上海国际问题研究院
SHANGHAI INSTITUTES FOR INTERNATIONAL STUDIES

清华大学战略与安全研究中心
CENTER FOR INTERNATIONAL STRATEGY AND SECURITY
TSINGHUA UNIVERSITY

北京大学互联网发展研究中心

International Rules of Artificial Intelligence: Trends, Domains and China's Role

Contributing Authors:

LU Chuanying

TIAN Li

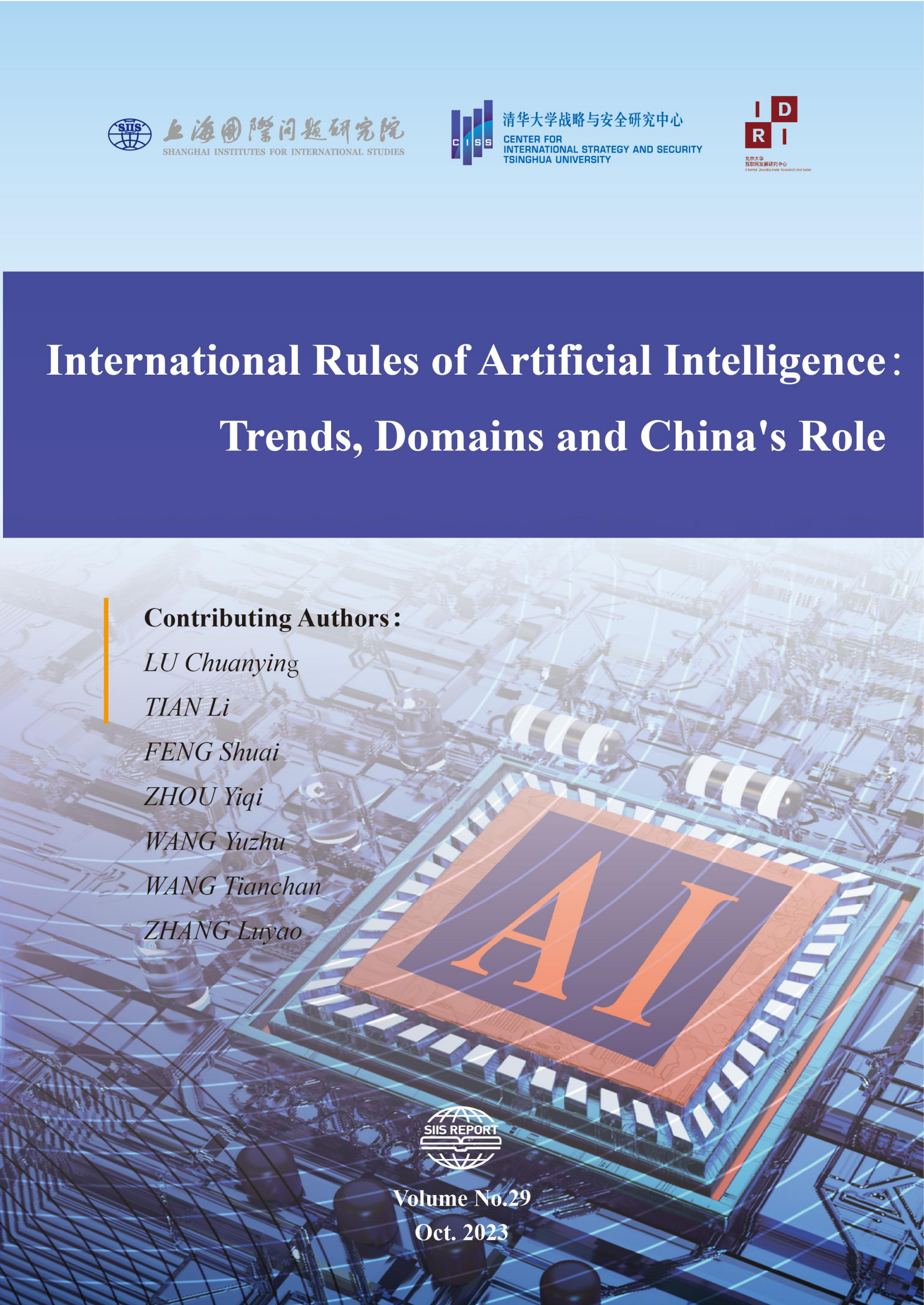
FENG Shuai

ZHOU Yiqi

WANG Yuzhu

WANG Tianchan

ZHANG Luyao



SIIS REPORT

Volume No.29

Oct. 2023

International Rules of Artificial Intelligence: Trends, Domains and China's Role

Editor-in-Chief & Producer
CHEN Dongxiao

Contributing Authors
LU Chuanying *TIAN Li* *FENG Shuai* *ZHOU Yiqi*
WANG Yuzhu *WANG Tianchan* *ZHANG Luyao*

Exec. Editor-in-Chief
LI Xin

Editor
CHEN Xue

SHANGHAI INSTITUTES FOR INTERNATIONAL STUDIES

清华大学战略与安全研究中心
CENTER FOR INTERNATIONAL SECURITY AND STRATEGY
TSINGHUA UNIVERSITY

北京大学互联网发展研究中心
Internet Development Research Institution

Publishing Team
ZHU Juhua
ZHANG Jun
GE Jieyi

Shanghai Institutes for International Studies
195-15 Tianlin Road, Xuhui District
Shanghai 200233, PRC
Tel/Fax: +86 21 64850100
http://www.siis.org.cn

This publication can be downloaded at no cost at SIIS website.

