllability; investigate engineering pitfalls and use their mitigation techniques and methods to assess typical associated threats and risks of artificial intelligence systems; investigate research to achieve robustness, resilience, reliability, accuracy, safety of human health and production technology activities, social politics, privacy and other methods of performance; investigate the source types of bias in artificial intelligence systems, with a minimum of targeting, including but not limited to statistical bias in artificial intelligence systems and artificial intelligence-assisted decision making.

China established the National Artificial Intelligence Standardization Group and the Expert Advisory Group to strengthen the development of artificial intelligence security standards. At present, China's artificial intelligence security standards mainly focus on the safety standards of a small number of application areas such as biometrics, intelligent networked cars, and supporting security standards such as big data security and privacy protection, and are related to the security or basic commonality of artificial intelligence. There are few standards such as safety reference

architecture, safety assessment, ethical design, safety requirements and assessment methods. In order to implement the task deployment of the "New Generation Artificial Intelligence Development Plan", strengthen the overall coordination and systematic research of the standardization work in the field of artificial intelligence, play a supporting role of standardization, and lead the role, under the guidance of relevant ministries and commissions, China Artificial Intelligence Industry Development Alliance, the country Industrial intelligence standardization group and expert consulting group and other industry organizations have been established to strengthen the research on the basic standards of artificial intelligence security and continue to deepen the security standardization work in the application field. China Communications Standards Association (CCSA) is developing the "Guidelines for the Safety Assessment of Artificial Intelligence Products, Applications and Services" and the "Safety Assessment Requirements for Artificial Intelligence Service Platforms".

(4) Construction of artificial intelligence security technology at home and abroad

In order to quickly and timely avoid and prevent artificial intelligence security risks, the security technology is an essential component of the overall management system, including safety supervision methods and safety protection measures, etc. Relying on safety supervision technology, it can timely discover safety problems in artificial intelligence products and applications, and take emergency measures to reduce the impact of safety problems. Relying on safety protection technology, it can enhance safety protection capability and improve the safety and reliability of artificial intelligence products and applications.

Both domestic and foreign governments attach great importance to the construction of artificial intelligence security supervision technology, and relevant ideas are reflected in the planning report. In October 2016, the UK House of Commons Science and Technology Committee published the "Robot Technology and Artificial Intelligence" report, calling for the government to regulate artificial intelligence. The British government tried to verify and confirm, the transparency of the decision system, the most biased minimization, privacy and right to know, accountability system and responsibility commitment, strengthen the control of artificial intelligence security.

China's "New Generation Artificial Intelligence Development Plan" proposes: Establishing and improving an open and transparent artificial intelligence supervision

system, and implementing a two-tier supervision structure with equal emphasis on design accountability and application supervision; implementing artificial intelligence algorithm design, product development and results, full process supervision of applications, etc. At present, the construction of artificial intelligence safety supervision technology means mainly rely on the enterprise itself. Since artificial intelligence is still in the early stage of industrial application, governments have adopted incentive policies that promote development, and have not yet formed a complete artificial intelligence safety supervision system. The government mainly strengthens the artificial intelligence safety supervision by regulating the enterprise and self-discipline of the industry. The relevant technical means mainly rely on the enterprise itself to carry out construction.

Artificial intelligence enterprises and network security enterprises are paying more and more attention to the construction of artificial intelligence security protection measures. With the continuous development of artificial intelligence technology and the continuous advancement of applications, artificial intelligence products and systems are gradually moving from laboratory to practical application. Artificial intelligence enterprises are committed to enhancing the maturity and reliability of their products and actively improving the safety protection capabilities of artificial intelligence. For example, Baidu attaches great importance to the safety of autonomous driving systems. The Apollo platform provides a complete security framework and system components based on isolation and trusted security systems, protecting against network intrusion, protecting user privacy and car information security. Security companies have increased security research on artificial intelligence software and hardware platforms, data security and algorithm design. 360 Security Research Institute has repeatedly discovered security vulnerabilities in artificial intelligence technology architecture, and researched and enhanced artificial intelligence security protection capabilities.

(5) Safety assessment of key applications of artificial intelligence at home and abroad

The evaluation of safety assessments for artificial intelligence products and applications covers a wide range of applications in various fields. Different safety assessment indicators, methods and requirements, artificial intelligence safety assessment work is still in the research and exploration stage, and it has not been fully developed. At present, the artificial intelligence industry in various countries mainly focuses on the

safety assessment of human intelligence pilot applications, with emphasis on autonomous driving, intelligent service robots, etc.

Safety test verification for autonomous driving is highly valued by countries but does not form uniform security standards and evaluation system. On September 20, 2016, the US Department of Transportation promulgated the "Federal Autopilot Vehicle Policy", the world's first proposed system for autonomous driving safety regulatory and safety assessment requirements, emphasizing safety as the first criterion, requiring that technological innovation must be Security features are guaranteed. By modifying the current road traffic laws and establishing automatic driving ethics standards, Germany clarifies that safety is the premise of automatic driving. It is stipulated that autonomous vehicles should install "black box" to record driving activities and ensure data security. On April 12, 2018, China's Ministry of Industry and Information Technology, Ministry of Public Security, and the Ministry of Communications jointly issued the "Intelligent Network Linked Vehicle Road Test Management Specification (Trial)", the main body of the intelligent networked vehicle on-road test, testing drivers and test vehicles, testing applications And regulations, test management, traffic violations and accident handling are clearly defined.

However, at present, countries lack safety-related safety standards. The safety assessment and evaluation work is mainly carried out by the enterprises themselves. The national policy documents are more restrictive of corporate behavior, and no third-party testing and certification institutions have been formed, and it is impossible to unify the safety evaluation of autonomous driving, so the future may restrict the work of the development of the industry.

The safety standards related to industrial robots are relatively complete, but the intelligent service robot safety standard system and evaluation capabilities have yet to be improved. Among the major international standards organizations, ISO robot testing standards fall into two categories: industrial robots and service robots. The TC299 undertakes robot standardization work. The focus is on safety and performance testing standards. Among them, ISO/TC299/WG2 completes the service robot field. The first safety standard - ISO13482: Among the major international standards organizations, ISO robot testing standards fall into two categories: industrial robots and service robots. The TC299 undertakes robot standardization work. The focus is on safety and performance

testing standards. Among them, ISO/TC299/WG2 completes the service robot field. The first safety standard - ISO13482: 2014 "Safety requirements for personal care robots"; ISO/TC299/WG3 focuses on industrial safety, the current standards for industrial robots have been relatively complete, the standards for ontology safety and integrated safety have been revised. In addition, IEC standardization work is mainly undertaken by TC59, TC61, TC62, TC116 technical committees. The standards are mainly related to the safety and performance of home service robots, the functional safety of industrial robots and the safety of medical robots. In March 2015, the National Development and Reform Commission, the Ministry of Industry and Information Technology, the National Standards Commission, the Certification and Accreditation Administration and other departments jointly guided the establishment of the National Robot Testing and Evaluation Center, and gradually carried out standard revision, testing services, certification services, etc. In January 2017, under the guidance of the above four ministries, the Robot Testing and Certification Alliance issued the "Household/Commercial Service Robot Safety and EMC Certification Implementation Rules", focusing on product safety and electromagnetic compatibility. At home and abroad, some progress has been made in the construction of intelligent robot safety assessment capabilities, more safety tests are carried out on the mechanical structure, electrical characteristics, system functions and other indicators of robots, and information data and intelligent algorithms related to artificial intelligence are Decision-making models and other security assessment evaluation capabilities are insufficient.

(6) Construction of artificial intelligence talents at home and abroad

Talent is the cornerstone of technology and industry development, and is the prerequisite for sustainable and healthy development of the industry. It is possible to ensure the safe and healthy development of artificial intelligence in China by increasing the education and training of artificial intelligence talents and forming a stable talent supply and a reasonable talent team.

China attaches great importance to the cultivation of artificial intelligence talents and achieves good results. In recent years, China's colleges and universities have actively laid out the construction of artificial intelligence-related disciplines and intensified personnel training. As of December 2017, a total of 71 universities across the

country have set up 86 secondary disciplines or interdisciplinary subjects around the field of artificial intelligence.