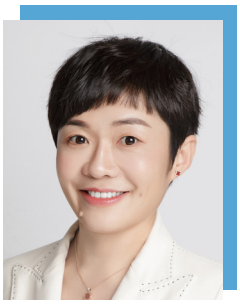
About the Authors

LU Chuanying

Deputy Director of the SIIS Institute for Public Policy and Innovation Studies

Secretary General of the SIIS Cyberspace International Governance Research Center

Nonresident Senior Fellow, CISS, Tsinghua University



TIAN Li

Deputy Dean of the School of New Media,

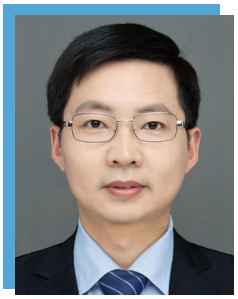
Director of the Internet Development Research Institution of Peking University

FENG Shuai

Assistant Director of the SIIS Institute for International Strategic and Security Studies

ZHOU Yiqi

Associate Senior Fellow of the SIIS Institute for Public Policy and Innovation Studies



WANG Yuzhu

Senior Fellow of the SIIS Institute for World Economy Studies

Secretary General of the SIIS Center for BRI and Shanghai Studies

WANG Tianchan

Intern of the SIIS Cyberspace International Governance Research Center

ZHANG Luyao

Intern of the SIIS Cyberspace International Governance Research Center

Contents

Introduction

Since the beginning of 2023, breakthroughs in artificial intelligence technology have once again become the focus of global attention. Driven by rapidly advancing generative artificial intelligence and breakthrough progress in large language models, humans seem to have begun to hit the threshold of general artificial intelligence technology. The imagination of an intelligent society around the leapfrog development of artificial intelligence technology has grabbed the attention of global social media and capital markets. However, the various security risks and social issues that may arise from artificial intelligence technology have also become the focus of public attention. In order to ensure that artificial intelligence technology always maintains sound development and that technological progress is aligned with human values, establishing rational international rules for the development of artificial intelligence has gradually become an important consensus worldwide.

On July 18, 2023, the United Nations Security Council held a high-level public meeting on Artificial Intelligence: Opportunities and Risks for International Peace and Security. It is the first-ever meeting of the United Nations Security Council on the theme of artificial intelligence. At the meeting, UN Secretary-General Guterres publicly called for effective governance of risks associated with artificial intelligence technology and proposed the establishment of a new regulatory body within the UN framework to promote global governance in the artificial intelligence field. International rules play an important guiding role, and governments and enterprises around the world all hope to participate in or even lead the agenda to establish international rules for artificial intelligence. It can be foreseen that there will be complex co-opetition around the issue of international rulemaking on artificial intelligence, and the resulting international rule system will also become an important variable affecting global artificial intelligence technology, industry, and safety.

Over the years, with the concerted efforts of all relevant parties, China has achieved fruitful results in developing artificial intelligence technology and industry. Artificial intelligence has become an important driving force for the high-quality development of the Chinese economy, and China has also become an important part of the global artificial intelligence landscape. China has accumulated

rich experience in artificial intelligence governance and can provide a valuable developing country perspective for the establishment of global rules on artificial intelligence. Therefore, China should proactively participate in the formulation of international rules for artificial intelligence. On the one hand, it should transform the Chinese experience in artificial intelligence governance into international rules. On the other hand, by participating in discussions on international rules, China can deepen its understanding of advanced governance ideas and better promote the safety and development of artificial intelligence in the country.

I. General Trend of International Rulemaking on Artificial Intelligence

The reason why the establishment of international rules for artificial intelligence has received wide attention from countries around the world is that artificial intelligence has become an important force affecting national security and development. The United States, China, EU nations and other major countries have successively released their national development strategies for artificial intelligence, and international competition surrounding artificial intelligence has become an important component of the current international agenda. Artificial intelligence and related derivative technologies have strong permeability and high empowerment capacity, and can establish deep links with many social production sectors. If their technological potential is fully unleashed, it may have a disruptive impact on the production and life of human society. Therefore, in the course of promoting the progress of artificial intelligence technology, it is imperative to pay full attention to the dynamic balance between technological development and social stability.

In this context, countries around the world are gradually recognizing the need to promote AI governance through rulemaking. Developed countries like the United States and European nations believe that taking the initiative in formulating artificial intelligence rules is an important part of maintaining their dominant position in the technological and industrial domains. They actively promote research on relevant issues and try to dominate the international agenda of rulemaking, attempting to improve mechanisms and establish rules at both domestic and international levels. However, compared to the more homogeneous domestic environment, the interaction between countries is more complex, the entanglement of interests between different entities is more prominent, and the cultures and values involved are more diverse. This also makes it more difficult to form consensus on international rules for artificial intelligence.

As for actual progress, international rules for artificial intelligence are still in the early stages of construction at present, and various entities are trying to promote all kinds of governance plans at different levels, which can be summarized into the following aspects:

Firstly, at the multilateral level, the United Nations actively leads the establishment of international rules for artificial intelligence governance. In June 2023, UN Secretary-General Guterres publicly stated his support for the establishment of an "International Atomic Energy Agency (IAEA)" for the field of artificial intelligence to regulate artificial intelligence technology. This means that the United Nations hopes to adjust its role and enhance its functions in the future process of formulating international rules for artificial intelligence and deepening its scope of action from the principles level to the mechanism level, so as to set up a more binding regulatory agency on artificial intelligence. This may become the basic structure for the establishment of future international rules for artificial intelligence. In addition, *the New Agenda for Peace* issued by the United Nations in July 2023 also specifically mentions the impact of artificial intelligence and lethal autonomous weapon systems on international peace and security, and recommends that member countries should urgently formulate national strategies on the responsible design, development and use of artificial intelligence; promote the design, development and use of artificial intelligence in the military field through multilateral and multi-party processes; and ultimately reach agreement on a global regulatory framework for artificial intelligence.

In recent years, multiple agencies under the United Nations have actively participated in the artificial intelligence governance agenda and achieved phased results. *The Recommendation on the Ethics of Artificial Intelligence*, which was issued in 2021 by the UNESCO Ad Hoc Expert Group for the Recommendation on the Ethics of Artificial Intelligence, is an important milestone in the establishment of international rules on artificial intelligence. This recommendation sets out eleven guiding principles for the development of artificial intelligence based on the common values of humanity and the basic position of sustainable development goals. It was jointly adopted by 193 member states of the United Nations. Since 2014, the United Nations has held multiple informal expert meetings under the Convention on Certain Conventional Weapons talks mechanism to discuss issues related to lethal autonomous weapon systems, with nearly 100 countries participating in related activities. In addition, in the activities and agendas of institutions such as the United Nations Institute for Disarmament Research and the United Nations Office on Drugs and Crime, issues such as the ethics and safety of lethal autonomous weapon systems and the political risks of deep forgery have also been included in their scopes of discussion, but no specific principles, proposals or international rules have been introduced.

Secondly, in the developed countries block, the United States, Britain and the Netherlands have intensified their leverage in attempting to monopolize dominance on international rules for artificial intelligence. The United States has done a lot of work in promoting the internationalization of its domestic rules. Guided by its strategic documents such as *The National Strategy for Critical and Emerging Technology*, the United States has taken "promoting cooperation with allies" and "building world leadership in emerging technologies" as its basic stance, and established a series of international

cooperation mechanisms for artificial intelligence. For example, in the military and defense domain, the United States has not only published the *Artificial Intelligence Strategy* within the NATO framework to promote its own stance on security and geopolitical competition, but also established the AI Partnership for Defense (AI PfD) and the new Australia-United Kingdom-United States Partnership (AUKUS) mechanism to coordinate the rules on the use of artificial intelligence in the defense and security domain. In the field of standards coordination, the United States has leveraged the QUAD Critical and Emerging Technology Working Group, the US-EU Trade and Technology Council (TTC), the US-India Artificial Intelligence (USIAI) Initiative and other mechanisms to bring together its key allies in the Asia Pacific, Indo Pacific and Western Europe to undertake technical cooperation and promote common standards for artificial intelligence technology. In the values domain, the United States has selectively listened to the praise and recognition of its ideology and values from partners of its "small circles" through mechanisms such as the Summit for Democracy and the Freedom Online Coalition's Task force on Artificial Intelligence and Human Rights (T-FAIR), which naturally confirm the rationality of its ideas and proactively promote their transition from regional ideas to global ones.

Britain, the Netherlands and other Western countries have also begun to form an "alliance of consensus" in the artificial intelligence field, hoping to bring the ability to dominate international rules for artificial intelligence under the control of a few developed countries. In February 2023, the Netherlands hosted the Responsible Artificial Intelligence in the Military Domain Summit in the Hague, with more than 80 countries from around the world invited to participate. The summit had three core objectives: raise the political importance of use of artificial intelligence by military leaders; mobilize and activate a wider range of stakeholders to participate; and share experience and practice. Although it claimed to seek to pool international consensus on artificial intelligence in the military domain, its actual inclusiveness was nevertheless influenced by the Western geopolitical agenda, and Russia was not invited to the summit. At the summit, the Chinese delegate elaborated on China's propositions for the artificial intelligence security governance field and put forward the governance principles of intelligence for good, people-oriented intelligence, and multilateralism. In March 2023, the British government released a white paper titled *A Pro-innovation Approach to AI Regulation*, proposing its own regulatory framework for artificial intelligence. In addition, Britain has announced that it will host the first Global Summit on Artificial Intelligence Safety within 2023. Although the conference may seem to be able to produce a positive effect on promoting the formulation of international rules for artificial intelligence under the professed goal of "bringing together countries to coordinate on AI safety standards and policies", British Prime Minister Sunak actually inadvertently revealed his country's true intention: trying to coordinate UK-US cooperation and consolidate their "joint leadership in the future technology field." Britain is trying to set up an international agency on artificial intelligence regulation in London, thereby controlling the international discourse on artificial intelligence rulemaking.

Thirdly, developing countries are still in a disadvantageous position in establishing international rules for artificial intelligence. Developing countries as a whole are still in the early stages of international rulemaking for artificial intelligence. On the one hand, this is related to the fact that artificial intelligence in many developing countries is still in a backward stage of development. On the other hand, it also has a lot to do with the lack of motivation and awareness among developing

countries to participate in the formulation of international rules. In the face of Western countries' continuous strengthening of the international rules system for artificial intelligence, developing countries should quickly awaken, intensify their participation, and contribute their respective wisdom. As a representative of non-Western and developing countries, China is one of the few countries that has proactively speaking out its voice in the field of international rules for artificial intelligence. China has not only actively participated in the work of the United Nations on international rules for artificial intelligence, but also proactively promoted and strengthened cooperation in the field of international rules for artificial intelligence in such international organizations as the G20, BRICS countries, and the Shanghai Cooperation Organization. Moreover, China has also been proactively promoting cooperation among developing countries in rulemaking in the artificial intelligence field. For example, China has held various forms of meetings and cooperation with ASEAN countries and Middle Eastern countries around the issue of artificial intelligence, and successfully established tacit understanding and norms of interaction.

Fair and rational international rules need to balance the positions and voices of both Western and non-Western countries. Against the current background of imbalanced representation, non-Western countries need to work more closely together, be proactive in international rulemaking, and strive for their own interests. A truly rational system of international rules should be inclusive rather than exclusive, and respect the broadest national interests and concerns rather than just focusing on the interests of a small number of groups. The various safety risks that artificial intelligence may bring are challenges that humanity must face together. The reasonable demands of non-Western countries in the field of artificial intelligence cannot be ignored, and the wisdom of non-Western countries on governance is also worth learning from.

II. Domains Covered by International Rules for Artificial Intelligence

Artificial intelligence has become a core driving force for the development of the digital economy. Especially in the context of continuous technological breakthroughs in artificial intelligence-generated content (AIGC) and artificial general intelligence (AGI), the gap between the technical needs of artificial intelligence governance and the governance effectiveness of sovereign states is increasingly widening. At present, the formulation of international rules for artificial intelligence worldwide is still in its infancy, with international soft law and indirect regulation being their main form and content. However, as major countries and regional organizations around the world attach more importance to the establishment of regulations in the artificial intelligence field, a global competition of artificial intelligence regulation has actually begun. Various countries and regions have not only proactively drafted and introduced laws and regulations to regulate artificial intelligence technology, but also competed to introduce policy guidelines related to the development of artificial intelligence technology for the good. Such regulatory competition helps to promote the formation of a comprehensive consensus-based artificial intelligence governance framework globally. It also brings opportunities and challenges for China to strive to be a "first mover" in international rulemaking.

Based on the current status of development of artificial intelligence technology and a summary of experiences in artificial intelligence governance from various countries, international rules for artificial intelligence worldwide have formed relatively stable areas of coverage, with rulemaking in related fields to become the main pillar of the international rules.

(1) Ethical Issues

Ethical issues of artificial intelligence technology have long been a key area of international

rulemaking, whose aim is to ensure that technological progress will not deviate from the basic interests and moral principles of humanity by effectively regulating the ethics of artificial intelligence. Regarding the ethical issues, there have been many explorations of rules internationally, with some important international organizations and institutions setting out ethical principles and guidelines for artificial intelligence. Some countries have also begun to regulate ethical issues of technology in specialized fields.