In April 2018, in order to implement the requirements of the "New Generation Artificial Intelligence Development Plan", the Ministry of Education Taiwan's "Artificial Intelligence Innovation Action Plan for Colleges and Universities", actively promoted artificial intelligence innovation initiatives and disciplines, increased basic research efforts, and strengthened the construction of talent teams. Later, Tsinghua University, Nankai and other 985 universities established artificial intelligence research institutes. In July 2018, 26 universities including Tsinghua University, Nanda University and Xijiao University jointly signed the "Proposal on Setting up Artificial Intelligence Professionals". After the construction of the early talent team, the total number of artificial intellectual papers and the number of highly cited papers in China exceeded the United States in 2013, ranking first in the world. China has surpassed the United States and Japan in the number of artificial intelligence patents, making it the country with the largest number of artificial intelligence patents in the world[XXII].

There is still a big gap between the artificial intelligence talents in China and the United States and Britain. Although China's artificial intelligence talents have achieved great results in recent years, there is still a big gap compared with the developed countries in the United States and Britain. In terms of discipline construction, the cultivation system of artificial intelligence talents such as the United States and the United Kingdom is solid, and the research talents have significant advantages. At present, the academic ability in the field of artificial intelligence ranks among the top 20 schools in the world, with 14 in the United States, and the top eight seats are occupied by the United States[XXIII]. In terms of industrial talents, as of June 2017, the total number of US industrial talents is about It is twice as large as China. There are about 78,000 employees in 1078 artificial intelligence companies in the United States and about 39,000 employees in 592 companies in China[XXIV].

Moreover, the proportion of experienced artificial intelligence practitioners is significantly different from that of the United States (38.7% of employees with more than 10 years, and 71.5%[XXV] of the US).

(7) Ecological cultivation of artificial intelligence industry at home and abroad

The ecological development of China's artificial intelligence industry is uneven, and the basic links are weak. At present, China has a layout in the artificial intelligence base layer, technology layer and application layer, and has initially possessed a relatively complete artificial intelligence industry ecology. However, the ecological development of China's artificial intelligence industry is not balanced and cannot achieve the controllable development of ecology. **Application layer links**, due to China's advantages in the mobile Internet market and rich application scenarios, it provides convenient conditions for the development of artificial intelligence application layer. Domestic artificial intelligence industry investment and technology research mainly focus on the application layer link, automated driving, computer vision, speech recognition and other application areas have formed certain advantages, and even some industries are at the international leading level.

However, **the basic layer links**, due to the long investment cycle and high risk of earnings, the development of artificial intelligence base layer in China is relatively slow, lacking major original results, and there is a big gap in basic theory, core algorithms and key equipment, high-end chips, etc. Excellent entrepreneurial companies such as Cambrian, Horizon, etc., and have formed mature products, but compared with giants such as NVIDIA and Google, it is still difficult to obtain market competition. The overall industrial layout of the United States is leading the development of the global artificial intelligence industry. US giants such as Google, IBM, and Microsoft are stepping up the overall industrial intelligence layout. They are in a leading position in the artificial intelligence base layer, technology layer and application layer, especially in the field of algorithms and chips, which have accumulated strong technological innovation advantages in global artificial intelligence industry development. The Oxford University research report proposes the National Artificial Intelligence Potential Index (AIPI), which ranks second in the world in terms of industrial ecology, but scores only[XXVI] a quarter of the US.

5. Suggestions for AI Security Development

At present, China's national security and international competition situation is increasingly complicated, and it must be globally oriented. Address AI security at the national strategic level; engage in system layout and active planning; and insist on

accelerating technology and application innovation as the main line, and on improving legal and ethical standards as a safeguard. With regulatory norms acting as traction, vigorously promote standards construction, industry coordination, personnel training, international exchanges, publicity and education, etc. Comprehensively enhance the China's AI security capabilities; firmly grasp the strategic initiative of international competition in the new stage of AI development; foster competitiveness in the new advantages in the development of this new space; and effectively guarantee China's cyberspace security and stable economic and social development.