Among them, the representative achievements include the following: in November 2021, the United Nations Educational, Scientific and Cultural Organization (UNESCO) released the *Recommendation on the Ethics of Artificial Intelligence*, which is the first normative global ethical framework for artificial intelligence, while giving countries the responsibility to apply the framework at corresponding levels. In the European Union, the *Artificial Intelligence Act* proposed in April 2021 has entered its final stage after multiple rounds of consultations, which is highly likely to become the world's first artificial intelligence regulatory bill and therefore produce a demonstrative effect on rulemaking. The United States has made more detailed regulations on ethical issues related to artificial intelligence, such as the *Ethical Principles of Artificial Intelligence* released by the US Department of Defense in February 2020, and *The Principles of Artificial Intelligence Ethics for the Intelligence Community* and *The Artificial Intelligence Ethics Framework for the Intelligence Community* issued by the Office of the Director of National Intelligence in July 2020. In December 2021, China issued *The Position Paper on Regulating the Military Application of Artificial Intelligence*, calling on all parties to comply with national or regional ethical and moral guidelines for artificial intelligence. In November 2022, China released *The Position Paper on Strengthening Ethical Governance of Artificial Intelligence*, making specific prepositions from the aspects of technical regulation, research and development, use, and international cooperation and calling on all parties to uphold the idea of joint consultation, joint establishment and sharing, and promote international ethical governance of artificial intelligence.

Although these principles and guidelines differ in specific content, they all emphasize that artificial intelligence should conform to human values and interests, respect human dignity and rights, and ensure fairness, transparency, interpretability, trustworthiness, and accountability. Rulemaking in this field will safeguard the basic direction of development for artificial intelligence.

(2) International Security

The development of artificial intelligence technology can have a wide international impact and is of great significance for global security. From the perspective of international security, the combination of artificial intelligence and military weapons has already posed a practical challenge to international security.

At present, many organizations and institutions in the world are paying attention to and discussing the impact and challenges of artificial intelligence on militarization, nuclear weapons, and other aspects. For example, the United Nations Institute for Disarmament Research released in February 2023 a report entitled *Towards Responsible AI in Defense: A Mapping and Comparative Analysis of AI Principles Adopted by States*, which is the first phase result of the Towards Responsible AI in Defense Project. This project aims to establish key consensus on the research, design, development,

deployment, and use of responsible artificial intelligence systems, and review the status of application of responsible artificial intelligence in the defense domain. In 2021, the International Committee of the Red Cross released its *Position on Autonomous Weapon Systems* document in Geneva, which clearly stated that current military interests and investment trends have fully demonstrated that without establishing internationally recognized restrictive standards, future autonomous weapon systems may increasingly rely on artificial intelligence and machine learning software, raising concerns about their unpredictability at the design level. These organizations and institutions call for strengthened supervision and restriction on artificial intelligence in the military domain, so as to prevent artificial intelligence from triggering new conflicts and crises and maintain international peace and security. The security impact of artificial intelligence in different fields will continue to expand, and establishing corresponding international rules is of great significance.

The application of artificial intelligence technology in the nuclear domain has gradually attracted the attention of the international community. This is especially true of the impact of artificial intelligence on nuclear safety. In September 2020, the 64th International Atomic Energy Agency General Conference discussed the issue of artificial intelligence applications in the nuclear domain for the first time and demonstrated how nuclear technology empowered by artificial intelligence could benefit human health, water resources management, and nuclear fusion research. In August 2022, the International Atomic Energy Agency released a report entitled *Artificial Intelligence for Accelerating Nuclear Applications, Science and Technology*, introducing the application of artificial intelligence in nuclear science and technology, nuclear power, as well as nuclear safety, security verification safeguard and other domains. In June 2023, the article *How Artificial Intelligence Will Change Information and Computer Security in the Nuclear World was published in the IAEA Bulletin*, specifically emphasizing the many risks that the rapid development of artificial intelligence brings to nuclear safety. The article pointed out that malicious actors may use artificial intelligence to launch more advanced and more targeted attacks, or use it to disrupt the integrity of networks, systems, and sensitive information at nuclear and radioactive facilities. To address the challenge, the International Atomic Energy Agency has also developed a Coordinated Research Project (CRP) entitled *Enhancing Computer Security Incident Analysis at Nuclear Facilities* to support research on strengthening computer security. This project brings together representatives from 13 countries and is committed to improving the computer security capabilities of nuclear facilities, including detecting artificial intelligence technology abnormalities caused by cyberattacks. Since then, the nuclear safety issue of artificial intelligence has been officially included in the agendas of international mechanisms.

(3) Technical Safety

Right from the birth of artificial intelligence technology, how to ensure the safety of the technology itself has been the focus of attention in various countries. Simply put, the technical safety of artificial intelligence is concentrated at three levels: algorithm, data, and scenario. Establishing norms for technological development by setting relevant standards will constitute another important component of the international rules for artificial intelligence. Specifically, the development of artificial intelligence technology should adhere to the following principles: 1. scalable supervision, which ensures effective supervision of the development, use, and updating processes of artificial intelligence technology, and prevents its use for purposes that endanger human safety; 2. mechanism

interpretability, which helps people understand the decision-making process of artificial intelligence technology and thus better prevent and respond to the potential negative impacts of artificial intelligence technology; and 3. hazard capability test, which can help identify potential risks of artificial intelligence technology and take measures to reduce such risks. Currently, the international community has carried out considerable work around the three technical safety standards mentioned above, and many international organizations, regions, and countries have also begun to intensify efforts in the standards domain.

As for technical safety, some international standardization organizations and institutions have already been proactively developing and promoting safety standards and norms for artificial intelligence. In July 2022, the International Organization for Standardization (ISO) and the International Electrotechnical Commission (IEC) jointly released two new basic standards in the artificial intelligence domain, namely *Information Technology - Artificial Intelligence - Artificial Intelligence Concepts and Terminology* (ISO/IEC 22989:2012) and *Framework for Artificial Intelligence (AI) Systems Using Machine Learning (ML)*. These technical standard and framework are put forward to provide normative guarantees for global digital transformation and a systematic framework for the development of general artificial intelligence technology. In addition, the International Telecommunication Union (ITU) is also proactively involved in the establishment of artificial intelligence technology standards and norms, specifically responsible for standardization work related to artificial intelligence and machine learning through its ITU Telecommunication Standardization Sector (ITU-T) and ITU Radiocommunication Sector (ITU-R). In addition, the international standard *Framework of Artificial Intelligence Enhanced Telecom Operation and Management* (AITOM) developed by China Telecom under the leadership of the International Telecommunication Union was officially released in 2021, promoting China Telecom's smart telecom operation management technology and safety practices to the global industry in the form of standards to help the industry address the problem of artificial intelligence technology applications in telecommunications operations and management.