(1) Strengthen Indigenous [or "Independent"] Innovation and Achieve Breakthroughs on General Key Technologies

First, with a base of indigenous innovation, increase the introduction and absorption of technology, and achieve breakthroughs in the key basic technologies of AI. China's AI industry currently has an inverted triangle structure—heavy on applications, light on foundations—bringing many uncertainties for AI development. Therefore, it is necessary to start from research in key general technologies such as cloud computing, big data, and machine learning to resolve basic security risks. Stand independently and implement major technical research projects with the goal of secure and controllable development of key technologies such as sensors, smart chips, and basic algorithms. At the same time, increase technology introduction and conduct external technical cooperation with an open and pragmatic attitude, to achieve technology digestion, assimilation, and re-innovation. Relying on a development model combining indigenous innovation and technology introduction, formulate a secure and controllable development roadmap for key AI technologies, and solve the "stranglehold" problem in the foundational links of AI.

Second, increase research on AI security technology, and improve AI security protection capabilities. In view of the current situation, in which AI security research lags behind application research: aim at AI security issues and risk pain points; guide many parties to increase investment; vigorously support research institutes, AI enterprises, and cyber security enterprises to deepen AI security attack and defense technology research and build AI security attack and defense drill platforms; designing an all-round and integrated security protection technology architecture with independent intellectual property rights in the AI base layer, technology layer, and application layer. Accelerate

the development of security protection products, and explore and promote security best practices in key applications. Ensure the simultaneous advancement of AI security technology research and the application industrialization process.

(2) Improve Laws and Regulations; Formulate Ethics and Norms

First, establish and improve existing laws and regulations to deal with the issues of privacy security risks and subject liability brought about by AI. First, promote the construction of laws and regulations for the protection of personal information. At present, some of the existing laws and regulations in China already involve personal privacy protection, but the terms are more dispersed and cannot form a complete system. It is necessary to speed up unified legislation and draw on the relevant provisions and practical experiences of the European Union's General Data Protection Regulation to: promote formulation of the Personal Information Protection Law of the People's Republic of China; clarify the scope of personal information; protect a user's right to know; strengthen the responsibilities of data processors; handle the relationship between open data use and personal privacy protection in accordance with the law; ensure reasonable requirements for AI data resources are met; and prevent excessive use of personal information. Secondly, improve the current laws and regulations to clarify the subject liability issue. Current laws and regulations lack constraints on the application of AI products and systems. The liabilities and obligations in the design, production, sale, and use of artificial intelligence products and systems are not clearly defined. With regard to fairness and justice problems and accidents, potential violations of laws and regulations, and the resulting property damage, personal injury, and social harm that may be brought about by AI, there is a need for strengthened research, forward-looking legislation, and further clear subject constraints and division of responsibilities at the legal level.

Second, study and formulate ethical and moral norms to adapt to the social behavior model of human-computer symbiosis in the intelligent age. Guided by the government, relevant universities, research institutions, enterprises, etc.: establish AI ethics research institutions; strengthen the overall planning of AI ethics research; track the impact of AI on ethics, morals, and security risks; and build systematic ethical and moral norms to constrain AI research goals and directions, as well as the behavior of system designers and developers. Advocate and strengthen algorithmic ethics, improve the consistency of AI with human values, avoid an AI arms race between countries, ensure

that the entire society can share in the economic prosperity created byAI, and advance the healthy development of human society.

(3) Improve the Supervision System and Guide Industry Toward Healthy Development

First, improve the government's supervision system, and optimize the administrative framework. The increasing integration of AI with traditional industries will result in a large number of new forms and new models. The security risks of related industries have become intertwined and complicated. Supervision work involves multiple government departments, and the government supervision system should be optimized and improved according to actual developments. Enhance the support capabilities of regulatory technology. Conduct safety supervision pilots for the pioneering areas of intelligent recommendation, automatic driving, intelligent service robots, smart home, and other AI applications. At the same time, promote the timely adjustment of the governance structure of administrative bodies, while ensuring the impact of new technology development on the industry and society is within the controllable range, giving intelligent industrial technology and industrial innovation developments space to mature.

Second, constrain the market behavior of enterprises, and strengthen corporate self-disciplinary responsibilities. In the era of big data, AI enterprises (especially large Internet platforms) can access massive amounts of data. Learning from and using data involves multiple levels including personal privacy, public safety, and social governance. Therefore, enterprise platforms are important carriers of data, and the regulatory compliance and legality of their behavior is more important. The security of training data and algorithmic decision-making will directly affect user rights and interests and societal security. It is necessary to: increase the supervision of enterprises; define the boundaries of government and enterprise responsibility; guide enterprises to emphasize their own economic benefit while at the same time strengthening their sense of social responsibility; strengthen self-discipline and self-government; ensure the legality and security of data collection, storage, and circulation; and properly apply AI technology.