At the national and regional organizational levels, the formulation of artificial intelligence technology specifications and standards is also in full swing. Among them, the European Union, as the world's leading 'normative force', put forward ethical guidelines for artificial intelligence as early as April 2019, listing seven criteria for evaluating “trustworthy artificial intelligence”. On May 22, 2019, the Organization for Economic Cooperation and Development (OECD) adopted the *Recommendation for Artificial Intelligence*, which is the first intergovernmental standard in the artificial intelligence field, aimed at ensuring respect for human rights and democratic values through responsible management of artificial intelligence technology, and promoting the technological innovation of artificial intelligence. In 2022, the European Commission tabled the *Artificial Intelligence Liability Directive* motion, which further proposed an artificial intelligence safety liability determination mechanism to determine who should bear responsibility when artificial intelligence malfunctions or causes harm. In June 2023, the European Parliament adopted the draft *Artificial Intelligence Act*, which classifies risks in artificial intelligence systems, restricts deep forgery, and puts forward higher transparency requirements for generative artificial intelligence such as ChatGPT.

Besides, the United States, Britain and other countries have also undertaken some standardized

practical activities on the basic principles and specific standards of artificial intelligence technology safety. For example, in its first *National Artificial Intelligence Strategy* released on September 22, 2021, Britain called for domestically establishing cross-departmental standards to ensure the transparency of artificial intelligence algorithms, and internationally participating in the formulation of global standards for artificial intelligence and raising the government awareness of artificial intelligence technology safety. In its final research report officially released on March 2, 2021, the US National Security Commission on Artificial Intelligence (NSCAI) also emphasized the need to build up reasonable confidence in artificial intelligence systems, namely ensuring their solidity, robustness, and reliability. In addition, on May 23, 2023, the White House released the *National Artificial Intelligence R&D Strategic Plan*, with ensuring the safety of artificial intelligence systems being one of its strategic pillars.

It is worth noting that since 2020, discussions on AI value alignment in the academic and industrial circles have entered a heated stage. AI value alignment is considered one of the major challenges facing AI technology safety. Because the development of artificial intelligence technology has generated many uncertainties, how to ensure that artificial intelligence is aligned with human expectations and goals, ethical norms and values through "full-chain" governance from design to use has become a core concern at present. This is a matter of value alignment. Currently, leading AI companies such as IBM, Google and OpenAI have all released their own solutions and paths on the value alignment issue, emerging as a fresh force in establishing global standards artificial intelligence technology. In 2023, hundreds of world-leading technology researchers and developers, scholars, and business leaders, coordinated by the non-governmental organization Center for AI Safety, jointly signed an online statement, stating that "Mitigating the risk of extinction from AI should be a global priority alongside other societal-scale risks such as pandemics and nuclear war".

Overall, the current international standards and norms on the safety of artificial intelligence technology cover the terminology, framework, methods, ideas and many other aspects of artificial intelligence, and aim to improve the safety, reliability and quality of artificial intelligence. Rulemaking in this domain ensures that artificial intelligence technology always operates within the safety standards.

(4) Technological Innovation

Promoting the safe and effective development of artificial intelligence is the foothold of international rules for artificial intelligence. Ensuring the full unleash of technological element capabilities through establishing rules is also an important purpose of rulemaking. Therefore, it can be said that the development of technological innovation is a sufficient condition for pushing forward the formulation of international rules for artificial intelligence, as well as the goal and destination of international rulemaking.

At present, leading regions and countries in the world have begun to make layouts to promote and support the innovation and development of artificial intelligence technology. In the *White Paper on Artificial Intelligence* released in February 2020, the European Commission called for significantly raising the level of investment in the artificial intelligence research and innovation domain. In its *AI*

Strategy 2019, Japan vowed to develop artificial intelligence as a national strategic technology and promote the integration of artificial intelligence technology development with social applications. In the *National Strategy for Artificial Intelligence* issued in December 2019, the Republic of Korea set the goals of expanding artificial intelligence infrastructure, ensuring the mastery of competitive artificial intelligence technologies, innovating regulations and adjusting legal systems, among others. In May 2023, the White House released the third edition of the *National Artificial Intelligence Research and Development Strategic Plan*, following the first edition in 2016 and the second edition in 2019. The plan made investment in artificial intelligence technology research and development the top priority and stressed the supporting role of talent and international cooperation for technological innovation. China has also made plans and layouts for the development of artificial intelligence technology. For example, the *New Generation Artificial Intelligence Development Plan* issued by the State Council in 2017 was China's first strategic plan of systematic layout for the development and application of artificial intelligence technology. In July 2022, six ministries and commissions including the Ministry of Science and Technology jointly issued a circular entitled *Guiding Opinions on Accelerating Scenario Innovation and Promoting High-Quality Economic Development Through the High-Level Application of Artificial Intelligence*, providing systematic guidance for various localities and entities to accelerate the application of artificial intelligence scenarios, promote the technological upgrading and industrial growth of artificial intelligence through scenario innovation, and ultimately achieve high-quality economic development.

It can be seen that current efforts by the international community to promote artificial intelligence technology innovation are mainly focused on basic research, application development, talent cultivation, ethical norms, international cooperation, and so on. Through a series of strategic plans and policy measures, the countries and international organizations mentioned above have put big funds and resources into the innovation and development of artificial intelligence technology to cultivate talents and teams, build infrastructures and platforms, and promote cross-field and cross-border cooperation and exchange.

(5) Social Development

It is the common aspiration of countries around the world to fully utilize artificial intelligence technology to drive the comprehensive development of all sectors of society. However, relevant institutional guidance is very important for ensuring that the achievements of intelligent technology can successfully spill over to other social domain, From the perspective of development issues, there are currently some organizations and institutions in the world paying attention and responding to the impact and challenges of artificial intelligence on economic and social development. However, overall, there is still room for further improvement in the establishment of relevant systems.

Currently, the international mechanisms formed on the issues of artificial intelligence and social development are mainly concentrated in the multilateral domain. For example, at the opening ceremony of the 2023 World Artificial Intelligence Conference (WAIC 2023) held in Shanghai, China on July 6, 2023, the United Nations Industrial Development Organization (UNIDO) launched an initiative to establish the Global Alliance on Artificial Intelligence for Industrial and Manufacturing ("AIM-Global"). This groundbreaking initiative aims to unite national governments,

the private sector, and international organizations, as well as industry leaders focused on promoting responsible, sustainable, and inclusive applications of artificial intelligence technology to jointly create a shared, open, free, and safe digital future for all humanity. In addition, the World Economic Forum is concerned about the potential impact of artificial intelligence on human work mode. The report *Top 10 Emerging Technologies of 2023* released in June 2023 points out that AIGC technology, in addition to bringing significant development to society and the economy, will inevitably cause job losses. In the education sector, the World Bank released in September 2021 a report entitled *Steering Tertiary Education—Toward Resilient Systems that Deliver for Al*, introducing a series of the World Bank's practices in the development of effective, fair, efficient, and resilient higher education systems in the era of artificial intelligence.

At present, international organizations and institutions have mainly analyzed the impacts of artificial intelligence on employment, education, health, environment, and other aspects, and put forward some response suggestions and measures. However, more research and design are still needed if more comprehensive social development goals are to be achieved.

III. Chinese Values in the Course of Artificial Intelligence Governance

The Chinese government attaches great importance to the development of artificial intelligence technology and has paid close attention to the industrial and social changes which artificial intelligence technology may bring. The State Council has successively issued a series of government documents including the *New Generation Artificial Intelligence Development Plan*, which provide effective top-level designs for the development of artificial intelligence technology and industry. At the same time, China has also closely monitored the various risks and challenges that may arise from the development of artificial intelligence and proactively explored governance solutions suitable for its national conditions to ensure the healthy development of the artificial intelligence industry. As for governance, China has successively released important documents such as the *Principles of New Generation Artificial Intelligence Governance*, and its understanding of artificial intelligence governance has also gradually matured. After a period of practice and in-depth theoretical exploration, the unique values of the artificial intelligence governance rules with Chinese characteristics have gradually formed, which will also provide a solid core of thinking for China to participate in the establishment of international rules for artificial intelligence. For a long time, China has proactively participated in discussions on artificial intelligence governance within the United Nations framework, and successively submitted two position papers on regulating the military application of artificial intelligence and strengthening the ethical governance of artificial intelligence within the United Nations framework.

During the United Nations Security Council debate on artificial intelligence on July 18, 2023, China specifically proposed five principles on artificial intelligence governance, namely adhering to putting ethics first; adhering to safety and controllability; adhering to fairness and inclusiveness; adhering to openness and inclusiveness; and adhering to peaceful utilization. These five principles reflect

the philosophy of a community with a shared future for mankind in the artificial intelligence field, fully demonstrate China's values on artificial intelligence governance, and have important practical significance.

In summary, after a long period of research and practice, China has preliminarily formed the Chinese values on artificial intelligence governance and rulemaking, which will provide outstanding "Chinese wisdom" for the global agenda of artificial intelligence rulemaking and governance.

First, promoting the safe and orderly development of technology and industry is the fundamental orientation of China's values.

The idea "development is the top priority" constitutes the underlying logic of Chinese society in the era of reform and opening up, and it is also the underlying support of all values. China has always regarded the surging wave of artificial intelligence technology as a major strategic opportunity, aspiring to leverage technological progress and industrial development to drive the overall leap and leapfrog development of national competitiveness. Therefore, a basic direction in China's artificial intelligence governance is to reduce the unfavorable factors that hinder the development of artificial intelligence technology, promote the wide application of technology, and enable more production departments and the population to enjoy the dividends brought by technology.

Safety is a prerequisite for development. Without it, the achievements of development will be lost. The Chinese government attaches great importance to the issue of artificial intelligence safety and has proactively promoted the establishment of corresponding governance systems in response to potential safety risks that may arise from the development of artificial intelligence technology. For example, in the face of the rapid development of generative artificial intelligence, the Cyberspace Administration of China in conjunction with seven ministries and commissions has jointly issued the *Interim Measures for the Management of Generative Artificial Intelligence Services* on July 10, 2023, making China one of the first countries in the world to contribute to AIGC governance. These measures deeply reflect China's basic values and ideas on the artificial intelligence governance domain. On the one hand, through administrative measures, the document clarifies China's stance on proactively supporting the development of AIGC technology and promoting the creation of greater commercial and social value through the technology. On the other hand, the document also puts forward clear ethical and moral requirements for AI-generated contents and services, and specifies the principles of regulatory bodies and the division of responsibilities and rights. Faced with the challenges of emerging technologies, China adheres to an inclusive and cautious attitude in dealing with the safely issues brought about by new technologies, maintains the basic position of placing equal emphasis on development and safety, and systematically establish corresponding governance systems.

In China’s values on artificial intelligence governance, governance itself promotes the safe and orderly development of artificial intelligence technology and industry. Its purpose is to ensure the safety of artificial intelligence technology and enhance the common well-being of humanity through the establishment of relevant rules, while constantly maintaining the development process. Development and governance are inseparable, and governance serves the purpose of better development. Without

development, governance will be like water without a source or trees without roots. Therefore, promoting the establishment of relevant domestic rules should not come at the cost of sacrificing technological development and industrial progress. Rather, a reasonable balance must always be found between multiple dimensions such as development and safety, change and stability.

Second, building "responsible artificial intelligence" is the core connotation of China's values.

In the *Governance Principles for New Generation Artificial Intelligence* issued by the National New Generation Artificial Intelligence Governance Professional Committee, "responsible artificial intelligence" is described as a theme of the governance principles. Developing "responsible artificial intelligence" as an important criterion runs throughout the research, development and application processes of artificial intelligence, forming the core connotation of China's values on artificial intelligence governance.

Building "responsible artificial intelligence" requires that all work from basic research and development to practical application must be kept within the scope of safety and controllability and that all participating entities must maintain corresponding ethical and safety standards with a responsible attitude. As for ethics, it is necessary to ensure that artificial intelligence technology always conforms to human values, ethics and morals. Under the premise of ensuring social security and respecting human rights and interests, misuse must be avoided, and abuse and malicious use must be prohibited. At the same time, it is necessary to ensure that citizens can enjoy the fruits of technological progress in a fair manner, as well as to promote the transformation and upgrading of various industries and sectors, narrow regional gaps, enhance the adaptability of vulnerable groups, and strive to eliminate the digital divide. As for safety, it is necessary to continuously enhance the transparency, interpretability, reliability and controllability of artificial intelligence, and gradually meet the auditability, supervisability, traceability, trustworthiness and other requirements. As for data acquisition, it is necessary to respect and protect personal privacy, fully protect individuals' right to know and to choose, and prevent information and data abuse.