(4) Strengthen Standards as Guidance; Build a Security Assessment System.

First, formulate standards related to AI security to make up for existing gaps. Joint research institutes, technology companies, and third-party assessment agencies jointly promote the development of national standards, industry standards, and alliance standards

related to AI security, focusing on research related to AI training algorithms, decision models, and other related technical security requirements. Form a series of security standards for cybersecurity, data security, algorithm security, and application security of intelligent products and systems, as a unified reference for the security design and test verification of AI products, enhancing the security and reliability of AI products.

Second, guided by security standards, carry out security assessments and evaluation capacity-building. Guided by AI security standards, joint research institutions and technology companies jointly tackle security assessment and evaluation techniques for artificial intelligence products, applications, and services, and gradually accumulate knowledge resources such as security test sample libraries and knowledge libraries to form shared data sets. Develop a set of R&D test tools, build a public service platform for AI security testing and certification, establish an evaluation expert database and evaluation mechanisms, and realize the evaluation and evaluation capability of AI security. With technical means as a support, pragmatically avoid the problematic defects and security risks of AI products and applications.

(5) Promote Industry Collaboration; Promote Technology Security Applications.

First, promote collaboration between AI enterprises and cybersecurity enterprises to improve the depth of technology application and product maturity. AI enterprises have accumulated technology related to machine learning algorithms, and cybersecurity enterprises have data resources and security protection application scenarios, such as vulnerability databases and incident databases. Promote deeper cooperation between AI enterprises and cybersecurity enterprises. Leverage existing cybersecurity knowledge base resources to undertake data analysis and feature learning to improve the cybersecurity protection product self-defense capabilities such as vulnerability discovery, threat warning, and attack detection, and iteratively upgrade and optimize product maturity in application scenarios to jointly promote the deep application of AI technology in the field of cyber and information security.

Second, promote cooperation between AI enterprises and public security enterprises, expand the application of technology, and enhance social governance capabilities. AI is gradually developing into a new universal technology, which promotes the transformation of traditional industries through automation and intelligentization. In the field of public security, mature general technologies in AI, such as computer vision,

speech recognition and synthesis, and natural language processing, have started to be applied to fields such as security monitoring, data investigation, public opinion management, etc. Vigorously promote the coordinated development between AI enterprises and traditional public safety enterprises; jointly explore the needs of integration; aim at the development pain points of the profession; form integrated solutions; accelerate the application on the scene; and promote the widespread application of AI in public security, such as in smart public security, intelligent transportation, and smart finance; to improve the level of intelligentization of national social governance.

(6) Increase Personnel Training; Improve the Job Skills of Personnel

First, strengthen the construction of a talent corps in AI technology and industry, and reduce the risks of talent shortage for the development of the industry. First, based on school education, vigorously implement the relevant documents of the Ministry of Education such as the "Artificial Intelligence Innovation Action Plan for Higher Educational Institutions." Add AI-related majors in qualified universities, enlarge recruitment quotas, strengthen professional education and vocational education, and provide personnel with AI thinking, skills, and human-machine collaborative operation capabilities. Fund key university development labs and innovation centers to increase research talent training. Second, increase enterprise training, and aim at the current shortage of AI talent, encouraging AI technology enterprises to establish training institutions or jointly build laboratories with schools, to conduct technical and applied research, and to cultivate available talent in practice. Third, strengthen the introduction of foreign talent, formulate talent policies to introduce special talent, support universities or enterprises to introduce world-class leading talent, directly set up R&D centers abroad, and absorb local talent there for our own use. Encourage industry acquisitions and enterprise use of capital to retain or acquire teams from foreign companies with core technologies.

Second, optimize the personnel training system, improve the job skills of personnel, and reduce the unemployment risks caused by AI. For social changes in employment caused by the development of AI industry, first of all, the specializations of universities and vocational schools, etc., should be dynamically adjusted, and the recruitment quotas for majors that can be replaced should be gradually reduced or even eliminated to ensure that students can apply what they have learned and prevent

"graduation into unemployment." Secondly, encourage currently working people to establish a lifelong learning outlook; improve on-the-job training and re-employment training systems; update the employment skills of currently working people through multifarious training; promote higher-quality employment for currently working people; and reduce the social impact of the unemployment risk caused by AI.

(7) Strengthen International Exchanges; Address Common Security Risks

First, strengthen technical research cooperation, resolve the current stage's bottlenecks in AI technology, and promote the mature development of AI. For current deep learning technology bottlenecks, such as poor robustness against sample, non-explainability, incomplete information, and weak adaptability to uncertain environments, international cooperation in technical research can be carried out by setting up research centers abroad and organizing international technical exchanges. Track the latest technological achievements, jointly strengthen research on new technologies such as transfer learning and brain-like learning, solve security hazards and regulatory problems such as algorithmic black boxes and algorithmic discrimination, enhance the robustness and security of AI decision-making, and promote a move from specialized intelligence toward general intelligence in AI.

Second, actively participate in the formulation of standards to jointly address the security issues and ethical impacts of AI. Actively participate in the ISO/IEC JTC1 SC27 and SC42 data security, privacy protection, problem responsibility determination, and trustworthy AI standards development work. Closely track the IEEE P7000 series of AI security- and ethics- related standards; strengthen exchange with the major world standardization organizations ISO, IEC, ITU, ETSI, NIST, etc.; establish exchange mechanisms with advanced countries and leading enterprises; share governance experience; and promote the continued and secure development of AI in China. At the same time, the governments of the world's major countries should establish an AI development exchange and dialogue mechanism, seek cooperation and win-win amidst competition, formulate AI ethics and moral standards that are generally observed by the international community, avoid the malicious applications of AI technology, and effectively guarantee that AI truly benefits humanity.

(8) Increase Social Propaganda and Scientifically Handle Security Issues

First, carry out propaganda and education to strengthen the awareness of security protection. AI technology is inherently neutral, but people could potentially apply it maliciously. For example, machine learning can be used for personal information mining to quickly obtain private information; and information content synthesis technology can enrich online scams and make them more confusing. Therefore, for new security incidents using AI technology, it is necessary to strengthen publicity, publicize the causes, carry out education on security measures for all people, cultivate citizen awareness of privacy protection and fraud potential, and reduce personal property losses and adverse social effects caused by malicious AI application.

Second, strengthen the guidance of public opinion and establish a proper development concept. Although the development of AI technology has achieved remarkable results at present, there are still some common problems. Many industrial applications, such as automatic driving, intelligent robots, etc., are in the stage of exploration and experimentation, and they are immature and may lead to security incidents. In view of the security incidents exposed by the current AI technology in the industrial application process, we should strengthen the proper direction of public opinion propaganda, reduce social anxiety, guide people to properly view the security issues in the development of new technologies, and create openness for the development of AI technology and industrial advancement with relaxed social environment.

Notes

[I] Master Candidate in Political Science at Federal University of Pernambuco (UFPE). E-mail: nvbittencourt@gmail.com.
[II] MasterCandidate in Political Science at Federal University of Pernambuco (UFPE). E-mail: karllagodoy@gmail.com.
[III] YING, Fu (傅莹). **An Analysis of the Influence of Artificial Intelligence on International Relations (人工智能对国际关系的影响初析).** Tsinghua University Press Journal Center.Quarterly Journal of International Politics, 2019, 04(01): 1-18, 2019.
[IV] Laskai, L. and Webster, G. Translatio: Chinese Expert Group Offers 'Governance principles for Responsible AI'. Available at: <https://www.newamerica.org/cybersecurity-initiative/digichina/blog/translation-chinese-expert-group-offers-governance-principles-responsible-ai/>. Accessed: 8 Nov 2019.
[V] Translated by Nathália Vivian Bittencourt and Karla Godoy da Costa Lima.
[VI] Artificial intelligence is an area of computer science. It mainly solves problems such as computer perception, reasoning and behavior.
[VII] Artificial intelligence is a discipline of knowledge - how to express knowledge and how to acquire knowledge and use knowledge.
[VIII] Artificial intelligence is a partial simulation of human intelligence.