Artificial intelligence developers, users, and other relevant parties should aim to build "responsible artificial intelligence", and strictly abide by laws, regulations, ethics, morals, standards and norms with a high sense of social responsibility and self-discipline. An artificial intelligence accountability mechanism needs to be established to specify the responsibilities of developers, users, beneficiaries, and so on. Only governance rules based on this value can maximally reflect China's national will.

Third, adhering to open exchange and diversified collaboration is the fundamental spirit of China's values.

The development of artificial intelligence technology and industry is a systematic undertaking that requires extensive exchange and collaboration, which include the division of labor and collaboration among multiple entities, as well as full exchange and resources sharing worldwide. From its long practices, China fully understands the importance of open exchange and diversified collaboration for the development of artificial intelligence technology, which constitute the basic spirit of China's values on artificial intelligence governance.

Diversified collaboration refers to that the development of artificial intelligence technology requires full exchanges across disciplines, fields, regions, and borders. It is necessary to proactively promote coordinated interaction among international organizations, government departments, research institutions, educational institutions, enterprises, social organizations, and the public in the development and governance of artificial intelligence. In the process of collaboration and interaction, it is necessary to adopt a fair attitude to repay all contributors, uphold a fair spirit to safeguard the rights and interests of stakeholders, fully unleash the potential of all entities, and provide equal opportunities.

Open exchange emphasizes that in the artificial intelligence field, countries around the world are at different stages of development and there are various unique application scenarios in the development of artificial intelligence. This requires collaborative research and dialogue on the basis of exchange and communication, continuous promotion of international cooperation, as well as efforts to form a widely recognized international framework, standards and norms on artificial intelligence governance under the premise of fully respecting the principles and practices of artificial intelligence governance in various countries.

The development of artificial intelligence is not a matter for one country or one place. Isolation and beggar-thy-neighbor behaviors are not rational ways to develop artificial intelligence. Only by fully respecting the principles and practices of artificial intelligence governance in various countries and promoting the formation of an internationally recognized framework, standards and norms on artificial intelligence governance can the common well-being of humanity be enhanced.

In summary, the global development of artificial intelligence has now entered a new stage, and technological progress is profoundly changing human and social life. Humanity is entering an all-new world. Faced with the all-new situation, how to ensure the safe, reliable, controllable development of artificial intelligence technology and constantly keep it on the track of improvement has become a challenge that China and the rest of the world must face together. On the basis of theoretical research and practical integration, China has gradually formed governance ideas and values with Chinese characteristics. Under the guidance of these ideas and values, China will continuously improve intelligent technological means, optimize management mechanisms, improve governance systems, use them as the anchor point of its thinking and participate in the formulation of international rules. Based on these efforts, China will promote the healthy and orderly development of new generation artificial intelligence, better coordinate the relationship between development and governance, ensure that artificial intelligence boosts sustainable economic, social and ecological progress, and build a community with a shared future for humanity in the artificial intelligence field together.

IV. China's Path to Participate in the Formulation of International Rules on Artificial Intelligence

In order to drive the sustainable development of economic and social life and the continuous improvement of artificial intelligence technology for the good, China should follow the guidance of the idea of a community with a shared future for mankind to proactively participate in the formulation of international rules for artificial intelligence showcase and promote the Chinese values and experience in the artificial intelligence field in a way that is understandable and acceptable to the international community, and actively put forward its own plan for establishing international rules for artificial intelligence, striving to gain the initiative in international rulemaking. To this end, this report proposes four major principles for China to participate in the formulation of international rules for artificial intelligence, namely AI Governance by UN. which is to pursue international governance with the United Nations as the main stage; AI for Development, which takes promoting global economic and social development as the critical goal; AI for Stability, which takes establishing stable big power mechanisms as the basic safeguard; and AI Implication by Private Sector, which develops artificial intelligence application solutions based on the knowledge, experience and resources of the private sector. From these principles, a unique path for China to participate in the establishment of international rules for artificial intelligence will be formed, which is to "achieve multilateral and stable development of AI for all people (MSDP), with the United Nations as the center".

Firstly, taking multilateralism platforms represented by the United Nations as the fundamental foothold. As the most representative and authoritative international organization in the world, the United Nations plays an irreplaceable leading role in global digital governance and rulemaking.

In the first open debate on artificial intelligence held by the United Nations Security Council, Secretary-General Guterres emphasized that all parties must work together and strive to develop artificial intelligence to bridge the social, digital and economic divide, rather than further widening the distance between people. To this end, China should proactively respond to the call of the UN Secretary-General and formulate relevant rules within the framework of the United Nations through extensive international consultations, while ensuring the reasonable development interests of developing countries. For example, under the leadership of the Office of the Secretary General's Special Envoy on Technology, China has proactively participated in the process of reaching agreement on the *Global Digital Compact* at the 2024 United Nations Summit of the Future and will jointly formulate a legally binding agreement with various parties on the prohibition of use of fully autonomous weapon systems by 2026. China also needs to continue to support the work of UNESCO and other organizations on the ethics and standards of artificial intelligence and promote the implementation and supervision of the *Recommendation on the Ethics of Artificial Intelligence worldwide*.

Secondly, proactively participating in the formulation of international rules for artificial intelligence and making greater contributions to bridging the gap in developing countries' participation in artificial intelligence governance. Currently, there is a significant digital divide in the global artificial intelligence governance domain. Due to various technological and political reasons, many developing countries have clearly not made their voices on artificial intelligence governance heard. Their internal demands cannot be effectively expressed, and often they have to passively accept the various standards set by developed countries, making it difficult for them to deeply share the dividends of artificial intelligence development. Therefore, China as the world's largest developing country plays a special role in the process of international rulemaking for artificial intelligence. China is not only the main representative for developing countries to participate in the formulation of artificial intelligence rules, but also a key force in reversing the lack of voice of developing countries on artificial intelligence governance. China needs to proactively coordinate the positions of developing countries, effectively understand their core concerns about relevant international rules, and organically integrate their positions with the Chinese stance, fully express these positions in the process of international rulemaking, and provide an effective path for developing countries to participate in the establishment of a global artificial intelligence governance system.