[IX] Fredrikson M, Lantz E, Jha S, et al. Privacy in pharmacogenetics: an end- to- end case study of personalized warfarin dosing
[X] Fredrikson M, Jha S, Ristenpart T. Model inversion attacks that exploit confidence information and basic countermeasures

[XI] Source: ProPublica
[XII] Joy Buolamwini, Timnit Gebru,《Gender Shades: Intersectional Accuracy Disparities inCommercial Gender Classification》

[XIII] Biggio B, Corona I, Maiorca D, et al. Evasion attacks against machine learning at test time
[XIV] Nguyen A M, Yosinski J, Clune J. Deep neural networks are easily fooled: high confidence predictions for
unrecognizable images
[XV] Li Kaifu, "Artificial Intelligence", Wang Haifeng, "The Road to Artificial Intelligence in China", etc.
[XVI] Russia's "2025 advanced military robot technology equipment research and development special comprehensive plan"
[XVII] Source: China Electronics Society
[XVIII] China's security industry "13th Five-Year Plan" (2016-2020) Development Plan
[XIX] Billion Euro think tank: "2018 China AI+ Security Industry Development Research Report"
[XX] Zhou Songqing, "Comparative Study on the Legal Regulation of Internet Public Opinion Monitoring in China and the United States"
[XXI] Source: Seagate Technology
[XXII] Tsinghua University, "2018 China Artificial Intelligence Development Report"
[XXIII] Tencent Research Institute, 2017 Global Artificial Intelligence Talent White Paper
[XXIV] Tencent Research Institute, "Comprehensive Interpretation of the Development of Artificial Intelligence Industry in China and the United States"
[XXV XXV] LinkedIn, Global AI Talent Report
[XXVI] Oxford University, Deciphering Chinese AI Dream

7 References

BARDIN, Laurence. **Análise de conteúdo (Content Analysis)**. Lisboa: Edições 70, 1979.

CELLARD, A. **A análise documental.** In: POUPART, J. et al. A pesquisa qualitativa: enfoques epistemológicos e metodológicos. Petrópolis, Vozes, 2008.

CHINA, Ministry of Education. ***Action Plan for Artificial Intelligence Innovation in Colleges and Universities [高等学校人工智能创新行动计划].*** *April 2018. Available at:*
<http://www.moe.gov.cn/srcsite/A16/s7062/201804/t20180410_332722.html>.Accesse d: 8 nov 2019.

CHINA, Ministry of Science and Technology. China Academy for Information and Communications Technology (CAICT)**. 'Artificial Intelligence Security White Paper' [人工智能安全白皮书].** September, 2018.

http://www.caict.ac.cn/kxyj/qwfb/bps/201809/t20180918_185339.htm.> Accessed: 9 October 2019. (Own english translation available in appendix)

CHINA, Ministry of Science and Technology. National New Generation Artificial Intelligence Governance Expert Committee**. 'New Generation of AI Governance Principles - Develop responsible artificial intelligence'.** June, 2019. Available at: <https://mp.weixin.qq.com/s/JWRehPFXJJz_mu80hlO2kQ> Accessed: 9 October 2019.

CHINA, *State Council the People's Republic of China.* ***New Generation Artificial Intelligence Development Plan [新一代人工智能发展规划].*** July, 2017. Available at <https://flia.org/wp-content/uploads/2017/07/A-New-Generation-of-Artificial-Intelligence-Development-Plan-1.pdf> Translated by Flora Sapio (FLIA Scholar), Weiming Chen (FLIA Research Assistant), and Adrian Lo (FLIA Research Intern). Accessed: 8 nov 2019.

CHINA, State Council the People's Republic of China. ***Made in China 2025.*** *2015.* ***Available at: <http://www.cittadellascienza.it/cina/wp-content/uploads/2017/02/IoT-one-Made-in-China-2025.pdf>*** Accessed: 8 nov 2019.

COWLS, J. and FLORIDI, L. **Prolegomena to a White Paper on an Ethical Framework for a Good AI Society**. 2018. Available at https://ssrn.com/abstract=3198732 or http://dx.doi.org/10.2139/ssrn.3198732 Accessed 08 nov 2019.