Thirdly, prioritizing the promotion of economic and social development in the Global South. Fundamentally speaking, in order to bridge the digital divide and enable countries around the world to share the development dividends brought by artificial intelligence technology, it is necessary to fully promote the Global South to engage in deep cooperation in artificial intelligence and encourage countries of the Global South to get deeply involved in the development of the artificial intelligence industry. Therefore, China should prioritize the promotion of the development and application of artificial intelligence technology by countries of the Global South, and proactively engage in technological and industrial cooperation with important countries and regions. In particular, China

should pay attention to potential markets such as Southeast Asia and the Middle East, proactively try different cooperation plans, accumulate beneficial experience in the development and governance of artificial intelligence from the perspective of the Global South, actively promote the participation of Global South countries in the global artificial intelligence industrial and value chains, and enhance their comprehensive influence and discourse in the practical process of international rulemaking. In addition, China also has rich practices and cases in utilizing artificial intelligence to promote economic and social development, such as smart cities, smart healthcare, smart education, and smart agriculture. China should proactively share with countries of the Global South its knowledge, experience and resources in the artificial intelligence field and assist in the building of economic and social development and governance systems.

Fourthly, taking the establishment of artificial intelligence stability mechanisms as an opportunity to promote major country cooperation. The current situation of international rulemaking on artificial intelligence is complex, and some countries hope to promote the consensus of their small circles and promote it to the rest of the world in the form of international rules, so as to gain the dominant power in rulemaking. Despite fierce competition and bargaining, rulemaking itself remains significant for the development of artificial intelligence technology. In February 2023, the Summit on Responsible Artificial Intelligence in the Military Domain jointly hosted by the Netherlands and the Republic of Korea was held in the Hague. More than 60 countries, including the United States and China, signed a joint statement calling for the responsible development and use of artificial intelligence in the military domain. As a representative of developing countries, China also needs, while upholding the interests of developing countries, to proactively engage in dialogue and cooperation with leading technology countries such as the United States, European countries and Japan to seek consensus and prevent miscalculations, and establish multi-level dialogue mechanisms accordingly to continuously update ideas of governance. For example, China can promote the establishment of multiple-faceted artificial intelligence safety and stability dialogue mechanisms, and establish coordinated, complementary, and collaborative artificial intelligence cooperation networks at the regional level through the European Union, APEC, and other organizations to boost technological exchange and industrial collaboration in artificial intelligence. At the bilateral level, China can establish dialogue mechanisms on artificial intelligence safety and stability with the United States, Britain and other technology powers to promote bilateral exchange and cooperation in the research and application of artificial intelligence.

Fifthly, prioritizing the development of artificial intelligence application solutions based on the knowledge, experience, and resources of the private sector. As an important force in the global development of artificial intelligence, China has a large number of entities of various types related to the development of artificial intelligence technology and industry, including numerous technology and industry regulatory departments, industry associations, internet enterprises, universities, research institutions and social science-related research departments, all of which have made varying degrees of contributions to the development of artificial intelligence in the country. Big technology enterprises play a crucial role in promoting the development of artificial intelligence technology

that cannot be replaced by other entities. They are the main owners of various intelligent technology development elements, the main drivers of cutting-edge intelligent technology development, and of course, the main beneficiaries of the artificial intelligence industry. Therefore, big internet companies are more enthusiastic about for the establishment of systems of rules, and they are also most sensitive to and most familiar with the potential risks of artificial intelligence. For this reason, in the process of formulating international rules for artificial intelligence, it is necessary to pay full attention to the views of Chinese artificial intelligence enterprises, care about their core positions and demands, and actively encourage them to participate in international dialogues at the multinational corporation level and become the main force in international rulemaking.

About SIIS

Founded in 1960, the Shanghai Institutes for International Studies (SIIS) is a government- affiliated high-caliber think tank dedicated to informing government decision-making by conducting policy-oriented studies in world politics, economics, foreign policy, and international security. SIIS maintains intensive and extensive exchanges and cooperation with research institutions at home and abroad, bolstering China's international influence and soft power.

SIIS boasts an authorized size of 106 full-time research fellows and staff, including 60 percent senior fellows. SIIS was ranked one of the top ten Chinese think tanks for the long time being. SIIS comprises six institutes and six research centers, namely, the institute for global governance studies, the institute for foreign policy studies, the institute for world economic studies, the institute for international strategic studies, the institute for comparative politics and public policy, the institute for Taiwan, Hong Kong & Macao Studies, the center for American studies, the center for Asia-Pacific Studies, the center for Russian and Central Asian Studies, the center for West Asia and Africa studies, the center for European studies, and the center for maritime and polar studies. SIIS has also set up six in-house research platforms, i.e., the research base on people's diplomacy of Shanghai, center for the study of Chinese diplomatic theory and practice, center for world politics and political parties, center for China-South Asia cooperation, center for BRI and Shanghai studies, and center for international cyber governance. In addition, SIIS is an institutional member of the Shanghai International Strategic Studies Association and the Shanghai International Relations Association.

About CISS

Established on November 7, 2018, the Center for International Security and Strategy (CISS) of Tsinghua University is an university-level research institution in the field of international security and strategy. With Madame FU Ying, former Vice Foreign Minister of China, being its Founding Chair, CISS aims to remain current with the changes in global dynamics, offering policy suggestions for decision-makers through researches on topics of foreign affairs, international relations, and security and strategic studies. It strives to promote, elucidate, and spread China's perspectives and policy views by carrying out various forms of knowledge exchanges and cooperation internationally, aiming to enhance the international community's understanding of China and improve Tsinghua's global influence in the fields of international relations and strategic studies.

The Center has an Academic Committee that serves as the academic advisory body, and below consists of 6 divisions including the Research Project on the US and Europe, the Research Project on Global Governance, the Research Project on Eurasia, the Research Project on AI and International Security, the "CISS Youth " Research Exchange Program, and the Secretariat of China Forum.

About IDRI

The Internet Development Research Institution（IDRI）of Peking University is a research institution established by Peking University to keep up with the times, serve national cyber development strategy, and promote knowledge innovation and cultural exchanges in the Internet field. Centered on the development of the Internet, IDRI is committed to problem oriented, social oriented, and future oriented research. Relying on the academic advantages and social influence of Peking University, IDRI is committed to carrying out interdisciplinary integrated research combining academic, policy and practical research, and also providing relevant research, consultation and training services for government agencies, enterprises, and social organizations, aiming to promote innovation in research, so as to better utilize the Internet for the benefit of society and human welfare.

Research fields of IDRI include:

Internet industry development, Internet governance, Internet ethics, Internet and society, Internet and communication, digital protection and development of minors, etc.

上海国際问题研究院
SHANGHAI INSTITUTES FOR INTERNATIONAL STUDIES

Shanghai Institutes for International Studies
195-15 Tianlin Road, Xuhui District
Shanghai 200233, PRC
Tel/Fax: +86 21 64850100
www.siis.org.cn