JANZ, Nicole. **Bringing the Gold Standard into the Classroom: Replication in University Teaching.** International Studies Perspectives, doi: 10.1111/insp.12104, p. 1-16, 2015.

KANIA, E., PETERSON, D, WEBSTER, G and LASKAI, L. **Translation: Key Chinese Think Tank's "AI Security White Paper" (Excerpts).** DigiChina Project. New merica Org. Available at:<https://www.newamerica.org/cybersecurity-initiative/digichina/blog/translation-key-chinese-think-tanks-ai-security-white-paper-excerpts/>. 2018. Accessed: 9 nov 2019.

KING, G. **Replication.** PS: Political Science and Politics, v. 28, p. 443-499. Available at: <https://gking.harvard.edu/files/replication.pdf >. Accessed 08 nov 2019.

MEARSHEIMER, J.J. **The Tragedy of Great Power Politics**. New York: W.W. Norton & Company, 2001

MINGHAO Zhao. **'Is a New Cold War Inevitable? Chinese Perspectives on US – China Strategic Competition'.** The Chinese Journal of International Politics, Volume 12, Issue 3, August 2019, 371– 394, 2019. Available at: <https://doi.org/10.1093/cjip/poz010>. Accessed: 9 October 2019.

OLABUENAGA, J. I. R., & ISPIZÚA, M. A. **La descodificación de la vida cotidiana: metodos de investigacion cualitativa.** Bilbao: Universidade de Deusto, 1989.

WEBER, R.P. **Basic Content Analysis**. Newbury Park, CA: Sage Publications. 1990

SIRUI, Z. **'China Bets Big on AI: Summary of Central Government Policies'**. EqualOcean, 26 July, 2019. Available at: <https://equalocean.com/ai/20190726-china-betting-big-on-ai-summary-of-central-government-policies> Accessed: 9 October 2019.

The New Generation Artificial Intelligence Governance Expert Committee. **'A New Generation of Artificial Intelligence - Governance Principles for Responsible AI'** [新一代人工智能治 理原则——发展负责任的人工智能], 2019. Available at: https://mp.weixin.qq.com/s/JWRehPFXJJz_mu80hlO2kQ. (Accessed: 9 Oct)

XUETONG, Yan. **From Keeping a Low Profile to Striving for Achievement.** The Chinese Journal of International Politics, 153–184, 2014. Available at: < https://academic.oup.com/cjip/article/7/2/153/438673> Accessed: 08 Nov.

XUETONG, Yan. **Political Leadership and Power Redistribution.** The Chinese Journal of International Politics, 1-26, 2016. Available at: <https://academic.oup.com/cjip/article/9/1/1/2365941> Accessed: 08 Nov.

WANG Y. and CHEN, D. **Rising Sino-U.S. Competition in Artificial Intelligence.** China Quarterly of International Strategic Studies, Vol. 4, No. 2, 241–258, 2018. Available at: <https://www.worldscientific.com/doi/abs/10.1142/S2377740018500148> Accessed: 05 november 2019

WU, Shellen. **China: How science made a superpower.** Nature 574, 25-28, 2019. Avalable at: <https://www.nature.com/articles/d41586-019-02937-2>. Accessed: 08 nov 2019.

YING, F. (傅莹). **An Analysis of the Influence of Artificial Intelligence on International Relations (人工智能对国际关系的影响初析).** Tsinghua University Press Journal Center.Quarterly Journal of International Politics, 04(01): 1-18, 2019.

Available at: < http://qjip.tsinghuajournals.com/article/2019/2096-1545/101393D-2019-1-102.shtml>. Accessed: 8 Nov 2019.

ZHANG, Feng. **The Tsinghua Approach and the Inception of Chinese Theories of International Relations.** The Chinese Journal of International Politics, Vol. 5, 73–102, 2012. Available at: <http://fengzhang.net/wp-content/uploads/2017/05/The-Tsinghua-Approach.pdf>. Accessed: 8 nov 2019.

ZHANG ,Y. e CHANG, T. **Constructing a Chinese School of International Relations: Ongoing Debates and Sociological Realities**. London, UK: Routledge, 350 pgs, 2016.